Identifying Relevant Evidence for Systematic Reviews and Review Updates

Amal H. Alharbi

Faculty of Engineering

Department of Computer Science

University of Sheffield

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Faculty of Engineering December 2020

I would like to dedicate this thesis to the loving memory of my mother *Eidah Albahith* and my father *Hamed Alharbi* ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Amal H. Alharbi

December 2020

Acknowledgements

During my PhD journey, I have faced many obstacles and challenges. There are special people whose presence has helped me keep going and get through all the hard times. I want to take the opportunity to thank them.

First of all, I wish to express my deepest gratitude to my supervisor, Dr Mark Stevenson. This work would not have been possible without your invaluable support. Thank you very much for your continued support, motivation, patience and wonderful leadership. You helped me through my PhD journey by your inspiring guidance and valuable discussions. I learnt a lot from you.

I want to thank my beautiful family. Words cannot express how grateful I am to my mother-in-law, my father-in-law, my sister and my brothers. Thank you for your continued prayers and support. Thank you for your unlimited love. Thank you for your trust in me.

I wish to express my special gratitude to my beloved husband, Engr Waleed Alahmadi. Regardless of the distance that separated us during my study, you were always with me. Thank you for all the support and love.

My lovely daughter, my little princess, Tala. She was only two years old when I started the PhD, and she grew up in Sheffield. Your beautiful smile removes the trouble of a long day. Your love has been always giving me the strength to complete this journey and success.

Special thanks for my dearest friends Rehab and Areej. Your friendship, support and love helped me during this journey. I'm thankful to have shared beautiful moments with you.

Finally, I take the opportunity to thank King Abdulaziz University for their financial support.

THE UNIVERSITY OF SHEFFIELD

Abstract

Faculty of Engineering

Department of Computer Science

Doctor of Philosophy

Identifying Relevant Evidence for Updating Systematic Reviews

By Amal H. Alharbi

Systematic reviews identify, assess and synthesise the evidence available to answer complex research questions. They are essential in healthcare, where the volume of evidence in scientific research publications is vast and cannot feasibly be identified or analysed by individual clinicians or decision makers. However, the process of creating a systematic review is time consuming and expensive. The pace of scientific publication in medicine and related fields also means that evidence bases are continually changing and review conclusions can quickly become out of date. Therefore, developing methods to support the creating and updating of reviews is essential to reduce the workload required and thereby ensure that reviews remain up to date.

This research aims to support systematic reviews, thus improving healthcare through natural language processing and information retrieval techniques. More specifically, this thesis aims to support the process of identifying relevant evidence for systematic reviews and review updates to reduce the workload required from researchers.

This research proposes methods to improve studies ranking for systematic reviews. In addition, this thesis describes a dataset of systematic review updates in the field of medicine created using 25 Cochrane reviews. Moreover, this thesis develops an algorithm to automatically refine the Boolean query to improve the identification of relevant studies for review updates.

The research demonstrates that automating the process of identifying relevant evidence can reduce the workload of conducting and updating systematic reviews.

Table of Contents

List of Figures **xv**

List of Tables **xix**

List of Abbreviations **xxiii**

1 Introduction **1**

1.1 Background 3

1.1.1 Systematic Review 3

1.1.2 Systematic Review Stages 5

1.1.3 Systematic Review Updates 7

1.1.4 Challenges in Developing Systematic Reviews 8

1.2 Research Aim and Objectives 10

1.3 Thesis Contributions 10

1.4 Published Work 11

1.5 Thesis Outline 12

2 Systematic Review of the Literature **15**

2.1 Introduction 15

2.2 Framing the Research Question 16

2.2.1 Questions 16

2.2.2 Inclusion and Exclusion Criteria 17

2.3 Identifying Relevant Studies 17

2.3.1 Boolean Search 17
2.3.2 Studies Screening 20
2.4 Results 21
2.4.1 Q1. Which NLP/IR techniques have been proposed to support the screening process? 22
2.4.2 Q2. What are the datasets used? Are they publicly available? 31
2.4.3 Q3. How are those techniques evaluated? 39
2.4.4 Q4. Which techniques are applied in the screening stage of the review update process? 44
2.5 Limitations 46
2.6 Summary 47

3 Query Adaptation to Improve Ranking 53
3.1 Introduction 53
3.2 Approach 1: Query Terms and Medical Subject Headings 55
3.2.1 Boolean Query 55
3.2.2 The Medical Subject Headings (MeSH) Hierarchy 57
3.2.3 Dataset 59
3.2.4 Experiments 60
3.2.5 Evaluation Metrics 65
3.2.6 Results and Discussion 65
3.3 Approach 2: Lexical Statistics 70
3.3.1 Log-Likelihood 71
3.3.2 Chi-Squared 72
3.3.3 Odds-Ratio 72
3.3.4 Experiments 73
3.3.5 Results and Discussion 75
3.4 Approach 3: Relevance Feedback 82
3.4.1 Rocchio's Algorithm 83
3.4.2 Experiments 84

3.4.3 Results and Discussion 85
3.5 Summary 87

4 A Dataset of Systematic Review Updates 89

4.1 Introduction 89
4.2 Dataset Configuration 90
4.2.1 Boolean Query 92
4.2.2 Included and Excluded Studies 94
4.2.3 Update History 96
4.3 Dataset Characteristics 97
4.4 Experiments 101
4.4.1 Approaches 101
4.4.2 Results and Discussion 102
4.5 Summary 107

5 Boolean Query Refinement to Improve the Identification of Relevant Studies 109

5.1 Introduction 109
5.2 Method 110
5.2.1 Step One: Boolean Query Transformation 111
5.2.2 Step Two: Boolean Query Selection 115
5.3 Dataset 116
5.4 Experiments 117
5.4.1 Approach 1: Baseline 117
5.4.2 Approach 2: Query Refinement 117
5.4.3 Approach 3: Oracle 118
5.5 Evaluation Metrics 118
5.6 Results and Discussion 119
5.7 Summary 126

6 Conclusion and Future Directions 127

6.1 Summary of the Thesis 127

6.2 Future Directions 129

Appendix A Search Results 133

Appendix B CLEF Datasets Characteristics 215

References 219

List of Figures

1.1 The process of conducting a systematic review. 5
1.2 Flow diagram describing the study selection process. 9

2.1 Inclusion and exclusion criteria. 17
2.2 Search query to retrieve studies from PMC. 18
2.3 Flow diagram for studies selection process. 21
2.4 Distribution of included studies by year. The peak in 2017 is due to the CLEF task. 22
2.5 Distribution of included studies by journal. 22
2.6 Distribution of included studies by approach. 23
2.7 Distribution of included studies by dataset used. 31
2.8 Example Cochrane reviews used in CLEF2018 training dataset (Theron et al., 2016). 38
2.9 Distribution of included studies by evaluation measure. 39
2.10 Example of Area Under the ROC Curve. 43
2.11 Ranking effectiveness example. 44
2.12 Distribution of included studies by type. 45

3.1 Example queries from Cochrane reviews (Nisenblat et al., 2016; Williams et al., 2013). 56
3.2 Example of an exploded MeSH including the narrower subject headings in the `Dementia` tree hierarchy. 58

3.3 Example of MeSH terms for an article from PubMed (Nakagawa et al., 1977). Major MeSH terms are denoted by asterisks '`*`'. 59
3.4 Example portion of a Boolean query (Van de Vrie et al., 2019) (a), sample of terms extracted from the Boolean query (b), sample of pre-processed MeSH headings extracted from the Boolean query (c) and the pre-processed MeSH headings for "`exp Dementia/`" (d). 62
3.5 AP scores for each review using baseline and different approaches for CLEF2017 test dataset. 68
3.6 AP scores for each review using baseline and different approaches for CLEF2018 test dataset. 69
3.7 Example of Baseline query (a) and expanded queries (b-d) generated by adding top five terms generated form each lexical statistic. 75
3.8 AP scores for each review in CLEF2017 test dataset using baseline and the three lexical statistics. 78
3.9 AP scores for each review in CLEF2018 test dataset using baseline and the three lexical statistics. 79
3.10 AP scores for each review in the CLEF2017 test dataset using baseline and Relevance Feedback. 86
3.11 AP scores for each review in the CLEF2018 test dataset using baseline and Relevance Feedback. 86

4.1 Forest plot diagram from review CD002733 (Kopsaftis et al., 2018) 91
4.2 Example portion of a Boolean query (Hughes et al., 2007) in (a) the OVID format and (b) its translation into the single-line PubMed format. This portion of the query contains three clauses, and the last clause represents the combining results of clause 1 and 2 in a disjunction (OR). 92
4.3 Example of a Boolean query (Campbell and Strippoli, 2017) which has a mistake in the lines numbers: the last line (no. 9) combines the results of lines 5 to 8 and ignores the first four lines of the query. 93
4.4 Example of studies with available PMID (*highlighted*). 94

4.5 Example of search query generated from title and publication year for an article without a PMID. 95
4.6 The result of searching PubMed using the query in Figure 4.5. 95
4.7 An example of version history information available with Cochrane review (Hughes et al., 2007). 97
4.8 The structure of the update dataset. 100
4.9 AP scores for each review using Baseline Query, Relevance Feedback and Chi-Squared. 105
4.10 Review Background (a), Boolean Query (b) and the top 100 terms with highest lexical statistics Chi-Squared scores (c) for review CD004214 (New et al., 2011). 106

5.1 Example of query expansion applied to the first clause of the Boolean query of review CD005025 (a) by adding up to five terms generated by Log-Likelihood (b) and the full list of transformations that can be added to the clause (c). 114
5.2 Example of query reduction for review CD005025 (Mahtani and Perera, 2011).115
5.3 Weighted Average Recall scores for the various approaches. The baseline approach is included to allow comparison. 123
5.4 Weighted Average Precision scores for the various approaches. The baseline approach is included to allow comparison. 123
5.5 Example of the original Boolean query for review CD002064 (Beauverd et al., 2012) (a) and the transformed Boolean query after nine iterations (b) with highlighted lines representing the clauses transformed by the algorithm. . . 125

List of Tables

2.1 Characteristics of the Drugs dataset. 32

2.2 Characteristics of COPD and Proton datasets. 33

2.3 Characteristics of CLEF datasets. 33

2.4 CLEF2017 training dataset characteristics. 34

2.5 CLEF2017 test dataset characteristics. 35

2.6 CLEF2018 training dataset characteristics. 36

2.7 CLEF2018 test dataset characteristics. 37

2.8 Confusion Matrix. 40

2.9 Summary of the information extracted from the 63 included studies. 49

2.9 Summary of the information extracted from the 63 included studies. 50

2.9 Summary of the information extracted from the 63 included studies. 51

3.1 Set of OVID and PubMed Boolean query operators and restriction fields with their meanings. 57

3.2 OVID and PubMed common query restriction fields and whether we consider each one a term or a MeSH. 60

3.3 Results of making use of different information from the Boolean query and studies for CLEF2017 and CLEF2018 test datasets. The best performance among all methods is in boldface. 66

3.4 Contingency table for computing lexical statistics. 71

3.5 Lexical statistic results for CLEF2017 and CLEF2018 test datasets. Values in boldface denote the best result achieved by each lexical statistic and the underlined values represent the best results among all three lexical statistics. 77
3.6 Top 20 terms based on different lexical statistics scores derived from CLEF2017 training dataset. 80
3.7 Top 20 terms based on different lexical statistics scores derived from CLEF2018 training dataset. 81
3.8 MAP scores over a range of values for the weighting parameters β and γ using the training dataset of CLEF2018. 85
3.9 Relevance Feedback results for the CLEF2017 and CLEF2018 test datasets. . 85

4.1 OVID restriction fields and their equivalent in PubMed format. 93
4.2 List of the 25 systematic reviews with the Boolean query type, the total number of studies returned by the query (Total) and the number included following the *Abstract* and *Content* screening stages. The average (unweighted mean) number of studies is shown in the bottom row. Note that for the updated review, the number of included studies in the table indicates only the new studies that were added during the update. 99
4.3 Performance ranking abstracts for updated reviews at (a) abstract and (b) content levels. Results are computed across all reviews at the abstract level (25 reviews) and only across reviews in which a new article was added in the updated version for the content level (19 reviews). Values in boldface denote the best result achieved among approaches. Superscript * and †in MAP indicate that the corresponding method significantly outperformed the Baseline with $p < 0.001$ and $p < 0.05$, respectively 103
4.4 Performance ranking abstracts for updated reviews by adding different numbers of top terms for each lexical statistic. Values in boldface denote the best result achieved by each lexical statistic, and the underlined values represent the best results among all three lexical statistics. 104

5.1 Recall and Precision results for each review in the update dataset. Values in boldface denote results improved when compared with the baseline. 121
5.2 Analysis of transformation types used in each method in the query refinement approach and oracle. The numbers represent how many times each transformation has been used through all iterations. 124

B.1 CLEF2019 DTA test dataset characteristics. 215
B.2 CLEF2019 Interventions training dataset characteristics. 216
B.3 CLEF2019 Interventions test dataset characteristics. 217
B.4 CLEF2019 Prognosis test dataset characteristics. 217
B.5 CLEF2019 Qualitative test dataset characteristics. 217

List of Abbreviations

AutoTAR Autonomous Technology Assisted Review

BMC Bio Medical Central

CLEF Conference and Labs of the Evaluation Forum

COPD Chronic Obstructive Pulmonary Disease

DTA Diagnostic Test Accuracy

HealTAC Healthcare Text Analytics Conference

IR Information Retrieval

JAMIA Journal of the American Medical Informatics Association

MERS Middle East Respiratory Syndrome

MeSH Medical Subject Heading

NLP Natural Language Processing

NLTK Natural Language Toolkit

PMC PubMed Central

PMID PubMed Identification

SARS Severe Acute Respiratory Syndrome

SVM Support Vector Machine

TBL Transformation-Based Learning

WHO World Health Organization

WSS Work Saved over Sampling

Chapter 1

Introduction

The volume of publications that appear in the field of medicine on a daily basis is rapidly increasing. Consequently, healthcare researchers, clinicians and policy-makers are deluged with this uncontrollable amount of information, including evidence from health research (Chalmers, 2000; Masic et al., 2008). A common scenario in clinical practice is when a physician finds in one study that drug A is recommended to treat a particular disease X, while in another research, they find that drug B is also advised as a treatment for the same disease (Avery et al., 2013; Rahmner et al., 2012; Slawson and Reed, 2009). In such a scenario, they become confused about which treatment is more effective to prescribe. They need strong evidence to make a decision. It is impractical to rely on the results of one or two studies to make decisions. Furthermore, with the growing number of published articles, it is not possible for a physician to read all up-to-date published evidence, assess it and take decisions. It requires time and skill to deal with this amount of information when searching the literature to find and interpret evidence and apply it to clinical practice (McGowan and Sampson, 2005; Sutherland, 2004).

In the past decades, many studies have been published that do not lead to single, clear and practicable results to be followed. In 1979, Archie Cochrane, a British physician in whose honour the well-known Cochrane Collaboration was named, wrote:

> "It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials." (Cochrane, 1979)

Cochrane realised that healthcare specialists and clinicians who need to make healthcare decisions needed reliable reviews of the available evidence. Systematic reviews attempt to identify, synthesise and summarise all available evidence to answer a specific research question. A good systematic review is a significant addition to the medical literature. They help healthcare researchers and clinicians benefit from the large amount of information available to make healthcare decisions. In addition, they help correct common misconceptions. As an example, vitamin C supplementation had been proposed for preventing and treating common cold since the 1930s. However, in 1998, a systematic review was conducted and found that vitamin C supplementation has no benefit in avoiding the cold but can lightly reduce the duration of the cold symptoms (Hemila and Chalker, 2013).

In recent years, the increase of the volume of medical publications has made the process of summarising the available evidence difficult for individuals (i.e. healthcare researchers, clinicians and policy-makers). Therefore, the need for systematic reviews to summarise the evidence has become more urgent (Bastian et al., 2010). A recent example is the novel COVID-19 virus, which first emerged in Wuhan, China, in late December 2019. By 20th May 2020, the total confirmed cases were 4,789,205, and the total worldwide deaths were 318,789 (Johns Hopkins University, 2020). COVID-19 is a new virus which has no vaccine yet. Accordingly, recognising vaccine and treatment choices as quickly as possible is essential for the reaction to the COVID-19 outbreak. However, vaccines and medications used to prevent and treat diseases from the same family (e.g. Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS)) may be useful to develop a therapy for COVID-19. Having a systematic review of all available evidence about past medications for these viruses could be useful to develop a treatment for the new COVID-19, which would save the lives of many people. A PubMed search for "SARS" and "MERS" returns about 9,000 and more than 14,000 articles, respectively. This makes the process of conducting such a review laborious and time consuming, and with the wide spread of COVID-19, it is worthwhile to obtain tools to facilitate the interpretation of this information as quickly as possible.

While the very existence of systematic reviews is important, it is yet more important for them to stay up to date as evidence changes. This is, however, challenging in a field such as medicine where thousands of publications appear on a daily basis (Pain, 2016). Developing methods to support the updating of reviews is important to reduce the workload required and thereby ensure that reviews remain valuable and useful.

This thesis seeks to improve the process of conducting and updating systematic reviews, thus improving healthcare. It aims to apply Natural Language Processing (NLP) and Information Retrieval (IR) techniques to facilitate the process of finding relevant evidence.

This chapter provides background information about systematic reviews including the process of conducting reviews and review updates. It also outlines the main challenges in conducting reviews. The rest of this chapter details the aims and objectives of the research described in this thesis and summarises the main contributions.

1.1 Background

1.1.1 Systematic Review

The Cochrane Handbook for Systematic Reviews of Interventions defines systematic review as:

> "A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made (Antman et al., 1992; Oxman, 1993)." (Green et al., 2011).

The Cochrane Collaboration is one the key producers of medical systematic reviews. Its library contains over 8,300 reviews[1] which fall into five main categories (About Cochrane Reviews, 2019):

1. **Intervention reviews:** These reviews mainly assess the benefits and harms of interventions used in healthcare and health policy. An example is a review that assesses the effects of physical exercise training in patients with chronic kidney disease and determines how the exercise programme should be designed to have an effect on the fitness of the patients (Heiwe et al., 2011).

2. **Diagnostic Test Accuracy reviews (DTA):** These reviews assess the accuracy of a diagnostic test when used to detect a particular disease. An example of a DTA review is a review that assesses the accuracy of serum-based markers for Down's syndrome screening in the first trimester (Alldred et al., 2015).

3. **Methodology reviews:** These reviews explore issues associated with the process of conducting systematic reviews and clinical trials. For example, they compare searching manually to identify randomised trials with using electronic searching (Hopewell et al., 2007).

4. **Qualitative reviews:** These address questions related to healthcare interventions other than effectiveness by synthesising qualitative evidence. For example, a review that identifies and synthesises qualitative studies exploring women's views and experiences of attending antenatal care and Healthcare providers' views and experiences of providing antenatal care (Downe et al., 2019).

5. **Prognosis reviews:** These reviews address the probable course or future outcome(s) of individuals with a specific health problem (i.e. diseases or conditions). An example is a review that finds whether protease activity is an independent prognostic factor for the healing of venous leg ulcers (Westby et al., 2018).

[1]At the date of writing this Thesis, May 2020

1.1.2 Systematic Review Stages

Systematic reviews often progress in a number of stages, as shown in Figure 1.1 (Boland et al., 2014; Gough et al., 2012a; Khan et al., 2003; Tsafnat et al., 2014). Below, the steps are described in further detail.

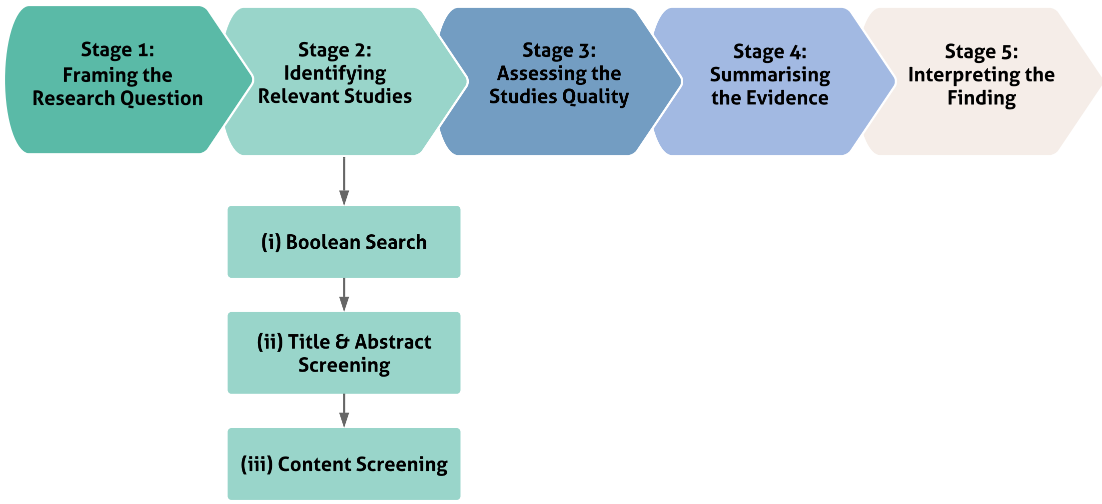

Figure 1.1: The process of conducting a systematic review.

Stage 1: Framing the Research Question

In the beginning, experts identify the objective of the review and define a clearly formulated research question, for example "for deep vein thrombosis is D-dimer testing or ultrasound more accurate for diagnosis?" (Nisio et al., 2007). In this stage, researchers prepare the review protocol. In addition, they declare the inclusion/exclusion criteria (eligibility criteria) to decide about the studies of interest to be included in the review. The eligibility criteria should not be too narrow or too broad because this will lead to an inefficient screening process (Jahan et al., 2016).

Stage 2: Identifying Relevant Studies

Reviewers perform an exhaustive search to find potentially eligible studies that match the predefined criteria. This stage usually consists of multiple steps (Kanoulas et al., 2017):

(i) **Boolean Search:** Experts construct a Boolean query designed to identify all evidence relevant to the review question. The Boolean query is used to search medical databases such as MEDLINE, EMBASE and CENTRAL.

(ii) **Title and Abstract Screening:** Only the title and abstract of the studies retrieved by the Boolean search are examined by experts to identify those that are potentially relevant for inclusion in the review. It is common for the majority of studies to be removed from consideration during this step (O'Mara-Eves et al., 2015; Sampson et al., 2011).

(iii) **Content Screening:** The full document is then retrieved for any study that has been identified as being relevant at the previous step. These are then examined in a second round of expert screening to form the final decision about their relevance to the review.

The screening processes (i.e. steps ii and iii) are usually performed by two reviewers independently of each other to assess studies for inclusion. This is done to avoid systematic errors, missing studies and the risk of bias in study selection (Gough et al., 2012a; Waffenschmidt et al., 2019).

Stage 3: Assessing Study Quality

The systematic review should minimise bias. The Cochrane Handbook for Systematic Reviews of Interventions defines bias as "a systematic error, or deviation from the truth, in results or inferences" (Green et al., 2011). Bias leads to errors in the review results, therefore, reviewers assess the risk of bias and the quality of all the included studies. This includes the assessment of the data and results extracted from included studies. They should be correct, valid and free of bias.

Stage 4: Summarising the Evidence

In this stage, researchers analyse, summarise and present the findings of the included studies. This includes presenting the main findings statistically in a simple tabular format

(Lewis and Clarke, 2001) and meta-analysis of the results (for quantitative systematic reviews) (Deeks et al., 2019), or non-statistically (for qualitative systematic reviews) (Noyes et al., 2019).

Stage 5: Interpreting the Finding and Drawing Conclusions

In the final stage, reviewers interpret the findings and publish the conclusion of the systematic review. This includes describing the strengths and weaknesses of the included studies and indicating future directions to strengthen the review.

1.1.3 Systematic Review Updates

As new evidence is completed and published, a systematic review may become out-of-date. Systematic reviews need to be updated to include the most recent evidence to continue to be useful. The process that is applied to update a systematic review is similar to the one used to create a new review (Elkins, 2018). A search query is run to identify studies published since the previous version of the review and the resulting studies are screened in a two-stage process: *title and abstract screening* and *content screening*. If any new relevant studies are found, then the data is extracted and integrated into the review. The review's findings are also updated if the evidence is found to have changed from the previous version.

Deciding when a review should be updated is a challenging problem since there is no commonly agreed approach on when this should happen (Elliott et al., 2017; Garner et al., 2016). A review can be updated at any point after it has been created (or already updated) and while the process should ideally be carried out whenever new evidence becomes available, the effort required makes this impractical. A common practice is to update reviews after a certain period has passed, for example, the Cochrane Collaboration recommends that reviews should be updated every two years (Moher and Tsertsvadze, 2006). Cochrane's Living Evidence Network has recently started developing living systematic reviews with the aim to produce evidence that is both reliable and up to date. The approach of this network is based on reviewing evidence frequently (normally monthly) and if any new evidence is identified then it is included in the review immediately. However, it is unclear whether this effort is sustainable (Elliott et al., 2017). The Agency for Healthcare Research and Quality suggests that reviews be updated depending on need, priority and the availability of new evidence (Lucenteforte et al., 2018).

1.1.4 Challenges in Developing Systematic Reviews

A range of challenges face developers of systematic reviews. The process of creating a systematic review is time consuming. A single review often requires from six months to more than two years of effort by expert reviewers (Chandler et al., 2019; Cohen et al., 2010; Karimi et al., 2010). The reviewers need to perform an extensive search and evaluation of the literature to find all the studies relevant for inclusion. Typically, this requires a group of reviewers manually investigating thousands of studies that have resulted from database searches (Rathbone et al., 2015).

The screening stages are one of the most time-consuming parts of this process since an experienced reviewer takes at least 30 seconds to review an abstract, and this time can be substantially longer for complex topics (Wallace et al., 2010). The problem is made more acute by the fact that the search queries used for systematic reviews are designed to maximise recall, with precision a secondary concern, while the volume of medical publications increases rapidly.

Figure 1.2 represents an example of a screening process for a Cochrane systematic review entitled: *"Optic nerve head and fibre layer imaging for diagnosing glaucoma"* (Michelessi et al., 2015). The total number of studies retrieved from the database is 9,322. In addition, ten studies were identified through other resources. The number of studies reduced to 7,306 after removing duplicated studies. After the first screening step, more than 94% of the studies were excluded because they were clearly not relevant to the review. The studies that passed the first screening were then screened as a full text to decide if they were relevant or not. In the end, only 1.45% of the studies retrieved were included in the systematic review.

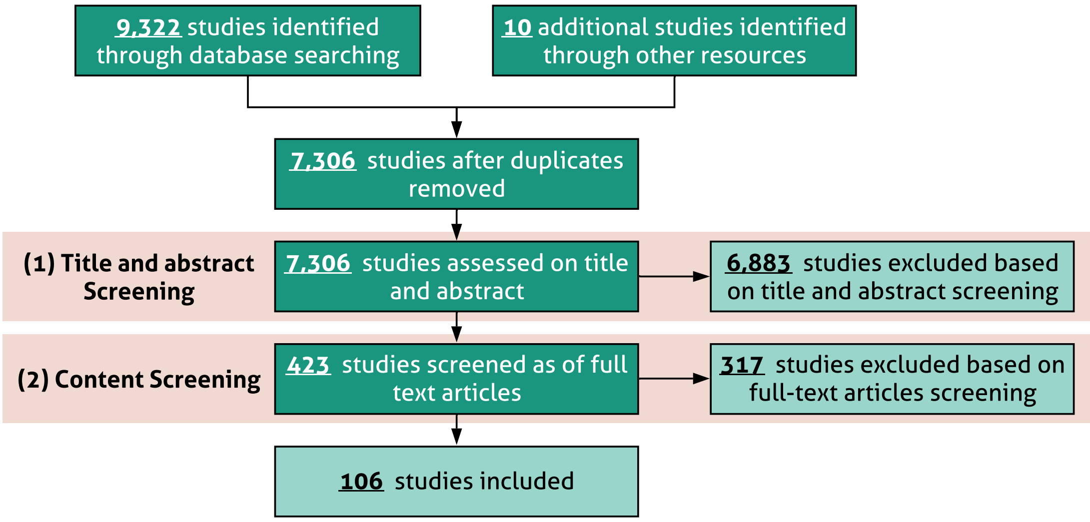

Figure 1.2: Flow diagram describing the study selection process.

Moreover, systematic reviews are costly; they require the cognitive efforts of the reviewers, who are usually experts and physicians and as such, their time is expensive. The cost of conducting a single systematic review could reach a quarter of a million U.S. dollars (McGowan and Sampson, 2005).

The pace of scientific publication in medicine and related fields also means that evidence bases are continually changing and review conclusions can quickly become out of date (Bastian et al., 2010). In fact, it has been estimated that 7% of systematic reviews in the medical field are already out of date by the time of publication; and almost a quarter (23%) two years after they have appeared (Shojania et al., 2007). Reliance on review conclusions based on out-of-date evidence increases the risk of recommendations for practice that are sub-optimal and potentially harmful to patients. With over eight thousand reviews being produced per year (Page et al., 2016), keeping them up to date presents a formidable challenge.

Therefore, there is a need to develop methods to support the process of creating and updating systematic reviews to reduce the workload required from researchers and ensure the reviews are consistent with current evidence.

1.2 Research Aim and Objectives

The main objective of this thesis is to support the process of developing systematic reviews. This research gives particular attention to systematic review updates that are of significant importance but not sufficiently addressed (see Section 2.4.4).

More specifically, this thesis aims to support the process of identifying relevant evidence for systematic reviews and review updates (i.e. Stage 2 in Section 1.1.2) to reduce the workload required from researchers and ensure the reviews are valuable and up to date.

Based on this objective, this thesis addresses the following question:

"How can NLP/IR techniques be used to reduce the workload required by researchers when identifying relevant studies for systematic reviews and review updates?".

This question leads to the following sub-questions:

RQ1. How can studies be ranked so that the potentially relevant ones appear as early in the ranking as possible?

RQ2. Can the feedback from reviewer(s) be used to improve studies rankings?

RQ3. Can the rankings for systematic review updates be improved by making use of information about the original review, such as search strategy and feedback from reviewers?

RQ4. Is it possible to generate Boolean search queries for review updates that are more effective than the one used for the original review?

1.3 Thesis Contributions

The main contributions of this thesis are:

1. The exploration of the application of lexical statistics and relevance feedback techniques to improve the ranking of studies for inclusion in systematic reviews.

2. The construction of an update dataset, which is the first publicly available dataset for the purpose of evaluating approaches to improve the identification of relevant studies for systematic review updates.

3. The development and evaluation of re-ranking techniques based on relevance feedback which use information available from the original reviews (i.e. Boolean query and relevance judgements) to improve ranking studies for systematic review updates.

4. The development and evaluation of an algorithm that automatically refines the original Boolean query to improve the identification of relevant studies for the update process.

1.4 Published Work

Work described in this thesis has been published in the following peer-reviewed conferences and journals:

1. Amal Alharbi and Mark Stevenson. Ranking abstracts to identify relevant evidence for systematic reviews: The University of Sheffield's approach to CLEF eHealth 2017 Task 2 , In Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, Dublin, Ireland, September 11-14 2017.

2. Amal Alharbi, William Briggs, and Mark Stevenson. Retrieving and Ranking Studies for Systematic Reviews: University of Sheffield's Approach to CLEF eHealth 2018 Task 2. In Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, Avignon, France, September 10-14 2018.

3. Amal Alharbi and Mark Stevenson. A Dataset of Systematic Review Updates. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19), Paris, France, July 21-25 2019.

4. Amal Alharbi and Mark Stevenson. Improving Ranking for Systematic Reviews Using Query Adaptation, In Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2019. Lecture Notes in Computer Science, vol 11696. Springer.

5. Amal Alharbi and Mark Stevenson. Ranking Studies for Systematic Reviews Using Query Adaptation: University of Sheffield's Approach to CLEF eHealth 2019 Task 2, In Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, Lugano, Switzerland, September 9-12 2019.

6. Amal Alharbi and Mark Stevenson. Refining Boolean Queries to Identify Relevant Studies for Systematic Review Updates. Journal of the American Medical Informatics Association, Volume 27, Issue 11, November 2020, Pages 1658–1666.

7. Amal Alharbi and Mark Stevenson. Using Natural Language Processing and Information Retrieval to improve screening for systematic reviews and their updates: a systematic review of the literature. (To be submitted)

1.5 Thesis Outline

The remainder of this thesis is organised as follows:

Chapter 2 describes a systematic literature review conducted to explore the use of NLP/IR techniques to facilitate the screening process for systematic reviews. The review also pays attention to the studies that tackle the problem of updating systematic reviews. It addresses four main questions: (Q1) Which NLP/IR techniques have been proposed to support the screening process?, (Q2) Which datasets are used? Are they publicly available?, (Q3) How are those techniques evaluated? and finally (Q4) Which techniques are applied in the screening stage of the review update process? The chapter starts by describing the steps followed in conducting the review then presents the answers to the questions formulated for the review.

Chapter 3 explores the use of different query adaptation approaches to improve the ranking of studies for the creation of a systematic review. Three main approaches are described. The first investigates which information from the Boolean query and studies is helpful for improving ranking. The second approach examines the use of lexical statistics in the domain of systematic reviews and how they can be used to identify terms that distinguish relevant studies from others. The final approach applies a relevance feedback method using the Rocchio algorithm. To evaluate approaches, two datasets consist of DTA reviews are used. Results further prove the ability of NLP to improve studies ranking.

Chapter 4 describes the process of creating a dataset containing 25 intervention reviews from the Cochrane collaboration. This dataset can be used to support the development of approaches to automate the updating process. This chapter also investigates the use of query expansion to reduce workload during the screening stage of a review update by making use of information from the original review. Two main approaches from Chapter 3 are applied to exploit this information: lexical statistics and relevance feedback. Results show that the relevance judgements from the original review can help to improve study selection for systematic review updates.

Chapter 5 presents and evaluates a novel algorithm proposed to automatically refine the Boolean query used in a review to improve the identification of relevant studies for review updates. An iterative algorithm is proposed to generate query variants by applying a set of transformations including operator substitution, query expansion and query reduction. These are assessed using information about which studies were included in the original review and the most effective transformation is chosen to update the query. The best query produced by the algorithm is then used to retrieve studies for the review update. The dataset described in Chapter 4 is used for evaluation. The proposed algorithm proves to be useful to help in the identification of relevant information among the growing volume of medical literature.

Chapter 6 concludes by summarising the work presented in this thesis and identifies some issues for future work.

Chapter 2

Systematic Review of the Literature

2.1 Introduction

The screening stages are one of the most time consuming parts of the process of conducting systematic reviews since an experienced reviewer takes at least 30 seconds to review an average abstract and substantially longer if the topic is complex (Wallace et al., 2010) (see Section 1.1.4). A significant number of previous studies have demonstrated the usefulness of text mining techniques to reduce the workload involved in the systematic review screening stages (see Section 2.4). Text mining can help to accelerate the process of conducting systematic reviews by automatically classifying studies as relevant or non-relevant. In addition, text mining can be used to rank the studies retrieved by the search so that those most likely to be relevant are listed at the top.

Whilst systematic reviews are particularly associated with the medical domain, they can be used to answer a question for any area of research (Gough et al., 2012b). For example, a previous systematic review of using text mining for study identification in systematic reviews conducted by O'Mara-Eves et al. (2015). This review aimed to identify the state of the art concerning the use of text mining for study identification in systematic reviews focusing on non-technical issues. They evaluated 44 studies and concluded that using text mining techniques reduced workload required to carry out reviews. However, the review by O'Mara-Eves et al. needs to be updated since it was conducted five years ago.

This chapter describes a systematic literature review conducted by using standard systematic review process (see Section 1.1.2). The aim of this review is to collect and summarise the studies on using NLP/IR techniques to facilitate the screening process for systematic reviews. The goal of this is to identify and understand the state of the art NLP/IR techniques in this field and to identify the research gaps. The review also pays attention to the studies that tackle the problem of updating systematic reviews. The review extends the one performed by O'Mara-Eves et al. by including studies published until 2020 and by focusing on systematic review updates.

This chapter provides the reader with an example of a systematic review which is the main focus of this thesis. It starts by describing the steps followed when conducting the review. Then, presents the answers to the questions formulated for the review. In addition, it lists the summary points obtained from the review's results.

2.2 Framing the Research Question

2.2.1 Questions

The aim of this review is to explore the use of NLP/IR techniques to facilitate the screening process for systematic reviews. The main research question of this review is:
Are NLP/IR techniques helpful in improving the screening stage of a systematic review?
Specifically, we seek to answer the following questions:

Q1. *Which NLP/IR techniques have been proposed to support the screening process?*

Q2. *Which datasets are used? Are they publicly available?*

Q3. *How are those techniques evaluated?*

Q4. *Which techniques are applied in the screening stage of the review update process?*

2.2.2 Inclusion and Exclusion Criteria

Figure 2.1 presents the inclusion and exclusion criteria for the studies to be included in the systematic review.

(a) Inclusion Criteria
(1) Study published between 2005 and 2020 (see Section 2.3.1)
(2) Study about creation of new/update systematic review
(3) Study focuses on (semi-)automation of the screening stage by using NLP/IR techniques
(4) Study from a peer-reviewed source

(b) Exclusion Criteria
(1) Study does not focus on the screening stage of systematic reviews/review updates
(2) Study does not include sufficient information about dataset used and/or evaluation applied
(3) Survey study
(4) Study not from acceptable peer-reviewed sources such as books, abstracts and panels
(5) Study not written in English

Figure 2.1: Inclusion and exclusion criteria.

2.3 Identifying Relevant Studies

2.3.1 Boolean Search

An electronic search was conducted in October 2020 using the Boolean query presented in Figure 2.2 to retrieve studies relevant to the review. The query was derived from O'Mara-Eves et al. (2015) and expanded by adding more terms and search filters to express what constitutes relevant information to our review (see Section Search Terms). Below, the search is described in detail.

```
(‘‘text mining’’ OR ‘‘literature mining’’ OR ‘‘machine learning’’ OR
‘‘machine-learning’’ OR ‘‘automation’’ OR ‘‘semi-automation’’ OR ‘‘semi-automated’’
OR ‘‘automated’’ OR ‘‘automating’’ OR ‘‘text classification’’ OR ‘‘text classifier’’
OR ‘‘text categorization’’ OR ‘‘text categorizer’’ OR ‘‘classify* text’’ OR
‘‘category* text’’ OR ‘‘support vector machine’’ OR SVM OR ‘‘Natural Language
Processing’’ OR ‘‘active learning’’ OR ‘‘text clusters’’ OR ‘‘text clustering’’
OR ‘‘clustering tool’’ OR ‘‘text analysis’’ OR ‘‘textual analysis’’ OR ‘‘data
mining’’ OR ‘‘term recognition’’ OR ‘‘word frequency analysis’’ OR ‘‘automated’’ OR
‘‘Clinical Research Evidence’’ OR ‘‘Text Categorization’’ OR ‘‘Biomedical Document
Classification’’) AND (‘‘systematic review*’’ OR ‘‘article retrieval’’ ‘‘document
retrieval’’ OR ‘‘citation retrieval’’ OR ‘‘retrieval task’’ OR ‘‘identify*
articles’’ OR ‘‘identify* citations’’ OR ‘‘identify* documents’’ OR ‘‘citation
screening’’ OR ‘‘document screening’’ OR ‘‘article screening’’ OR ‘‘citation
management’’ OR ‘‘review management’’ OR ‘‘evidence synthesis’’ OR ‘‘research
synthesis’’ OR ‘‘evidence review’’ OR ‘‘research review’’ OR ‘‘comprehensive
review’’ OR ‘‘reference scanning’’ OR ‘‘Clinical Research Evidence’’ OR "update*"
OR ‘‘systematic review* update*’’) AND ( "BMC Public Health"[Journal] OR
‘‘Journal of the American Medical Informatics Association :  JAMIA’’[Journal]
OR ‘‘Journal of Biomedical Informatics"[Journal] OR ‘‘J Am Med Inform
Assoc’’[Journal] OR "BMC Bioinformatics"[Journal] OR "BMC Med Inform Decis
Mak"[Journal] OR ‘‘AMIA Annu Symp Proc’’[Journal] OR ‘‘Int J Comput Biol Drug
Des’’[Journal] OR ‘‘Healthc Inform Res’’[Journal] OR ‘‘J Biomed Inform’’[Journal]
OR ‘‘Genet Med’’[Journal] OR ‘‘Syst Rev’’[ Journal]) AND ("2005"[Publication
Date] :  "2019"[Publication Date])
```

Figure 2.2: Search query to retrieve studies from PMC.

Search limits

The following search limits were applied to the search:

Time limit

The search time frame was set between 2005 and 2020. This time frame was chosen based on Jonnalagadda and Petitti (2013) who stated that the first use of text mining techniques to support the screening process in systematic review occurred in 2005.

Database limit

To make the search results manageable, the search was limited to PubMed Central[1] (PMC). PMC is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM). PMC searches in many journals and indexes 6.1 Million articles(PMC Overview, 2020).

Journal limit

The search was limited to the following specific journals that are relevant to systematic reviews and NLP/IR techniques:

- Bio Medical Central (BMC):
  - BMC Systematic Reviews
  - BMC Public Health
  - BMC Bioinformatics
  - BMC Medical Informatics and Decision Making
- Journal of the American Medical Informatics Association (JAMIA)
- AMIA Annual Symposium Proceedings
- Journal of Biomedical Informatics
- International Journal of Computational Biology and Drug Design (IJCBDD)
- Journal of Healthcare Informatics Research - Springer
- Genetics in Medicine

Search Terms

The keywords in the Boolean query (see Figure 2.2) are derived from O'Mara-Eves et al. (2015). More terms were added to the query such as "Biomedical Document Classification"

[1] https://www.ncbi.nlm.nih.gov/pmc/

and “Text Categorization”. Furthermore, the term “update” was added to include studies which tackle the use of NLP/IR techniques for the update process. In addition, the search was limited to specific journals listed above by using the `[journal]` filter. Finally, the `[Publication Date]` filter was used to limit the search to the period from 2005 to 2020.

Running this Boolean query on PMC retrieved 653 studies from seven journals (without duplicates). In addition, 80 studies were identified from different resources (papers known to author from manual searches or recommend by colleagues (12), papers from O'Mara-Eves et al. (2015) (44) and CLEF working notes (24)). The full list of all search results can be found in Appendix A.

2.3.2 Studies Screening

Studies retrieved from the electronic search were already de-duplicated by PMC. The information about the studies (title, authors, year, journal, abstract and publication date) were stored in Mendeley Desktop[2] in addition to studies identified from external resources. Then, any duplication found in the combined studies set was removed by Mendeley. At the beginning of the screening process, the title and abstract of the retrieved studies were manually screened to find those that were most likely to be relevant. After the first screening, 623 studies that were clearly not relevant were excluded (88% of the studies). The remaining 85 studies were further screened and assessed against the inclusion/exclusion criteria (see Figure 2.1). This resulted in 63 studies being judged as relevant and included in the final systematic review, representing 9% of the total studies screened. Figure 2.3 shows the flow diagram for the screening process. All the studies were screened by a single Researcher (the author).

[2]Mendeley Ltd, version 1.19.9, 2019

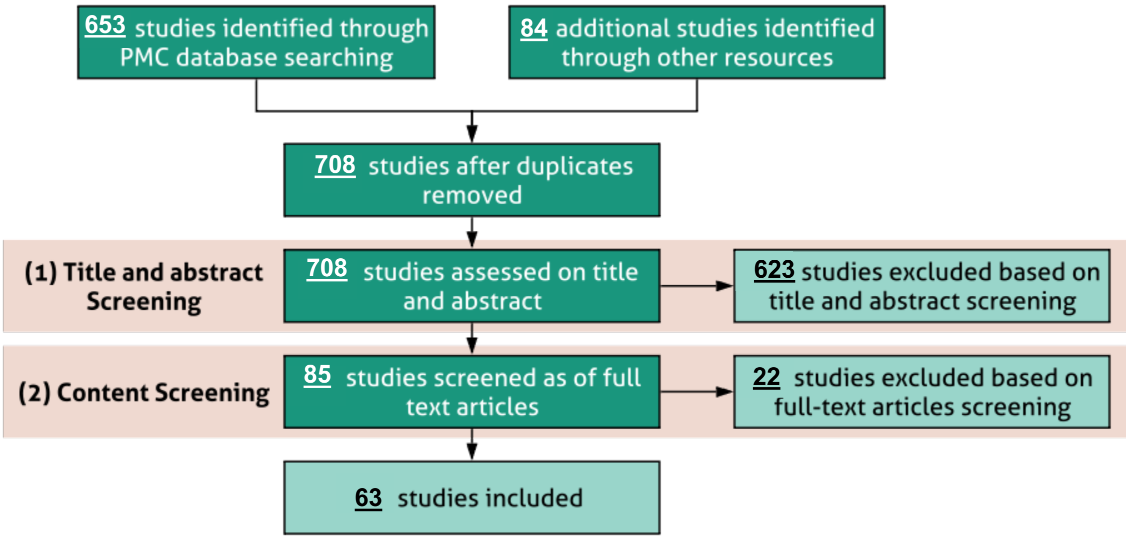

Figure 2.3: Flow diagram for studies selection process.

2.4 Results

This section presents the answers to the questions formulated for the systematic review (see Section 2.2.1). Figure 2.4 shows the number of relevant studies by publication year. The 63 studies used in the systematic review were published between 2005 and 2020. It can be seen that the majority of the studies were published in 2017, followed by 2018 when a new CLEF task was defined on 2017 (see Section 2.4.1). Earlier, in 2010 and 2012, there was an increase of attention shown to this domain with five articles published each year. Figure 2.5 presents the distribution of included studies by journal. However, not considering the studies from CLEF working notes, most of the included studies were from the JAMIA journal (10 studies) and Elsevier journal (8 studies) .

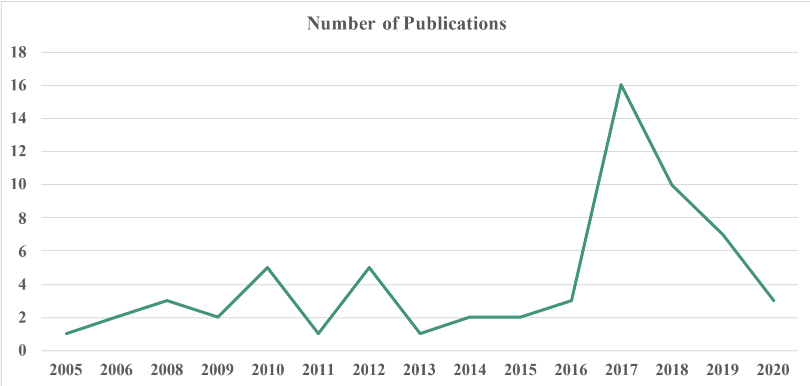

Figure 2.4: Distribution of included studies by year. The peak in 2017 is due to the CLEF task.

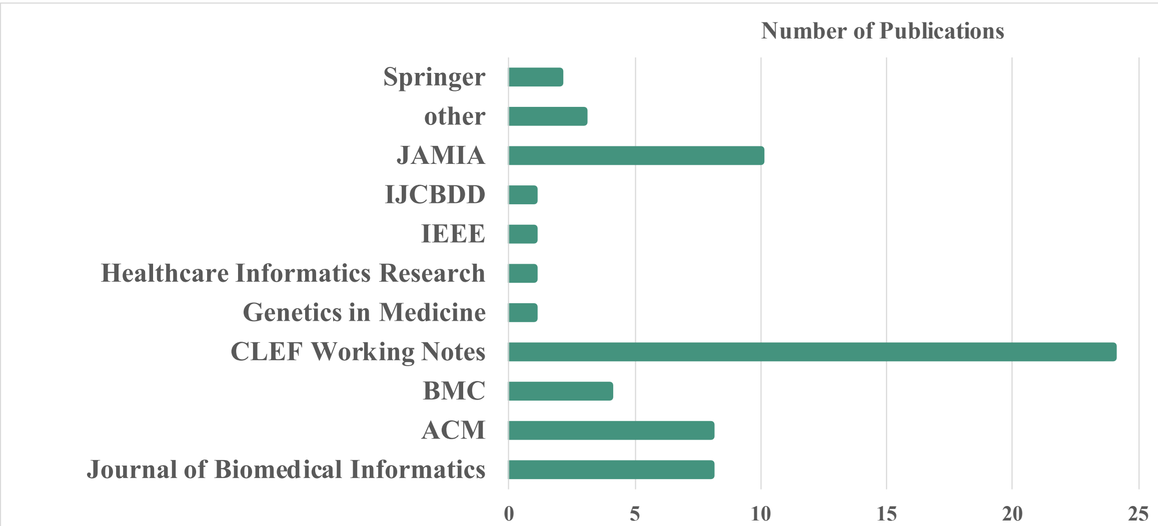

Figure 2.5: Distribution of included studies by journal.

2.4.1 Q1. Which NLP/IR techniques have been proposed to support the screening process?

The literature describes a range of approaches to support the screening process. Figure 2.6 presents the distribution of the included studies based on the NLP/IR approach applied. As can be seen from the figure, the approaches can be divided into three main categories: classification, ranking and active learning. Text classification aims to classify a set of documents using a predefined set of classes. It has many applications in the real world, such as spam filtering, news categorisation and search engines (Agarwal and Mittal, 2014). Ranking or work prioritisation aims to prioritise the studies retrieved from a search so the ones most likely to be relevant are listed at the top of the rank list (Karimi et al., 2010). Active learning can be used in both classification and ranking. Active learning is an iterative process whereby performance is improved by incrementally obtaining labelled data (Settles, 2012). It starts with an initial sample of labelled data to learn from. Then, it carefully selects a number of instances and asks the expert to assign labels for them. After that, it learns from the results and adapts its new knowledge and chooses other instances for the experts to label. This process continues until it reaches a stopping criterion (O'Mara-Eves et al., 2015; Settles, 2010).

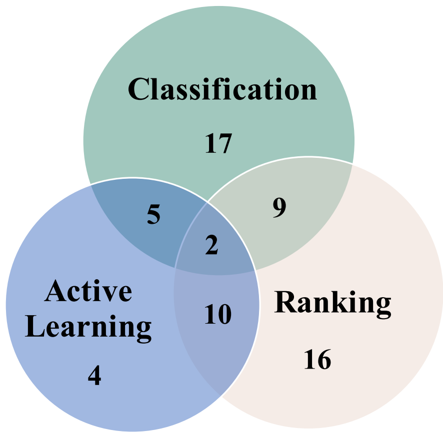

Figure 2.6: Distribution of included studies by approach.

The majority of the included studies (27%) applied classification without other approaches. However, some studies applied more than one approach: 15% applied both classification and ranking, 16% used both ranking and active learning, and 8% applied classification and active learning. Few studies considered all three categories of NLP/IR approaches: classification, ranking and active learning (3%).

Text Classification

Many of the included studies have attempted to use text classification to automate studies screening. One of the first attempts to use NLP/IR for internal medicine was made by

Aphinyanaphongs et al. (2005). They investigated its use to improve information retrieval in evidence-based medicine. They proposed an algorithm that automatically classifies internal medicine articles as either high or low quality. They applied several feature spaces and tested different text classification techniques (Support Vector Machine (SVM), Naïve Bayes and Boosting). They found that the best result was obtained by using SVM with the title, Medical Subject Heading (MeSH) terms and publication type features.

In subsequent years, many NLP/IR approaches have been proposed, Cohen et al. (2006) applied machine learning algorithms to drug-related topics. They aimed to reduce the experts' effort needed at the screening stage by removing as many non-relevant studies as possible. They used a voting perceptron based classifier and applied a bag-of-words method to the title, abstract, MeSH terms and publication type. Their proposed system demonstrates good results for reducing the workload during screening to 11 systematic reviews out of 15 reviews. The number of studies required to be screened manually was decreased by 50% or more for three of the 15 systematic reviews. Later the same year, Cohen (2006) applied re-sampling with SVM classifier by using SVM on TREC 2005 Genomics track data.

Yu et al. (2008) used SVM to classify human genetic association literature in PubMed, achieving 97.5% recall and 31.9% precision.

Kilicoglu et al. (2009) proposed a model to recognise medical publications to assist in evidence-based medicine. They applied different classifiers (SVM, Naïve Bayes, and Boosting). Experiments demonstrated that using a high-quality gold standard with advanced classification can improve selecting medical publications. The model achieved 61.5% recall and 73.7% precision.

Matwin et al. (2010) reported that using a Factorised version of the Complement Naïve Bayes classifier and the Weight Engineering techniques is better than using a voting perceptron-based system proposed by Cohen et al. (2006), achieving an enhancement of 15% over the average of WSS (see Section 2.4.3). They experimented on the drug dataset prepared by Cohen (2006) and applied different features, including the title, abstract, MeSH terms and publication type. They used word frequencies to represent the abstract and used binary representations to represent the MeSH and publication type.

Cohen et al. (2010) proposed a system to classify a list of studies from 18 systematic reviews topics based on SVM^{light}. It classifies samples based on the signed-margin distance. The studies with a large positive margin distance are classified as strongly relevant. On the other hand, studies with a very negative margin are classified as strongly non-relevant. The proposed system achieved a high AUC (see Section 2.4.3), with a mean of 0.89 across all topics.

Frunza et al. (2010) proposed a model based on a pre-question text classification. They used Complement Naïve Bayes classifier to exploit the question in the systematic review protocol.

Kim and Choi (2012) and Kim and Choi (2014) used SVM classifier to enhance the process of choosing relevant studies in evidence-based medicine. They applied this method to systematic review datasets on procedures and drugs. The SVM classifier was trained on the combination of studies included and commonly excluded. In Kim and Choi (2012), performance was enhanced by 15% for procedure topics and 11% for drug topics. In Kim and Choi (2014) the mean AUC was 0.95 for procedure topics and 0.84 for drug topics. They concluded that using a combination of included and commonly excluded studies is more effective than a combination of included and excluded studies.

Bian et al. (2017) applied a high impact Naïve Bayes classifier to classifying high impact studies for clinical decision support. They tested several features, including bibliometrics, MEDLINE metadata, such as MeSH terms and publication type, and social media exposure. The main limitation of this work is that they used a single dataset to train the classifier.

The SVM classifier has been consistently shown to work well on classification of biomedical texts (Aphinyanaphongs et al., 2005; Cohen, 2008; Cohen et al., 2010; Kilicoglu et al., 2009; Kim and Choi, 2012, 2014; Martinez et al., 2008; Miwa et al., 2014; Wallace et al., 2010, 2012a). In addition, other algorithms used in automating screening stage include Naïve Bayes, neural network, K-nearest neighbour and decision tree (Aphinyanaphongs et al., 2005; Bian et al., 2017; Hashimoto et al., 2016; Kilicoglu et al., 2009; Matwin et al., 2010).

Ranking

A number of studies used ranking to improve the identification of relevant evidence. Cohen et al. (2009) proposed a method to improve the performance of ranking for 24 systematic drug class reviews. They explored the use of combination training data: training from topic-specific data and training on data from other topics. They found that using hybrid data is better than training the SVM on topic-specific annotated data only with the AUC improving by 20%. However, Karimi et al. (2010) found that using ranking by itself is not helpful in terms of high recall. Therefore, they proposed a hybrid system consisting of ranking and Boolean querying.

Lee and Sun (2018) proposed a method to improve ranking by using a seed-driven document. They assumed that at least one relevant document is known before the screening start, i.e., the seed document. This document is used to form a query and ranking the candidate documents. Experiments showed the effectiveness of the proposed method to reduce workload by experts achieving an enhancement of 15% over WSS.

Scells et al. (2020) proposed an extension to coordination level matching, by exploiting the query-document relationship with rank fusion. They applied their method on CLEF2017 and CLEF2018 collections. The model performed statically significantly better than the state-of-the-art PubMed ranker in term of MAP for CLEF2018 dataset.

Zuccon et al. (2020) proposed a query variation sampling methods for training learning to rank models to rank queries. The results show that query sampling methods do directly impact the ability of a learning to rank models to effectively identify good query variations. Thus, selecting appropriate query sampling methods is a key problem for the automatic reformulation of effective Boolean queries for systematic review literature search.

Much research has been devoted to using ranking with classification for the purpose of workload reduction when conducting a systematic review. Below, this research is described.

Martinez et al. (2008) proposed a two-stage ranking search system based on ranked queries and re-ranking using text classification to restrict the results to high-quality studies. They applied their system to the 15 systematic reviews from the drug class previously used by Cohen et al. (2006) and examined two feature sets: one consisting of abstract and references and another - of abstract, references and MeSH terms. Their proposed system is beneficial for most systematic reviews, with an average WSS of 34.3 when using MeSH terms.

Cohen (2008) constructed a classification system based on work prioritisation and evaluated three feature combinations to classify studies based on SVM^{light}. He applied these to the 15 systematic reviews from the drug class mentioned earlier. Using a binary representation for all features, the study found that the best scoring result of the three feature combinations was that of a combination of unigram and n-gram, with a length of two, extracted from the title, abstract and MeSH terms. The worst scoring combination was that of unigram, MeSH terms and UMLS CUI (Unified Medical Language System Concept Unique Identifier). Cohen demonstrated that work prioritisation with the use of the MeSH feature can enhance the efficiency of conducting a systematic review.

Bui et al. (2015) developed an unsupervised system to retrieve studies based on query expansion and ranking. Query expansion aims to extend the original query by adding related terms to create a query that is more likely to retrieve relevant studies (Baeza-Yates and Ribeiro-Neto, 2011). Using query expansion improved recall to 80.2% while the precision decreased by 0.2% compared with a default PubMed search. For the ranking, they proposed the clinical research scoring approach using three dimensions: MeSH majority, study design, and journal ranking. They compared their ranking system with two systems: using machine learning (specifically, classification) and PubMed default sort (by relevance). The best average precision and recall was achieved by using the clinical research scoring approach.

Cohen et al. (2015) developed a method to rank and predict the relevance of studies which they applied in a randomized controlled trial to a large set of MEDLINE articles. They applied the SVM classifier with features that included the abstract and MeSH terms for the article. They reported a high AUC (0.973) but the method missed a number of relevant studies (5%).

Scells and Zuccon (2018) and Scells et al. (2019) introduced an approach to improving Boolean queries used for study identification in systematic reviews. The query used in the review was iteratively altered by applying a set of transformations such as replacing logical operators and field restrictions. They found that the modified queries generated by this approach improved upon those used in the original review. The best modified queries were identified using classifiers and learning to rank methods. Their approaches produced queries with higher precision and F1 scores than the original query but not improved recall. Their method was used to demonstrate that it was possible to improve the Boolean query used for the original review.

Active Learning

Many studies have used active learning to reduce the workload involved in systematic reviews. Wallace et al. (2010) developed a semi-automated approach to reducing the number of studies need to be manually screened by reviewers. Their approach was based on an SVM with different feature spaces. The classifier was trained using an active learning strategy. It chooses the study to be screened next, rather than sampling at random. The researchers applied their method to three systematic review datasets. As a result, the number of studies that needed to be screened manually was reduced by 40% to 50%. Two years later, Wallace et al. (2012a) designed an online tool "Abstrackr" for facilitating the screening stage of systematic reviews. The system was based on active learning, whereby the classification model interacts with reviewers. The tool is open source and free to use; it has been used to assist in 50 systematic reviews (as of the date of publication, 2012).

Tomassetti et al. (2011) developed a semi-automated method for screening studies using linked data and text mining. Their model was able to reduce the workload by 20% compared with the manual screening.

Jonnalagadda and Petitti (2013) proposed a system that uses simple relevance feedback from reviewers to modify the query. The system presents the initial document to the reviewer and asks them to classify it as relevant or not relevant. Then the query is modified and the next document is presented to the reviewer based on the modified query. Their system was able to reduce the number of studies need to be screened manually by 6%-30% with recall of 95%.

Zhang et al. (2016) proposed a method to accurately and efficiently find the number of relevant studies in a collection for a certain topic (the volume estimation problem). Their proposed system is based on active learning and sampling. First, active learning is used to find the knee point in the effort versus recall gain curve where all the easy to find relevant studies are identified, then a sampling technique is applied to locate the number of relevant studies in the rest of the collection. They explored three sampling strategies: Negative binomial sampling, Horowitz-Thompson estimator, and stratified sampling (Tillé, 2006). They found the Negative binomial sampling to be more accurate than the other sampling methods. They demonstrated that using active learning with sampling strategies can help in the volume estimation problem.

Kontonatsios et al. (2017) developed a semi-automated system that uses active learning to enhance classification of studies. They used two vector space representations: (a) bag-of-words and (b) a data-dependent, spectral embedding. They applied their system to the COPD and Proton Beam datasets (see Section 2.4.2) in addition to four public health[3] systematic reviews. They found that results in clinical and biomedical domains show consistent improvements.

Miwa et al. (2014) applied active learning for both clinical medicine and public health data. They addressed the problem of imbalanced data where the number of examples in one class is very small relative to the number in another class (e.g. the number of relevant documents vs the number of non-relevant documents). They applied a weighting method whereby they assigned a greater weight to relevant studies than to non-relevant studies. They demonstrated that using the weighting certainty method is the most promising approach for active learning with imbalanced data.

[3]http://eppi.ioe.ac.uk/cms/.

Based on Miwa et al. (2014), Hashimoto et al. (2016) proposed a model to enhance active learning text classification used in the screening stage. They used neural network based vector space model (paragraph vectors) to find similarity between studies. They applied k-means clustering algorithm. Their proposed system outperformed the work done by Miwa et al. (2014) with a 1%-15% improvement for WSS@95.

Conference and Labs of the Evaluation Forum (CLEF) 2017-2019

During the last three years (2017-2019), the CLEF eHealth forum ran a task on systematic reviews that aimed to support the screening phase by (semi)automatically ranking the studies by relevance to the review (Kanoulas et al., 2017, 2018, 2019). The task focused on the effectiveness of ranking during the first phase of screening: *title and abstract screening*. Participants were provided with two datasets: a training dataset and a test dataset (see Section 2.4.2). For each dataset, a list of studies retrieved from a Boolean query was provided. The participants were asked to rank the studies in an efficient way.

The task attracted a wide range of participants from the text mining community. Figure 2.5 shows that a large number of included studies are from CLEF working notes. In 2017, 15 groups participated in the task. The following year, 2018, seven groups participated. In 2019, only three groups participated in the task. In general, participants applied both supervised and unsupervised approaches. A variety of ranking algorithms were used, including BM25 (Alharbi and Stevenson, 2019b; Hollmann and Eickhoff, 2017a; Kalphov and Azzopardi, 2017; Nunzio, 2019; Wu et al., 2018), relevance feedback (Alharbi et al., 2018; Anagnostou and Vlahavas, 2017; Hollmann and Eickhoff, 2017a; Minas et al., 2018; Norman et al., 2017; Nunzio et al., 2017, 2018; Wu et al., 2018; Yu and Menzies, 2017), continuous active learning (Li and Kanoulas, 2019; Nunzio, 2019; Yu and Menzies, 2017) and learning to rank (Anagnostou and Vlahavas, 2017; Chen et al., 2017; Minas et al., 2018). In addition, some participants used a classification algorithm to classify the search results as relevant or non-relevant and then rank them. The algorithms used included random forest (Altena and Olabarriaga, 2017; Chen et al., 2017; Scells et al., 2017a), SVM (Anagnostou and Vlahavas, 2017; Minas et al., 2018; Yu and Menzies, 2017), logistic regression (Norman et al., 2017, 2018) and neural network (Lee, 2017; Norman et al., 2018). Furthermore, a stopping criteria was applied by some participants to provide a threshold to stop the ranking (i.e. to suggest a stopping point beyond which the reviewer does not need to screen the studies) (Cormack and Grossman, 2018, 2017b; Li and Kanoulas, 2019; Nunzio, 2019).

Results from these exercises demonstrated that automating the screening stage of systematic review can be efficient in identifying most, if not all, relevant studies with less effort and time than manual screening.

2.4.2 Q2. What are the datasets used? Are they publicly available?

Well-defined datasets are needed to evaluate NLP/IR techniques. Researchers usually prefer to use a public dataset so they can compare the performance of their models with the results achieved by other researchers. This section explores the datasets used in the literature to evaluate NLP/IR techniques for study screening in systematic reviews. In addition, it presents the characteristics of those datasets. Figure 2.7 shows the distribution of the included studies by the dataset used.

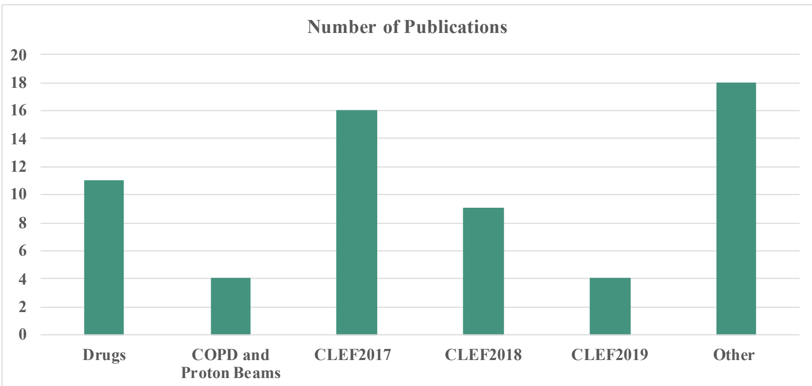

Figure 2.7: Distribution of included studies by dataset used.

Drugs Dataset

An early dataset was interpreted by the U.S. Agency for Healthcare Research and Quality, (AHRQ)[4] consisting of 15 systematic reviews on drug-related topics with associated data,

[4]https://www.ahrq.gov/

including Boolean queries and search results. This dataset was first assembled by Cohen et al. (2006) and has been extensively used in the literature (19% of included studies used this dataset). Table 2.1 provides an overview of the drugs dataset. The dataset is publicly available from the systematic drug class review Gold Standard website[5]. The file includes the drug name, a list of PubMed identification numbers (PMIDs) and the relevance judgement.

Table 2.1: Characteristics of the Drugs dataset.

Drug Systematic Review	No. of Studies	No. of Relevant Abstract
ACEInhibitors	2,544	183 (7.19%)
ADHD	851	84 (9.87%)
Antihistamines	310	92 (29.68%)
AtypicalAntipsychotics	1,120	363 (32.41%)
BetaBlockers	2,072	302 (14.58%)
CalciumChannelBlockers	1,218	279 (22.91%)
Estrogens	368	80 (21.74%)
NSAIDs	393	88 (22.39%)
Opioids	1,915	48 (2.51%)
OralHypoglycemics	503	139 (27.63%)
ProtonPumpInhibitors	1,333	238 (17.85%)
SkeletalMuscleRelaxants	1,643	34 (2.07%)
Statins	3,465	173 (4.99%)
Triptans	671	218 (32.49%)
UrinaryIncontinence	327	78 (23.85%)
Total	**18,733**	**2,399 (12.81%)**

Proton and COPD Datasets

The Proton and Chronic Obstructive Pulmonary Disease (COPD) datasets have been used in 7% of the studies. These datasets were reported by Wallace et al. (2010). The Proton Beam dataset is derived from a systematic review of particle beam radiation therapies for cancer by Tufts Medical Centre Evidence-based Practice Centre (Terasawa et al., 2009). The COPD systematic review was created in 2010. Its main purpose was to examine the outcomes of interval versus continuous training on peak oxygen uptake, peak power, six-minute walk test distance and health-related quality of life in individuals with COPD

[5]https://dmice.ohsu.edu/cohenaa/systematic-drug-class-review-data.html

(Castaldi et al., 2009). These datasets are publicly available as XML files containing the title, abstract and MeSH terms for each study[6]. Table 2.2 provides a summary of the information about these datasets.

Table 2.2: Characteristics of COPD and Proton datasets.

Dataset	No. of Studies	No. of Relevant Abstract
COPD	1,524	196 (12.86%)
Proton	4,751	243 (5.11%)

CLEF datasets

CLEF datasets were provided by the CLEF organiser (Kanoulas et al., 2017, 2018, 2019). They were used in 44% of the included studies. Tables 2.3-2.7 present the characteristics of CLEF datasets. Each dataset is divided into a training dataset and a test dataset. Each of these contains a number of reviews. In 2017 and 2018, all reviews focused on Diagnostic Test Accuracy (DTA). The 2019 dataset includes four types of reviews: DTA, Intervention, Prognosis and Qualitative reviews. However, the training split of the CLEF2018 dataset is a subset of CLEF2017 dataset and the DTA training set of CLEF2019 represents CLEF2018 dataset.

Table 2.3: Characteristics of CLEF datasets.

		Number of reviews		
Year	**Review Type**	**Training**	**Test**	**Total**
CLEF2017	**DTA**	20	30	50
CLEF2018	**DTA**	42	30	72
CLEF2019	**DTA**	72	8	80
	Intervention	20	20	40
	Prognosis	0	1	1
	Qualitative	0	2	2
Total number of distinct reviews				**123**

[6]https://static-content.springer.com/esm/art%3A10.1186%2F1471-2105-11-55/MediaObjects/12859_2009_3512_MOESM1_ESM.ZIP

Table 2.4: CLEF2017 training dataset characteristics.

Review	No. of Studies	No. of Relevant Abstract
CD010438	3,250	39 (1.39%)
CD011984	8,192	454 (16.19%)
CD008643	15,083	11 (0.39%)
CD009944	1,181	117 (4.17%)
CD007427	1,521	123 (4.39%)
CD009593	14,922	78 (2.78%)
CD011549	12,705	2 (0.07%)
CD011134	1,953	215 (7.67%)
CD008686	3,966	7 (0.25%)
CD011975	8,201	619 (22.08%)
CD009323	3,881	122 (4.35%)
CD009020	1,584	162 (5.78%)
CD011548	12,708	113 (4.03%)
CD010409	43,363	76 (2.71%)
CD008054	3,217	274 (9.77%)
CD010771	322	48 (1.71%)
CD009591	7,991	144 (5.14%)
CD008691	1,316	73 (2.60%)
CD010632	1,504	32 (1.14%)
CD007394	2,545	95 (3.39%)
Total	**149,405**	**2,804 (1.88%)**

Table 2.5: CLEF2017 test dataset characteristics.

Review	No. of Studies	No. of Relevant Abstract
CD010775	241	11 (0.59%)
CD009786	2,065	10 (0.54%)
CD009579	6,455	138 (7.43%)
CD009925	6,531	460 (24.77%)
CD007431	2,074	24 (1.29%)
CD008803	5,220	99 (5.33%)
CD008782	10,507	45 (2.42%)
CD009647	2,785	56 (3.02%)
CD009135	791	77 (4.15%)
CD008760	64	12 (0.65%)
CD009519	5,971	104 (5.60%)
CD009372	2,248	25 (1.35%)
CD010276	5,495	54 (2.91%)
CD009551	1,911	46 (2.48%)
CD012019	10,317	3 (0.16%)
CD008081	970	26 (1.40%)
CD009185	1,615	92 (4.95%)
CD010339	12,807	114 (6.14%)
CD010653	8,002	45 (2.42%)
CD010542	348	20 (1.08%)
CD010896	169	6 (0.32%)
CD010023	981	52 (2.80%)
CD010772	316	47 (2.53%)
CD011145	10,872	202 (10.88%)
CD010705	114	23 (1.24%)
CD010633	1,573	4 (0.22%)
CD010173	5,495	23 (1.24%)
CD010386	626	2 (0.11%)
CD010783	10,905	30 (1.62%)
CD010860	94	7 (0.38%)
Total	**117,562**	**1,857 (1.58%)**

Table 2.6: CLEF2018 training dataset characteristics.

Review	No. of Studies	No. of Relevant Abstracts
CD007394	2,542	92 (3.62%)
CD007427	1,457	59 (4.05%)
CD008054	3,149	206 (6.54%)
CD008081	970	26 (2.68%)
CD008643	15,078	6 (0.04%)
CD008686	3,964	5 (0.13%)
CD008691	1,310	67 (5.11%)
CD008760	64	12 (18.75%)
CD008782	10,507	45 (0.43%)
CD008803	5,220	99 (1.90%)
CD009020	1,576	154 (9.77%)
CD009135	791	77 (9.73%)
CD009185	1,615	92 (5.70%)
CD009323	3,857	98 (2.54%)
CD009372	2,248	25 (1.11%)
CD009519	5,971	104 (1.74%)
CD009551	1,911	46 (2.41%)
CD009579	6,455	138 (2.14%)
CD009591	7,990	143 (1.79%)
CD009593	14,907	63 (0.42%)
CD009647	2,785	56 (2.01%)
CD009786	2,065	10 (0.48%)
CD009925	6,531	460 (7.04%)
CD009944	1,162	98 (8.43%)
CD010023	981	52 (5.30%)
CD010173	5,495	23 (0.42%)
CD010276	5,495	54 (0.98%)
CD010339	12,807	114 (0.89%)
CD010386	626	2 (0.32%)
CD010409	43,335	48 (0.11%)
CD010438	3,241	30 (0.93%)
CD010542	348	20 (5.75%)
CD010632	1,499	27 (1.80%)
CD010633	1,573	4 (0.25%)
CD010653	8,002	45 (0.56%)
CD010705	114	23 (20.18%)
CD011134	1,938	200 (10.32%)
CD011548	12,704	109 (0.86%)
CD011549	12,704	1 (0.01%)
CD011975	8,186	604 (7.38%)
CD011984	8,180	442 (5.40%)
CD012019	10,317	3 (0.03%)
Total	**241,670**	**3,982 (1.65%)**

Table 2.7: CLEF2018 test dataset characteristics.

Review	No. of Studies	No. of Relevant Abstract
CD008122	1,911	272 (14.23%)
CD008587	9,158	79 (0.86%)
CD008759	932	60 (6.44%)
CD008892	1,499	69 (4.60%)
CD009175	5,644	65 (1.15%)
CD009263	79,786	124 (0.16%)
CD009694	161	16 (9.94%)
CD010213	15,198	599 (3.94%)
CD010296	4,602	53 (1.15%)
CD010502	2,985	229 (7.67%)
CD010657	1,859	139 (7.48%)
CD010680	8,405	26 (0.31%)
CD010864	2,505	44 (1.76%)
CD011053	2,235	12 (0.54%)
CD011126	6,000	13 (0.22%)
CD011420	251	42 (16.73%)
CD011431	1,182	297 (25.13%)
CD011515	7,244	127 (1.75%)
CD011602	6,157	8 (0.13%)
CD011686	9,443	55 (0.58%)
CD011912	1,406	36 (2.56%)
CD011926	4,050	40 (0.99%)
CD012009	536	37 (6.90%)
CD012010	6,830	290 (4.25%)
CD012083	322	11 (3.42%)
CD012165	10,222	308 (3.01%)
CD012179	9,832	304 (3.09%)
CD012216	217	11 (5.07%)
CD012281	9,876	23 (0.23%)
CD012599	8,048	575 (7.14%)
Total	**218,496**	**3,964 (1.81%)**

For each review in the dataset, the following information is available (see Figure 2.8):

- Review (topic) ID
- Title of the review (written by Cochrane experts)
- A Boolean query using either OVID or PubMed syntax (manually constructed by Cochrane experts)

- Set of PMIDs returned by running the query in the MEDLINE database
- Relevance judgement at both abstract and content levels

Review ID: CD010705

Title: The diagnostic accuracy of the GenoType® MTBDRsl assay for the detection of resistance to second-line anti-tuberculosis drugs.

Boolean Query:

1. MTBDR*.ti,ab.
2. Genotype MTBDR*.ti,ab
3. OR/1-2
4. exp Tuberculosis, Pulmonary/
5. exp Tuberculosis, Multidrug-Resistant/
6. MDR-TB.ti,ab
7. XDR-TB.ti,ab
8. Mycobacterium tuberculosis/
9. TB.ti,ab
10. tuberculosis.ti,ab
11. OR/4-10
12. 3 AND 11

PMIDs:

24429319, 24197880, 24172155, 24098523, 24056651, 24046537, 24039735, 24029194, 23895665, 23883707, 23808160, 23782980, 23689727, 23658272, 23633684, 23467605, 23392466, 23383320, ...

Figure 2.8: Example Cochrane reviews used in CLEF2018 training dataset (Theron et al., 2016).

The total number of distinct reviews from CLEF is 123. All reviews are from the Cochrane library and they are publicly available from the CLEF organiser[7]. More details about the CLEF dataset are presented in Appendix B.

In addition to these datasets, there are others, including: a publicly available dataset of 94 Cochrane reviews[8] published by Scells et al. (2017b) (used by one study), TREC 2005 Genomics track dataset[9] (used by one study), TREC 2015 Total Recall Track dataset[10] (used by one study), reviews from AHRQ (used by two studies), update dataset (used by two studies) and different reviews form MEDLINE (used by 11 studies).

In general, as can be seen, all datasets which are publicly available were created for the purpose of conducting new systematic reviews.

2.4.3 Q3. How are those techniques evaluated?

Figure 2.9 shows the measures applied to evaluate the performance of the techniques applied in the included studies. The most commonly used measures are WSS (55.5%), AP (44.4%), recall (27%), precision (20.6%), AUC (14%) and F-score (9.5%). Below, each measure is described in detail.

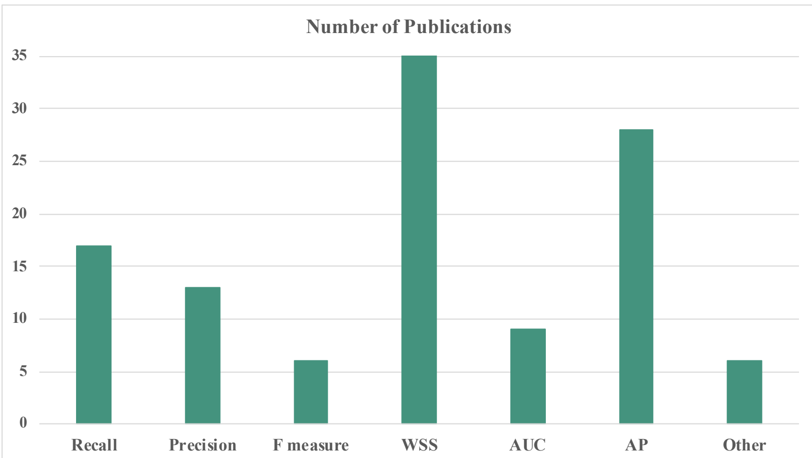

Figure 2.9: Distribution of included studies by evaluation measure.

[7] https://github.com/CLEF-TAR/tar
[8] available on https://github.com/ielab/SIGIR2017-PICO-Collection
[9] available on https://trec.nist.gov/data/t14_genomics.html
[10] https://trec.nist.gov/pubs/trec24/papers/Overview-TR.pdf

Table 2.8 presents the confusion matrix that will be used to define each measure. True Positives (TP) are those labels that were correctly predicted to be relevant. False Positives (FP) are those labels that were predicted to be relevant but were, in fact, non-relevant. True Negatives (TN) are those labels that were correctly predicted to be non-relevant. False Negatives (FN) are those labels that were predicted to be non-relevant but were, in fact, relevant.

Table 2.8: Confusion Matrix.

		annotated labels	
		Relevant	**Non-Relevant**
predicted labels	**Relevant**	TP	FP
	Non-Relevant	FN	TN

Recall and Precision

Recall and Precision are standard metrics widely used in IR (Baeza-Yates and Ribeiro-Neto, 2011). Recall (see Equation 2.1) is calculated as the number of correctly identified relevant studies divided by the total number of relevant studies in the collection.

$$recall = \frac{TP}{TP + FN} \tag{2.1}$$

Precision (see Equation 2.2) is calculated as the number of correctly identified relevant studies divided by the total number of retrieved studies.

$$precision = \frac{TP}{TP + FP} \tag{2.2}$$

In professional search tasks, such as patent retrieval and legal search, the reviewers' goal is to identify almost all of the publications reasonably related to the search topic, i.e., there is typically an emphasis on recall. It is required to achieve high recall at acceptable precision (high-recall task) (Kim et al., 2011; Shalaby and Zadrozny, 2019; Song et al., 2019). The nature of the search problem in systematic reviews is like the professional search, it requires high recall since the goal is to identify as many eligible studies as possible

(Carol et al., 2020). However, retrieving a large number of non-relevant studies increases the screening effort required by the reviewers and it is, hence, beneficial to ensure that the precision is as high as possible. Therefore, an evaluation of the trade-off between potentially missing studies and reducing burden is required. They allow reviewers to change the relative importance of these two metrics depending on priorities in a given review. These metrics include notably the F-score and Work Saved over Sampling, which are summarised below.

F-score

F-score is a single measure that trades off precision versus recall. It is calculated as follows:

$$F_\beta = \frac{(\beta^2+1) \times precision \times recall}{\beta^2 \times precision + recall} \quad \text{where} \quad \beta^2 \in [0, \infty] \tag{2.3}$$

It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. When $\beta > 1$, F becomes more recall-oriented and if $\beta < 1$, it becomes more precision oriented. When $\beta = 1$, F1 (see Equation 2.4) comes to be equivalent to the harmonic mean of both recall and precision (Manning et al., 2008a).

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{2.4}$$

Work Saved over Sampling (WSS)

WSS was introduced by Cohen et al. (2006) as a measure in the field of systematic reviews. Cohen et al. defined WSS as “the percentage of papers that meet the original search criteria that the reviewers do not have to read (because they have been screened out by the classifier)” (Cohen et al., 2006). WSS is calculated using Equation 2.5.

$$WSS = \frac{TN + FN}{N} - (1.0 - recall) \tag{2.5}$$

where N is the total number of studies. For example, the workload over sampling at 95% recall is defined as:

$$WSS@95\% = \frac{TN + FN}{N} - (0.05) \tag{2.6}$$

Area Under the ROC Curve (AUC)

AUC is used to evaluate the performance of machine learning algorithms. It measures the area underneath the ROC curve (ROC = Receiver Operating Characteristic curve). ROC plots two parameters (see Figure 2.10): The True Positive Rate and the False Positive Rate. The True Positive Rate (TPR) is a synonym for recall and is therefore defined as follows (Bradley, 1997):

$$TPR = recall = \frac{TP}{TP + FN} \tag{2.7}$$

False Positive Rate (FPR) is defined as follows:

$$FPR = \frac{FP}{FP + TN} \tag{2.8}$$

ROC is used to compare different classification threshold values for classification models. AUC is the area underneath this curve, the higher its numerical value, the better it is. It is used to compare the performance of different classification models or to find the probability that a given classification system works better than the baseline (random ordering).

However, in the case of imbalanced data where the number of negative (non-relevant) examples is much greater than the number of the positive (relevant) examples, it is better to use a recall-precision curve because precision does not include the TN in its calculation and is not affected by the class imbalance.

Mean Average Precision (MAP)

MAP is widely used in practice for evaluating the performance of a ranking system (Shobha and Rangaswamy, 2018). MAP for a set of reviews represents the mean of the average

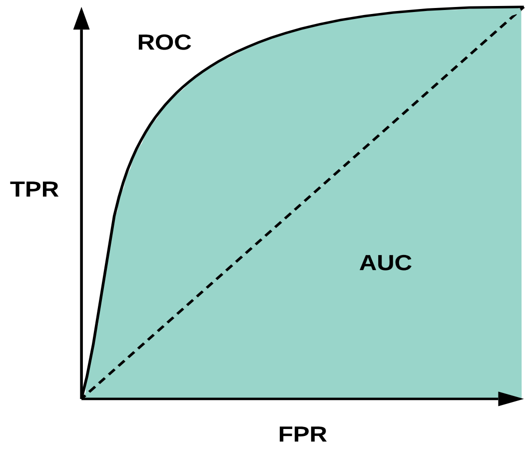

Figure 2.10: Example of Area Under the ROC Curve.

precision (AP) scores over all reviews. The AP of a single review is the average of the precision scores for each relevant study retrieved in the search result list (Ibrahim and Landa-Silva, 2017; Thom and Scholer, 2007). These are rank positions at which relevant studies are retrieved. MAP is computed as:

$$MAP = \frac{\sum_{i=1}^{T} AP}{T} \tag{2.9}$$

where T is the total number of reviews.

To exemplify, Figure 2.11 shows the ranking effectiveness for two reviews. The AP for the first review (a) is the average of precision scores at positions 1, 2, 4, 7 and 9 where relevant documents were retrieved: $(1 + 1 + 0.75 + 0.57 + 0.56)/5 = 0.78$. On the other hand, the AP for the second review (b) is the average of precision at positions 2, 4, 5 and 9: $(0.5 + 0.5 + 0.6 + 0.44)/4 = 0.51$. Therefore, the MAP for this example is $(0.78 + 0.51)/2 = 0.65$.

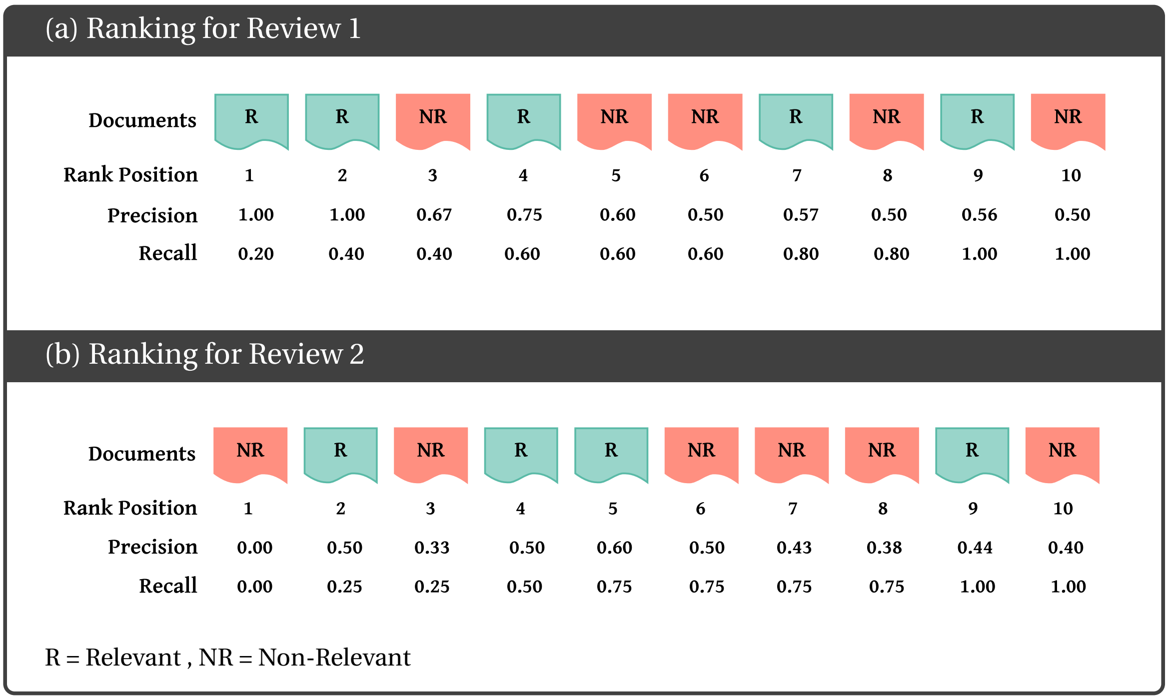

Figure 2.11: Ranking effectiveness example.

In addition to the measures discussed above, other measures were used by studies such as utility (5%), yield (5%) and burden (3.5%) (Miwa et al., 2014). These metrics are not explained in detail as they will not be used in this thesis.

2.4.4 Q4. Which techniques are applied in the screening stage of the review update process?

Most of the studies applied techniques to support the identification of studies for the creation of new reviews (85%). Figure 2.12 shows the distribution of studies by the type of the review (i.e. new review, update or both). Few researchers have addressed the problem of identifying the relevant studies for updating reviews - 10 (16%) of the total included studies. The update process was the main focus of 70% of these studies (Alharbi and Stevenson (2019c, 2020); Cohen et al. (2012); Dalal et al. (2012); Lerner et al. (2019); Surian et al. (2018); Wallace et al. (2012a)) while it constituted only a part of the study for the remaining 30% (Cohen (2008); Cohen et al. (2009); Khabsa et al. (2016)).

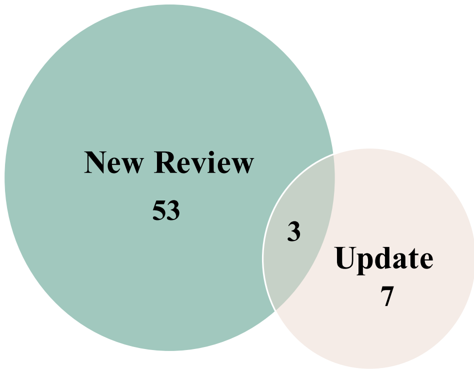

Figure 2.12: Distribution of included studies by type.

Most of these studies (71%) applied classification (SVM and random forest), one study applied ranking, and one study applied both.

Wallace et al. (2012b) used SVM to reduce the workload involved in the screening process for the systematic review update. They used the relevance judgement of the original review to train a classifier. Wallace et al. used four systematic reviews to validate their proposed approach and found that the classifier reduced the workload by 70%-90%.

Cohen et al. (2012) designed a classification algorithms that gives an alert when new evidence is available to schedule a review update. They used nine drug therapy systematic reviews from MEDLINE to evaluate their approach. However, their system did not achieve a reliably high recall or precision.

Dalal et al. (2012) used labelled datasets for two systematic reviews from MEDLINE to train a classifier. Then, they used the classifier to predict which studies should be included in a simulated update. The system was able to reduce workload by 50%. However, recall decreased.

Surian et al. (2018) used a matrix factorisation approach to identify relevant studies for a systematic review update. The main limitation of this work was the authors' focus on drug interventions in type 2 diabetes. This means the results cannot be generalised to other domains.

Alharbi and Stevenson (2019c) used an update dataset consists of 25 intervention reviews (see Chapter 4). They aimed to improve study selection for systematic review updates by using information produced during the screening stage of the original review. They applied relevance feedback and the screening effort required to identify all relevant studies (100% recall) was reduced by 63.5%.

Lerner et al. (2019) developed an algorithm for automatically screening abstracts when updating living network meta-analysis. They applied word embeddings and logistic regression. The algorithm achieved 95-100% recall and decreased the workload by 53%.

Alharbi and Stevenson (2020) proposed a method to automatically refine Boolean queries for the study selection stage of systematic review updates (see Chapter 5). The proposed approach generates a set of transformed queries using three methods: operator substitution, query expansion and query reduction. The best query is then selected using an objective function that considers both recall and precision. The method improves the original query both in terms of recall and precision. It produces queries that are able to identify relevant studies that would not be retrieved using the query from the original review.

Most of these studies evaluated their approaches using simulations of the update process by using “time-slicing". For example, by assuming that the included studies appeared in the three years before the review publication date, like in the case of studies included in the updated version of the review by Khabsa et al. (2016). An exception is a work that used update information for nine drug therapy systematic reviews by Cohen et al. (2012), but this dataset is not publicly available.

2.5 Limitations

The main limitation of this review is that the search was restricted to only one database. The reason was to make the search results manageable. However, PMC was chosen to conduct the search which allowed to specify the search limits. Seven journals were chosen which are highly related to the domain of this review (i.e NLP/IR techniques and systematic review screening). PMC also produced a de-duplicated list of results.

A further limitation is that this review was performed by only two researchers while systematic reviews are usually conducted by a team.

In addition, this review was focusing on techniques which have been applied with systematic reviews to support screening process. However, there are many recent techniques in NLP that have not been considered in the review because they were not used in the field of systematic reviews, e.g. language models such as BERT and BioBERT (Lee et al., 2020). This offers plenty of scope for future work, e.g. it would be interesting to apply and evaluate these recent techniques in the domain of systematic reviews.

2.6 Summary

This chapter described a systematic review conducted to explore the use of NLP/IR techniques in facilitating the screening process for systematic reviews. It also paid attention to the studies that tackle the problem of updating systematic reviews. PMC was used to search for relevant studies across seven journals. After screening, from 554 studies,62 were found fulfilled the inclusion/exclusion criteria. The information was extracted and the results obtained were summarised.

Although it is difficult to have a single conclusion across all included studies due to the differences in methodologies implemented, datasets used and evaluation metrics applied, the overall picture shows that NLP/IR techniques are useful to improve the screening process. In particular, classification can reduce the workload by excluding non-relevant studies. On the other hand, ranking can help in identifying relevant studies earlier which will help researchers to gain more knowledge about the inclusion criteria and therefore accelerate the process of conducting systematic reviews. Few studies pay attention to an important problem which is the update process.

The results obtained from this systematic review can be summarised in the following points:

- Using NLP/IR techniques can improve the process of identifying relevant evidence and reduce the workload required from experts.

- The most commonly applied classifier is SVM.
- Classification can reduce workload by reducing the number of articles which need to be screened (increase precision) but in most studies this results in missing relevant articles (decrease recall).
- The majority of studies applied techniques for the creation of new reviews, a limited number of studies tackled the problem of identifying relevant evidence for the update process.
- The most commonly used datasets to evaluate techniques are drugs and CLEF.
- There is no publicly available dataset to evaluate techniques for the update process (except the one constructed by the Author - See Chapter 4).
- Current proposed systems are able to reduce workload by reducing the number of studies which need to be manually screened by 30%-70%. However, this reduction is usually achieved at the expense of reduction of recall by 5%.

Table 2.9 presents a summary of information extracted from the 63 included studies.

Table 2.9: Summary of the information extracted from the 63 included studies.

Author	Year	Type	Dataset Source	Approach	Evaluation Measures
Aphinyanaphongs et al.	2005	NR	MEDLINE	Classification	AUC
Cohen et al.	2006	NR	AHRQ	Classification	R, P, F, WSS
Cohen	2006	NR	TREC	Classification	NU
Yu et al.	2008	NR	MEDLINE	Classification	R, P, S
Cohen	2008	NR and UR	AHRQ	Classification	AUC
Martinez et al.	2008	NR	AHRQ	Ranking and Classification	AUC, WSS
Kilicoglu et al.	2009	NR	MEDLINE	Classification	R, P, F, AUC
Cohen et al.	2009	NR and UR	AHRQ	Ranking and Classification	AUC
Matwin et al.	2010	NR	AHRQ	Classification	WSS
Wallace et al.	2010	NR	AHRQ	Active Learning and Classification	Y, B
Cohen et al.	2010	NR	AHRQ	Classification	AUC
Karimi et al.	2010	NR	AHRQ	Ranking	R, P
Frunza et al.	2010	NR	MEDLINE	Classification	R, P, F, WSS
Tomassetti et al.	2011	NR	MEDLINE	Active Learning and Classification	WR
Wallace et al.	2012	NR	MEDLINE	Active Learning and Classification	WR
Wallace et al.	2012	UR	MEDLINE	Classification	R, P
Kim and Choi	2012	NR	nHTA and AHRQ	Classification	MP
Cohen et al.	2012	UR	MEDLINE	Classification	R, P, F
Dalal et al.	2012	UR	MEDLINE	Classification	R, P, WR
Jonnalagadda and Petitti	2013	NR	AHRQ	Active Learning	R, WSS
Kim and Choi	2014	NR	nHTA and AHRQ	Classification	AUC
Miwa et al.	2014	NR	AHRQ	Active Learning and Classification	U, C, AUC
Bui et al.	2015	NR	MEDLINE	Ranking and query expanding	R, P
Cohen et al.	2015	NR	MEDLINE	Ranking and Classification	AUC, F, A, AP
Hashimoto et al.	2016	NR	AHRQ	Active Learning and Classification	Y, B, WSS
Zhang et al.	2016	NR	TREC2015	Active Learning	Y
Khabsa et al.	2016	NR and UR	AHRQ	Classification	R, WSS
Kontonatsios et al.	2017	NR	AHRQ	Active Learning	AU
Cormack and Grossman	2017	NR	CLEF2017	Active learning and Ranking	WSS, AP
Anagnostou and Vlahavas	2017	NR	CLEF2017	Ranking and Classification	WSS, AP
Nunzio et al.	2017	NR	CLEF2017	Active learning and Ranking	WSS, AP
Alharbi and Stevenson	2017	NR	CLEF2017	Ranking	WSS, AP

Table 2.9: Summary of the information extracted from the 63 included studies.

Author	Year	Type	Dataset Source	Approach	Evaluation Measures
Norman et al.	2017	NR	CLEF2017	Active learning and Ranking	WSS, AP
Hollmann and Eickhoff	2017	NR	CLEF2017	Active learning and Ranking	WSS, AP
Kalphov and Azzopardi	2017	NR	CLEF2017	Active learning and Ranking	WSS, AP
Scells et al.	2017	NR	CLEF2017	Ranking	WSS, AP
Yu and Menzies	2017	NR	CLEF2017	Ranking and Classification	WSS, AP
Chen et al.	2017	NR	CLEF2017	Ranking	WSS, AP
Lee	2017	NR	CLEF2017	Ranking and Classification	WSS, AP
Singh et al.	2017	NR	CLEF2017	Ranking	WSS, AP
Altena and Olabarriaga	2017	NR	CLEF2017	Ranking and Classification	WSS, AP
Singh and Thomas	2017	NR	CLEF2017	Ranking	WSS, AP
Bian et al.	2017	NR	MEDLINE	Classification	MAP
Scells and Zuccon	2018	NR	MEDLINE	Ranking and Classification	R, P, F, WSS
Surian et al.	2018	UR	MEDLINE	Ranking	R, WSS, MR
Minas et al.	2018	NR	CLEF2018	Active learning and Ranking	WSS, AP
Norman et al.	2018	NR	CLEF2018	Active learning, Ranking and Classification	WSS, AP
Wu et al.	2018	NR	CLEF2018	Ranking	WSS, AP
Cohen and Smalheiser	2018	NR	CLEF2018	Active learning and Ranking	WSS, AP
Alharbi et al.	2018	NR	CLEF2018	Active learning and Ranking	WSS, AP
Nunzio et al.	2018	NR	CLEF2018	Ranking	WSS, AP
Cormack and Grossman	2018	NR	CLEF2018	Active learning and Ranking	WSS, AP
Lee and Sun	2018	NR	CLEF2017	Ranking	AP
Alharbi and Stevenson	2019	NR	CLEF2019	Ranking	WSS, AP
Nunzio	2019	NR	CLEF2019	Ranking	WSS, AP
Alharbi and Stevenson	2019	UR	Cochrane	Active learning and Ranking	WSS, AP
Li and Kanoulas	2019	NR	CLEF2019	Ranking	WSS, AP
Scells et al.	2019	NR	MEDLINE	Ranking and Classification	R, P, F, WSS
Lerner et al.	2019	UR	MEDLINE	Classification	R
Alharbi and Stevenson	2019	NR	CLEF2018/19	Active learning and Ranking	WSS, AP
Scells et al.	2020	NR	CLEF2017/18	Ranking	MAP
Zuccon et al.	2020	NR	Cochrane	Ranking	R, P
Alharbi and Stevenson	2020	UR	Cochrane	Query Refinement	R, P

Table 2.9: Summary of the information extracted from the 63 included studies.

Author	Year	Type	Dataset Source	Approach	Evaluation Measures

New Review (NR), Update Review (UR)
Cochrane Collaboration and the Agency for Healthcare Research and Quality (AHRQ), New Health Technology Assessment (nHTA),
Total Recall (TREC), Recall (R), Precision (P), F1 (F1), Work Reduced (WR), Specificity (S), Yield (Y), Burden (B), Mean Percentage (MP),
Accuracy (A), Utility (U), Average Utility (AU), Area Under The Curve (AUC), Normalised Utility (NU), Coverage (C), Median Rank (MR)

Chapter 3

Query Adaptation to Improve Ranking

3.1 Introduction

As we have seen in Chapter 2, a considerable amount of literature has demonstrated the effectiveness of ranking in the systematic review screening process. O'Mara-Eves et al. (2015) and Cohen (2008) highlighted many benefits of ranking studies for a systematic review in terms of workflow efficiency and reducing the burden of abstract screening. One is that reviewers acquire a better understanding of the inclusion criteria earlier in the process because they find more examples of relevant studies faster than it would otherwise be the case. It also allows analysis of document content to start earlier than it can happen when studies are screened at random. This can be a significant benefit because accessing the content of studies induces the content screening process, since the majority of relevant studies should be identified early. On the other hand, in reviews with searches resulting in a very large number of studies, it would be particularly useful to review the studies in order of their likely importance. In this case, the remaining studies can be screened in the following months by less experienced reviewers.

Ranking also helps in reducing the workload required by the researches by increasing the rate (or speed) of screening. Instead of screening a large unordered set of documents to assist their eligibility, under ranking, the most relevant documents should tend to appear early in the list. This means that study screening and selection can be better focused and take less time to complete compared with conventional manual screening and thus reduce the workload (Karimi et al., 2010; Olofsson et al., 2017).

Query adaptation techniques have been widely applied in IR and considered as a promising approach to improve retrieval performance (Carpineto and Romano, 2012). Query adaptation is the process of reformulating a given query with the aim of improving retrieval performance (Shobha and Rangaswamy, 2018). One approach to query adaptation is query expansion where the original query is extended by adding related terms. Related terms can be identified using unstructured data (e.g. text documents) or structured data (e.g. ontology) (Bai and Nie, 2008; Shen et al., 2006).

Several studies have applied query adaptation for medical domain search. For example, Díaz-Galiano et al. (2009) used a medical ontology (MeSH) to expand the query to improve the retrieval system. They applied their system on the ImageCLEFmed dataset and showed that results improved. Abdoaziz et al. (2016) applied linear combinations of different query expansion techniques by finding synonyms and re-weighting original query terms. Their proposed model improved performance (MAP) by 21.06% compared with the baseline. Furthermore, previous work on the refinement and generation of Boolean queries for other types of professional searches, such as prior art search, has been proposed. Kim et al. (2011) designed a Boolean query suggestion technique in which a decision tree was learned from pseudo-relevant documents and then used to generate queries. Graf et al. (2010) developed a method for automatically generating queries for prior art search by analysing the distribution of terms among topic-relevant documents. Harris et al. (2014) presented an interactive Boolean search system which helps the user to create a Boolean search query. The interactive system suggests semantically similar search terms to the user.

This chapter explores the use of different query adaptation approaches to improve studies ranking for the creation of systematic reviews. It aims to apply three main approaches. In the first approach (Section 3.2), it investigates which information from the Boolean query and studies is helpful for improving ranking. In the second approach (Section 3.3), it explores the applications of lexical statistics techniques in the domain of systematic review and how they can be used to improve ranking studies. In the final approach (Section 3.4), it applies a relevance feedback method using the Rocchio's algorithm. For each approach, it presents the method, datasets, evaluation measures and results.

3.2 Approach 1: Query Terms and Medical Subject Headings

This section investigates which information from the review and studies can help to improve ranking for systematic reviews. Information available from the review includes the title of the review, the Boolean query and the list of studies retrieved from the search. On the other hand, the information available from each study are the title, abstract, as well as a list of MeSH terms.

This section first gives a brief overview of Boolean queries and MeSH. Then, explains the different methods proposed to improve ranking: the use of Boolean query terms, the use of query MeSH terms, the use of query Explode MeSH and the use of article Major MeSH terms. It compares the results obtained from theses methods with a baseline system which represents the common scenario with many systematic review projects where the studies are evaluated in the order they are retrieved without any prioritisation.

3.2.1 Boolean Query

Candidate studies for inclusion in systematic reviews are identified using Boolean queries constructed by domain experts. These queries are designed to optimise recall as reviews aim to identify and assess all relevant evidence. Boolean queries in the reviews used throughout this Thesis are created for either the OVID or PubMed interfaces to the MEDLINE database of medical literature. Figure 3.1 shows examples of two different formulations for the Boolean queries of two Cochrane reviews.

Queries are often complex, consisting of multiple lines and including operators such as AND, OR and NOT, in addition to advanced operators such as wildcard, explosion and

(a) OVID format Query for review CD009591

```
1.  exp magnetic resonance imaging/ OR ultrasonography/ OR exp Imaging,
Three-Dimensional/ OR exp radiography/
2.  ultraso$.tw.  OR magnetic resonance imaging.tw.  OR MRI.tw.  OR imag$.tw.
3.  diagnos$.tw.
4.  1 OR 2 OR 3
...
...
9.  (animals not (humans and animals)).sh.
10.  8 not 9
```

(b) PubMed format Query for review CD008643

```
"Medical History Taking"[mesh] OR history[tw] OR "red flag"[tw] OR "red
flags" OR Physical examination[mesh] OR "physical examination"[tw]
OR "function test"[tw] OR "physical test"[tw] OR ((clinical[tw]
OR clinically[tw]) AND (diagnosis[tw] OR sign[tw] OR signs[tw] OR
significance[tw] OR symptom*[tw] OR parameter*[tw] OR assessment[tw] OR
finding*[tw] OR evaluat*[tw] OR indication*[tw] OR examination*[tw]))
...
```

Figure 3.1: Example queries from Cochrane reviews (Nisenblat et al., 2016; Williams et al., 2013).

truncation (Karimi et al., 2010). Furthermore, restriction fields are used to specify the search (e.g. using .ab. to search for terms appear in abstract only and .sh. to search for MeSH terms only). Table 3.1 provides a list of the operators and restriction fields that can be used to create OVID and PubMed format queries.

OVID queries usually consist of multiple lines (clauses), which are numbered so they can be referenced. For example, in Figure 3.1(a), line 4 combines the results of lines 1, 2 and 3 in a disjunction (OR).

Table 3.1: Set of OVID and PubMed Boolean query operators and restriction fields with their meanings.

Query Format	Operator/Restriction Field	Meaning
OVID and PubMed	AND	conjunction, include all search terms
	OR	disjunction , include at least one of the search terms
	NOT	exclude search terms
OVID	.tw.	term appears in title or abstract
	.ti.	term appears in title
	.ab.	term appears in abstract
	.ti,ab.	term appears in title or abstract
	/ or .mp.	MeSH terms
	.sh.	MeSH Subheadings
	.af.	term appears in any field
PubMed	[Text Word] or [tw]	term appears in title, abstract or MeSH
	[Title] or [ti]	term appears in title
	[Title/Abstract] or [tiab]	term appears in title or abstract
	[mesh] or [mh]	MeSH terms
	[sh]	MeSH Subheadings
	[All filelds] or [All]	term appears in any field

3.2.2 The Medical Subject Headings (MeSH) Hierarchy

The MeSH thesaurus was created by the National Library of Medicine[1] and is used to describe the subject of each article in MEDLINE. It is used to support indexing and searching for biomedical articles. There are over 27,000 main MeSH terms representing concepts found in the biomedical literature (Dhammi and Kumar, 2014). They are arranged by subject in a hierarchy known as the MeSH Tree Structures, which can be used to expand or narrow down the search. Each main MeSH term has a number of subheadings to describe a specific aspect of a concept.

Term Explosion

In a systematic review, it is common to add MeSH terms and subheadings to the Boolean query to assist in subject searches. However, this can be done in different ways: the use of the Explode function `'exp'` with the MeSH term (e.g. `exp Dementia/`) or by only including the MeSH term without Explode function (e.g. `Dementia/`). The Explode function `'exp'` searches for the main MeSH terms and automatically includes all its

[1] https://www.nlm.nih.gov/

narrower terms (subheadings) (PubMed Tutorial, 2017). For example, the Boolean query of review CD009786 includes the line: `exp Dementia/` (Van de Vrie et al., 2019). When this query is run on MEDLINE, it will retrieve all articles indexed with the `Dementia` MeSH and/or with the narrower subject headings in the `Dementia` tree hierarchy (underlined text in Figure 3.2).

All MeSH Categories
↳ Diseases Category
↳ Nervous System Diseases
↳ Central Nervous System Diseases
↳ Brain Diseases
↳ Dementia
↳ <u>AIDS Dementia Complex</u>
<u>Alzheimer Disease</u>
<u>Aphasia, Primary Progressive</u>
<u>Creutzfeldt-Jakob Syndrome</u>
<u>Dementia, Vascular</u>
<u>Dementia, Multi-Infarct</u>

Figure 3.2: Example of an exploded MeSH including the narrower subject headings in the `Dementia` tree hierarchy.

Major MeSH Terms

Each article in MEDLINE has a list of MeSH terms that describe the article subject (Newman et al., 2009). Moreover, asterisks `'*'` on MeSH terms and subheadings indicate that they are the major topics of this particular article, usually obtained from the title and the statement of purpose of the article. Figure 3.3 shows an example of an article from PubMed. As can be seen, this article has a list of MeSH terms, two of which are major:

```
Deoxycholic Acid/analogs & derivatives*
Cholelithiasis/drug therapy*
```

Major MeSH terms help to identify the subject of an article when it has no abstract (Medical Subject Headings (MeSH®) in MEDLINE®/PubMed®: A Tutorial, 2012).

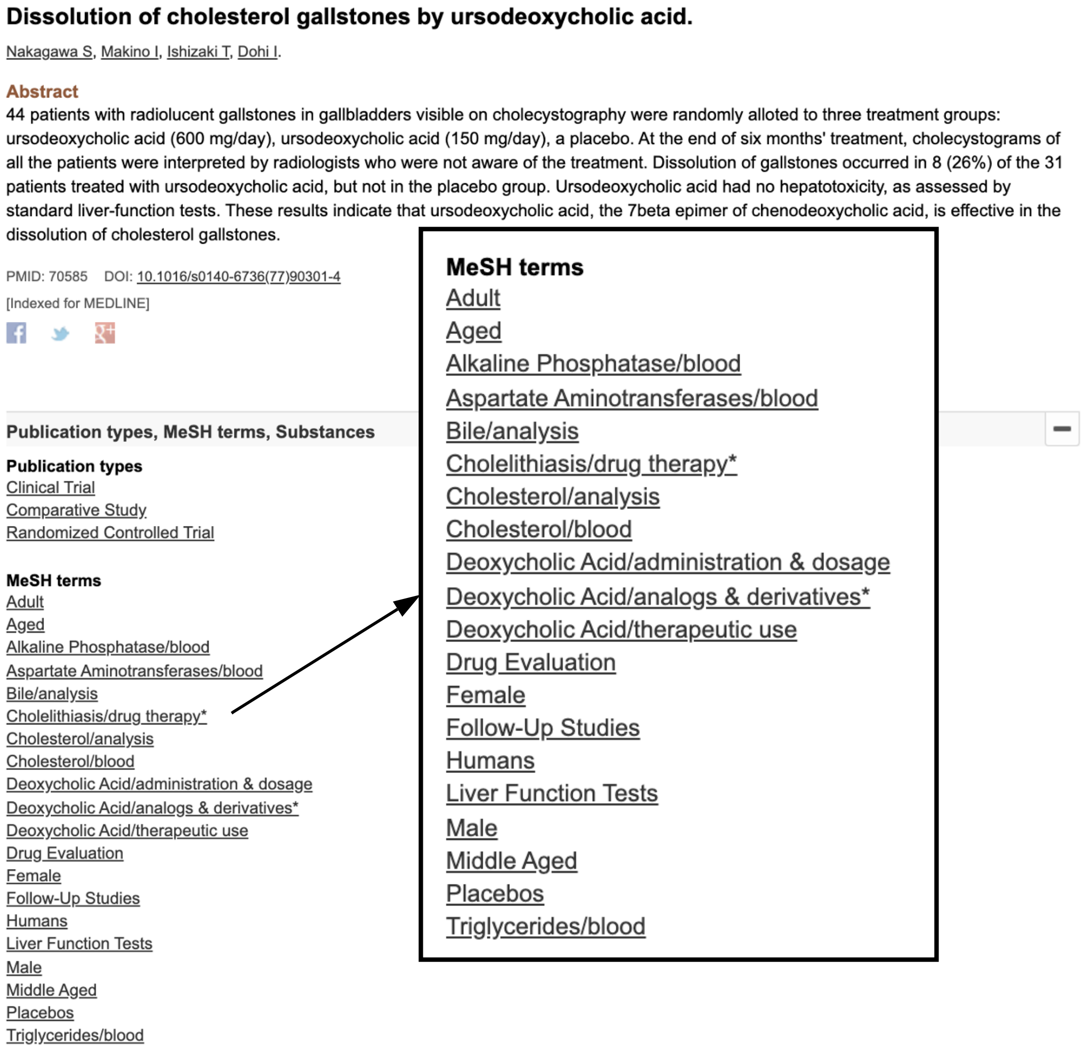

Figure 3.3: Example of MeSH terms for an article from PubMed (Nakagawa et al., 1977). Major MeSH terms are denoted by asterisks '∗'.

3.2.3 Dataset

The CLEF2017 and CLEF2018 datasets described in Section 2.4.2 were used. The CLEF2017 dataset contained 266,967 abstracts divided into training and test sets containing 20 and 30 reviews, respectively. The CLEF2018 dataset contains 460,165 abstracts and is divided into a training dataset consisting of 42 reviews and a test dataset of 30 reviews. All reviews are DTA reviews. For each review, the dataset contains review id, review title, Boolean query and a list of PMIDs retrieved by the query (see Figure 2.8). Some of the Boolean queries are in the OVID format and others are in the PubMed format. Throughout this chapter, these datasets will be used.

3.2.4 Experiments

A number of experiments were conducted to explore the use of a variety of information from the review and studies to improve ranking. Our methods make use of three pieces of information from each review in the dataset: (1) review title, (2) terms extracted from the Boolean query and (3) MeSH terms extracted from the Boolean query (including the exploded MeSH terms). This information is extracted from the Boolean query automatically using a simple parser designed to interpret both OVID and PubMed style queries[2].

The terms were extracted from the Boolean query based on the restriction fields. For example, $.ti,ab.$ is extracted as a term, whereas $.sh.$ is considered a MeSH term. Table 3.2 shows common query restriction fields and whether we consider each one a term or a MeSH. Terms and MeSH terms modified by certain operators (e.g. `NOT`) were not extracted.

Table 3.2: OVID and PubMed common query restriction fields and whether we consider each one a term or a MeSH.

Query Format	Restriction Field	Name	Term or MeSH ?
OVID	/ or .mp.	MeSH term	MeSH
	.sh.	MeSH Subheading	MeSH
	.af.	All Fields	Term
	.tw.	Text Word	Term
	.ti.	Title	Term
	.ti,ab.	Title/Abstract	Term
PubMed	[mesh] or [mh]	MeSH term	MeSH
	[sh]	MeSH Subheading	MeSH
	[All Fields] or [All]	All Fields	Term
	[Text Word] or [tw]	Text Word	Term
	[Title] or [ti]	Title	Term
	[Title/Abstract] or [tiab]	Title/Abstract	Term

[2]The approach was implemented using Python v3.6

Some MeSH terms (e.g. `Spine`) are also standard English words that could appear as a term in an abstract. To avoid false matches, all MeSH terms extracted from a query were prefixed with the string `MeSH`. In addition, MeSH terms were pre-processed to remove white spaces and punctuation (e.g. `Lumbar vertibrae` becomes `MeSHLumbarvertibrate`).

Exploded MeSH terms, indicated with the prefix `'exp'` in the Boolean query, were identified using a simple parser. Subheadings of the exploded terms were identified by querying the MeSH vocabulary tree[3]. These MeSH terms and subheadings were prefixed with the string `MeSH`. Figure 3.4 shows an example of a Boolean query and the list of terms, pre-processed MeSH terms and `'exp'` MeSH extracted by the parser.

The studies returned by the Boolean query for each review in the dataset are defined by their PMIDs. All PMIDs provided with the review were downloaded from PubMed[4]. The text of the title, abstract and MeSH terms for each article were extracted and the MeSH terms pre-processed using the same approach that had been applied to the Boolean query.

Pre-processing was applied to both the studies and information extracted from the review. The text was tokenised[5], converted to lower-case, stop words[6]/punctuation removed and the remaining tokens stemmed[7].

A Vector Space Model (VSM) was used for retrieval. We chose `tf.idf` which has been used in the field of biomedical text retrieval in a significant number of previous and recent studies, for example, Bashir et al. (2020); Jabri et al. (2018); Lerner et al. (2019); Scells et al. (2017b, 2020); Surian et al. (2018). The information extracted from the review and each of the studies was converted into `tf.idf` weighted vectors in a high dimensional vector space. The similarity between the review (R) and each article (a) was then generated by computing the angle between the review vector ($\vec{R}$) and the article vector ($\vec{a}$) using the cosine similarity function (Baeza-Yates and Ribeiro-Neto, 2011):

[3] ftp://nlmpubs.nlm.nih.gov/online/mesh/.
[4] The `Entrez` package from `biopython.org` was used.
[5] NLTK's `tokenize` package was used for tokenisation.
[6] The list of stop words provided by Scikit-learn (http:scikit-learn.org/stable/) was used.
[7] NLTK's `LancasterStemmer` package was used for stemming.

(a) Boolean Query

```
1.  exp Dementia/
2.  Cognition Disorders/
3.  (alzheimer$ or dement$).ti,ab.
4.  ((cognit$ or memory or cerebr$ or mental$) adj3 (declin$
or impair$ or los$ or deteriorat$ or degenerat$ or complain$
or disturb$ or disorder$)).ti,ab.
5.  (forgetful$ or confused or confusion).ti,ab.
6.  MCI.ti,ab.
7.  ACMI.ti,ab.
8.  ARCD.ti,ab.
9.  SMC.ti,ab.
10.  CIND.ti,ab.
11.  BSF.ti,ab.
12.  Positron-Emission Tomography/
13.  disease progression/
...
...
```

(b) Terms extracted from the query

```
alzheimer , dement , cognit , memory , cerebr , mental , declin ,
impair , los , deteriorat , degenerat , complain , disturb , MCI ,
disorder , forgetful , confused , confusion , ACMI , ARCD , SMC , CIND ,
BSF , ...
```

(c) Pre-processed MeSH headings

```
MeSHDementia , MeSHCognitionDisorders , MeSHdiseaseprogression
MeSHPositronEmissionTomography , ...
```

(d) Pre-processed ‘`exp`’ MeSH headings for “`exp Dementia/`”

```
MeSHAIDSDementiaComplex , MeSHAlzheimerDisease , MeSHAphasia
, MeSHPrimaryProgressive , MeSHCreutzfeldtJakobSyndrome ,
MeSHDementiaVascula , MeSHMultiInfarct
```

Figure 3.4: Example portion of a Boolean query (Van de Vrie et al., 2019) (a), sample of terms extracted from the Boolean query (b), sample of pre-processed MeSH headings extracted from the Boolean query (c) and the pre-processed MeSH headings for “`exp Dementia/`” (d).

$$similarity(R,a) = \cos(\theta) = \frac{\vec{R}\cdot\vec{a}}{|\vec{R}|\times|\vec{a}|} = \frac{\sum_{i=1}^{n} R_i \times a_i}{\sqrt{\sum_{i=1}^{n} R_i^2} \times \sqrt{\sum_{i=1}^{n} a_i^2}} \tag{3.1}$$

where $\vec{R}\cdot\vec{a}$ represents the dot product of the two vectors, $|\vec{R}|$ and $|\vec{a}|$ represent the length of the review and article vectors, respectively. Another approach that can be applied with the VSM is the dense embeddings. For example, the newly introduced BioBERT, which is a domain-specific language representation model pre-trained on large-scale biomedical corpora (Lee et al., 2020). Because of the time limit, we left this for future work.

Studies were ranked based on the scores generated by the cosine similarity function[8]. Studies at the top of the ranking list are those closer (more similar) to the review vector which are more likely to be relevant to the review.

Four approaches were explored in addition to a baseline system. For each of the four approaches, studies were ordered based on the cosine similarity scores (Equation 3.1). Below, each approach is explained.

Baseline

For the baseline system, the list of studies was randomly ordered with the intention of representing the scenario in which the results of the Boolean query are simply evaluated in the order in which they are retrieved without any attempt to identify those most likely to be relevant. This situation simulates common practice within many systematic review projects in which reviewers examine each of the retrieved studies in turn. The score of each study is calculated using the following equation:

$$score = \frac{t-r+1}{t} \tag{3.2}$$

where t is the total number of studies returned by the Boolean query and r the study's rank in the random ordering.

[8]Scikit-learn's `TfidfVectorizer` and `linear_kernel` packages were used for these steps

Since this approach was the starting point on the retrieval problem for systematic reviews that this thesis is addressing, we selected this simple baseline. It provides a useful first step and allows us to better understand the problem in order to apprise us of the best way to approach it. It gives an idea of how effective the NLP techniques compared to what is happening in the common practice within many systematic review projects in the real world.

Query Terms

In this experiment, only the review title and the terms extracted from the Boolean query were used when calculating the similarity between the review and each study. For this experiment, $\vec{R}$ represents terms from review's title and Boolean query, while $\vec{a}$ represents terms from the study's title and abstract.

Query MeSH

This experiment used only terms from review's title and MeSH terms extracted from the Boolean query (terms extracted from the query were not used). For this experiment, $\vec{R}$ represents terms from review's title and MeSH terms from Boolean query, while $\vec{a}$ represents terms from the study's title, terms from abstract and MeSH terms.

Query Exploded MeSH

This experiment explored the use of the `'exp'` function (see Section 3.2.2) in the Boolean query. It evaluated the performance of ranking when each explode MeSH in the query is replaced by all the subheadings of this MeSH. For this experiment, $\vec{R}$ represents terms from review's title, terms from Boolean query, MeSH terms from Boolean query as well as the subheading for each of these MeSH retrieved from the MeSH tree. Meanwhile, $\vec{a}$ represents terms from the study's title, terms from abstract and MeSH terms.

Article Major MeSH

The aim of this experiment was to investigate the use of major MeSH terms (see Section 3.2.2) associated with each article in search results. We examined whether the use of major MeSH terms only is more beneficial than the use of the whole MeSH list retrieved with the article.

A simple parser was used to identify Major MeSH terms for each study. After that, the cosine similarity calculated where $\vec{R}$ represents terms from review's title and terms and MeSH terms from Boolean query. Meanwhile, $\vec{a}$ represents terms from the study's title, terms from abstract and Major MeSH terms only.

3.2.5 Evaluation Metrics

For the evaluation, AP, MAP, WSS@100 and WSS@95 which have been described in Section 2.4.3 were used. Theses are the most commonly used metrics when evaluating approaches to study identification for systematic reviews, e.g. Cohen et al. (2006); Kanoulas et al. (2017); Suominen et al. (2018).

3.2.6 Results and Discussion

Table 3.3 shows the results of applying the different proposed methods: using Query Terms, using Query MeSH, using Query Exploded MeSH and using Article Major MeSH. As expected, all of the implemented methods outperform the simple baseline approach where the studies are randomly ranked for both CLEF2017 and CLEF2018 datasets.

Table 3.3: Results of making use of different information from the Boolean query and studies for CLEF2017 and CLEF2018 test datasets. The best performance among all methods is in boldface.

	CLEF2017			CLEF2018		
Approach	**MAP**	**WSS@100**	**WSS@95**	**MAP**	**WSS@100**	**WSS@95**
Baseline	0.045	3.90%	3.10%	0.051	2.30%	2.90%
Query Terms	**0.218**	38.50%	49.30%	**0.224**	**37.70%**	50.60%
Query MeSH	0.158	30.30%	42.30%	0.184	33.80%	45.80%
Query Exp MeSH	0.199	36.60%	47.00%	0.207	32.60%	49.00%
Article Major MeSH	0.187	**38.60%**	**49.50%**	0.213	37.40%	**52.70%**

The best performance in terms of MAP score was achieved by including only terms from the Boolean query. The MAP score improved by 17.3% compared to the baseline for both CLEF2017 and CLEF2018 datasets. This method is also close to the best result for the WSS@100 and WSS@95 scores. The screening effort required to identify all relevant studies (100% recall) is reduced by a third and for identifying 95% of the relevant studies, it is reduced by almost a half for both CLEF2017 and CLEF2018 datasets.

Results suggest that including terms extracted from the Boolean query is beneficial. However, the usefulness of MeSH terms extracted is less clear. The MAP score decreases when these are used instead of query terms (e.g. compare Query Terms and Query MeSH). On the other hand, using Query Exploded MeSH is more beneficial than using Query MeSH or Article Major MeSH in terms of MAP while Article Major MeSH is more beneficial in terms of reducing workload.

Figures 3.5 and 3.6 show the results of AP for each review in the test dataset using the four proposed methods against the baseline on CLEF2017 and CLEF2018 datasets, respectively. From the figures, it can be seen that the AP for all the reviews was significantly improved. In addition, it is apparent that there is a variation of the AP scores between reviews. A possible explanation for this might be the different percentage of relevant studies among the retrieved results. For example, for Review CD010772, the relevant studies represent 14.87% of retrieved documents. In contrast, in review CD009786, the relevant studies represent only 0.48% of the retrieved documents.

The results of these experiments demonstrated that even straightforward ranking techniques provide potential benefit to systematic reviews by ensuring that studies more likely to be relevant are placed higher in the rankings. The review title and terms extracted from the Boolean query were found to be the most useful pieces of information in improving studies ranking.

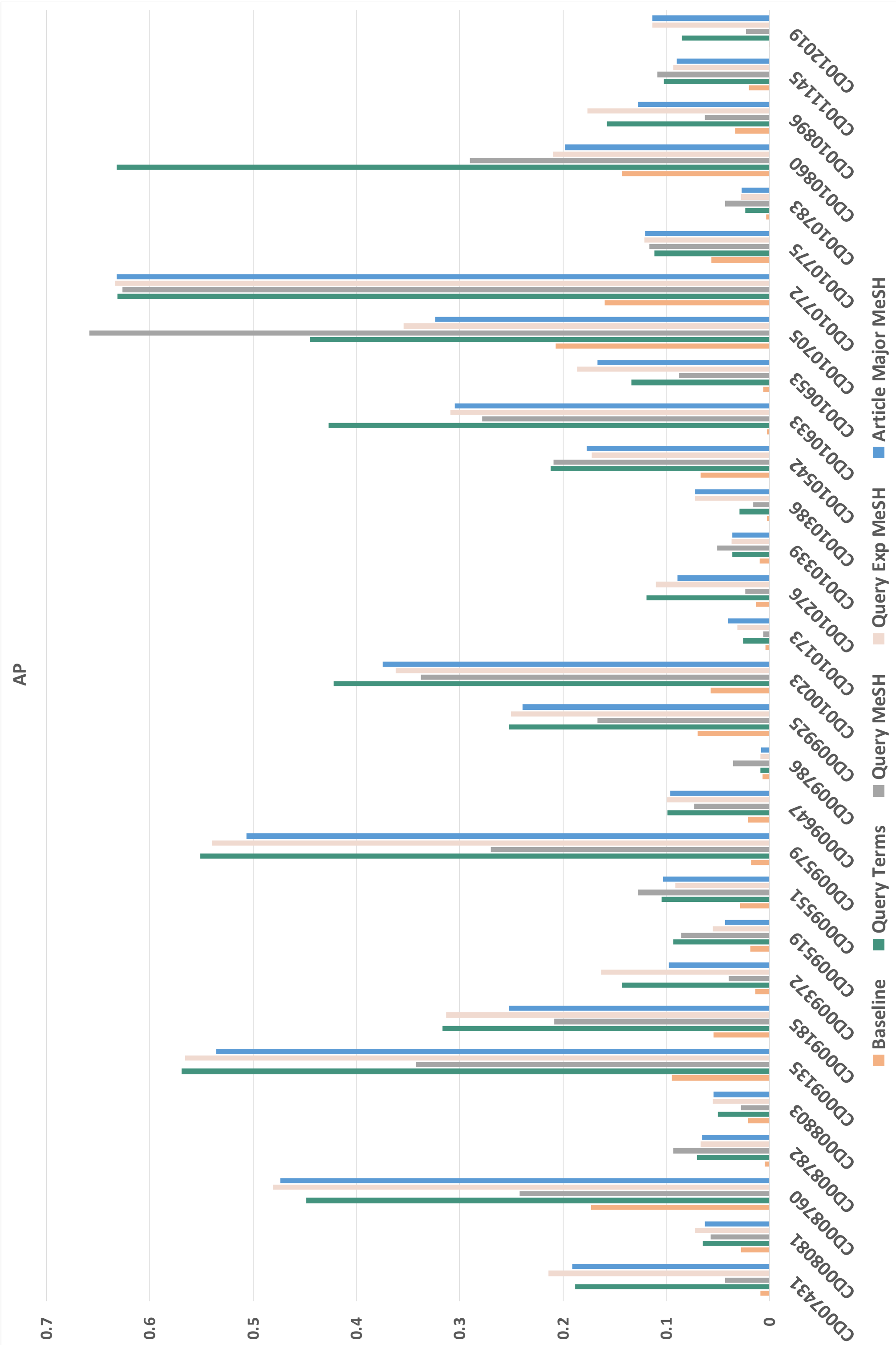

Figure 3.5: AP scores for each review using baseline and different approaches for CLEF2017 test dataset.

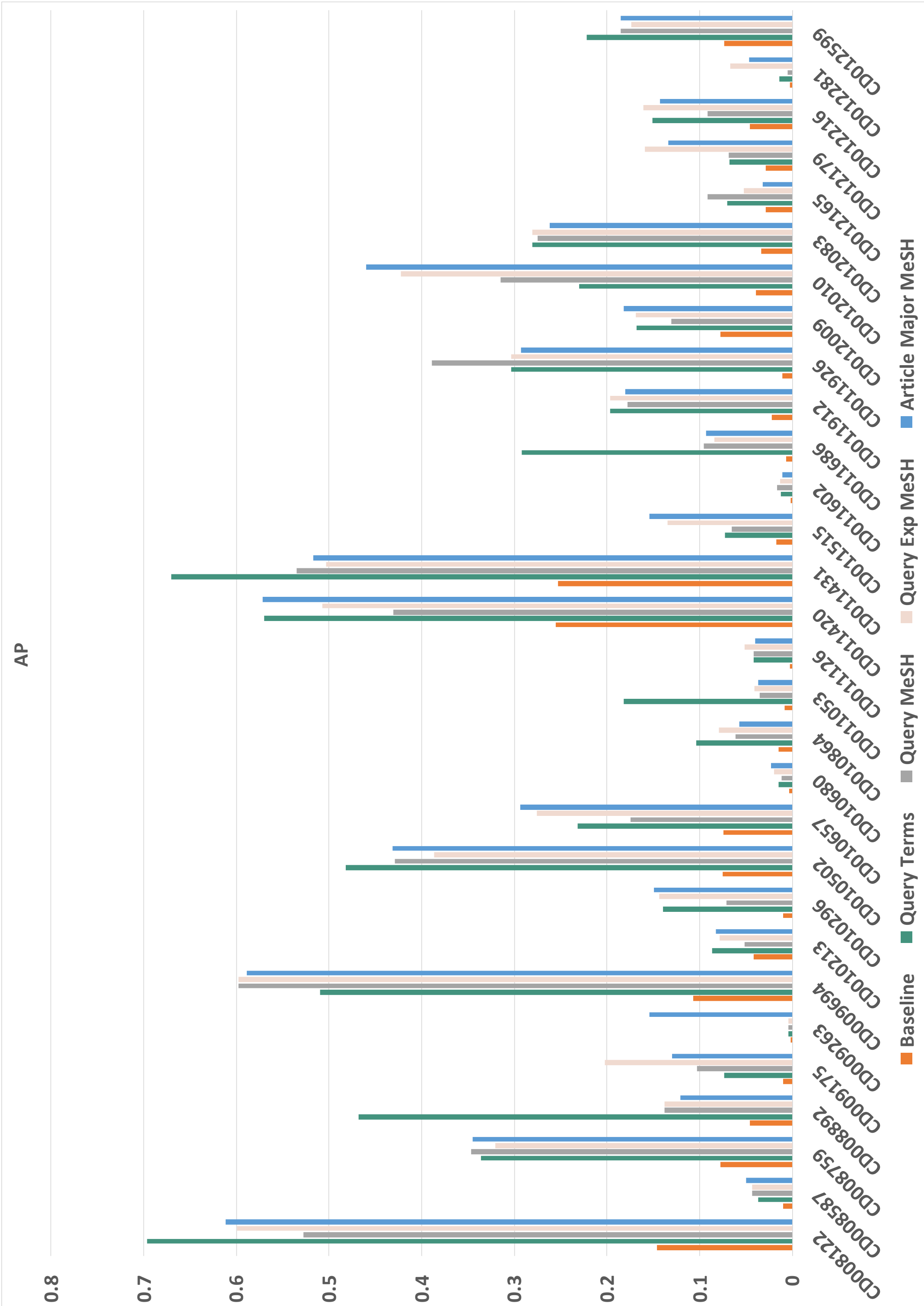

Figure 3.6: AP scores for each review using baseline and different approaches for CLEF2018 test dataset.

3.3 Approach 2: Lexical Statistics

This approach hypothesises that there are terms which distinguish the studies that are likely to be included in a specific type of review (e.g. DTA review) from other literature. Expanding the Boolean query with those terms may help find the most relevant studies. This section explores the use of lexical statistics to derive a list of key terms that indicate evidence relevant to a specific type of Cochrane reviews.

Keyword analysis has increasingly been used in applied linguistics in recent years (Pojanapunya and Todd, 2018). A keyword refers to a lexical item which occurs with unusual frequency, either with a significantly higher or lower frequency in a target text or corpus, when compared to a reference corpus. The majority of previous studies using keyword analysis have used the Log-Likelihood or Chi-Square statistics (Bestgen, 2018; Pojanapunya and Todd, 2018; Rayson, 2019). These statistics are used to identify the key terms that are characteristic of a sub-corpus (Dunning, 1993; Rayson, 2008).

Chi-Squared was first applied in a corpus analysis context by Hofland and Stig (1982) to compare word frequencies in corpora of one million words of American English (the Brown Corpus) with one million words of British English (the LOB Corpus). On the other hand, Log-Likelihood was first brought to the attention of the corpus community by Dunning (1993) for collocation analysis. In 2018, Pojanapunya and Todd (2018) argued that the Log-Likelihood and Odds-Ratio statistics produce different keywords applicable to research focusing on different purposes.

This approach set out to investigate the usefulness of these most widely applied lexical statistics (i.e. Log-Likelihood, Chi-Squared and Odds-Ratio) to identify terms characteristic of studies likely to be relevant for DTA reviews. We analysis the keywords list generated by each statistic and find which statistic is the most appropriate to improve the retrieval performance for systematic reviews.

The relevant studies are treated as a target sub-corpus with the aim to identify the terms that characterise it compared with the comparative sub-corpus of non-relevant studies, so that they can be used to adapt the query.

In general, the steps involved in conducting keywords analysis can be summarised as follows (Pojanapunya and Todd, 2018; Rayson, 2012). First, the corpus partitioned into two sub-corpora: the target sub-corpus and the comparative sub-corpus. In this approach, the target sub-corpus contains the relevant studies, and the comparative sub-corpus contains the non-relevant studies. Second, a terms list is generated for both sub-corpora including the terms and their frequency. In this step, a minimum threshold value may be assigned. In previous research, the threshold was usually set to 2, 4, 5 or 10 (Pojanapunya and Todd, 2018; Scott and Tribble, 2006). This step involves the use of a contingency table that is created for each term (see Table 3.4). It encodes information about the frequency with which the term appears in each sub-corpus. For example, O_{rel} represents the number of times the term occurs within the entire set of relevant studies and N_{rel}, the sum of the occurrences of all terms in the relevant set.

Table 3.4: Contingency table for computing lexical statistics.

	Relevant	**Non-relevant**
Frequency of term	O_{rel}	O_{nonRel}
Total tokens	N_{rel}	N_{nonRel}

In the third step, for each term in the list, the statistical scores are calculated (e.g. Log-Likelihood, Chi-Squared or Odds-Ratio). Finally, the terms on the list are re-sorted based on the statistical scores generated. The terms on the top of the list are the most likely ones to differentiate the two sub-corpora.

Below, we explain how each lexical statistic is computed.

3.3.1 Log-Likelihood

Log-Likelihood depends on the comparison of the relative frequencies of a particular term in a sub-corpus. Based on Table 3.4, Log-Likelihood for a single term is calculated as follows (Pojanapunya and Todd, 2018; Rayson, 2008):

$$Log\text{-}Likelihood = 2 \times \left(O_{rel} \times \ln \frac{O_{rel}}{E_{rel}} + O_{nonRel} \times \ln \frac{O_{nonRel}}{E_{nonRel}} \right) \tag{3.3}$$

where O_{rel} and O_{nonRel} are the observed frequency of the term in different subsets of the collection (e.g. relevant and non-relevant studies). E_{rel} and E_{nonRel} are the term's expected frequencies calculated as:

$$E_{rel} = N_{rel} \times \frac{O_{rel} + O_{nonRel}}{N_{rel} + N_{nonRel}} \quad , \quad E_{nonRel} = N_{nonRel} \times \frac{O_{rel} + O_{nonRel}}{N_{rel} + N_{nonRel}} \tag{3.4}$$

where N_{rel} and N_{nonRel} represent sub-corpus size.

Terms are assigned high Log-Likelihood scores for a particular corpus when their observed frequency is (much) higher than the expected frequency. In other words, a high Log-Likelihood score implies that a term has a more significant relative frequency differentiation between the two sub-corpora.

3.3.2 Chi-Squared

Chi-Squared is used to compare frequencies of a term across two sub-corpora. In relation to Table 3.4, Chi-Squared for each term is calculated as:

$$Chi\text{-}Squared = \frac{(O_{rel} - E_{rel})^2}{E_{rel}} + \frac{(O_{nonRel} - E_{nonRel})^2}{E_{nonRel}} \tag{3.5}$$

where O_{rel} and O_{nonRel} are the observed values and E_{rel} and E_{nonRel} are the expected values calculated using Equation 3.4.

3.3.3 Odds-Ratio

Odds-Ratio is the lexical statistic most commonly applied for keyword analysis and terms identification (Ghani et al., 2005; Pojanapunya and Todd, 2018). Unlike Log-Likelihood and Chi-Squared, Odds-Ratio depends on the absolute frequencies of a term in a sub-corpus. The Odds-Ratio for each term is calculated as:

$$Odds\text{-}Ratio = \frac{O_{rel} \times (N_{nonRel} - O_{nonRel})}{O_{nonRel} \times (N_{rel} - O_{rel})} \tag{3.6}$$

where O_{rel} and O_{nonRel} are the frequency counts of the term in the relevant and non-relevant sub-corpus and N_{rel} and N_{nonRel} are the total number of terms in each of these sub-corpora.

Odds-Ratio scores are heavily influenced by the terms that have very low frequencies. In other words, the terms which occur only once may appear at the top of the ranked scores. For this reason, it is important to exclude terms with frequency of occurrence below a minimum threshold.

3.3.4 Experiments

A number of experiments were conducted using the three lexical statistics and results compared against a baseline system.

Baseline

To evaluate the proposed hypothesis, the best method applied in Section 3.2 was chosen as a baseline system for comparison (i.e. using query terms). In the baseline system, the studies were ranked by comparing each study against review title and terms extracted from the Boolean query. `tf.idf` weighted vectors were used to represent information obtained from the review and studies, then ranked the studies based on their cosine similarity scores (Equation 3.1).

Lexical Statistics

The Log-Likelihood, Chi-Squared and Odds-Ratio statistics were used to derive lists of terms that indicate evidence relevant to DTA reviews as described in Sections: 3.3.1, 3.3.2 and 3.3.3. The training sets from CLEF 2017 and CLEF 2018 datasets were partitioned into relevant and non-relevant studies depending upon whether the study was included in the systematic review. Terms that occurred fewer than ten times were excluded since it is difficult to generate reliable statistics for these rare terms and, also, they are unlikely to be useful for identifying relevant studies. Setting the minimum frequency threshold at ten is popular, for example, Culpeper (2009); Pojanapunya and Todd (2018); Scott and Tribble (2006).

After computing the lexical statistics for each term in every review, the average for each statistic for each term across all the reviews in the training dataset was computed as:

$$Avg\_statistic(t_i) = \frac{\sum_{j=1}^{T} statistic_j t_i}{T} \tag{3.7}$$

where $statistic_j$ represents the statistic (Log-Likelihood, Chi-Squared or Odds-Ratio) for the term t_i in review j and T is the total number of reviews in the training portion of the dataset (20 for the CLEF2017 dataset and 42 for the CLEF2018 dataset).

For each lexical statistic, the terms with the highest scores were identified and added to the query for each review in the test portion of the dataset. Different numbers of top terms with the highest scores were examined. These included five, ten and twenty top scores. The studies in the test dataset were ranked by matching terms from the review title and terms from the expanded queries against those in the study's title and abstract using cosine similarity measure (Equation 3.1).

Figure 3.7 shows an example of baseline query in addition to expanded queries by adding the top five terms generated from lexical statistics.

(a) Baseline Query

```
lung , pulmonary , neoplasm , cancer , carcinoma, adenocarcinoma
, angiosarcoma , chrondosarcoma , sarcoma , teratoma , lymphoma ,
blastoma , microcytic , tumour , tumor , nsclc , fdg , fludeoxyglucose
, fluorodeoxyglucose , depreotide , positron , photon , scintillation ,
emission , tomograph , cgc , pet , spect , neotect , neospect , neotec
```

(b) Lexical statistic: Log-Likelihood

```
lung , pulmonary , neoplasm , cancer , carcinoma, adenocarcinoma
, angiosarcoma , chrondosarcoma , sarcoma , teratoma , lymphoma ,
blastoma , microcytic , tumour , tumor , nsclc , fdg , fludeoxyglucose
, fluorodeoxyglucose , depreotide , positron , photon , scintillation ,
emission , tomograph , cgc , pet , spect , neotect , neospect , neotec ,
sensitivity , predictive , gonadotropin , hcp , false
```

(c) Lexical statistic: Chi-Squared

```
lung , pulmonary , neoplasm , cancer , carcinoma, adenocarcinoma
, angiosarcoma , chrondosarcoma , sarcoma , teratoma , lymphoma ,
blastoma , microcytic , tumour , tumor , nsclc , fdg , fludeoxyglucose
, fluorodeoxyglucose , depreotide , positron , photon , scintillation ,
emission , tomograph , cgc , pet , spect , neotect , neospect , neotec ,
mtbrif , vulva , inguinfemoral , Xpert , groin
```

(d) Lexical statistic: Odds-Ratio

```
lung , pulmonary , neoplasm , cancer , carcinoma, adenocarcinoma
, angiosarcoma , chrondosarcoma , sarcoma , teratoma , lymphoma ,
blastoma , microcytic , tumour , tumor , nsclc , fdg , fludeoxyglucose
, fluorodeoxyglucose , depreotide , positron , photon , scintillation ,
emission , tomograph , cgc , pet , spect , neotect , neospect , neotec ,
vulva , mtbrif , Xpert , inguinfemoral , geneXpert
```

Figure 3.7: Example of Baseline query (a) and expanded queries (b-d) generated by adding top five terms generated form each lexical statistic.

3.3.5 Results and Discussion

Table 3.5 shows the results of the experiments. The lower part of the table shows the results that were obtained when different numbers of terms with the highest scores were added to each query using different statistics (i.e. Log-Likelihood, Chi-Squared and Odds-Ratio).

Retrieval performance improved when the additional terms were added to the queries, and this improvement was consistent across evaluation metrics for both CLEF2017 and CLEF2018 datasets.

It is apparent from this table that the best performance was achieved by using Log-Likelihood. The MAP improved by almost 1.5% for the CLEF2017 dataset and by 3.5% for the CLEF2018 datasets compared with the baseline system. In addition, the WSS@100 slightly improved by 0.4% for CLEF2017 and by 3.7% for CLEF2018 and WSS@95 improved by 1.4% and by 3.9% for CLEF2017 and CLEF2018, respectively. The performance on CLEF2018 is better than CLEF2017 dataset, this may be down to the number of reviews in the training part of the dataset where the list of terms was derived. The number of reviews in the CLEF2018 training dataset is more than double the number of reviews in the CLEF2017 training dataset (see Section 3.2.3).

Furthermore, it can be seen from the results that enriching the query with more key terms generally improved retrieval performance. For example, adding the top 20 terms using Log-Likelihood improved the MAP by 3.5% while adding only 5 terms improved the MAP by 2% for CLEF2018 dataset. An exception noticed for CLEF2017 dataset when using Log-Likelihood, the MAP decreased by adding the top 10 terms comparing by adding 5 terms only; and then improved by adding the top 20 terms. By analysing the list of terms with the highest scores derived from CLEF2017 training dataset (see Table 3.6), we noticed that the list includes a term which is very specific to certain review (i.e. the 8^{th} term "vulva"). Obviously, this term has caused a decrease in retrieval performance.

Figures 3.8 and 3.9 show the AP results obtained from adding top twenty terms generated form each lexical statistic for the CLEF2017 and CLEF2018 test datasets, respectively. For the CLEF2017 dataset, when using Log-likelihood, the performance of 28 (93.33%) of the reviews improved based on AP compared against the baseline. In addition, when using Chi-Squared and Odds-Ratio, the AP improved for 27 (90%) and 25 (83.33%) of the reviews, respectively. On the other hand, for the CLEF2018 dataset, the AP of almost 80% of the reviews improved when using either Log-likelihood, Chi-Squared or Odds-Ratio as compared against the baseline.

Table 3.5: Lexical statistic results for CLEF2017 and CLEF2018 test datasets. Values in boldface denote the best result achieved by each lexical statistic and the underlined values represent the best results among all three lexical statistics.

		(a) CLEF2017 Dataset			**(b) CLEF2018 Dataset**		
Lexical Statistic	**Terms**	**MAP**	**WSS@100**	**WSS@95**	**MAP**	**WSS@100**	**WSS@95**
Baseline	-	0.218	38.50%	49.30%	0.224	37.70%	50.60%
Log-Likelihood	5	0.232	<u>**38.90%**</u>	**50.70%**	0.244	38.90%	52.50%
	10	0.227	38.00%	49.70%	0.251	40.70%	53.50%
	20	<u>**0.233**</u>	38.40%	**50.70%**	<u>**0.259**</u>	<u>**41.40%**</u>	**54.50%**
Chi-Squared	5	0.214	<u>**38.90%**</u>	49.00%	0.232	38.00%	51.50%
	10	**0.230**	<u>**38.90%**</u>	50.70%	0.242	39.60%	53.00%
	20	**0.230**	<u>**38.90%**</u>	<u>**50.80%**</u>	**0.253**	**40.90%**	<u>**54.70%**</u>
Odds-Ratio	5	0.214	<u>**38.90%**</u>	49.00%	0.221	37.70%	50.50%
	10	0.214	38.80%	48.90%	0.231	38.00%	51.50%
	20	<u>**0.233**</u>	<u>**38.90%**</u>	**50.60%**	**0.252**	**39.80%**	**54.10%**

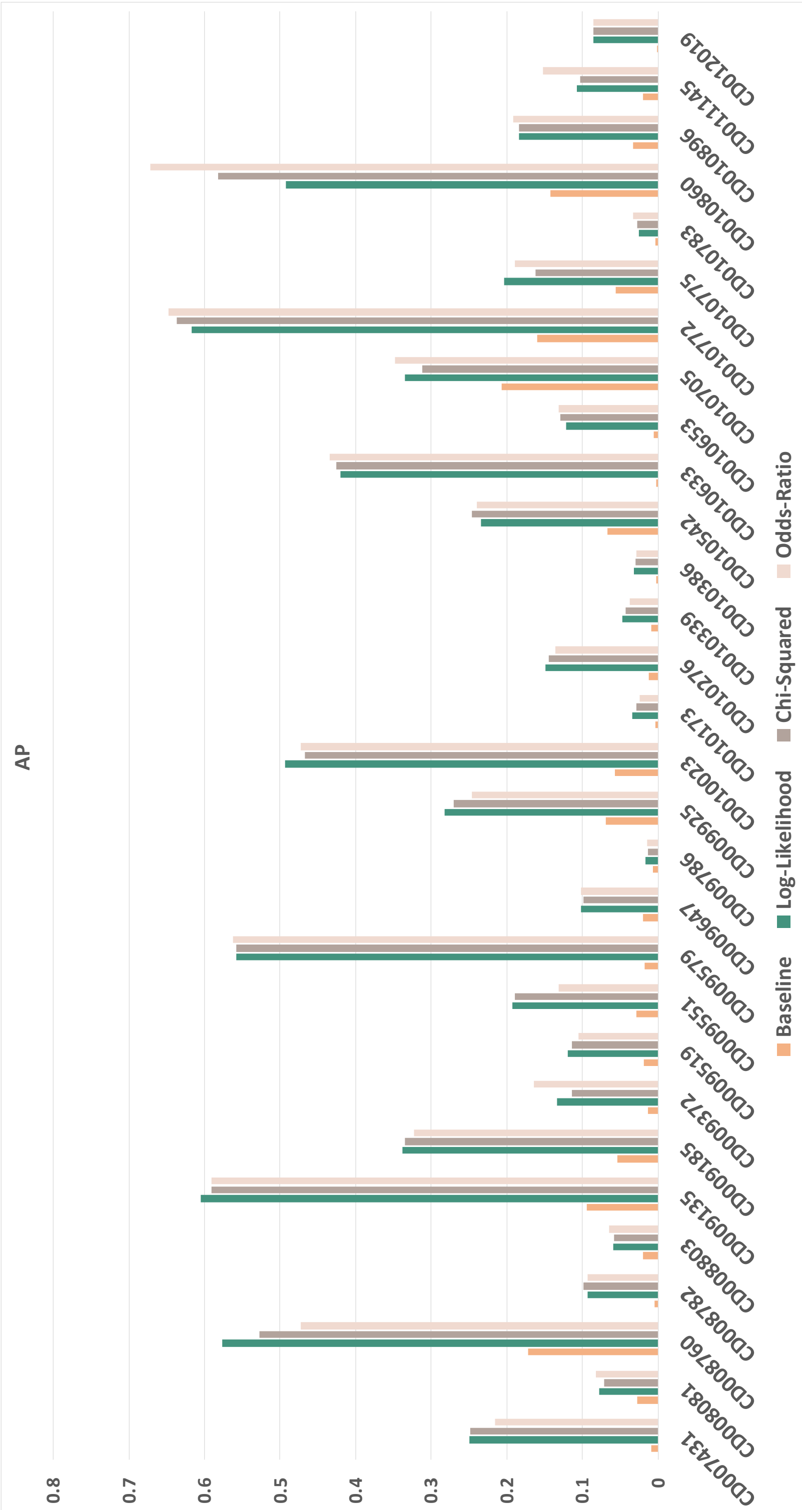

Figure 3.8: AP scores for each review in CLEF2017 test dataset using baseline and the three lexical statistics.

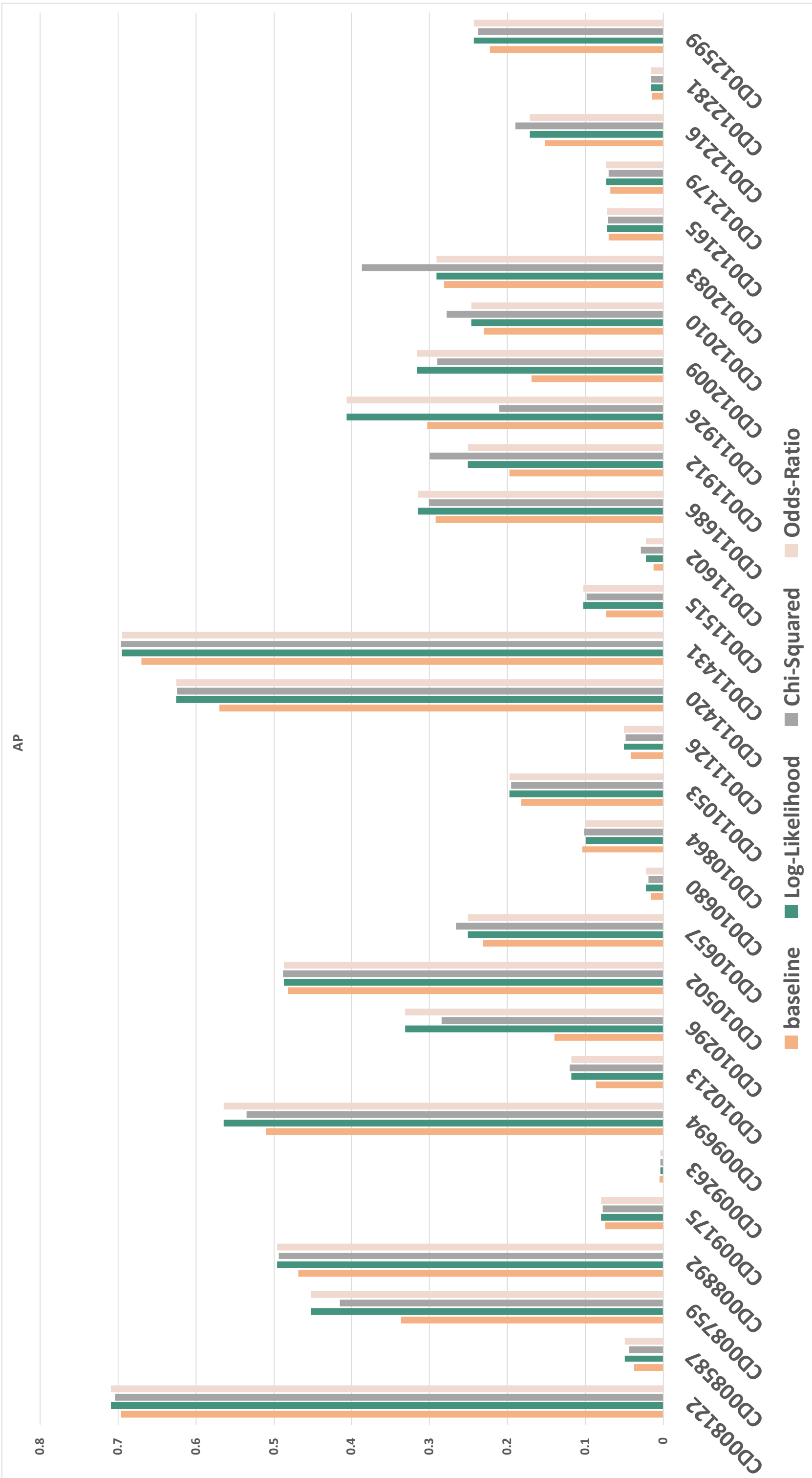

Figure 3.9: AP scores for each review in CLEF2018 test dataset using baseline and the three lexical statistics.

Tables 3.6 and 3.7 show the twenty terms with the highest scores derived from the CLEF2017 and CLEF2018 training datasets, respectively. We noticed that the top terms identified by the lexical statistics include ones that are highly indicative of the subjects discussed in DTA reviews, for example "sensitivity", "predictive" and "positive" are terms which relate to accuracy of a medical test.

Table 3.6: Top 20 terms based on different lexical statistics scores derived from CLEF2017 training dataset.

	Log-Likelihood		Chi-Squared		Odds-Ratio	
	Term	**Score**	**Term**	**Score**	**Term**	**Score**
1	sensitivity	58.25	mtb rif	468.11	vulva	306.15
2	predictive	41.68	vulva	461.16	mtb rif	242.33
3	gonadotropin	38.56	inguinofemoral	333.33	Xpert	142.73
4	hcg	32.74	Xpert	197.38	inguinofemoral	34.05
5	false	31.09	groin	126.17	geneXpert	33.73
6	mtb rif	31.05	sensitivity	112.44	cepheid	29.44
7	positive	30.31	cepheid	101.52	siln	28.86
8	vulva	29.35	geneXpert	85.34	sentinel	18.78
9	protein	28.69	predictive	79.85	dcbe	17.06
10	fetoprotein	28.05	inguine	66.45	blunt	12.47
11	value	27.88	gonadotropin	62.65	groin	7.82
12	alpha-fetoprotein	27.60	false	56.79	sensitivity	7.69
13	negative	26.81	hcg	55.33	midline	5.74
14	detect	25.75	midline	55.19	neoplasm	5.73
15	alpha	25.05	negative	48.28	jelly	5.60
16	prospect	24.85	dye	45.29	predictive	5.56
17	subunit	24.80	fetoprotein	44.18	vulvectomy	5.22
18	MoM	24.10	prospect	43.63	trauma	5.04
19	Xpert	23.95	alpha-fetoprotein	43.50	prehospital	4.71
20	alpha-fetoproteins-analyse	23.37	positive	43.11	specificity	4.51

It is also interesting to note that several of the terms that appear in this list are also used in standard filters for DTA reviews that have been developed to support information professionals searching for relevant literature (White et al., 2001). For example: "sensitiv.mp." , "predictive value.mp." , "accurac.tw." are filters which are used to increase the sensitivity and specificity in retrieving DTA studies (Haynes and Wilczynski, 2004).

However, we also note that the list also includes terms that appear to be specific to particular DTA reviews (e.g. "gonadotropin"). The CLEF 2017 training dataset contains only 20 reviews and CLEF2018 contains 42 reviews (including a subset of CLEF2017 dataset), and if a particular term proves to be very important for a small set of reviews, then its overall score can be high enough for it to be included in this list. Comparing the lists of terms generated by the different lexical statistics, it can be clearly noted that there is a high similarity between the Log-Likelihood and Chi-Squared lists. The similarity between the two lists is 65% for the CLEF2017 dataset and 60% for the CLEF2018 dataset. This similarity was expected since both Log-Likelihood and Chi-Squared are probability statistics while Odds-Ratio is an effect size statistic (Manning and Schütze, 1999; Pojanapunya and Todd, 2018). Furthermore, comparison analyses by Chujo and Utiyama (2006) and Culpeper (2009) show that Log-Likelihood and Chi-Squared produce very similar rankings of keywords.

Table 3.7: Top 20 terms based on different lexical statistics scores derived from CLEF2018 training dataset.

	Log-Likelihood		**Chi-Squared**		**Odds-Ratio**	
	Term	**Score**	**Term**	**Score**	**Term**	**Score**
1	sensitivity	92.48	mtb rif	386.04	Xpert	217.30
2	predictive	59.88	vulva	306.95	mtb rif	195.29
3	gonadotropin	57.81	inguinofemoral	185.00	silng	46.08
4	protein	49.34	sensitivity	175.71	cepheid	24.13
5	hcg	48.63	Xpert	162.36	sentinel	23.69
6	false	47.76	predictive	113.87	inguinofem	23.52
7	positive	47.30	gonadotropin	94.75	geneXpert	23.35
8	value	43.73	cepheid	94.31	sensitivity	17.64
9	fetoprotein	43.10	false	86.35	dcbe	17.18
10	alpha-fetoprotein	42.29	hcg	83.15	diagnose	14.10
11	prospect	40.71	groin	82.91	blunt	13.86
12	detect	40.26	geneXpert	73.26	paty	13.45
13	negative	39.34	prospect	71.30	impair	12.25
14	alpha	38.97	protein	71.24	predictive	10.84
15	screening	38.12	negative	70.80	mild	10.41
16	blood	36.95	positive	68.59	specificity	8.72
17	MoM	36.29	fetoprotein	68.12	turbo	8.71
18	alpha-fetoproteins-analyse	36.10	alpha-fetoprotein	66.84	blind	8.40
19	beta	35.67	strip	64.65	female	8.08
20	subunit	35.33	MoM	64.60	value	8.03

Taken together, theses results indicate that lexical statistics can be used to identify terms characteristic of studies likely to be relevant for DTA reviews. Results demonstrate that enriching the query with additional Key terms, generated from an independent set of reviews, provide information about the types of studies that are likely to be relevant for DTA reviews, independently of their specific review. The experiments demonstrate that including general information about the type of publication that is likely to be of relevance for a systematic review can improve retrieval performance. The best performance is achieved using the Log-Likelihood statistic.

3.4 Approach 3: Relevance Feedback

This approach explores the use of relevance feedback to improve studies ranking for systematic review. Relevance feedback is widely applied to improve information retrieval performance and has proven to be a successful approach (Azad and Deepak, 2019; Ruthven and Lalmas, 2003). The process of adapting the query using relevance feedback operates as follows (Baeza-Yates and Ribeiro-Neto, 2011; Manning et al., 2008a). First, the user generates an initial query and submits it to the IR system. A set of documents is retrieved by the system based on the user query. Then, the user labels each document returned as relevant or non-relevant. After that, the query is adapted based on the relevance judgements provided by the user. Finally, the adapted query is used by the system to retrieve relevant documents. This process may be applied once or more times until the user is satisfied with the search results.

The query can be adapted by re-weighting query terms or by adding or removing terms to/from the query based on the relevance judgements. The well-known Rocchio's algorithm (Baeza-Yates and Ribeiro-Neto, 2011) is used to modify the Boolean query representation for the experiments described in this section.

The following subsections describe the application of the Rocchio's algorithm, explain the experiment conducted, and discuss the results obtained.

3.4.1 Rocchio's Algorithm

Rocchio's algorithm is used to reformulate a query by enriching it with additional terms weighted using information about the relevance of the documents it returned (Baeza-Yates and Ribeiro-Neto, 2011; Shobha and Rangaswamy, 2018). First, using relevance judgements (that may be provided by a user), the documents are partitioned into two sets: positive (relevant) set D_{rel} and negative (non-relevant) set D_{nonRel}. Then, the query and the documents are represented in a vector space model. After that, the centroid vector of each set is computed as:

$$\vec{\mu}(D) = \frac{1}{|D|} \sum_{\forall \vec{d}_i \in D} \vec{d}_i \tag{3.8}$$

where $|D|$ is the number of documents in the set D and $\vec{d}_i$ is a weighted term vector associated with document i. Based on Equation 3.8, the updated query is calculated as follows:

$$\vec{q}_m = \alpha \vec{q} + \beta \vec{\mu}(D_{rel}) - \gamma \vec{\mu}(D_{nonRel}) \tag{3.9}$$

$$\vec{q}_m = \alpha \vec{q} + \frac{\beta}{|D_{rel}|} \sum_{\forall \vec{d}_i \in D_{rel}} \vec{d}_i - \frac{\gamma}{|D_{nonRel}|} \sum_{\forall \vec{d}_i \in D_{nonRel}} \vec{d}_i \tag{3.10}$$

where $\vec{q_m}$ is the modified query vector and $\vec{q}$ is the original query vector. D_{rel} is the set of relevant documents among the documents retrieved and $|D_{rel}|$ is the number of documents in D_{rel}. D_{nonRel} is the set of non-relevant documents among the documents retrieved and $|D_{nonRel}|$ is the number of documents in D_{nonRel}. α, β and γ are weighting parameters. α specifies the importance of the initial query $\vec{q}$. On the other hand, the higher the value of β, the more $\vec{q}$ moves toward the centroid of the relevant documents. The higher the value of γ, the more $\vec{q}$ moves away from the centroid of the non-relevant documents.

3.4.2 Experiments

Baseline

As for the lexical statistics approach, the best method from Section 3.2 was used as the baseline system. Studies were ranked by comparing each study against the review title and terms extracted from the Boolean query using the cosine similarity measure (Equation 3.1).

Relevance Feedback

In this approach, studies were ranked using a simple `tf.idf` weighted cosine similarity measure comparing each study with terms extracted from the Boolean query. After that, relevance judgements from the 10% top-ranked studies (up to a maximum of 1,000) were divided into relevant and non-relevant sets. The centroid of each set was then calculated using Equation 3.8. The query was reformulated using Rocchio's algorithm (Equation 3.10). The remaining studies (90%) were re-ranked using the updated query vector $\vec{q}_m$ (i.e. each study was compared with the terms extracted from the modified query).

In most IR system that use Rocchio's algorithm they set $\beta > \gamma$, where the positive feedback is more valuable than the negative feedback (Manning et al., 2008b). According to Baeza-Yates and Ribeiro-Neto (2011); Manning et al. (2008b), reasonable values might be $\alpha = 1$, $\beta = 0.75$, and $\gamma = 0.25$. However, in systematic reviews, there are very few positive (relevant) documents compared to the negative (non-relevant) documents. Thus, a range of values for the weighting parameters β and γ were explored by conducting experiments on the training dataset[9]. Table 3.8 shows the performance of the Rocchio using 64 combinations of the weighting parameters β and γ on the training dataset. As can be seen from the table, the performance of the approach is better when the value of γ is slightly greater than β (i.e. giving greater weight for the negative (non-relevant) documents). From the experiments, it was found that the best results were achieved by setting $\beta = 1$ and $\gamma = 1.5$[10].

[9]The values selected from the set: 0.25, 0.5, 075, 1, 1.25, 1.5, 1.75, 2.

[10]The value of α was set to 1 as proposed by Rocchio(Baeza-Yates and Ribeiro-Neto, 2011)

Table 3.8: MAP scores over a range of values for the weighting parameters β and γ using the training dataset of CLEF2018.

β \ γ	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**
0.25	0.214	0.215	0.212	0.211	0.210	0.210	0.209	0.209
0.5	0.212	0.214	0.216	0.216	0.213	0.212	0.211	0.210
0.75	0.211	0.213	0.216	0.216	0.216	0.214	0.213	0.212
1	0.210	0.212	0.214	0.214	0.217	**0.218**	0.216	0.214
1.25	0.210	0.211	0.212	0.215	0.215	0.216	0.216	0.215
1.5	0.209	0.211	0.212	0.213	0.215	0.215	0.216	0.216
1.75	0.209	0.210	0.211	0.212	0.214	0.215	0.215	0.216
2	0.209	0.210	0.211	0.212	0.213	0.214	0.215	0.214

3.4.3 Results and Discussion

Results are shown in Table 3.9. Retrieval performance improved for all metrics when using relevance feedback compared with the baseline. The MAP improved by 2.5% and 1.4% for CLEF2017 and CLEF2018, respectively. The WSS@100 and WSS@95 improved by 4.7% and 4.3% for CLEF2017 and by 6.4% and 10.2% for CLEF2018. The results also demonstrate that this approach outperforms the lexical statistics approach (see Table 3.5). On the other hand, a higher MAP score for the CLEF2018 dataset is obtained using lexical statistics.

Table 3.9: Relevance Feedback results for the CLEF2017 and CLEF2018 test datasets.

	(a) CLEF2017 Dataset			**(b) CLEF2018 Dataset**		
Approach	**MAP**	**WSS@100**	**WSS@95**	**MAP**	**WSS@100**	**WSS@95**
Baseline	0.218	38.50%	49.30%	0.224	37.70%	50.60%
Relevance Feedback	**0.243**	**43.20%**	**55.70%**	**0.238**	**42.00%**	**60.80%**

Figures 3.10 and 3.11 show the AP results for each review using relevance feedback approach for both CLEF2017 and CLEF2018 datasets. Applying relevance feedback approach improved AP for 80% of CLEF2017 and 90% of CLEF2018 reviews.

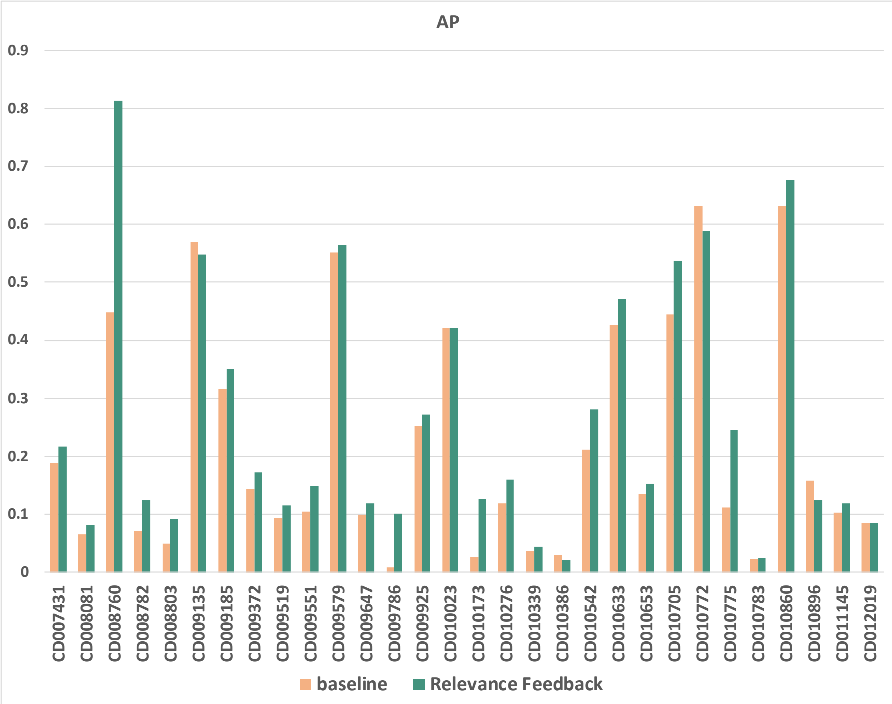

Figure 3.10: AP scores for each review in the CLEF2017 test dataset using baseline and Relevance Feedback.

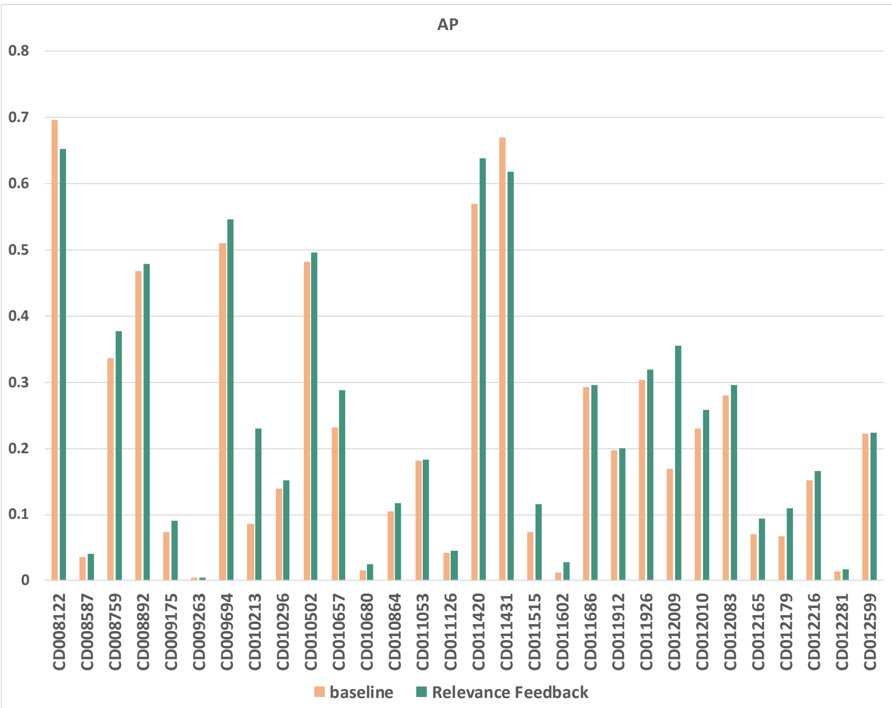

Figure 3.11: AP scores for each review in the CLEF2018 test dataset using baseline and Relevance Feedback.

3.5 Summary

This chapter explored the use of different query adaptation approaches to improve studies ranking for the screening stage of systematic review creation. First, it investigated the use of information available from Boolean query and studies retrieved. Experiments showed that using the terms extracted from the Boolean query is more beneficial than using MeSH terms either from query or studies. The query term approach was the best approach for the CLEF2017 task with no relevance feedback and it was later used by Lee and Sun (2018) as their baseline system.

Second, this chapter investigated the application of lexical statistics techniques in the domain of systematic reviews. It explored the use of lexical statistics to identify terms that characterise DTA reviews. Experiments showed that this approach improves the retrieval performance and reduces workload. The best performance was achieved by adding 20 terms when using Log-Likelihood. Some of the terms generated are used to describe DTA reviews.

Finally, this chapter applied the Rocchio's algorithm in the domain of systematic review and results showed that this approach is useful for improving retrieval performance.

In summary, the results in this chapter provide a further demonstration of the benefits of ranking to reduce the workload required from experts when conducting systematic reviews. We investigated the use of terms and MeSH terms extracted from the Boolean query. As future work, we will consider exploring possible performance improvements by incorporating additional information such as publication types, citation counts or h-index of authors.

Chapter 4

A Dataset of Systematic Review Updates

4.1 Introduction

As discussed in Chapter 2, a significant number of previous studies have demonstrated the usefulness of NLP/IR techniques to reduce the workload involved in the systematic review screening process for new reviews. Updating systematic reviews is a significant problem but one which has largely been overlooked. Developing methods to support the updating of reviews is therefore required to reduce the workload required and thereby ensure that reviews remain up to date. However, only a few previous studies have explored the use of NLP/IR techniques to support the problem of updating reviews (see Sections 2.4.1 and 2.4.4). A possible explanation is the lack of available datasets that can be used to evaluate such techniques. In the majority of cases, this work has been evaluated against simulations of the update process (see Section 2.4.4).

As shown in Chapter 2, no accessible dataset focuses on the problem of identifying studies for inclusion in a review update. The problem is subtly different from the identification of studies for inclusion in a new review because relevance judgements are available (from the original review) which have the potential to improve performance. A suitable dataset for this problem would include the list of studies considered for inclusion in both the original and updated reviews, together with a list of the studies that were actually involved in each review. In response, this chapter provides a valuable dataset with the aim of evaluating automated methods applied to the problem of identifying relevant evidence for updating systematic reviews. This is the first resource made available for this purpose.

This chapter describes the process of constructing the update dataset, the criteria of selecting the reviews and the characteristics of the dataset. In addition, this chapter explores the use of two approaches from the previous chapter (i.e. lexical statistics and relevance feedback) to improve studies ranking for systematic review updates.

4.2 Dataset Configuration

The dataset is constructed using systematic reviews from the Cochrane Database of Systematic Reviews[1], a standard source of evidence to inform healthcare decision-making. Intervention reviews - that are, reviews which assess the effectiveness of a particular healthcare intervention for a disease (see Section 1.1.1) - are the most common type of reviews carried out by Cochrane. All the reviews selected for the dataset are published intervention systematic reviews as these are the most popular reviews in Cochrane library.

Several criteria were taken into consideration when selecting the reviews to be included in the dataset. One significant aspect is that Cochrane reviews may be withdrawn from the library for different reasons. However, only a version of the review may be withdrawn, not the overall review. Review withdrawal may occur due to a severe error in the review, which may result in harm to patients or populations. Review versions may also be withdrawn when included studies are removed from publication (i.e. the article is no longer available), which may lead to an error in the review analysis and conclusion (Harriet MacLehose, 2018). Reviews included in the dataset must have been available in an original and updated version (i.e. an updated version of the review has been published) and not withdrawn from the Cochrane library.

In addition, to allow meaningful experiments to be conducted, reviews included in the dataset were restricted to ones for which at least one relevant article identified during the abstract screening stage for the update.

[1]https://www.cochranelibrary.com/cdsr/about-cdsr.

Moreover, the reviews included in the dataset should contain a forest plot. The forest plot diagram represents the results of the systematic review graphically and shows the findings of individual studies that address the same issue (Lewis and Clarke, 2001). Figure 4.1 shows a forest plot diagram from review CD002733 entitled: "*Influenza vaccine for patients with chronic obstructive pulmonary disease*". The diagram illustrates a summary of the findings of two studies (listed on the leftmost column). These studies evaluated the impact of using influenza vaccinations in people with chronic obstructive pulmonary disease and the ability to reduce illness and death. The left side of the vertical line shows the studies that favoured the vaccine, and the right side shows the studies that favoured the placebo. In this example, both studies favoured the vaccine (treatment).

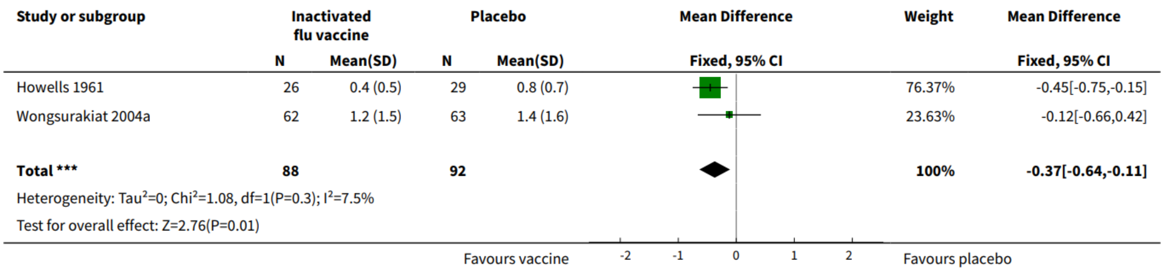

Figure 4.1: Forest plot diagram from review CD002733 (Kopsaftis et al., 2018)

Reviews with forest plots were selected in order to benefit from the summary statistical information of the included studies presented in the forest plot. This information can help determine whether a new included article will change the conclusion of the systematic review update. However, this requirement added an additional restriction on reviews considered for inclusion in the dataset.

These restrictions limited the number of suitable reviews that could be identified for inclusion in the dataset. In the end, a set of 25 published intervention systematic reviews which satisfied the criteria were selected for inclusion in the dataset.

A Python script was developed which applied the process of constructing the dataset automatically and extracted the following information from each review: (1) review title, (2) Boolean query, (3) set of included and excluded studies (for both the original and updated versions), (4) update history (including publication date and URL of original and updated versions) and (5) MeSH keywords.

The process is now described in more detail.

4.2.1 Boolean Query

Boolean queries in the reviews included in the dataset are created for either the OVID or PubMed interfaces to the MEDLINE database of medical literature. For ease of processing, each OVID query was automatically converted to a single-line PubMed query using a Python script created specifically for this purpose (see Figure 4.2). The script first reads the query line by line. For each line (clause), the script translates all the restriction fields form the OVID format to the PubMed format, as shown in Table 4.1. However, the OVID restriction field $.ab.$ and the adjacency ADJ are not supported by PubMed; therefore, the closest equivalents were used, which are $[tiab]$ and AND, respectively. After all the restriction fields in all clauses have been transformed, the script converts the query to single-line query.

(a) Multi-line query in OVID format

```
1.  endometriosis/
2.  (adenomyosis OR endometrio$).tw.
3.  OR/1-2
```

(b) One-line PubMed translation

```
endometriosis[Mesh:NoExp] OR adenomyosis[Text Word] OR
endometrio*[Text Word]
```

Figure 4.2: Example portion of a Boolean query (Hughes et al., 2007) in (a) the OVID format and (b) its translation into the single-line PubMed format. This portion of the query contains three clauses, and the last clause represents the combining results of clause 1 and 2 in a disjunction (OR).

Table 4.1: OVID restriction fields and their equivalent in PubMed format.

Name	OVID	PubMed
MeSH Term	**exp** *Mesh term* **/**	*"Mesh term"*[**Mesh**]
MeSH Term	**exp** *Mesh term* **.mp.**	*"Mesh term"*[**Mesh**]
MeSH Subheading	**exp** *Mesh term* **.sh.**	*"Mesh term"*[**sh**]
Unexploded MeSH Term	*Mesh term* **/**	*"Mesh term"*[**Mesh:NoExp**]
Major MeSH Term	**exp** **Mesh term* **/**	*"Mesh term"*[**Majr**]
All Fields	*term***.af.**	*term*[**All Fields**] or *term*[**All**]
Text Word	*term***.tw.**	*term*[**Text Word**] or *term*[**tw**]
Title	*term***.ti.**	*term*[**ti**] or *term*[**Title**]
Title/Abstract	*term***.ti,ab.**	*term*[**tiab**] or *term*[**Title/Abstract**]
Publication Date	*date***.dp.**	*date*[**dp**] or *date*[**Date - Publication**]

During the construction process, some of the OVID queries were found to contain numbering errors. For example, the Boolean query for review CD004679 (see Figure 4.3) includes only the combination of lines 5 to 8 and ignores the first four lines. That leads to dropping part of the query when running it in the search engine. Therefore, for accuracy, only systematic reviews that have correctly numbered queries were considered for inclusion in the final dataset. A Python script was created specifically for this purpose.

```
1.  exp Peritoneal Dialysis/
2.  peritoneal dialysis.tw.
3.  (PD OR CAPD OR CCPD OR APD).tw.
4.  OR/1-3
5.  Peritonitis/
6.  peritonitis.tw.
7.  Catheter-Related Infections/
8.  infection*.tw.
9.  OR/5-8
```

Figure 4.3: Example of a Boolean query (Campbell and Strippoli, 2017) which has a mistake in the lines numbers: the last line (no. 9) combines the results of lines 5 to 8 and ignores the first four lines of the query.

4.2.2 Included and Excluded Studies

For each version of the reviews (original and updated), the dataset includes a list of all the studies that were included after each stage of the screening process (abstract and content). The set of studies included after the content level screening is a subset of those included after abstract screening and represents the studies included in the updated review.

Included and excluded studies are listed in the dataset as PMIDs that make it straightforward to access details about the publication. If the PMID for an article was listed in the systematic review in the Cochrane library (which accounted for a majority of cases), then it was extracted and used for the dataset. Figure 4.4 shows an example of included studies from review CD000523 (Handoll and Pearce, 2012); as can be seen, the information contains the PMIDs for these particular studies.

Atkin DM, Bohay DR, Slabaugh P, Smith BW. Treatment of ulnar shaft fractures: A prospective, randomised study. Orthopedics 1995;18(6):543-7. [MEDLINE: 1995406131]

Gebuhr P, Holmich P, Orsnes T, Soelberg M, Krasheninnikoff M, Kjersgaard AG. Isolated ulnar shaft fractures: Comparison of treatment by a functional brace and long-arm cast. Journal of Bone and Joint Surgery - British Volume 1992;74(5):757-9. [MEDLINE: 1992406976]

Figure 4.4: Example of studies with available PMID (*highlighted*).

On the other hand, some of the studies in the Cochrane library do not include PMIDs. In this case, there are two possibilities: the article is not indexed by MEDLINE, or the article is indexed by MEDLINE, but the PMID is not provided in the Cochrane library. When the PMID was missing, then the title of the article and year of publication were extracted for use in forming a query that was used to search PubMed[2] (see Figure 4.5). However, the search usually retrieves either just one record or no records at all. If the entire text of the title, publication year and volume of the retrieved record match the details listed in the Cochrane library, then the PMID of that article is used. Figure 4.6 shows the

[2]https://www.ncbi.nlm.nih.gov/pubmed/.

single record retrieved from running the query from Figure 4.5 in PubMed. As shown, the information of this record matches the information listed in the Cochrane library (see Figure 4.6). Therefore, this PMID was added to the dataset.

Article as it appears in the Cochrane Library:

Claesson B, Bergquist C. Clinical experience treating endometriosis with nafarelin. The Journal of Reproductive Medicine 1989;34 Suppl(12):1025-8.

Article Title: Clinical experience treating endometriosis with nafarelin.

Publication Year: 1989

Search Query:

```
clinical[Title] AND experience[Title] AND treating[Title]
AND endometriosis[Title] AND nafarelin[Title] AND 1989[Date -
Publication]
```

Figure 4.5: Example of search query generated from title and publication year for an article without a PMID.

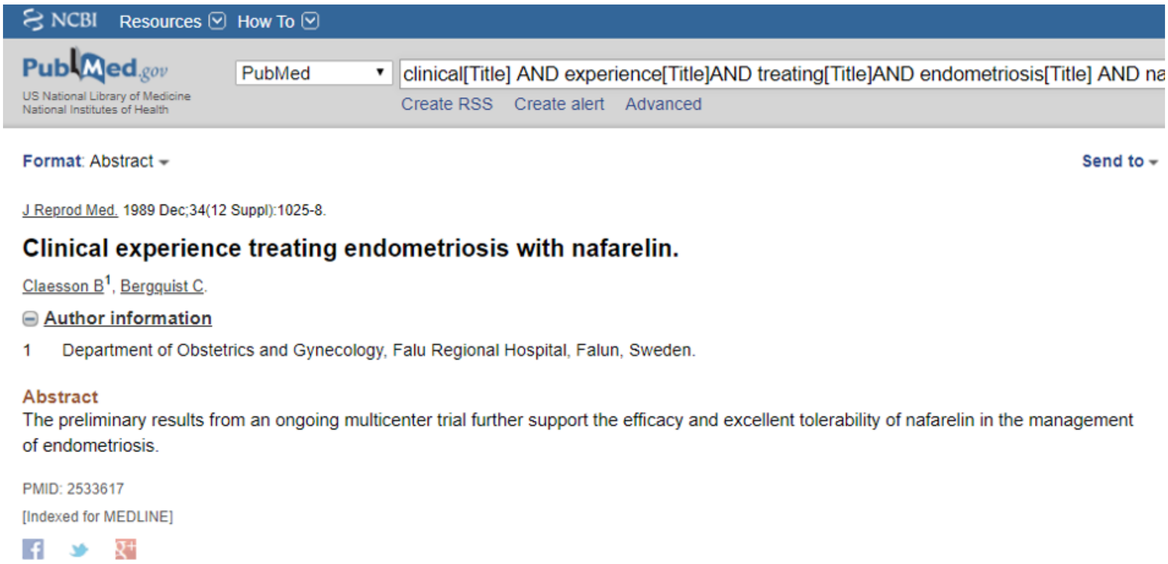

Figure 4.6: The result of searching PubMed using the query in Figure 4.5.

The search was restricted for an exact match of the title, publication year and volume to avoid including wrong studies in the dataset. However, if the search retrieved no records, this indicate that either the article was not indexed by MEDLINE or that there was a misspelled term in the title which led the exact matching to fail. In this case, a maximum edit distance of three was used and the retrieved records were manually examined. Given the two strings $S1$ and $S2$, the edit distance $d(S1, S2)$ is the minimum number of edit operations needed to transfer $S1$ into $S2$ (Navarro, 2001). For example, given the article's title on PubMed $S1$: "*The* ***effectivenss*** *of danazol on subsequent fertility in minimal endometriosis*" and the article's title in the Cochrane library $S2$: "*The* ***effectiveness*** *of danazol on subsequent fertility in minimal endometriosis*", then the edit distance $d(S1, S2) = 1$ (i.e. one operation is needed to transform $S1$ into $S2$ by inserting e into the term ***effectiveness)***.

This mapping process was performed using a Python script, and the record was retrieved using the `Entrez` package from Biopython (`biopython.org`).

To evaluate the mapping process, five systematic reviews were randomly selected, and the PMIDs of the included studies were manually examined. In total, 120 studies were available in the Cochrane library for these reviews. However, 43.33% of these studies did not have PMIDs. On the other hand, 29% of studies without PMIDs were not indexed by MEDLINE (as a result, they were not added to the dataset). The remaining studies, which represent 71% of the total studies, were indexed by MEDLINE. After the mapping process was completed, 92% of these studies were retrieved correctly without the need for using edit distance or manual examination. Only three studies needed manual examination.

4.2.3 Update History

Details of the date of publication of each version (original and update) were also extracted and included in the dataset. As an example, Figure 4.7 shows the version history for review CD000155 entitled: "*Ovulation suppression for endometriosis*". This review was first published in July 2003, then it was updated four years later (July 2007). This information can help to know which period was covered when conducting the search for the original review as well as the updated version of the review.

Version history

Title	Stage	Authors	Version	Publication Date
Ovulation suppression for endometriosis	Review	Edward Hughes, Julie Brown, John J Collins, Cindy Farquhar, Donna M Fedorkow, Patrick Vanderkerchove	https://doi.org/10.1002/14651858.CD000155.pub2	18 July 2007
Ovulation suppression for endometriosis	Review	Edward Hughes, Donna M Fedorkow, John Collins, Patrick Vandekerckhove	https://doi.org/10.1002/14651858.CD000155	21 July 2003

Figure 4.7: An example of version history information available with Cochrane review (Hughes et al., 2007).

4.3 Dataset Characteristics

Descriptive statistics for the 25 systematic reviews that form the dataset are shown in Table 4.2. It is worth drawing attention to the small number of studies included after the initial abstract screening stage. From a range of 1 to 46 studies, the average number of included studies for the update based on *abstract screening* is 7. On the other hand, the average number of included studies based on *content screening* is 3 from a range of 0 to 13 studies. Note that for the updated review, the number of included studies in the table lists only the new studies that were added during the update process.

The total number of studies retrieved from the search for the original reviews ranged from 36 to 41,675. On the other hand, for updated reviews, the number of studies retrieved ranged from 9 to 6,720. Furthermore, 88% of the reviews used the PubMed format Boolean query and the remaining 12% were in OVID format. The length of the Boolean query varied between 5 and 52 lines.

To make the dataset reusable by other researchers, it was published in two formats: text files and a `pickle` file. In the text file format, the files include the following information for each review: (1) review title, (2) Boolean query, (3) list of PMIDs for studies included in the original review, (4) list of PMIDs for studies included in the updated review, (5) publication date of the original and updated review and (6) list of MeSH keywords associated with the review.

Table 4.2: List of the 25 systematic reviews with the Boolean query type, the total number of studies returned by the query (Total) and the number included following the *Abstract* and *Content* screening stages. The average (unweighted mean) number of studies is shown in the bottom row. Note that for the updated review, the number of included studies in the table indicates only the new studies that were added during the update.

		Original Review			Updated Review		
Review	**Query Type**	**Total**	**Abstract**	**Content**	**Total**	**Abstract**	**Content**
CD000155	OVID	397	42	14	101	6	4
CD000160	OVID	433	7	6	1,980	1	1
CD000523	OVID	34	6	3	18	1	1
CD001298	OVID	1,384	22	15	1,020	17	13
CD001552	OVID	2,082	2	2	844	2	2
CD002064	OVID	38	2	2	9	1	0
CD002733	PubMed	13,778	30	10	6,109	6	6
CD004069	OVID	951	5	2	771	9	7
CD004214	OVID	57	5	2	21	4	1
CD004241	OVID	838	25	9	193	5	3
CD004479	OVID	112	6	1	153	4	3
CD005025	OVID	1,524	43	8	1,309	46	4
CD005055	OVID	648	8	4	353	3	0
CD005083	OVID	462	46	16	107	9	2
CD005128	OVID	25,873	5	4	5,820	9	3
CD005426	OVID	6,289	13	8	1,413	3	0
CD005607	PubMed	851	11	7	103	2	1
CD006839	OVID	239	8	6	93	3	3
CD006902	OVID	290	18	6	106	10	5
CD007020	OVID	348	47	4	47	4	3
CD007428	OVID	157	7	3	190	9	3
CD008127	PubMed	5,460	7	0	6,720	2	1
CD008392	OVID	5,548	15	5	1,095	2	0
CD010089	OVID	41,675	22	10	4,514	4	0
CD010847	OVID	571	15	1	111	6	0
Average		4,402	17	6	1,335	7	3

For the second format, the dataset was provided as a `pickle` file which is ready to use for programming. Figure 4.8 shows the structure of the dataset. In addition to the review title, query, MeSH keywords and list of included studies, the dataset contains beneficial information such as the abstract and the metadata (i.e. PMID, title and list of MeSH terms) for each record retrieved from the search. The list of terms and MeSH terms extracted from the Boolean query for each review were also included.

The dataset is available from https://github.com/Amal-Alharbi/Systematic_Reviews_Update.

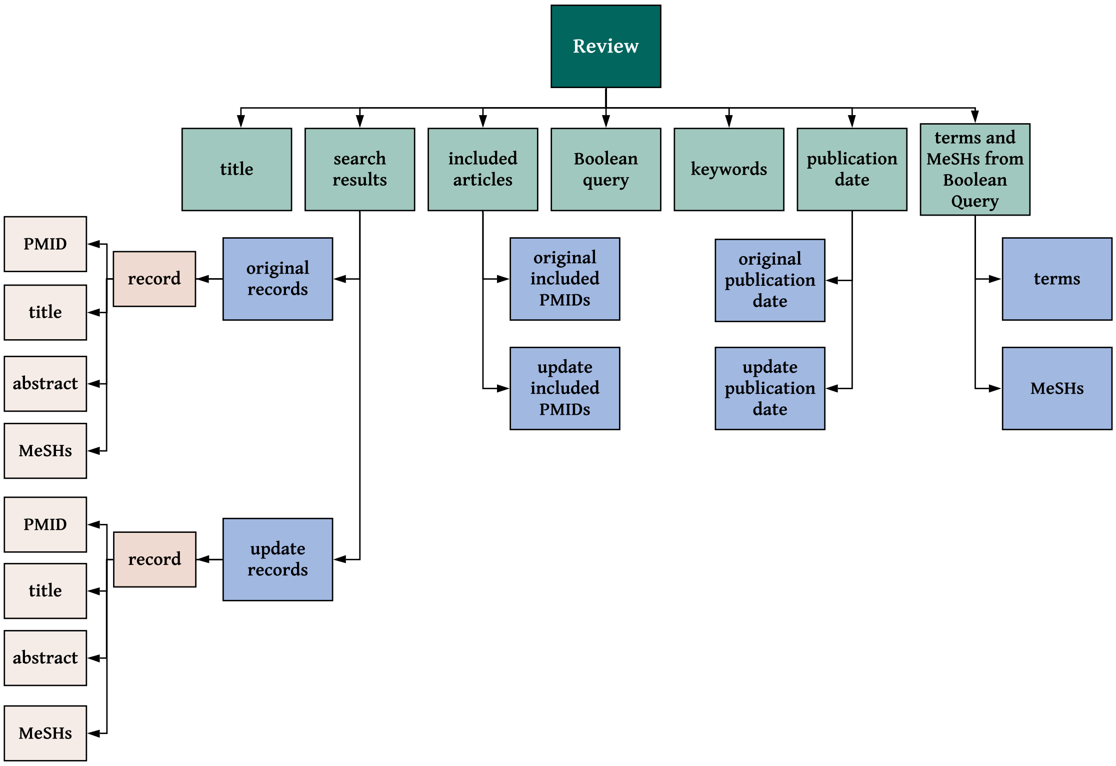

Figure 4.8: The structure of the update dataset.

4.4 Experiments

Experiments were conducted to establish baseline performance figures for the dataset. The aim is to reduce workload in the screening stage of the review update by ranking the list of studies retrieved by the Boolean query. Experiments were carried out to explore performance at both the abstract and content screening levels. The collection was created by using the Boolean query to search MEDLINE using the `Entrez` package from Biopython (biopython.org). The list of studies included after the abstract screening was used as the relevance judgements for abstract level evaluation and the list of studies included after the content screening was used for content level evaluation.

4.4.1 Approaches

Baseline

The best method from Section 3.2 in the previous chapter (i.e. using query terms) was applied for the baseline system. BM25 (Baeza-Yates and Ribeiro-Neto, 2011) was used to rank the set of studies returned from the Boolean query for the review update. BM25 has been widely used as a baseline model for text retrieval tasks in a significant number of previous and recent studies, for example, Hollmann and Eickhoff (2017b); Scells et al. (2020); Trotman and Lilly (2020); Zeng and Sakai (2019). Also, BM25 has been used as a baseline by CLEF organiser for CLEF2019/2020 eHealth task on systematic reviews which motivated us to apply it in our approach (Kanoulas et al., 2019; Suominen et al., 2020)

Pre-processing was applied to both the review title and terms extracted from the Boolean query. The text was tokenised[3] and converted to lower case, stop-words[4]/punctuation were removed and the remaining tokens were stemmed[5].

[3]The Natural Language Toolkit (NLTK) `tokenise` package was used for tokenisation.

[4]The PubMed stop-words list was used https://www.ncbi.nlm.nih.gov/books/NBK3827/table/pubmedhelp.T.stopwords/.

[5]The NLTK `LancasterStemmer` package was used for stemming.

Lexical Statistics

Section 3.3 showed that lexical statistics could help to identify terms that characterise a specific type of review which results in improved studies ranking. In the context of the systematic review update, we hypothesise that certain terms distinguish the studies that are likely to be included in reviews from other literature. Expanding the Boolean query with those terms may help to find the most relevant studies.

Lexical statistics were used to derive lists of terms that indicate evidence relevant to each review. For each review, the original version dataset was partitioned into relevant and non-relevant studies depending on whether the article was included in the systematic review. Three lexical statistics were used, which were introduced in Section 3.3: Log-Likelihood, Chi-Squared and Odds-Ratio.

For each lexical statistic, the top hundred terms with the highest scores were identified and added to the query for each review in the update portion of the dataset. The studies in the update dataset were ranked by matching terms from the expanded queries against those in the studies using a BM25.

Relevance Feedback

Relevance feedback was applied to exploit the information about which studies are suitable for inclusion from the original review. Rocchio's algorithm (see Section 3.4) was used to reformulate the baseline query by making use of relevance judgements derived from the original review.

Content screening judgements (included and excluded studies) were used for the majority of reviews. Abstract screening judgements were used if these were not available; i.e. no studies remained after content screening.

4.4.2 Results and Discussion

Results of the experiments are shown in Table 4.3. As expected, performance improved when lexical statistics or relevance feedback was used. Using lexical statistic outperformed the baseline for all metrics. Among the three lexical statistics, Chi-Squared achieved the best performance. Enriching the query by terms generated form Chi-Squared improved the MAP by 19% and reduced the workload by 67.50% to identify all relevant studies (100% recall) based on the abstract level and by 76.5% at the content level. On the other hand, using relevance feedback improved the MAP by 20%, and the screening effort required to identify all relevant studies (100% recall) was reduced by 63.5% at the abstract level and 74.9% at the content level. This demonstrates that making use of information from the original review can improve article selection for review updating. The best performance was achieved using Chi-Squared and relevance feedback.

Table 4.3: Performance ranking abstracts for updated reviews at (a) abstract and (b) content levels. Results are computed across all reviews at the abstract level (25 reviews) and only across reviews in which a new article was added in the updated version for the content level (19 reviews). Values in boldface denote the best result achieved among approaches. Superscript * and †in MAP indicate that the corresponding method significantly outperformed the Baseline with $p < 0.001$ and $p < 0.05$, respectively .

(a) abstract level (25 reviews)			
Approach	**MAP**	**WSS@100**	**WSS@95**
Baseline	0.213	56.60%	51.70%
Log-Likelihood	0.375*	66.00%	70.60%
Chi-Squared	0.404*	**67.50%**	**72.50%**
Odds-Ratio	0.329 †	65.20%	69.80%
Relevance Feedback	**0.413***	63.50%	58.80%
(b) content level (19 reviews)			
Approach	**MAP**	**WSS@100**	**WSS@95**
Baseline	0.260	70.50%	65.50%
Log-Likelihood	0.260	65.50%	70.50%
Chi-Squared	**0.426 †**	**76.50%**	**81.50%**
Odds-Ratio	0.364 †	72.80%	77.80%
Relevance Feedback	0.382 †	74.90%	69.90%

Table 4.4 shows the results obtained when different numbers of terms with the highest scores were added to each query using various statistics. The performance improved when more terms were added (e.g. compare adding a hundred terms against five terms). The MAP increased by 16.2% for Log-Likelihood, 19% for Chi-Squared and 11.6% for Odds-

Ratio. On the other hand, the screening effort required to identify all relevant studies reduced when more terms were added.

Table 4.4: Performance ranking abstracts for updated reviews by adding different numbers of top terms for each lexical statistic. Values in boldface denote the best result achieved by each lexical statistic, and the underlined values represent the best results among all three lexical statistics.

Lexical Statistics	No. of Terms	MAP	WSS@100	WSS@95
Log-Likelihood	**5**	0.296	62.00%	57.20%
	10	0.318	64.10%	59.30%
	20	0.356	67.50%	63.10%
	30	0.327	68.60%	64.20%
	50	0.336	**71.90%**	**67.30%**
	100	**0.375**	70.60%	66.00%
Chi-Squared	**5**	0.273	59.40%	54.60%
	10	0.280	59.60%	54.60%
	20	0.328	63.10%	58.20%
	30	0.335	64.80%	59.80%
	50	0.328	69.00%	64.10%
	100	<u>**0.404**</u>	<u>**72.50%**</u>	<u>**67.50%**</u>
Odds-Ratio	**5**	0.266	58.70%	53.90%
	10	0.278	58.90%	54.00%
	20	0.284	59.70%	54.80%
	30	0.300	61.40%	56.60%
	50	0.322	66.40%	61.60%
	100	**0.329**	**69.80%**	**65.20%**

Figure 4.9 shows the results of AP scores for all 25 reviews. Relevance feedback improved AP for 23 (92%) of the reviews. There were also four reviews where the AP score for relevance feedback was 1, indicating that the approach reduced work required by up to 99.9%. On the other hand, using Chi-Squared improved AP scores for 22 reviews (88%) with three reviews having an AP score of 1.

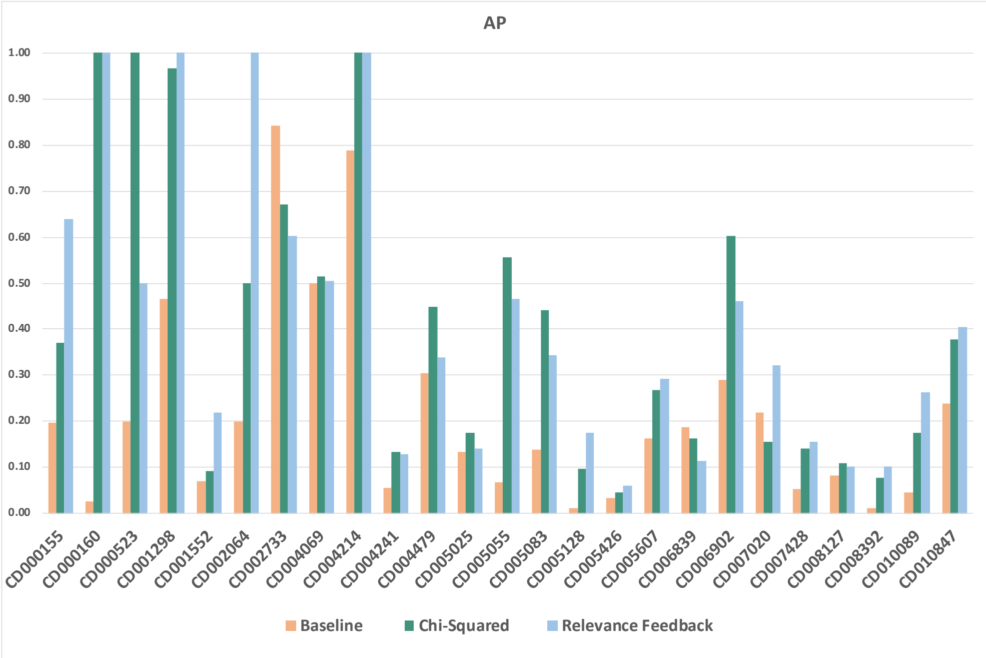

Figure 4.9: AP scores for each review using Baseline Query, Relevance Feedback and Chi-Squared.

Figure 4.10 shows the background, Boolean query and the top 100 scored terms extracted by applying Chi-Squared on the original dataset of review CD004214 entitled *"Transfer of preterm infants from incubator to open cot at lower versus higher body weight"*. The main objective of this review is to assess the effect of infants' body weight and temperature control of a system when moving infants from an incubator to an open cot. We noticed that the top terms generated by Chi-Squared include those about weight, for example "weight", "birth-weight", "body-weight", "fat" and "gm (gram)". In addition, the list includes several terms which describe temperature, such as "thermoregulatory", "unheated", "therm" and "stable". From the review's background (see Figure 4.10(a)), we can see that temperature is considered an essential criterion for this specific review. However, the "temperature" term is mentioned without any synonym in the original Boolean query (see Figure 4.9(b), line 6). Adding the top 100 terms to the query increased the AP for this particular review to 100%, which indicates that all the relevant studies were retrieved (see Figure 4.9).

Taken together, these results provide important insights into the value of information available from the original review and how this information can improve the retrieval performance for updating the review. The relevance judgements of the original review can help to find terms which characterise the studies related to a specific review. Extracting these terms using lexical statistics or using relevance feedback has a significant impact on improving the retrieval performance.

(a) Review Background

A key criterion for discharging preterm infants home from nurseries is their ability to maintain temperature once transferred from incubators to open cots. The timing of transfer is important given the preterm infant's immature thermoregulatory mechanisms.

(b) Boolean Query

1. infant, newborn
2. intensive care units, neonatal
3. 1 AND 2
4. cot OR crib OR isolette OR incubator
5. 3 AND 4
6. temperature OR body temperature OR skin temperature
7. 5 AND 6
8. 7 AND weaning

(c) Top 100 terms with highest lexical statistics Chi-Squared scores

additional, wean, crib, grow, four, project, unheated, subcutaneous, awhonn, some, accumbens, rear, gross, skin-fold, dramatically, fast, deposit, efficacy, pair, intention , strata, stratum, lost, likewise, substantial, outline, compos, headquarter, base, protocol, canada, detail, step-wise, apart, body-weight, confirm, adopt, practice, healthy, energy, similar, gain, prematurity, rat, discharge, implication, aware, therm, total, open, research, recruit, fat, cool, return, utilization, pilot, fourteen, abdominal, intake, evidence, entry, success, manus-crib, member, expert, occasion, discuss, subsequence, thermoregulatory, regard, principle, consecutive, sex, future, reach, safe, therm, ear, fact, state, provide, birth-weight, form, group, weight, eight, stable, six, base, number, equal, two, cohort, process, gm, point, drop, thirty, 24hour

Figure 4.10: Review Background (a), Boolean Query (b) and the top 100 terms with highest lexical statistics Chi-Squared scores (c) for review CD004214 (New et al., 2011).

4.5 Summary

This chapter described a dataset containing 25 intervention reviews from the Cochrane collaboration. This dataset was constructed to support in the development of approaches to automate the updating process. The title, Boolean query and relevance judgements for both the original and the updated versions are included for each systematic review. The dataset is publicly available and ready for use.

In addition, this chapter described experiments conducted to improve the ranking of studies for systematic review updates by using lexical statistics and relevance feedback techniques explained in Chapter 3. Results demonstrated that information from the original review could be used to improve article selection for systematic review updates. Comparing the results with the previous chapter, we found that the performance using Chi-Squared was the best in case of the update dataset, while in Chapter 3, the performance of Log-Likelihood was the best when using CLEF dataset. One important difference between the two experiments is that in the CLEF dataset, the top terms were derived from the training part which contains a variety of reviews. That means the generated list contains terms based on different reviews. While in the case of update dataset, top terms for a review were derived using the original version of the same review. Therefore, for the CLEF dataset, we explored to add up to 20 terms only, while for the update dataset, we examined to add up to 100 terms.

Chapter 5

Boolean Query Refinement to Improve the Identification of Relevant Studies

5.1 Introduction

The previous chapter described a dataset for systematic review updates. In addition, experiments showed the usefulness of using original review relevance judgements to improve ranking studies for review updates. To further improve the retrieval performance, this chapter explores the use of query refinements and their ability to generate improved Boolean queries to retrieve studies for review updates with the aim of reducing the workload of researchers when conducting review updates. As we have seen in Section 2.4.1, previous work on the refinement of Boolean queries for systematic reviews (Scells and Zuccon, 2018) demonstrated that it is possible to improve the Boolean query used for an original review. However, they did not explore the refinement of queries for review updates. In addition, previous work on the refinement and generation of Boolean queries for other types of professional searches, such as prior art search has been discussed in Section 3.1.

This chapter proposes an algorithm that aims to improve the identification of relevant studies for a systematic review update by automatically adapting the Boolean query using information produced during the screening stage of the original review. An iterative algorithm is proposed to generate query variants by applying a set of transformations including operator substitution, query expansion and query reduction. These are assessed using information about which studies were included in the original review and the most effective transformation is chosen to update the query. The best query produced by the algorithm will be used to retrieve studies for the review update.

5.2 Method

The proposed approach is outlined in Algorithm 1. It starts with the Boolean query used for the original review. As described in Section 3.2.1, the Boolean queries used in systematic review are often complex, consist of multiple lines and include advanced operators. A set of transformed queries is generated by applying a range of transformations (e.g. operator substitution, query expansion and query reduction) to the original query. Each transformed query is then evaluated using the relevance judgements produced for the original review and the best transformation is selected. The process is then repeated by applying the transformations to the newly selected query and evaluating the transformed queries produced. The process continues until the best transformed query is no better than the query from the previous iteration (i.e. the query cannot be improved using this process).

This approach can be considered as an example of Transformation-Based Learning (TBL). TBL is an automatic machine learning technique which has been applied for many linguistics tasks such as part-of-speech tagging (Brill, 1992; Corston-Oliver and Gamon, 2003). The fundamental idea behind TBL is to begin with some simple solution to the problem (in our approach: start with the original Boolean query), and apply transformations (in our approach: three types of transformations including operator substitution, query expansion and query reduction) - at each step, the transformation which results in the largest benefit is selected and applied to the problem (in our approach: the transformed query that produces the highest score is chosen for the next iteration). The algorithm stops when the selected transformation does not modify the data in enough places, or there are no more transformations to be selected (in our approach: stop when the query cannot be further improved) (Brill, 1992, 1995; Ngai and Florian, 2001).

Below, the individual steps of our proposed approach are described in further detail.

5.2.1 Step One: Boolean Query Transformation

In the first step, the algorithm applies a set of query transformations to generate new queries from the current one. Three types of transformation are proposed.

(a) **Operator Substitution.**

This transformation replaces one query operator with another. For example, disjunction with conjunction:

`(blind$ OR mask$).ti.` → `(blind$ AND mask$).ti.`

or alters a restriction field:

`(blind$ OR mask$).ti.` → `(blind$ OR mask$).ti,ab.`

In the second example, `.ti,ab.` indicates that the terms are searched in both the title and the abstract, rather than just in the title.

A set of useful operator substitution transformations was identified during preliminary experiments: `.tw.`→`.ti.`, `.tw.`→`.ti,ab.`, `.ti,ab.`→`.ti.`, `.ti,ab.`→`.tw.`, `.ti.`→`.tw.`, `.ti.`→`.ti,ab.`, `.ab.`→`.ti,ab.`, `.ab.`→`.ti.`, `.sh.`→`*`, `AND`→`OR` and `OR`→`AND` (See Table 3.1 for the meanings of these OVID query operators and field restrictions). Some of these transformations were used in previous work (Scells and Zuccon, 2018): logical operator replacement (`AND`→`OR` and `OR`→`AND`) and four field restrictions (`.ti,ab.`→`.ti.`, `.ti.`→`.ti,ab.`, `.ab.`→`.ti,ab.` and `.ab.`→`.ti.`). The remaining transformations were developed for this research. Additional transformation types were also explored but not found to improve performance, including three field restriction transformations: `.af.`→`.ti,ab.`, `.af.`→ `.ti.` and `.af.`→ `.tw.`.

(b) **Query Expansion.**

This transformation adds new elements to the query. Lexical statistics are used to

Algorithm 1: Automatic improvement of Boolean query

```
Input  : Boolean query from original review (q), set of query transformations (T)
         and original review's relevance judgements (R_orig)
Output : Updated Boolean query (q*)
q* ← q
while True do
    // Step one: Boolean Query Transformation
    // Generate set of updated queries by applying all possible
    // transformations
    Q̂ ← {}
    for t in T do
        for clause c in q* do
            if t can be applied to c then
                Q̂ ← Q̂ ∪ t(q*_c) // where t(q*_c) denotes transformation t
                // applied to clause c of q*
            end
        end
    end
    // Step two: Boolean Query Selection
    // Evaluate each transformed query and select the highest
    // scoring for the next iteration
    for q̂ in Q̂ do
        Compute f(q̂|R_orig) // Where f is some scoring function based
        // on R_orig
    end
    q' = argmax_{q̂∈Q̂} f(q̂|R_orig)
    // if performance of the best new query is the same as the
       base
    // query then the query cannot be improved
    if f(q'|R_orig) ≤ f(q*|R_orig) then
        break
    end
    q* ← q'
end
return q*
```

identify terms that discriminate relevant studies and these are added to the query. Log-likelihood statistic, which achieved the best result among the lexical statistics defined in Section 3.3.1, is applied to the set of studies retrieved for the original review (partitioned into relevant and non-relevant sub-corpora) and the score for each term is computed using Equation 3.3.

Log-likelihood scores are used to identify the five terms that are most closely associated with the relevant studies. Only the top five terms were selected to make the number of transformations produced more manageable. These terms are then used to form a set of transformations in which the terms are added to a query clause using the logical `OR` operator and `.tw.` as the restriction field. Terms are either added individually or the top n, producing nine transformations of this type: add 1st term, add 2nd term, add 3rd term, add 4th term, add 5th term, add 1st and 2nd terms, add 1st to 3rd terms, add 1st to 4th terms and add all 5 terms.

For example, the terms `packaging, blister, pack, calendar` and `medication` are identified as the top five terms that identify relevant studies for the review CD005025 entitled *"Reminder packaging for improving adherence to self-administered long-term medications"* (Mahtani and Perera, 2011). Figure 5.1 shows the possible transformations that can be applied to the first clause of the Boolean query for this review.

(a) First clause from the Boolean query

```
Reminder Systems/
```

(b) Top five terms

```
packaging, blister, pack, calendar, medication
```

(c) Possible transformations

```
T1:  Reminder Systems/ OR packaging.tw.
T2:  Reminder Systems/ OR blister.tw.
T3:  Reminder Systems/ OR pack.tw.
T4:  Reminder Systems/ OR calendar.tw.
T5:  Reminder Systems/ OR calendar.tw.
T6:  Reminder Systems/ OR packaging.tw.  OR blister.tw.
T7:  Reminder Systems/ OR packaging.tw.  OR blister.tw.  OR pack.tw.
T8:  Reminder Systems/ OR packaging.tw.  OR blister.tw.  OR pack.tw.
OR calendar.tw.
T9:  Reminder Systems/ OR packaging.tw.  OR blister.tw.  OR pack.tw.
OR calendar.tw.  OR medication.tw.
```

Figure 5.1: Example of query expansion applied to the first clause of the Boolean query of review CD005025 (a) by adding up to five terms generated by Log-Likelihood (b) and the full list of transformations that can be added to the clause (c).

(c) **Query Reduction.**

The final transformation method reduces the query by deleting a clause from it. Figure 5.2 shows an example of query reduction by removing the second clause from the Boolean query of the review CD005025(Mahtani and Perera, 2011).

(a) Original query

```
1.  Reminder Systems/
2.  exp Patient Compliance/
3.  Treatment Refusal/
```

(b) Transformed query

```
1.  Reminder Systems/
2.  Treatment Refusal/
```

Figure 5.2: Example of query reduction for review CD005025 (Mahtani and Perera, 2011).

The transformed queries produced during each iteration differ from the query selected during the previous iteration by a single clause. A total of 21 transformation types are used, leading to up to 21 $\times c$ transformed queries being produced during each iteration (where c is the number of clauses in the query selected during the previous iteration). However, this is an upper bound value because not all transformation types are applicable to all clauses. For example, the operator substitution `.tw.`→`.ti,ab.` cannot be applied to a clause that does not contain the `.tw.` restriction field.

5.2.2 Step Two: Boolean Query Selection

The set of transformed queries generated during step one is evaluated by assessing the queries against the relevance judgements produced for the original review. Each transformed query is run against MEDLINE and the list of studies it retrieves is returned. The query is then assessed using the following function which favours improvements in recall over improvements in precision:

$$f(\hat{q}) = recall(\hat{q}|R_{orig}) \times 100 + precision(\hat{q}|R_{orig}) \tag{5.1}$$

where $\hat{q}$ is the transformed query and R_{orig} the relevance judgements from the original review. Recall and precision are calculated as follows:

$$recall(\hat{q}|R_{orig}) = \frac{\text{Number of relevant studies in } R_{orig} \text{ retrieved by } \hat{q}}{\text{Total number of relevant studies in } R_{orig}} \tag{5.2}$$

$$precision(\hat{q}|R_{orig}) = \frac{\text{Number of relevant studies in } R_{orig} \text{ retrieved by } \hat{q}}{\text{Total number of studies retrieved by } \hat{q}} \tag{5.3}$$

As can be seen from Equation 5.1, the objective function always assigns a higher score to a query that produces an improvement in recall compared to one that improves precision. This is due to the nature of the search problem in systematic reviews where high recall is important since the goal is to identify all potentially relevant studies. However, retrieving a large number of non-relevant studies increases the screening effort required by the reviewers and it is therefore beneficial to ensure that the precision of queries is as high as possible.

The transformed query that produces the highest score is then chosen for the next iteration. If there are multiple queries with the same highest score then one is chosen at random. If there is no difference between performance of the highest scoring query and the query from the previous iteration then the algorithm stops.

5.3 Dataset

Experiments in this chapter were carried out using the intervention reviews from the update dataset (see Section 4.2). The reviews with an OVID-format query were selected (22 reviews). For each review, the majority of the included PMIDs were identified using the Boolean query but additional studies were often identified using alternative techniques such as hand searching key journals and examination of the lists of references of the included studies. The gold standard dataset includes all the relevant studies which are available on PubMed regardless of whether they were identified using the Boolean query or by other methods. Therefore, the query used for the review may not achieve full recall since it is possible it does not retrieve all studies included in the review or an update.

PMIDs included after abstract level screening were used for the experiments since the goal of this research is to develop queries that are applied to databases of scientific abstracts, such as PubMed, and for some reviews, only very few studies are included after content level screening.

5.4 Experiments

Experiments were carried out to explore performance of the method proposed in this chapter. Below, the three main approaches which were applied are described.

5.4.1 Approach 1: Baseline

A baseline approach was implemented which used the Boolean query from the original review to retrieve studies for the updated review without any transformation. The original Boolean query was run against MEDLINE and the set of studies that match the query retrieved. The aim of this approach was to assess performance when the query developed for the original review is re-used for the update, which is common practice within the systematic review community (Chandler and Cumpston, 2019).

5.4.2 Approach 2: Query Refinement

This approach employs the method proposed in Section 5.2. To assess the effectiveness of each transformation type defined in Section 5.2.1, four experiments were conducted: (1) using operator substitution, (2) using query expansion, (3) using query reduction and (4) using all the three transformation types defined in Section 5.2.1.

In this approach, for all the four experiments, the relevance judgements from the original review (i.e. information about the included/excluded studies) were used to select the best transformed query at each iteration, information which was readily available for complete systematic reviews since the results of the Boolean query are manually screened and reported in the review.

In each experiment, transformed queries were run against MEDLINE using the `Entrez` package from `biopython.org` to retrieve studies for the updated version of the review. Publication dates were used to identify studies published since the previous version of the review. To run the queries against MEDLINE, the OVID-format Boolean queries were converted to a single-line PubMed-format query as described in Section 4.2.1.

5.4.3 Approach 3: Oracle

An oracle approach was also implemented that is similar to the proposed method (see Section 5.2) with the exception that performance of the transformed query was assessed using the relevance judgements for the updated review (R_{update}) rather than for the original, i.e. using the following objective function:

$$f(\hat{q}) = recall(\hat{q}|R_{update}) \times 100 + precision(\hat{q}|R_{update}) \tag{5.4}$$

The oracle approach represents an unrealistic scenario since it has access to the relevance judgements for the updated review. However, it provides context for the results of the proposed method by placing an upper bound on the results that are possible by transforming queries for review updates.

5.5 Evaluation Metrics

For the evaluation, recall and precision which have been described in Section 2.4.3 were used. These are the most commonly used metrics in evaluating approaches for IR systems. However, since the aim is to develop improved queries which can be used to support review updating, approaches were evaluated using the set of studies included in the update as a gold standard list of relevant studies. This information was not available to the proposed approach, which only made use of the information about the studies considered for inclusion in the original review.

5.6 Results and Discussion

Results are shown in Table 5.1. Recall and precision scores are shown for each approach, both for each review individually and averaged across all reviews. Averages are weighted by the number of studies in each topic to place more weight on reviews where there are larger numbers of studies to be screened. The iteration of the algorithm that produced the final query is also shown for each method in the query refinement approach and the oracle approach. This information is not included for the baseline which is simply the unmodified query from the original review.

Considering average performance, all the four query refinement methods produce queries that improve upon those used for the original review (baseline) both in terms of recall and precision (except for query expansion method where the precision is lower than the baseline). The best performance was achieved by applying all transformation types. The increase in recall (10.3%) represents a marked increase in the number of relevant studies that are identified for review updates. Interestingly, the recall score achieved by this method (0.669) is close to the score achieved by the oracle approach (0.691) which represents the upper bound of the possible score. Although the precision of the queries produced by this approach is still low (0.7%), it is more than double the precision obtained using the original queries, thereby halving the set of studies that need to be considered during the expensive manual screening process. More generally, using queries produced by this approach led to increased recall for seven of the 22 reviews and the same recall for another 14. Recall reduced for a single review (CD007428), from 0.667 to 0.556. There were 9 relevant studies for this review and this change represented a single document having been missed. In addition, precision increased for 13 reviews without reducing recall.

Results of other query refinement methods indicate that using only one type of transformation generally produces queries that are more effective than the original query but the improvement is much smaller than using all types of transformations, indicating the importance of using different types of query transformations. Query expansion transformations are able to achieve recall almost as high as when all three transformation types are combined (an increase of 10% against the baseline), but at the expense of precision.

The reduction in precision caused by using query expansion leads to more search results being retrieved. This method improved the recall for six of the 22 reviews while the recall of the remaining reviews did not change. It is also notable from Table 5.1 that the algorithm performed one to four iterations only, that is substantially fewer compared with other approaches. Perhaps surprisingly, applying only the simple query reduction transformations is more effective than applying operator substitution transformations, leading to improvements in both precision and recall. However, recall drops for more reviews when only a single transformation type is used compared with all types: two for query reduction and three for operator substitution.

Performance of the oracle method demonstrates the challenge of developing high precision queries while also maintaining recall. The best possible recall achieved represented an improvement of 12.5% compared with the baseline. In addition, the precision can be improved, reaching four times that achieved using the original query.

Table 5.1: Recall and Precision results for each review in the update dataset. Values in boldface denote results improved when compared with the baseline.

			Query Refinement														
	Baseline		1. Operator Substitution			2. Query Expansion			3. Query Reduction			4. All Transformations			Oracle		
Review	**Recall**	**Precision**	**Recall**	**Precision**	**iter.**	**Recall**	**Precision**	**iter.**	**Recall**	**Precision**	**iter.**	**Recall**	**Precision**	**iter.**	**Recall**	**Precision**	**iter.**
CD000155	0.3333	0.0182	0.3333	0.0012	13	0.3333	0.0145	3	0.3333	0.0007	4	0.3333	**0.0206**	8	0.3333	**0.0667**	11
CD000160	1.0000	0.0005	1.0000	**0.0008**	4	1.0000	0.0003	2	1.0000	**0.0034**	13	1.0000	**0.0108**	15	1.0000	**0.0047**	12
CD000523	1.0000	0.0526	1.0000	0.0233	8	1.0000	0.0500	2	1.0000	0.0238	8	1.0000	**0.0909**	8	1.0000	**1.0000**	5
CD001298	0.0000	0.0000	**0.5882**	**0.0003**	13	**0.5882**	**0.0006**	3	**0.2353**	**0.0054**	20	**0.5882**	**0.0010**	17	**0.5882**	**0.0154**	2
CD001552	1.0000	0.0021	0.5000	**0.0152**	11	1.0000	0.0021	1	1.0000	**0.0039**	11	1.0000	**0.0043**	12	1.0000	**0.0444**	10
CD002064	1.0000	0.0833	1.0000	**0.1429**	6	1.0000	**0.0909**	2	0.0000	0.0000	9	1.0000	**1.0000**	9	1.0000	**0.5000**	5
CD004069	0.8889	0.0068	0.8889	0.0068	1	0.8889	0.0068	1	0.8889	0.0068	1	0.8889	0.0068	1	**1.0000**	**0.0089**	5
CD004214	0.0000	0.0000	**0.5000**	**0.0004**	3	**0.5000**	**0.0010**	2	0.0000	0.0000	1	**0.5000**	**0.0010**	2	**0.5000**	**0.0010**	2
CD004241	0.6000	0.0116	0.6000	0.0005	15	0.6000	0.0052	4	0.6000	0.0002	12	0.6000	0.0022	20	0.6000	**0.1765**	6
CD004479	0.7500	0.0189	0.7500	0.0149	3	0.7500	0.0001	2	0.7500	0.0013	2	0.7500	0.0013	2	0.7500	**0.0211**	2
CD005025	0.4130	0.0139	0.3913	0.0018	9	**0.5652**	0.0014	4	0.2609	0.0028	16	**0.6304**	0.0017	18	**0.7391**	0.0008	12
CD005055	0.6667	0.0033	0.6667	**0.0114**	5	0.6667	0.0016	2	**1.0000**	0.0001	4	0.6667	**0.0102**	7	**1.0000**	**0.0083**	6
CD005083	0.2222	0.0160	0.2222	0.0160	1	**0.5556**	0.0098	2	**0.5556**	0.0025	3	**0.5556**	0.0025	3	**0.5556**	**0.0403**	11
CD005128	0.5556	0.0007	0.5556	**0.0036**	7	0.5556	0.0007	1	0.5556	**0.0066**	7	0.5556	**0.0066**	7	0.5556	**0.0066**	6
CD005426	0.0000	0.0000	0.0000	0.0000	1	0.0000	0.0000	2	0.0000	0.0000	1	0.0000	0.0000	11	0.0000	0.0000	1
CD006839	0.6667	0.0204	0.6667	**0.1000**	7	0.6667	0.0204	1	0.6667	**0.1818**	8	0.6667	**0.2000**	9	**1.0000**	**0.3333**	11
CD006902	0.5000	0.0365	0.4000	0.0019	9	**0.8000**	0.0014	2	**0.6000**	0.0074	6	**0.8000**	0.0014	4	**0.8000**	0.0014	4
CD007020	0.2500	0.0156	0.2500	0.0083	3	0.2500	0.0147	2	0.2500	0.0052	2	0.2500	**0.0192**	3	0.2500	**0.0233**	4
CD007428	0.6667	0.0270	0.6667	**0.0435**	5	0.6667	0.0221	2	**0.7778**	0.0104	4	0.5556	**0.0568**	11	**0.7778**	**0.0538**	7
CD008392	1.0000	0.0014	1.0000	**0.0065**	7	1.0000	0.0013	2	1.0000	**0.0136**	10	1.0000	**0.0135**	10	1.0000	**0.0159**	12
CD010089	0.5000	0.0004	0.5000	**0.0008**	7	**0.7500**	**0.0005**	2	**0.7500**	**0.0009**	4	**0.7500**	**0.0021**	5	**0.7500**	**0.0021**	5
CD010847	0.6667	0.0755	0.6667	**0.1081**	5	0.6667	0.0755	1	0.6667	0.0154	3	**0.8333**	0.0007	11	**1.0000**	0.0003	8
Weighted Average	**0.566**	**0.003**	**0.571**	**0.004**	**7**	**0.666**	**0.002**	**2**	**0.641**	**0.005**	**7**	**0.669**	**0.007**	**9**	**0.691**	**0.012**	**7**

Figures 5.3 and 5.4 show the average weighted recall and precision scores for each iteration among the various approaches. The figures show the maximum number of iterations applied by each method (e.g. 12 for the oracle approach), although it is worth noting that the number of iterations applied to an individual review may be lower (e.g. see Table 5.1). Overall, improvements in recall (compared with the baseline) appear to be generated during the first iteration, while subsequent iterations help to improve precision. The effect is particularly pronounced for the oracle approach but can still be observed for other approaches. The best approach (i.e. using all transformation types) performed 20 iterations. The highest recall produced by this approach was at iteration six, then it slightly dropped, while the best precision was obtained at iteration 11 and remained constant until the algorithm stopped.

Table 5.2 shows an analysis of the transformation types used by the various approaches. The table indicates the number of times each transformation was selected to generate the modified query. As can be seen from the table, the transformation type applied most frequently by the best approach (i.e. using all transformation types) and oracle was remove line. The frequent use of this transformation may be explained by the fact removing lines from queries makes them less restrictive, and the objective function used to score queries prefers ones that maximise recall (i.e. less restrictive). On the other hand, the transformation types preferred most frequently by the operator substitution method were `OR`→`AND`, `.tw.`→`.ti.` and `.ti,ab.`→`.ti.` . All of these transformations lead to more restrictive queries thereby increasing the possibility of missing relevant studies. This is reflected in the low recall achieved using this method (see Table 5.1).

The original Boolean query is returned by the algorithm when the approach is unable to identify a transformation that improves performance. This happened for one review when the best approach (i.e. using all transformation types) and the oracle approach were used, for three reviews when using the operator substitution and query reduction methods, and for five reviews when using the query expansion method.

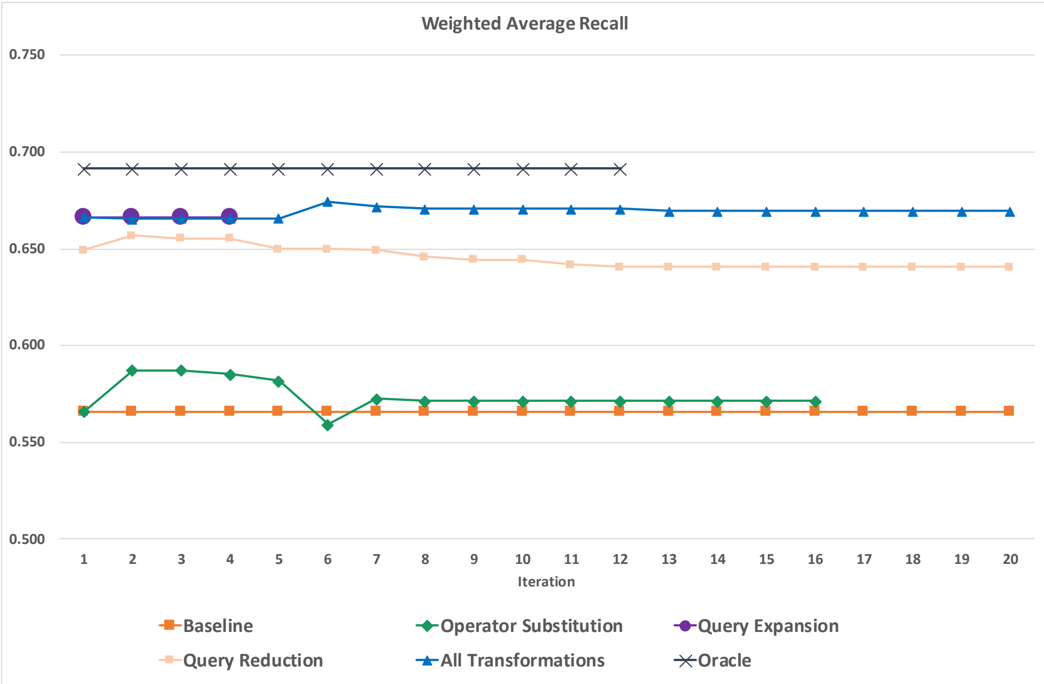

Figure 5.3: Weighted Average Recall scores for the various approaches. The baseline approach is included to allow comparison.

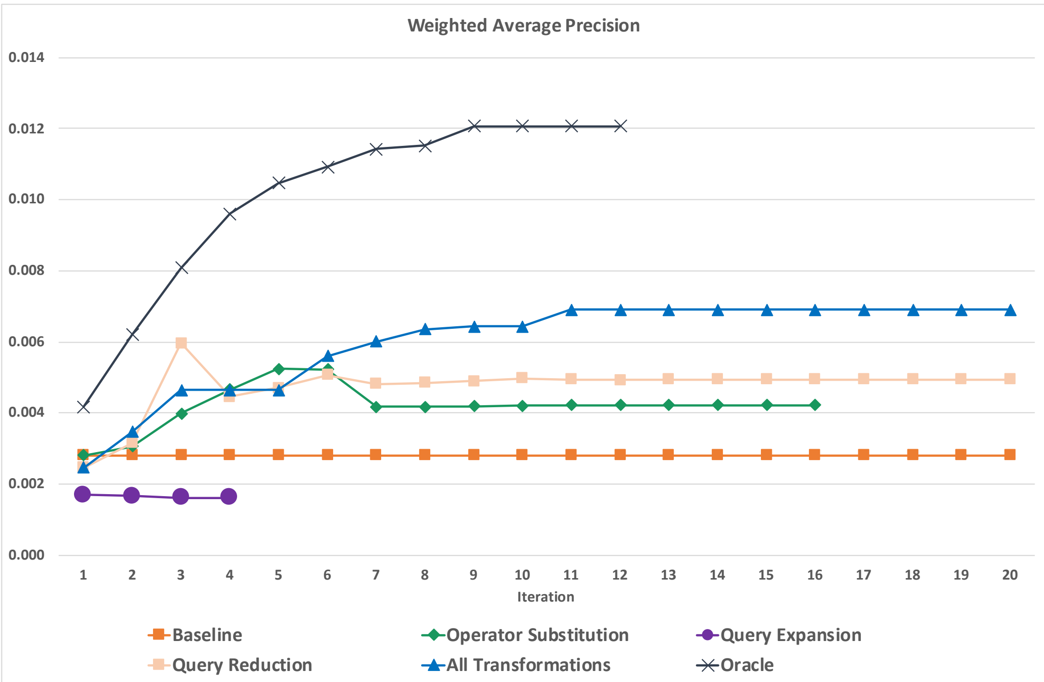

Figure 5.4: Weighted Average Precision scores for the various approaches. The baseline approach is included to allow comparison.

Table 5.2: Analysis of transformation types used in each method in the query refinement approach and oracle. The numbers represent how many times each transformation has been used through all iterations.

		Query Refinement				
Transformation Category	**Transformation Type**	**(1) Operator Substitution**	**(2) Query Expansion**	**(3) Query Reduction**	**(4) All Transformations**	**Oracle**
Operator Substitution	.tw.→.ti.	42	-	-	6	8
	.tw.→.tiab.	1	-	-	2	2
	.ab.→.ti.	0	-	-	0	0
	.ab.→.ti,ab.	0	-	-	0	0
	.ti,ab.→.ti.	36	-	-	16	19
	.ti,ab.→.tw.	0	-	-	0	1
	.ti.→.tw.	3	-	-	0	0
	.ti.→.ti,ab.	1	-	-	0	0
	AND→OR	11	-	-	2	1
	OR→AND	46	-	-	1	2
	.sh.→*	0	-	-	0	1
Query Expansion	1st top term	-	14	-	5	1
	2nd top term	-	7	-	3	1
	3rd top term	-	6	-	5	5
	4th top term	-	10	-	3	0
	5th top term	-	1	-	2	3
	1st & 2nd top terms	-	0	-	1	0
	1st to 3rd top terms	-	2	-	0	0
	1st to 4th top terms	-	0	-	0	0
	1st to 5th top terms	-	0	-	0	0
Query Reduction	remove line	-	-	146	146	102
Total		**140**	**40**	**146**	**192**	**146**

Figure 5.5 shows an example of a baseline Boolean query used for an original review and the transformed query produced by the best proposed method (i.e. using all transformation types). For this review, the algorithm ran for nine iterations with two types of transformations selected: operator substitution (use .ti. restriction for clauses 4,8 and 16) and query reduction (removal of clauses 1, 2, 3, 5 and 7). The transformed query improved precision by 92% without any reduction in recall.

Taken together, the results of the experiments indicate that Boolean query transformations can improve the retrieval performance for the review update in terms of recall and precision. The proposed algorithm can produce queries that retrieve more relevant studies and reduce the workload required by researchers by half.

(a) Original Query

1 randomized controlled trial.pt.
2 controlled clinical trial.pt.
3 randomized.ab.
4 placebo.ab.
5 drug therapy.fs.
6 randomly.ab.
7 trial.ab.
8 groups.ab.
9 or/1-8
10 exp animals/ not humans.sh.
11 9 not 10
12 exp Motor Neuron Disease/
13 (moto$1 neuron$1 disease$1 or moto?neuron$1 disease$1).mp.
14((Lou Gehrig$1 adj5 syndrome$1) or Lou Gehrig$1) adj5 disease).mp.
15 Charcot disease.tw.
16 amyotrophic lateral sclerosis.tw.
17 or/12-16
18 Insulin-Like Growth Factor I/
19 (rhIGF-1 or rhigf or rhigf-1 or insulin-like).mp.
20 11 and 17 and 19

(b) Transformed Query

~~1 randomized controlled trial.pt.~~
~~2 controlled clinical trial.pt.~~
~~3 randomized.ab.~~
4 placebo.ti.
~~5 drug therapy.fs.~~
6 randomly.ab.
~~7 trial.ab.~~
8 groups.ti.
9 or/1-8
10 exp animals/ not humans.sh.
11 9 not 10
12 exp Motor Neuron Disease/
13 (moto$1 neuron$1 disease$1 or moto?neuron$1
disease$1).mp.
14((Lou Gehrig$1 adj5 syndrome$1) or Lou Gehrig$1)
adj5 disease).mp.
15 Charcot disease.tw.
16 amyotrophic lateral sclerosis.ti.
17 or/12-16
18 Insulin-Like Growth Factor I/
19 (rhIGF-1 or rhigf or rhigf-1 or insulin-like).mp
20 11 and 17 and 19

Figure 5.5: Example of the original Boolean query for review CD002064 (Beauverd et al., 2012) (a) and the transformed Boolean query after nine iterations (b) with highlighted lines representing the clauses transformed by the algorithm.

5.7 Summary

This chapter proposed a novel algorithm to automatically refine Boolean queries for the study selection stage of systematic review updates. The algorithm extended previous work in two important ways. Firstly, it is applied to the problem of generating queries for review updates and makes use of information about which studies were included/excluded from the original review to guide the query modification. Secondly, it extends the set of query transformations introduced in literature and demonstrates that the new transformation leads to generation of more effective queries.

The proposed algorithm generates a set of transformed queries using three methods: operator substitution, query expansion and query reduction. The best query is then selected using an objective function that considers both recall and precision. The method improves the original query both in terms of recall and precision. It produces queries that are able to identify relevant studies that would not be retrieved using the query from the original review.

Results demonstrated that information available from the original review, particularly the relevance judgements, can be used to produce queries that are more effective than the ones used for the original review. The algorithm proposed in this chapter has the potential to assist researchers conducting updates of systematic reviews by supporting them to produce queries that both identify more relevant studies and reduce the number of what needs to be screened, thereby reducing the workload required to ensure that reviews remain up to date.

Chapter 6

Conclusion and Future Directions

Systematic reviews are essential in healthcare where the volume of evidence in scientific research publications is vast and cannot feasibly be identified or analysed by individual clinicians or decision makers. However, the process of creating a systematic review is time consuming and expensive. The problem of identifying relevant evidence is a significant part of the effort required by researchers to produce and update systematic reviews.

This thesis aimed to support systematic reviews through NLP/IR techniques. The particular focus of this study was to improve the process of identifying relevant evidence to reduce the workload required from researchers and ensure that the reviews are consistent with current evidence. This research gave particular attention to systematic review updates, which are of significant importance but the process of creating them has not been sufficiently addressed in previous work.

This chapter summarises the work presented throughout this thesis and indicates possible points for future directions.

6.1 Summary of the Thesis

Chapter 2 presented a systematic literature review of NLP/IR techniques used to facilitate the screening process for systematic reviews and review updates. The review focused on four main questions: (Q1) Which NLP/IR techniques have been proposed to support the screening process?, (Q2) Which datasets are used? Are they publicly available?, (Q3) How are those techniques evaluated? and finally (Q4) Which techniques are applied in the screening stage of the review update process? The review showed that NLP/IR techniques are beneficial to improve the screening process and reduce the workload required from researchers. In addition, the review demonstrated that the majority of work applied techniques for the creation of new reviews, while only a limited number of studies tackled the problem of identifying relevant evidence for review updates.

Chapter 3 explored the use of various query adaptation methods to improve studies ranking for systematic reviews. The chapter addressed **RQ1: How can studies be ranked so that the potentially relevant ones appear as early in the ranking as possible?** and **RQ2: Can the feedback from reviewer(s) be used to improve these rankings?** Three main approaches were explored. The first examined which information from the Boolean query is most helpful for ranking the studies. Results demonstrated that the review title and terms extracted from the Boolean query were found to be the most useful pieces of information. The second approach explored the use of lexical statistics to identify terms that distinguish relevant studies from others. The experiments demonstrated that including general information about the type of publication that is likely to be of relevance for a systematic review can improve retrieval performance. The best performance was achieved using the Log-Likelihood statistic. The final approach applied the Rocchio algorithm, and demonstrated that information contained in judgements about document relevance could improve the ranking of studies.

Chapter 4 introduced a dataset containing 25 intervention reviews from the Cochrane Collaboration and applied approaches from the previous chapter to it. The dataset is publicly available and ready for use to support the development of approaches to automate the updating process. The chapter addressed **RQ3: Can the rankings for systematic review updates be improved by making use of information about the original review, such as search strategy and feedback from reviewers?** by conducting experiments on the update dataset using lexical statistics and relevance feedback. Results demonstrated that the significant amount of knowledge about which studies are suitable from the original review (relevance judgements) can help to improve study selection for systematic review updates.

Chapter 5 developed and evaluated a novel algorithm to automatically refine Boolean queries to improve the identification of relevant studies for review updates. The chapter addressed **RQ4: Is it possible to generate Boolean search queries for review updates that are more effective than the one used for the original review?**. Experiments were carried out using the update dataset from the previous chapter. The proposed algorithm generates a set of transformed queries using three methods: operator substitution, query expansion and query reduction. The best query is then selected using an objective function that considers both recall and precision. The method improves the original query both in terms of recall and precision. It produces queries that are able to identify relevant studies that would not be retrieved using the query from the original review. Results demonstrated that information available from the original review, particularly the relevance judgements, can be used to produce queries that are more effective than the ones used for the original review. The algorithm has the potential to assist researchers conducting updates of systematic reviews by supporting them to produce queries that both identify more relevant studies and reduce the number of studies that need to be screened, thereby reducing the workload required to ensure that reviews remain up to date.

6.2 Future Directions

The work in this thesis can be further extended in different ways:

- The work presented in Chapter 5 can be extended as follows:
  - The objective function (Equation 5.1) was defined to favour recall due to the nature of the search problem in systematic reviews where high recall is important since the goal is to identify all potentially relevant studies. Although the function proved its effectiveness in improving the performance of retrieving studies in terms of recall and precision, it would be interesting to further

expand the work by investigating the effect of using a different balance of recall and precision by using a different objective function. One important point that should be taken into consideration when selecting the objective function is that systematic review is considered a High-Recall task (Carol et al., 2020). As discussed in Section 2.4.3, High-Recall Retrieval problem is one of the fundamental tasks for many applications such as patent retrieval, legal search and medical search; the reviewers' goal is to identify almost all of the publications reasonably related to the search topic, i.e., there is typically an emphasis on recall. Missing one relevant study might cause an enormous risk and it is highly undesirable. In systematic reviews, missing relevant studies could threaten the validity of the review, and, at worst, means the review could mislead (Garner et al., 2016; Waffenschmidt et al., 2019).

  – The experiments presented in Section 5.4 demonstrated that the proposed algorithm can improve the retrieval performance for the review update. However, the experiments were carried out on one type of systematic review (i.e. intervention reviews) since suitable datasets are not available for other review types (i.e. datasets that include information about both the original and updated versions of the review). In future work, it would be interesting to develop a dataset containing other review types, e.g. Diagnostic Test Accuracy reviews, to determine how the performance of the algorithm is affected by using different review types.

- This thesis demonstrated the usefulness of NLP/IR techniques in improving the identification of relevant evidence for systematic review updates. From here, another research direction of potential interest could focus on determining whether the new evidence is going to change the conclusion of the review. This would be very useful to prioritise the publication of updated reviews (i.e. the updates that change the original conclusion would be published first). The advantage is to help healthcare specialists and clinicians who need to make more conscious decisions about healthcare to reach reliable reviews of the recently available evidence. As discussed in Chapter 4, forest plots provide statistical information about studies included in the systematic review. It would be beneficial to use this information to predict whether the new evidence is going to change the review conclusion or not.

Appendix A

Search Results

Below is the full list of search results retrieved by running the Boolean query (see Figure 2.2) on PMC for the systematic review of the literature presented in Chapter 2.

1: Long-term efficacy of interventions for actinic keratosis: protocol for a systematic review and network meta-analysis Theresa Steeb, Markus V. Heppt, Lars Becker, Christoph Kohl, Lars E. French, Carola Berking Syst Rev. 2019; 8: 237. Published online 2019 Oct 11. doi: 10.1186/s13643-019-1156-8 PMCID: PMC6788027

2: Automatic discovery of 100-miRNA signature for cancer classification using ensemble feature selection Alejandro Lopez-Rincon, Marlet Martinez-Archundia, Gustavo U. Martinez-Ruiz, Alexander Schoenhuth, Alberto Tonda BMC Bioinformatics. 2019; 20: 480. Published online 2019 Sep 18. doi: 10.1186/s12859-019-3050-8 PMCID: PMC6751684

3: SIPsmartER delivered through rural, local health districts: adoption and implementation outcomes Kathleen J. Porter, Donna Jean Brock, Paul A. Estabrooks, Katelynn M. Perzynski, Erin R. Hecht, Pamela Ray, Natalie Kruzliakova, Eleanor S. Cantrell, Jamie M. Zoellner BMC Public Health. 2019; 19: 1273. Published online 2019 Sep 18. doi: 10.1186/s12889-019-7567-6 PMCID: PMC6751747

4: Towards pixel-to-pixel deep nucleus detection in microscopy images Fuyong Xing, Yuanpu Xie, Xiaoshuang Shi, Pingjun Chen, Zizhao Zhang, Lin Yang BMC Bioinformatics. 2019; 20: 472. Published online 2019 Sep 14. doi: 10.1186/s12859-019-3037-5 PMCID: PMC6744696

5: Measuring rank robustness in scored protein interaction networks Lyuba V. Bozhilova, Alan V. Whitmore, Jonny Wray, Gesine Reinert, Charlotte M. Deane BMC Bioinformatics. 2019; 20: 446. Published online 2019 Aug 28. doi: 10.1186/s12859-019-3036-6 PMCID: PMC6714100

6: Analysis of disease profile, and medical burden by lead exposure from hospital information systems in China Han Song, Jianchao Liu, Zipeng Cao, Wenjing Luo, Jing-Yuan Chen BMC Public Health. 2019; 19: 1170. Published online 2019 Aug 27. doi: 10.1186/s12889-019-7515-5 PMCID: PMC6712603

7: Breaking barriers: using the behavior change wheel to develop a tailored intervention to overcome workplace inhibitors to breaking up sitting time Samson O. Ojo, Daniel P. Bailey, Marsha L. Brierley, David J. Hewson, Angel M. Chater BMC Public Health. 2019; 19: 1126. Published online 2019 Aug 16. doi: 10.1186/s12889-019-7468-8 PMCID: PMC6697980

8: Collective intelligence in medical decision-making: a systematic scoping review Kate Radcliffe, Helena C. Lyson, Jill Barr-Walker, Urmimala Sarkar BMC Med Inform Decis Mak. 2019; 19: 158. Published online 2019 Aug 9. doi: 10.1186/s12911-019-0882-0 PMCID: PMC6688241

9: Public health effects of gambling – debate on a conceptual model Tiina Latvala, Tomi Lintonen, Anne Konu BMC Public Health. 2019; 19: 1077. Published online 2019 Aug 9. doi: 10.1186/s12889-019-7391-z PMCID: PMC6688345

10: Comparing drug safety of hepatitis C therapies using post-market data Jing Huang, Xinyuan Zhang, Jiayi Tong, Jingcheng Du, Rui Duan, Liu Yang, Jason H. Moore, Cui Tao, Yong Chen BMC Med Inform Decis Mak. 2019; 19(Suppl 4): 147. Published online 2019 Aug 8. doi: 10.1186/s12911-019-0860-6 PMCID: PMC6686214

11: Understanding how and why de-implementation works in health and care: research protocol for a realist synthesis of evidence Christopher Burton, Lynne Williams, Tracey Bucknall, Stephen Edwards, Denise Fisher, Beth Hall, Gill Harris, Peter Jones, Matthew Makin, Anne McBride, Rachel Meacock, John Parkinson, Jo Rycroft-Malone, Justin Waring Syst Rev. 2019; 8: 194. Published online 2019 Aug 5. doi: 10.1186/s13643-019-1111-8 PMCID: PMC6683493

12: Narrative Review for Exploring Barriers to Readiness of Electronic Health Record Implementation in Primary Health Care Sandra Hakiem Afrizal, Achmad Nizar Hidayanto, Putu Wuri Handayani, Meiwita Budiharsana, Tris Eryando Healthc Inform Res. 2019 Jul; 25(3): 141–152. Published online 2019 Jul 31. doi: 10.4258/hir.2019.25.3.141 PMCID: PMC6689507

13: Prediction and Staging of Hepatic Fibrosis in Children with Hepatitis C Virus: A Machine Learning Approach Nahla H. Barakat, Sana H. Barakat, Nadia Ahmed Healthc Inform Res. 2019 Jul; 25(3): 173–181. Published online 2019 Jul 31. doi: 10.4258/hir.2019.25.3.173 PMCID: PMC6689505

14: Comparative effectiveness trial comparing MyPlate to calorie counting for mostly low-income Latino primary care patients of a federally qualified community health center: study design, baseline characteristics Lillian Gelberg, Melvin W. Rico, Dena R. Herman, Thomas R. Belin, Maria Chandler, Evangelina Ramirez, Stephanie Love, William J. McCarthy BMC Public Health. 2019; 19: 990. Published online 2019 Jul 24. doi: 10.1186/s12889-019-7294-z PMCID: PMC6651946

15: Risk factors for hearing loss in children: a systematic literature review and meta-analysis protocol Bénédicte Vos, Dorie Noll, Marie Pigeon, Marlene Bagatto, Elizabeth M. Fitzpatrick Syst Rev. 2019; 8: 172. Published online 2019 Jul 17. doi: 10.1186/s13643-019-1073-x PMCID: PMC6637473

16: Clarifying differences between reviews within evidence ecosystems David Gough, James Thomas, Sandy Oliver Syst Rev. 2019; 8: 170. Published online 2019 Jul 15. doi: 10.1186/s13643-019-1089-2 PMCID: PMC6631872

17: Toward systematic review automation: a practical guide to using machine learning tools in research synthesis Iain J. Marshall, Byron C. Wallace Syst Rev. 2019; 8: 163. Published online 2019 Jul 11. doi: 10.1186/s13643-019-1074-9 PMCID: PMC6621996

18: Mapping evidence on tuberculosis active case finding policies, strategies, and interventions for tuberculosis key populations: a systematic scoping review protocol Desmond Kuupiel, Vitalis Bawontuo, Tivani P. Mashamba-Thompson Syst Rev. 2019; 8: 162. Published online 2019 Jul 10. doi: 10.1186/s13643-019-1098-1 PMCID: PMC6617702

19: A systematic review and meta-analysis in the effectiveness of mobile phone interventions used to improve adherence to antiretroviral therapy in HIV infection Reshma Shah, Julie Watson, Caroline Free BMC Public Health. 2019; 19: 915. Published online 2019 Jul 9. doi: 10.1186/s12889-019-6899-6 PMCID: PMC6617638

20: Usability and acceptability of four systematic review automation software packages: a mixed method design Gina Cleo, Anna Mae Scott, Farhana Islam, Blair Julien, Elaine Beller Syst Rev. 2019; 8: 145. Published online 2019 Jun 20. doi: 10.1186/s13643-019-1069-6 PMCID: PMC6587262

21: A question of trust: can we build an evidence base to gain trust in systematic review automation technologies? Annette M. O'Connor, Guy Tsafnat, James Thomas, Paul Glasziou,

Stephen B. Gilbert, Brian Hutton Syst Rev. 2019; 8: 143. Published online 2019 Jun 18. doi: 10.1186/s13643-019-1062-0 PMCID: PMC6582554

22: Newspaper coverage before and after the HPV vaccination crisis began in Japan: a text mining analysis Tsuyoshi Okuhara, Hirono Ishikawa, Masafumi Okada, Mio Kato, Takahiro Kiuchi BMC Public Health. 2019; 19: 770. Published online 2019 Jun 17. doi: 10.1186/s12889-019-7097-2 PMCID: PMC6580608

23: A systematic literature review of reported challenges in health care delivery to migrants and refugees in high-income countries - the 3C model Julia Brandenberger, Thorkild Tylleskär, Katrin Sontag, Bernadette Peterhans, Nicole Ritz BMC Public Health. 2019; 19: 755. Published online 2019 Jun 14. doi: 10.1186/s12889-019-7049-x PMCID: PMC6567460

24: ProteinNet: a standardized data set for machine learning of protein structure Mohammed AlQuraishi BMC Bioinformatics. 2019; 20: 311. Published online 2019 Jun 11. doi: 10.1186/s12859-019-2932-0 PMCID: PMC6560865

25: Study-based registers reduce waste in systematic reviewing: discussion and case report Farhad Shokraneh, Clive E. Adams Syst Rev. 2019; 8: 129. Published online 2019 May 30. doi: 10.1186/s13643-019-1035-3 PMCID: PMC6542007

26: Multi-domain semantic similarity in biomedical research João D. Ferreira, Francisco M. Couto BMC Bioinformatics. 2019; 20(Suppl 10): 246. Published online 2019 May 29. doi: 10.1186/s12859-019-2810-9 PMCID: PMC6538554

27: Challenges and solutions for instituting an efficient maintenance program for laboratory equipment in Central Asian, and developing world, countries Riza Ikranbegiin, George Schmid, David Hoos, Andrew Young, Phyllis Della-Latta, Paul Spearman, Artur Ramos, Bereket Alemayehu, Begaiym Achmetova, Gulzhan Nauryzova, Adilya Albetkova BMC Public Health. 2019; 19(Suppl 3): 476. Published online 2019 May 10. doi: 10.1186/s12889-019-6782-5 PMCID: PMC6696665

28: Machine learning to help researchers evaluate biases in clinical trials: a prospective, randomized user study Frank Soboczenski, Thomas A. Trikalinos, Joël Kuiper, Randolph G. Bias, Byron C. Wallace, Iain J. Marshall BMC Med Inform Decis Mak. 2019; 19: 96. Published online 2019 May 8. doi: 10.1186/s12911-019-0814-z PMCID: PMC6505190

29: Automated assessment of biological database assertions using the scientific literature Mohamed Reda Bouadjenek, Justin Zobel, Karin Verspoor BMC Bioinformatics. 2019; 20: 216. Published online 2019 Apr 29. doi: 10.1186/s12859-019-2801-x PMCID: PMC6489365

30: Risk of bias judgements and strength of conclusions in meta-evidence from the Cochrane Colorectal Cancer Group John Delaney, Rebecca Cui, Alexander Engel Syst Rev. 2019; 8: 90. Published online 2019 Apr 8. doi: 10.1186/s13643-019-1001-0 PMCID: PMC6452506

31: Using an intervention mapping approach to develop prevention and rehabilitation strategies for musculoskeletal pain among surgeons Tina Dalager, Anne Højmark, Pernille Tine Jensen, Karen Søgaard, Lotte Nygaard Andersen BMC Public Health. 2019; 19: 320. Published online 2019 Mar 18. doi: 10.1186/s12889-019-6625-4 PMCID: PMC6423851

32: Editorial: Systematic review automation thematic series Joseph Lau Syst Rev. 2019; 8: 70. Published online 2019 Mar 11. doi: 10.1186/s13643-019-0974-z PMCID: PMC6410513

33: Using decision fusion methods to improve outbreak detection in disease surveillance Gaëtan Texier, Rodrigue S. Alldoji, Loty Diop, Jean-Baptiste Meynard, Liliane Pellegrin, Hervé Chaudet BMC Med Inform Decis Mak. 2019; 19: 38. Published online 2019 Mar 5. doi: 10.1186/s12911-019-0774-3Correction in: BMC Med Inform Decis Mak. 2019; 19: 81. PMCID: PMC6402142

34: Construction and application of service quality evaluation system in the preclinical research on cardiovascular implant devices Yongchun Cui, Fuliang Luo, Boqing Yang, Bin Li, Qi Zhang, Gopika Das, Guangxin Yue, Jiajie Li, Yue Tang, Xin Wang BMC Med Inform Decis Mak. 2019; 19: 37. Published online 2019 Feb 28. doi: 10.1186/s12911-019-0773-4 PMCID: PMC6396521

35: Who can you trust? A review of free online sources of “trustworthy” information about treatment effects for patients and the public Andrew D. Oxman, Elizabeth J. Paulsen BMC Med Inform Decis Mak. 2019; 19: 35. Published online 2019 Feb 20. doi: 10.1186/s12911-019-0772-5 PMCID: PMC6381637

36: Still moving toward automation of the systematic review process: a summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR) Annette M. O’Connor, Guy Tsafnat, Stephen B. Gilbert, Kristina A. Thayer, Ian Shemilt, James Thomas, Paul Glasziou, Mary S. Wolfe Syst Rev. 2019; 8: 57. Published online 2019 Feb 20. doi: 10.1186/s13643-019-0975-y PMCID: PMC6381675

37: CeModule: an integrative framework for discovering regulatory patterns from genomic data in cancer Qiu Xiao, Jiawei Luo, Cheng Liang, Jie Cai, Guanghui Li, Buwen Cao BMC Bioinformatics. 2019; 20: 67. Published online 2019 Feb 7. doi: 10.1186/s12859-019-2654-3 PMCID: PMC6367773

38: Maintaining relevance in HIV systematic reviews: an evaluation of Cochrane reviews Ingrid Eshun-Wilson, Shahista Jaffer, Rhodine Smith, Samuel Johnson, Paul Hine, Alberto Mateo,

Anne-Marie Stephani, Paul Garner Syst Rev. 2019; 8: 46. Published online 2019 Feb 7. doi: 10.1186/s13643-019-0960-5 PMCID: PMC6366015

39: Changes in school-day step counts during a physical activity for Lent intervention: a cluster randomized crossover trial of the Savior's Sandals David Kahan, Kent A. Lorenz, Eyad Kawwa, Andrew Rioveros BMC Public Health. 2019; 19: 141. Published online 2019 Feb 1. doi: 10.1186/s12889-019-6479-9 PMCID: PMC6359766

40: The role of icodextrin in peritoneal dialysis: protocol for a systematic review and meta-analysis Monika Becker, Stefanie Bühn, Jessica Breuing, Catherine A. Firanek, Simone Hess, Hisanori Nariai, Mark R. Marshall, James A. Sloand, Qiang Yao, Käthe Goossen, Dawid Pieper Syst Rev. 2019; 8: 35. Published online 2019 Jan 30. doi: 10.1186/s13643-019-0959-y PMCID: PMC6352378

41: Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error Alexandra Bannach-Brown, Piotr Przybyła, James Thomas, Andrew S. C. Rice, Sophia Ananiadou, Jing Liao, Malcolm Robert Macleod Syst Rev. 2019; 8: 23. Published online 2019 Jan 15. doi: 10.1186/s13643-019-0942-7 PMCID: PMC6334440

42: The state of research on cyberattacks against hospitals and available best practice recommendations: a scoping review Salem T. Argaw, Nefti-Eboni Bempong, Bruce Eshaya-Chauvin, Antoine Flahault BMC Med Inform Decis Mak. 2019; 19: 10. Published online 2019 Jan 11. doi: 10.1186/s12911-018-0724-5 PMCID: PMC6330387

43: Effectiveness of text messaging interventions on prevention, detection, treatment, and knowledge outcomes for sexually transmitted infections (STIs)/HIV: a systematic review and meta-analysis Darlene Taylor, Carole Lunny, Petra Lolić, Orion Warje, Jasmina Geldman, Tom Wong, Mark Gilbert, Richard Lester, Gina Ogilvie Syst Rev. 2019; 8: 12. Published online 2019 Jan 8. doi: 10.1186/s13643-018-0921-4 PMCID: PMC6323863

44: Image-based classification of plant genus and family for trained and untrained plant species Marco Seeland, Michael Rzanny, David Boho, Jana Wäldchen, Patrick Mäder BMC Bioinformatics. 2019; 20: 4. Published online 2019 Jan 3. doi: 10.1186/s12859-018-2474-x PMCID: PMC6318858

45: A CRISPR focus on attitudes and beliefs toward somatic genome editing from stakeholders within the sickle cell disease community Anitra Persaud, Stacy Desine, Katherine Blizinsky, Vence L.

Bonham Genet Med. 2019; 21(8): 1726–1734. Published online 2018 Dec 24. doi: 10.1038/s41436-018-0409-6 PMCID: PMC6606394

46: Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease: a feature selection ensemble combining stability and predictability Telma Pereira, Francisco L. Ferreira, Sandra Cardoso, Dina Silva, Alexandre de Mendonça, Manuela Guerreiro, Sara C. Madeira, for the Alzheimer's Disease Neuroimaging Initiative BMC Med Inform Decis Mak. 2018; 18: 137. Published online 2018 Dec 19. doi: 10.1186/s12911-018-0710-y PMCID: PMC6299964

47: Study protocol of a 4- parallel arm, superiority, community based cluster randomized controlled trial comparing paper and e-platform based interventions to improve accuracy of recall of last menstrual period (LMP) dates in rural Bangladesh Shumona Sharmin Salam, Nazia Binte Ali, Ahmed Ehsanur Rahman, Tazeen Tahsina, Md. Irteja Islam, Afrin Iqbal, Dewan Md. Emdadul Hoque, Samir Kumar Saha, Shams El Arifeen BMC Public Health. 2018; 18: 1359. Published online 2018 Dec 10. doi: 10.1186/s12889-018-6258-z PMCID: PMC6288958

48: Secondary research uses of residual newborn screening dried bloodspots: a scoping review Erin Rothwell, Erin Johnson, Naomi Riches, Jeffrey R. Botkin Genet Med. 2019; 21(7): 1469–1475. Published online 2018 Dec 10. doi: 10.1038/s41436-018-0387-8 PMCID: PMC6557682

49: Multi-View Graph Convolutional Network and Its Applications on Neuroimage Analysis for Parkinson's Disease Xi Zhang, Lifang He, Kun Chen, Yuan Luo, Jiayu Zhou, Fei Wang AMIA Annu Symp Proc. 2018; 2018: 1147–1156. Published online 2018 Dec 5. PMCID: PMC6371363

50: Data Extraction and Synthesis in Systematic Reviews of Diagnostic Test Accuracy: A Corpus for Automating and Evaluating the Process Christopher Norman, Mariska Leeflang, Aurélie Névéol AMIA Annu Symp Proc. 2018; 2018: 817–826. Published online 2018 Dec 5. PMCID: PMC6371350

51: Longitudinal Clustering of High-cost Patients' Spend Trajectories:Delineating Individual Behaviors from Aggregate Trends Andrew M Placona, Rich King, Fengjuan Wang AMIA Annu Symp Proc. 2018; 2018: 907–915. Published online 2018 Dec 5. PMCID: PMC6371335

52: Learning to Personalize from Practice: A Real World Evidence Approach of Care Plan Personalization based on Differential Patient Behavioral Responses in Care Management Records Pei-Yun S. Hsueh, Subhro Das, Chandramouli Maduri, Karie Kelly AMIA Annu Symp Proc. 2018; 2018: 592–601. Published online 2018 Dec 5. PMCID: PMC6371321

53: Improving breast cancer risk prediction by using demographic risk factors, abnormality features on mammograms and genetic variants Shara I. Feld, Kaitlin M. Woo, Roxana Alexandridis, Yirong Wu, Jie Liu, Peggy Peissig, Adedayo A. Onitilo, Jennifer Cox, C. David Page, Elizabeth S. Burnside AMIA Annu Symp Proc. 2018; 2018: 1253–1262. Published online 2018 Dec 5. PMCID: PMC6371301

54: Full-length title: NRPPUR database search and in vitro analysis identify an NRPS-PKS biosynthetic gene cluster with a potential antibiotic effect Shirley Fritz, Andriamiharimamy Rajaonison, Olivier Chabrol, Didier Raoult, Jean-Marc Rolain, Vicky Merhej BMC Bioinformatics. 2018; 19: 463. Published online 2018 Dec 3. doi: 10.1186/s12859-018-2479-5 PMCID: PMC6276269

55: HIV-related data among key populations to inform evidence-based responses: protocol of a systematic review Amrita Rao, Sheree Schwartz, Keith Sabin, Tisha Wheeler, Jinkou Zhao, James Hargreaves, Stefan Baral, on behalf of the Global.HIV Research Group Syst Rev. 2018; 7: 220. Published online 2018 Dec 3. doi: 10.1186/s13643-018-0894-3 PMCID: PMC6278072

56: Efficacy of hearing conservation education programs for youth and young adults: a systematic review Khalid M. Khan, Sylvanna L. Bielko, Marjorie C. McCullagh BMC Public Health. 2018; 18: 1286. Published online 2018 Nov 22. doi: 10.1186/s12889-018-6198-7 PMCID: PMC6249850

57: Primary care physicians' attitudes to the adoption of electronic medical records: a systematic review and evidence synthesis using the clinical adoption framework Amy O'Donnell, Eileen Kaner, Caroline Shaw, Catherine Haighton BMC Med Inform Decis Mak. 2018; 18: 101. Published online 2018 Nov 13. doi: 10.1186/s12911-018-0703-x PMCID: PMC6234586

58: What maximizes the effectiveness and implementation of technology-based interventions to support healthcare professional practice? A systematic literature review C Keyworth, J Hart, C J Armitage, M P Tully BMC Med Inform Decis Mak. 2018; 18: 93. Published online 2018 Nov 7. doi: 10.1186/s12911-018-0661-3 PMCID: PMC6223001

59: Partially systematic thoughts on the history of systematic reviews Mike Clarke Syst Rev. 2018; 7: 176. Published online 2018 Oct 27. doi: 10.1186/s13643-018-0833-3 PMCID: PMC6204283

60: Development and validation of the PEPPER framework (Prenatal Exposure PubMed ParsER) with applications to food additives Mary Regina Boland, Aditya Kashyap, Jiadi Xiong, John Holmes, Scott Lorch J Am Med Inform Assoc. 2018 Nov; 25(11): 1432–1443. Published online 2018 Oct 26. doi: 10.1093/jamia/ocy119 PMCID: PMC6213088

61: HIV testing within general practices in Europe: a mixed-methods systematic review Jessika Deblonde, Dominique Van Beckhoven, Jasna Loos, Nicole Boffin, André Sasse, Christiana

Nöstlinger, Virginie Supervie, HERMETIC Study Group BMC Public Health. 2018; 18: 1191. Published online 2018 Oct 22. doi: 10.1186/s12889-018-6107-0 PMCID: PMC6196459

62: Effective study selection using text mining or a single-screening approach: a study protocol Siw Waffenschmidt, Elke Hausner, Wiebke Sieben, Thomas Jaschinski, Marco Knelangen, Inga Overesch Syst Rev. 2018; 7: 166. Published online 2018 Oct 20. doi: 10.1186/s13643-018-0839-x PMCID: PMC6195713

63: A comprehensive regional neurochemical theory in depression: a protocol for the systematic review and meta-analysis of 1H-MRS studies in major depressive disorder Thomas Drago, Patrick W O'Regan, Ivan Welaratne, Shane Rooney, Aoife O'Callaghan, Marissa Malkit, Elena Roman, Kirk J Levins, Lauren Alexander, Denis Barry, Erik O'Hanlon, Veronica O'Keane, Darren William Roddy Syst Rev. 2018; 7: 158. Published online 2018 Oct 12. doi: 10.1186/s13643-018-0830-6 PMCID: PMC6182786

64: Antibiotic therapy for skin and soft tissue infections: a protocol for a systematic review and network meta-analysis Jessica J. Bartoszko, Dominik Mertz, Lehana Thabane, Mark Loeb Syst Rev. 2018; 7: 138. Published online 2018 Sep 11. doi: 10.1186/s13643-018-0804-8 PMCID: PMC6134765

65: HLBS-PopOmics: an online knowledge base to accelerate dissemination and implementation of research advances in population genomics to reduce the burden of heart, lung, blood, and sleep disorders George A. Mensah, Wei Yu, Whitney L. Barfield, Mindy Clyne, Michael M. Engelgau, Muin J. Khoury Genet Med. 2019 Mar; 21(3): 519–524. Published online 2018 Sep 10. doi: 10.1038/s41436-018-0118-1 PMCID: PMC6402952

66: PseUI: Pseudouridine sites identification based on RNA sequence information Jingjing He, Ting Fang, Zizheng Zhang, Bei Huang, Xiaolei Zhu, Yi Xiong BMC Bioinformatics. 2018; 19: 306. Published online 2018 Aug 29. doi: 10.1186/s12859-018-2321-0 PMCID: PMC6114832

67: WELCOME: improving WEight controL and CO-Morbidities in children with obesity via Executive function training: study protocol for a randomized controlled trial Tiffany Naets, Leentje Vervoort, Marijke Ysebaert, Annelies Van Eyck, Stijn Verhulst, Luc Bruyndonckx, Benedicte De Winter, Kim Van Hoorenbeeck, Ann Tanghe, Caroline Braet BMC Public Health. 2018; 18: 1075. Published online 2018 Aug 29. doi: 10.1186/s12889-018-5950-3 PMCID: PMC6116429

68: PRS-on-Spark (PRSoS): a novel, efficient and flexible approach for generating polygenic risk scores Lawrence M. Chen, Nelson Yao, Elika Garg, Yuecai Zhu, Thao T. T. Nguyen, Irina Pokhvisneva, Shantala A. Hari Dass, Eva Unternaehrer, Hélène Gaudreau, Marie Forest, Lisa M.

McEwen, Julia L. MacIsaac, Michael S. Kobor, Celia M. T. Greenwood, Patricia P. Silveira, Michael J. Meaney, Kieran J. O'Donnell BMC Bioinformatics. 2018; 19: 295. Published online 2018 Aug 8. doi: 10.1186/s12859-018-2289-9 PMCID: PMC6083617

69: Fusion of encoder-decoder deep networks improves delineation of multiple nuclear phenotypes Mina Khoshdeli, Garrett Winkelmaier, Bahram Parvin BMC Bioinformatics. 2018; 19: 294. Published online 2018 Aug 7. doi: 10.1186/s12859-018-2285-0 PMCID: PMC6081825

70: Single screen of citations with excluded terms: an approach to citation screening in systematic reviews Brittany U. Carter Syst Rev. 2018; 7: 111. Published online 2018 Jul 28. doi: 10.1186/s13643-018-0782-x PMCID: PMC6064612

71: Evaluating semantic relations in neural word embeddings with biomedical and general domain knowledge bases Zhiwei Chen, Zhe He, Xiuwen Liu, Jiang Bian BMC Med Inform Decis Mak. 2018; 18(Suppl 2): 65. Published online 2018 Jul 23. doi: 10.1186/s12911-018-0630-xCorrection in: BMC Med Inform Decis Mak. 2018; 18: 73. PMCID: PMC6069806

72: Relief-Based Feature Selection: Introduction and Review Ryan J. Urbanowicz, Melissa Meeker, William La Cava, Randal S. Olson, Jason H. Moore J Biomed Inform. 2018 Sep; 85: 189–203. Published online 2018 Jul 18. doi: 10.1016/j.jbi.2018.07.014 PMCID: PMC6299836

73: The development, implementation and evaluation of interventions to reduce workplace sitting: a qualitative systematic review and evidence-based operational framework Kelly Mackenzie, Elizabeth Such, Paul Norman, Elizabeth Goyder BMC Public Health. 2018; 18: 833. Published online 2018 Jul 4. doi: 10.1186/s12889-018-5768-z PMCID: PMC6033205

74: Reflections and aspirations: the journal after 5 years David Moher, Lesley A. Stewart, Paul Shekelle Syst Rev. 2018; 7: 87. Published online 2018 Jun 20. doi: 10.1186/s13643-018-0753-2 PMCID: PMC6011401

75: Information standards for recording alcohol use in electronic health records: findings from a national consultation Shamil Haroon, Darren Wooldridge, Jan Hoogewerf, Krishnarajah Nirantharakumar, John Williams, Lina Martino, Neeraj Bhala BMC Med Inform Decis Mak. 2018; 18: 36. Published online 2018 Jun 7. doi: 10.1186/s12911-018-0612-z PMCID: PMC5992754

76: Effectiveness of capacity building interventions relevant to public health practice: a systematic review Kara DeCorby-Watson, Gloria Mensah, Kim Bergeron, Samiya Abdi, Benjamin Rempel, Heather Manson BMC Public Health. 2018; 18: 684. Published online 2018 Jun 1. doi: 10.1186/s12889-018-5591-6 PMCID: PMC5984748

77: Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR) Elaine Beller, Justin Clark, Guy Tsafnat, Clive Adams, Heinz Diehl, Hans Lund, Mourad Ouzzani, Kristina Thayer, James Thomas, Tari Turner, Jun Xia, Karen Robinson, Paul Glasziou, On behalf of the founding members of the ICASR group Syst Rev. 2018; 7: 77. Published online 2018 May 19. doi: 10.1186/s13643-018-0740-7 PMCID: PMC5960503

78: Identifying Falls Risk Screenings Not Documented with Administrative Codes Using Natural Language Processing Vivienne J Zhu, Tina D Walker, Robert W Warren, Peggy B Jenny, Stephane Meystre, Leslie A Lenert AMIA Annu Symp Proc. 2017; 2017: 1923–1930. Published online 2018 Apr 16. PMCID: PMC5977708

79: Thinking Together: Modeling Clinical Decision-Support as a Sociotechnical System Mustafa I. Hussain, Tera L. Reynolds, Fatemeh E. Mousavi, Yunan Chen, Kai Zheng AMIA Annu Symp Proc. 2017; 2017: 969–978. Published online 2018 Apr 16. PMCID: PMC5977688

80: Classifying Clinical Trial Eligibility Criteria to Facilitate Phased Cohort Identification Using Clinical Data Repositories Amy Y. Wang, William J. Lancaster, Matthew C. Wyatt, Luke V. Rasmussen, Daniel G. Fort, James J. Cimino AMIA Annu Symp Proc. 2017; 2017: 1754–1763. Published online 2018 Apr 16. PMCID: PMC5977684

81: Classifying Acute Ischemic Stroke Onset Time using Deep Imaging Features King Chung Ho, William Speier, Suzie El-Saden, Corey W. Arnold AMIA Annu Symp Proc. 2017; 2017: 892–901. Published online 2018 Apr 16. PMCID: PMC5977679

82: Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak Xinning Gui, Yue Wang, Yubo Kou, Tera Leigh Reynolds, Yunan Chen, Qiaozhu Mei, Kai Zheng AMIA Annu Symp Proc. 2017; 2017: 820–829. Published online 2018 Apr 16. PMCID: PMC5977662

83: Framing Electronic Medical Records as Polylingual Documents in Query Expansion Edward W Huang, Sheng Wang, Doris Jung-Lin Lee, Runshun Zhang, Baoyan Liu, Xuezhong Zhou, ChengXiang Zhai AMIA Annu Symp Proc. 2017; 2017: 940–949. Published online 2018 Apr 16. PMCID: PMC5977599

84: Realizing drug repositioning by adapting a recommendation system to handle the process Makbule Guclin Ozsoy, Tansel Özyer, Faruk Polat, Reda Alhajj BMC Bioinformatics. 2018; 19: 136.

Published online 2018 Apr 12. doi: 10.1186/s12859-018-2142-1Correction in: BMC Bioinformatics. 2018; 19: 250. PMCID: PMC5898022

85: Measuring phenotype-phenotype similarity through the interactome Jiajie Peng, Weiwei Hui, Xuequn Shang BMC Bioinformatics. 2018; 19(Suppl 5): 114. Published online 2018 Apr 11. doi: 10.1186/s12859-018-2102-9 PMCID: PMC5907215

86: AceTree: a major update and case study in the long term maintenance of open-source scientific software Braden Katzman, Doris Tang, Anthony Santella, Zhirong Bao BMC Bioinformatics. 2018; 19: 121. Published online 2018 Apr 4. doi: 10.1186/s12859-018-2127-0 PMCID: PMC5885296

87: Effect of peer support interventions on cardiovascular disease risk factors in adults with diabetes: a systematic review and meta-analysis Sonal J. Patil, Todd Ruppar, Richelle J. Koopman, Erik J. Lindbloom, Susan G. Elliott, David R. Mehr, Vicki S. Conn BMC Public Health. 2018; 18: 398. Published online 2018 Mar 23. doi: 10.1186/s12889-018-5326-8 PMCID: PMC5865386

88: Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool Allison Gates, Cydney Johnson, Lisa Hartling Syst Rev. 2018; 7: 45. Published online 2018 Mar 12. doi: 10.1186/s13643-018-0707-8 PMCID: PMC5848519

89: Comparative assessment of strategies to identify similar ligand-binding pockets in proteins Rajiv Gandhi Govindaraj, Michal Brylinski BMC Bioinformatics. 2018; 19: 91. Published online 2018 Mar 9. doi: 10.1186/s12859-018-2109-2 PMCID: PMC5845264

90: Building protein-protein interaction networks for Leishmania species through protein structural information Crhisllane Rafaele dos Santos Vasconcelos, Túlio de Lima Campos, Antonio Mauro Rezende BMC Bioinformatics. 2018; 19: 85. Published online 2018 Mar 6. doi: 10.1186/s12859-018-2105-6 PMCID: PMC5840830

91: HiComet: a high-throughput comet analysis tool for large-scale DNA damage assessment Taehoon Lee, Sungmin Lee, Woo Young Sim, Yu Mi Jung, Sunmi Han, Joong-Ho Won, Hyeyoung Min, Sungroh Yoon BMC Bioinformatics. 2018; 19(Suppl 1): 44. Published online 2018 Feb 19. doi: 10.1186/s12859-018-2015-7Correction in: BMC Bioinformatics. 2018; 19: 170. PMCID: PMC5836828

92: GenIO: a phenotype-genotype analysis web server for clinical genomics of rare diseases Daniel Koile, Marta Cordoba, Maximiliano de Sousa Serro, Marcelo Andres Kauffman, Patricio

Yankilevich BMC Bioinformatics. 2018; 19: 25. Published online 2018 Jan 27. doi: 10.1186/s12859-018-2027-3 PMCID: PMC5787240

93: SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research Honghan Wu, Giulia Toti, Katherine I Morley, Zina M Ibrahim, Amos Folarin, Richard Jackson, Ismail Kartoglu, Asha Agrawal, Clive Stringer, Darren Gale, Genevieve Gorrell, Angus Roberts, Matthew Broadbent, Robert Stewart, Richard JB Dobson J Am Med Inform Assoc. 2018 May; 25(5): 530–537. Published online 2018 Jan 19. doi: 10.1093/jamia/ocx160 PMCID: PMC6019046

94: Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR) Annette M. O'Connor, Guy Tsafnat, Stephen B. Gilbert, Kristina A. Thayer, Mary S. Wolfe Syst Rev. 2018; 7: 3. Published online 2018 Jan 9. doi: 10.1186/s13643-017-0667-4 PMCID: PMC5759184

95: Utilizing random Forest QSAR models with optimized parameters for target identification and its application to target-fishing server Kyoungyeul Lee, Minho Lee, Dongsup Kim BMC Bioinformatics. 2017; 18(Suppl 16): 567. Published online 2017 Dec 28. doi: 10.1186/s12859-017-1960-x PMCID: PMC5751401

96: Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review Matthew J. Page, David Moher Syst Rev. 2017; 6: 263. Published online 2017 Dec 19. doi: 10.1186/s13643-017-0663-8 PMCID: PMC5738221

97: Prevalence and pattern of co-occurring musculoskeletal pain and its association with back-related disability among people with persistent low back pain: protocol for a systematic review and meta-analysis Cecilie K. Overaas, Melker S. Johansson, Tarcisio F. de Campos, Manuela L. Ferreira, Bard Natvig, Paul J. Mork, Jan Hartvigsen Syst Rev. 2017; 6: 258. Published online 2017 Dec 16. doi: 10.1186/s13643-017-0656-7 PMCID: PMC5732369

98: From public health genomics to precision public health: a 20-year journey Muin J. Khoury, M. Scott Bowen, Mindy Clyne, W. David Dotson, Marta L. Gwinn, Ridgely Fisk Green, Katherine Kolor, Juan L. Rodriguez, Anja Wulf, Wei Yu Genet Med. 2018 Jun; 20(6): 574–582. Published online 2017 Dec 14. doi: 10.1038/gim.2017.211 PMCID: PMC6384815

99: A common base method for analysis of qPCR data and the application of simple blocking in qPCR experiments Michael T. Ganger, Geoffrey D. Dietz, Sarah J. Ewing BMC Bioinformatics. 2017; 18: 534. Published online 2017 Dec 1. doi: 10.1186/s12859-017-1949-5 PMCID: PMC5709943

100: Expediting citation screening using PICo-based title-only screening for identifying studies in scoping searches and rapid reviews John Rathbone, Loai Albarqouni, Mina Bakhit, Elaine Beller, Oyungerel Byambasuren, Tammy Hoffmann, Anna Mae Scott, Paul Glasziou Syst Rev. 2017; 6: 233. Published online 2017 Nov 25. doi: 10.1186/s13643-017-0629-x PMCID: PMC5702220

101: Toward a comprehensive evidence map of overview of systematic review methods: paper 1—purpose, eligibility, search and data extraction Carole Lunny, Sue E. Brennan, Steve McDonald, Joanne E. McKenzie Syst Rev. 2017; 6: 231. Published online 2017 Nov 21. doi: 10.1186/s13643-017-0617-1 PMCID: PMC5698938

102: Sequence-based information-theoretic features for gene essentiality prediction Dawit Nigatu, Patrick Sobetzko, Malik Yousef, Werner Henkel BMC Bioinformatics. 2017; 18: 473. Published online 2017 Nov 9. doi: 10.1186/s12859-017-1884-5 PMCID: PMC5679510

103: Paramedic literature search filters: optimised for clinicians and academics Alexander Olaussen, William Semple, Alaa Oteir, Paula Todd, Brett Williams BMC Med Inform Decis Mak. 2017; 17: 146. Published online 2017 Oct 11. doi: 10.1186/s12911-017-0544-z PMCID: PMC5637081

104: Bayesian Unidimensional Scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations Lan Huong Nguyen, Susan Holmes BMC Bioinformatics. 2017; 18(Suppl 10): 394. Published online 2017 Sep 13. doi: 10.1186/s12859-017-1790-x PMCID: PMC5606221

105: Reviewing clinical guideline development tools: features and characteristics Soudabeh Khodambashi, Øystein Nytrø BMC Med Inform Decis Mak. 2017; 17: 132. Published online 2017 Sep 4. doi: 10.1186/s12911-017-0530-5 PMCID: PMC5584508

106: Parallel multiple instance learning for extremely large histopathology image analysis Yan Xu, Yeshu Li, Zhengyang Shen, Ziwei Wu, Teng Gao, Yubo Fan, Maode Lai, Eric I-Chao Chang BMC Bioinformatics. 2017; 18: 360. Published online 2017 Aug 3. doi: 10.1186/s12859-017-1768-8 PMCID: PMC5543478

107: A semi-supervised approach using label propagation to support citation screening Georgios Kontonatsios, Austin J. Brockmeier, Piotr Przybyła, John McNaught, Tingting Mu, John Y.

Goulermas, Sophia Ananiadou J Biomed Inform. 2017 Aug; 72: 67–76. doi: 10.1016/j.jbi.2017.06.018 PMCID: PMC5726085

108: Automatic Identification of High Impact Articles in PubMed to Support Clinical Decision Making Jiantao Bian, Mohammad Amin Morid, Siddhartha Jonnalagadda, Gang Luo, Guilherme Del Fiol J Biomed Inform. Author manuscript; available in PMC 2018 Sep 1.Published in final edited form as: J Biomed Inform. 2017 Sep; 73: 95–103. Published online 2017 Jul 26. doi: 10.1016/j.jbi.2017.07.015 PMCID: PMC5583030

109: Selecting and implementing overview methods: implications from five exemplar overviews Alex Pollock, Pauline Campbell, Ginny Brunton, Harriet Hunt, Lise Estcourt Syst Rev. 2017; 6: 145. Published online 2017 Jul 18. doi: 10.1186/s13643-017-0534-3 PMCID: PMC5516331

110: Recurrent Neural Networks for Classifying Relations in Clinical Notes Yuan Luo J Biomed Inform. 2017 Aug; 72: 85–95. Published online 2017 Jul 8. doi: 10.1016/j.jbi.2017.07.006 PMCID: PMC6657689

111: Is health coaching effective in changing the health status and behaviour of prisoners?—a systematic review protocol Nadja Almondes, Denise Downie, Ayse B. Cinar, Derek Richards, Ruth Freeman Syst Rev. 2017; 6: 127. Published online 2017 Jul 3. doi: 10.1186/s13643-017-0524-5 PMCID: PMC5496214

112: An investigation of the impact of using different methods for network meta-analysis: a protocol for an empirical evaluation Amalia (Emily) Karahalios, Georgia Salanti, Simon L. Turner, G. Peter Herbison, Ian R. White, Areti Angeliki Veroniki, Adriani Nikolakopoulou, Joanne E. Mckenzie Syst Rev. 2017; 6: 119. Published online 2017 Jun 24. doi: 10.1186/s13643-017-0511-x PMCID: PMC5483272

113: Predicting Mental Conditions Based on "History of Present Illness" in Psychiatric Notes with Deep Neural Networks Tung Tran, Ramakanth Kavuluru J Biomed Inform. Author manuscript; available in PMC 2018 Nov 1.Published in final edited form as: J Biomed Inform. 2017 Nov; 75 Suppl: S138–S148. Published online 2017 Jun 10. doi: 10.1016/j.jbi.2017.06.010 PMCID: PMC5705423

114: Evidence appraisal: a scoping review, conceptual framework, and research agenda Andrew Goldstein, Eric Venker, Chunhua Weng J Am Med Inform Assoc. 2017 Nov; 24(6): 1192–1203. Published online 2017 May 24. doi: 10.1093/jamia/ocx050 PMCID: PMC6259661

115: Impact of a workplace 'sit less, move more' program on efficiency-related outcomes of office employees Anna Puig-Ribera, Judit Bort-Roig, Maria Giné-Garriga, Angel M. González-

Suárez, Iván Martínez-Lemos, Jesús Fortuño, Joan C. Martori, Laura Muñoz-Ortiz, Raimon Milà, Nicholas D. Gilson, Jim McKenna BMC Public Health. 2017; 17: 455. Published online 2017 May 16. doi: 10.1186/s12889-017-4367-8 PMCID: PMC5434625

116: Developing a Framework for Digital Objects in the Big Data to Knowledge (BD2K) Commons: Report from the Commons Framework Pilots Workshop Kathleen M. Jagodnik, Simon Koplev, Sherry Jenkins, Lucila Ohno-Machado, Benedict Paten, Stephan C. Schurer, Michel Dumontier, Ruben Verborgh, Alex Bui, Peipei Ping, Neil J. McKenna, Ravi Madduri, Ajay Pillai, Avi Ma'ayan J Biomed Inform. Author manuscript; available in PMC 2018 Jul 1.Published in final edited form as: J Biomed Inform. 2017 Jul; 71: 49–57. Published online 2017 May 10. doi: 10.1016/j.jbi.2017.05.006 PMCID: PMC5545976

117: Intervention in prediction measure: a new approach to assessing variable importance for random forests Irene Epifanio BMC Bioinformatics. 2017; 18: 230. Published online 2017 May 2. doi: 10.1186/s12859-017-1650-8 PMCID: PMC5414143

118: Evaluation of a community-based hypertension improvement program (ComHIP) in Ghana: data from a baseline survey Peter Lamptey, Amos Laar, Alma J. Adler, Rebecca Dirks, Aya Caldwell, David Prieto-Merino, Ann Aerts, Neil Pearce, Pablo Perel BMC Public Health. 2017; 17: 368. Published online 2017 Apr 28. doi: 10.1186/s12889-017-4260-5 PMCID: PMC5410035

119: Health care public reporting utilization – user clusters, web trails, and usage barriers on Germany's public reporting portal Weisse-Liste.de Christoph Pross, Lars-Henrik Averdunk, Josip Stjepanovic, Reinhard Busse, Alexander Geissler BMC Med Inform Decis Mak. 2017; 17: 48. Published online 2017 Apr 21. doi: 10.1186/s12911-017-0440-6 PMCID: PMC5399803

120: Using classification models for the generation of disease-specific medications from biomedical literature and clinical data repository Liqin Wang, Peter J. Haug, Guilherme Del Fiol J Biomed Inform. Author manuscript; available in PMC 2018 May 1.Published in final edited form as: J Biomed Inform. 2017 May; 69: 259–266. Published online 2017 Apr 20. doi: 10.1016/j.jbi.2017.04.014 PMCID: PMC5509335

121: Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence Quan Nha Hong, Pierre Pluye, Mathieu Bujold, Maggy Wassef Syst Rev. 2017; 6: 61. Published online 2017 Mar 23. doi: 10.1186/s13643-017-0454-2 PMCID: PMC5364694

122: Embedding of Semantic Predications Trevor Cohen, Dominic Widdows J Biomed Inform. Author manuscript; available in PMC 2018 Apr 1.Published in final edited form as: J Biomed Inform. 2017 Apr; 68: 150–166. Published online 2017 Mar 8. doi: 10.1016/j.jbi.2017.03.003 PMCID: PMC5441848

123: Unsupervised ensemble ranking of terms in electronic health record notes based on their importance to patients Jinying Chen, Hong Yu J Biomed Inform. Author manuscript; available in PMC 2018 Apr 1.Published in final edited form as: J Biomed Inform. 2017 Apr; 68: 121–131. Published online 2017 Mar 4. doi: 10.1016/j.jbi.2017.02.016 PMCID: PMC5505865

124: Information needs for making clinical recommendations about potential drug-drug interactions: a synthesis of literature review and interviews Katrina M. Romagnoli, Scott D. Nelson, Lisa Hines, Philip Empey, Richard D. Boyce, Harry Hochheiser BMC Med Inform Decis Mak. 2017; 17: 21. Published online 2017 Feb 22. doi: 10.1186/s12911-017-0419-3 PMCID: PMC5322613

125: Digital health system for personalised COPD long-term management Carmelo Velardo, Syed Ahmar Shah, Oliver Gibson, Gari Clifford, Carl Heneghan, Heather Rutter, Andrew Farmer, Lionel Tarassenko, on behalf of the EDGE COPD Team BMC Med Inform Decis Mak. 2017; 17: 19. Published online 2017 Feb 20. doi: 10.1186/s12911-017-0414-8 PMCID: PMC5319140

126: Documenting research with transgender and gender diverse people: protocol for an evidence map and thematic analysis Zack Marshall, Vivian Welch, James Thomas, Fern Brunger, Michelle Swab, Ian Shemilt, Chris Kaposy Syst Rev. 2017; 6: 35. Published online 2017 Feb 20. doi: 10.1186/s13643-017-0427-5 PMCID: PMC5319144

127: Fit for purpose: perspectives on rapid reviews from end-user interviews Lisa Hartling, Jeanne-Marie Guise, Susanne Hempel, Robin Featherstone, Matthew D. Mitchell, Makalapua L. Motu'apuaka, Karen A. Robinson, Karen Schoelles, Annette Totten, Evelyn Whitlock, Timothy J. Wilt, Johanna Anderson, Elise Berliner, Aysegul Gozu, Elisabeth Kato, Robin Paynter, Craig A. Umscheid Syst Rev. 2017; 6: 32. Published online 2017 Feb 17. doi: 10.1186/s13643-017-0425-7 PMCID: PMC5316162

128: Modeling and Validating HL7 FHIR Profiles Using Semantic Web Shape Expressions (ShEx) Harold R. Solbrig, Eric Prud'hommeaux, Grahame Grieve, Lloyd McKenzie, Joshua C. Mandel, Deepak K. Sharma, Guoqian Jiang J Biomed Inform. Author manuscript; available in PMC 2018 Mar 1.Published in final edited form as: J Biomed Inform. 2017 Mar; 67: 90–100. Published online 2017 Feb 16. doi: 10.1016/j.jbi.2017.02.009 PMCID: PMC5502481

129: Comparative visualization of protein secondary structures Lucia Kocincová, Miroslava Jarešová, Jan Byška, Július Parulek, Helwig Hauser, Barbora Kozlíková BMC Bioinformatics. 2017; 18(Suppl 2): 23. Published online 2017 Feb 15. doi: 10.1186/s12859-016-1449-z PMCID: PMC5333176

130: Analyzing Structural Changes in SNOMED CT's Bacterial Infectious Diseases Using a Visual Semantic Delta Christopher Ochs, James T. Case, Yehoshua Perl J Biomed Inform. Author manuscript; available in PMC 2018 Mar 1.Published in final edited form as: J Biomed Inform. 2017 Mar; 67: 101–116. Published online 2017 Feb 12. doi: 10.1016/j.jbi.2017.02.006 PMCID: PMC5407457

131: Feasibility of Extracting Key Elements from ClinicalTrials.gov to Support Clinicians' Patient Care Decisions Heejun Kim, Jiantao Bian, Javed Mostafa, Siddhartha Jonnalagadda, Guilherme Del Fiol AMIA Annu Symp Proc. 2016; 2016: 705–714. Published online 2017 Feb 10. PMCID: PMC5333330

132: Automatic data source identification for clinical trial eligibility criteria resolution Chaitanya Shivade, Courtney Hebert, Kelly Regan, Eric Fosler-Lussier, Albert M. Lai AMIA Annu Symp Proc. 2016; 2016: 1149–1158. Published online 2017 Feb 10. PMCID: PMC5333255

133: Improving Endpoint Detection to Support Automated Systematic Reviews Ana Lucic, Catherine L. Blake AMIA Annu Symp Proc. 2016; 2016: 1900–1909. Published online 2017 Feb 10. PMCID: PMC5333237

134: Mental Status Documentation: Information Quality and Data Processes Charlene Weir, Bryan Gibson, Teresa Taft, Stacey Slager, Lacey Lewis, Nancy Staggers AMIA Annu Symp Proc. 2016; 2016: 1219–1228. Published online 2017 Feb 10. PMCID: PMC5333230

135: Differentiating Sense through Semantic Interaction Data T. Elizabeth Workman, Charlene Weir, Thomas C. Rindflesch AMIA Annu Symp Proc. 2016; 2016: 1238–1247. Published online 2017 Feb 10. PMCID: PMC5333208

136: Pragmatic trial of an intervention to increase human papillomavirus vaccination in safety-net clinics Maureen Sanderson, Juan R. Canedo, Dineo Khabele, Mary K. Fadden, Cynthia Harris, Katina Beard, Marilyn Burress, Helen Pinkerton, Cynthia Jackson, Tilicia Mayo-Gamble, Margaret K. Hargreaves, Pamela C. Hull BMC Public Health. 2017; 17: 158. Published online 2017 Feb 2. doi: 10.1186/s12889-017-4094-1 PMCID: PMC5290601

137: Simulated case management of home telemonitoring to assess the impact of different alert algorithms on work-load and clinical decisions Illapha Cuba Gyllensten, Amanda Crundall-Goode,

Ronald M. Aarts, Kevin M. Goode BMC Med Inform Decis Mak. 2017; 17: 11. Published online 2017 Jan 17. doi: 10.1186/s12911-016-0398-9 PMCID: PMC5240411

138: Physicians' Perception of Alternative Displays of Clinical Research Evidence for Clinical Decision Support–A Study with Case Vignettes Stacey L. Slager, Charlene R. Weir, Heejun Kim, Javed Mostafa, Guilherme Del Fiol J Biomed Inform. Author manuscript; available in PMC 2018 Jul 1.Published in final edited form as: J Biomed Inform. 2017 Jul; 71 Suppl: S53–S59. Published online 2017 Jan 13. doi: 10.1016/j.jbi.2017.01.007 PMCID: PMC5509533

139: GTB – an online genome tolerance browser Hashem A. Shihab, Mark F. Rogers, Michael Ferlaino, Colin Campbell, Tom R. Gaunt BMC Bioinformatics. 2017; 18: 20. Published online 2017 Jan 6. doi: 10.1186/s12859-016-1436-4 PMCID: PMC5219737

140: Bioinformatics and systems biology research update from the 15th International Conference on Bioinformatics (InCoB2016) Christian Schönbach, Chandra Verma, Peter J. Bond, Shoba Ranganathan BMC Bioinformatics. 2016; 17(Suppl 19): 524. Published online 2016 Dec 22. doi: 10.1186/s12859-016-1409-7 PMCID: PMC5259976

141: Accuracy of an Automated Knowledge Base for Identifying Drug Adverse Reactions EA Voss, RD Boyce, PB Ryan, J van der Lei, PR Rijnbeek, MJ Schuemie J Biomed Inform. Author manuscript; available in PMC 2018 Feb 1.Published in final edited form as: J Biomed Inform. 2017 Feb; 66: 72–81. Published online 2016 Dec 16. doi: 10.1016/j.jbi.2016.12.005 PMCID: PMC5316295

142: Web-based infectious disease surveillance systems and public health perspectives: a systematic review Jihye Choi, Youngtae Cho, Eunyoung Shim, Hyekyung Woo BMC Public Health. 2016; 16: 1238. Published online 2016 Dec 8. doi: 10.1186/s12889-016-3893-0 PMCID: PMC5146908

143: From comorbidities of chronic obstructive pulmonary disease to identification of shared molecular mechanisms by data integration David Gomez-Cabrero, Jörg Menche, Claudia Vargas, Isaac Cano, Dieter Maier, Albert-László Barabási, Jesper Tegnér, Josep Roca, on behalf of Synergy-COPD Consortia BMC Bioinformatics. 2016; 17(Suppl 15): 23–35. Published online 2016 Nov 22. doi: 10.1186/s12859-016-1291-3 PMCID: PMC5133493

144: Generating disease-pertinent treatment vocabularies from MEDLINE citations Liqin Wang, Guilherme Del Fiol, Bruce E. Bray, Peter J. Haug J Biomed Inform. Author manuscript; available in PMC 2017 Sep 7.Published in final edited form as: J Biomed Inform. 2017 Jan; 65: 46–57. Published online 2016 Nov 16. doi: 10.1016/j.jbi.2016.11.004 PMCID: PMC5588694

145: Exploring issues in the conduct of website searching and other online sources for systematic reviews: how can we be systematic? Claire Stansfield, Kelly Dickson, Mukdarut Bangpan Syst Rev. 2016; 5: 191. Published online 2016 Nov 15. doi: 10.1186/s13643-016-0371-9 PMCID: PMC5111285

146: Cochrane Rapid Reviews Methods Group to play a leading role in guiding the production of informed high-quality, timely research evidence syntheses Chantelle Garritty, Adrienne Stevens, Gerald Gartlehner, Valerie King, Chris Kamel, on behalf of the Cochrane Rapid Reviews Methods Group Syst Rev. 2016; 5: 184. Published online 2016 Oct 28. doi: 10.1186/s13643-016-0360-z PMCID: PMC5084365

147: Extractive text summarization system to aid data extraction from full text in systematic review development Duy Duc An Bui, Guilherme Del Fiol, John F. Hurdle, Siddhartha Jonnalagadda J Biomed Inform. Author manuscript; available in PMC 2017 Dec 1.Published in final edited form as: J Biomed Inform. 2016 Dec; 64: 265–272. Published online 2016 Oct 27. doi: 10.1016/j.jbi.2016.10.014 PMCID: PMC5362293

148: Using the Intervention Mapping protocol to develop a family-based intervention for improving lifestyle habits among overweight and obese children: study protocol for a quasi-experimental trial Tonje Holte Stea, Tommy Haugen, Sveinung Berntsen, Vigdis Guttormsen, Nina Cecilie Øverby, Kristin Haraldstad, Eivind Meland, Eirik Abildsnes BMC Public Health. 2016; 16: 1092. Published online 2016 Oct 18. doi: 10.1186/s12889-016-3766-6 PMCID: PMC5070224

149: School-based sexual health education interventions to prevent STI/HIV in sub-Saharan Africa: a systematic review and meta-analysis A. Sadiq Sani, Charles Abraham, Sarah Denford, Susan Ball BMC Public Health. 2016; 16: 1069. Published online 2016 Oct 10. doi: 10.1186/s12889-016-3715-4 PMCID: PMC5057258

150: Human resource information systems in health care: a systematic evidence review Aizhan Tursunbayeva, Raluca Bunduchi, Massimo Franco, Claudia Pagliari J Am Med Inform Assoc. 2017 May; 24(3): 633–654. Published online 2016 Oct 5. doi: 10.1093/jamia/ocw141 PMCID: PMC5391731

151: Searching and synthesising ‘grey literature’ and ‘grey information’ in public health: critical reflections on three case studies Jean Adams, Frances C. Hillier-Brown, Helen J. Moore, Amelia A. Lake, Vera Araujo-Soares, Martin White, Carolyn Summerbell Syst Rev. 2016; 5: 164. Published online 2016 Sep 29. doi: 10.1186/s13643-016-0337-y PMCID: PMC5041336

152: A hybrid computational strategy to address WGS variant analysis in >5000 samples Zhuoyi Huang, Navin Rustagi, Narayanan Veeraraghavan, Andrew Carroll, Richard Gibbs, Eric Boerwinkle, Manjunath Gorentla Venkata, Fuli Yu BMC Bioinformatics. 2016; 17(1): 361. Published online 2016 Sep 10. doi: 10.1186/s12859-016-1211-6 PMCID: PMC5018196

153: Veterans Like Me: Formative evaluation of a patient decision aid design Bryan Gibson DPT, Jorie Butler, Katherine Doyon, Lee Ellington, Bruce E. Bray, Qing Zeng J Biomed Inform. Author manuscript; available in PMC 2017 Aug 12.Published in final edited form as: J Biomed Inform. 2017 Jul; 71 Suppl: S46–S52. Published online 2016 Sep 10. doi: 10.1016/j.jbi.2016.09.007 PMCID: PMC5513765

154: A Part-Of-Speech Term Weighting Scheme for Biomedical Information Retrieval Yanshan Wang, Stephen Wu, Dingcheng Li, Saeed Mehrabi, Hongfang Liu J Biomed Inform. Author manuscript; available in PMC 2017 Oct 1.Published in final edited form as: J Biomed Inform. 2016 Oct; 63: 379–389. Published online 2016 Sep 1. doi: 10.1016/j.jbi.2016.08.026 PMCID: PMC5493484

155: Health coaching interventions for persons with chronic conditions: a systematic review and meta-analysis protocol Kasey R. Boehmer, Suzette Barakat, Sangwoo Ahn, Larry J. Prokop, Patricia J. Erwin, M. Hassan Murad Syst Rev. 2016; 5(1): 146. Published online 2016 Sep 1. doi: 10.1186/s13643-016-0316-3 PMCID: PMC5009492

156: Systematic review of brucellosis in Kenya: disease frequency in humans and animals and risk factors for human infection J. Njeru, G. Wareth, F. Melzer, K. Henning, M. W. Pletz, R. Heller, H. Neubauer BMC Public Health. 2016; 16(1): 853. Published online 2016 Aug 22. doi: 10.1186/s12889-016-3532-9 PMCID: PMC4994226

157: Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews Ian Shemilt, Nada Khan, Sophie Park, James Thomas Syst Rev. 2016; 5: 140. Published online 2016 Aug 17. doi: 10.1186/s13643-016-0315-4 PMCID: PMC4989498

158: Does education level affect the efficacy of a community based salt reduction program? - A post-hoc analysis of the China Rural Health Initiative Sodium Reduction Study (CRHI-SRS) Xin Wang, Xian Li, Ilonca Vaartjes, Bruce Neal, Michiel L. Bots, Arno W. Hoes, Yangfeng Wu BMC Public Health. 2016; 16: 759. Published online 2016 Aug 11. doi: 10.1186/s12889-016-3454-6 PMCID: PMC4982434

159: Eliciting women's cervical screening preferences: a mixed methods systematic review protocol Brianne Wood, Susan Rogers Van Katwyk, Ziad El-Khatib, Susan McFaul, Monica Taljaard,

Erica Wright, Ian D. Graham, Julian Little Syst Rev. 2016; 5: 136. Published online 2016 Aug 11. doi: 10.1186/s13643-016-0310-9 PMCID: PMC4982264

160: User survey finds rapid evidence reviews increased uptake of evidence by Veterans Health Administration leadership to inform fast-paced health-system decision-making Kim Peterson, Nicole Floyd, Lauren Ferguson, Vivian Christensen, Mark Helfand Syst Rev. 2016; 5: 132. Published online 2016 Aug 5. doi: 10.1186/s13643-016-0306-5 PMCID: PMC4974754

161: Born in Bradford's Better Start: an experimental birth cohort study to evaluate the impact of early life interventions Josie Dickerson, Philippa K. Bird, Rosemary R. C. McEachan, Kate E. Pickett, Dagmar Waiblinger, Eleonora Uphoff, Dan Mason, Maria Bryant, Tracey Bywater, Claudine Bowyer-Crane, Pinki Sahota, Neil Small, Michaela Howell, Gill Thornton, Melanie Astin, Debbie A. Lawlor, John Wright BMC Public Health. 2016; 16(1): 711. Published online 2016 Aug 4. doi: 10.1186/s12889-016-3318-0 PMCID: PMC4996273

162: Development of an open-source web-based intervention for Brazilian smokers – Viva sem Tabaco H. P. Gomide, H. S. Bernardino, K. Richter, L. F. Martins, T. M. Ronzani BMC Med Inform Decis Mak. 2016; 16: 103. Published online 2016 Aug 2. doi: 10.1186/s12911-016-0339-7 PMCID: PMC4970282

163: Topic detection using paragraph vectors to support active learning in systematic reviews Kazuma Hashimoto, Georgios Kontonatsios, Makoto Miwa, Sophia Ananiadou J Biomed Inform. 2016 Aug; 62: 59–65. doi: 10.1016/j.jbi.2016.06.001 PMCID: PMC4981645

164: Protocol of the impact of alternative social assistance disbursement on drug-related harm (TASA) study: a randomized controlled trial to evaluate changes to payment timing and frequency among people who use illicit drugs Lindsey Richardson, Allison Laing, M-J Milloy, Russ Maynard, Bohdan Nosyk, Brandon Marshall, Eric Grafstein, Patricia Daly, Evan Wood, Julio Montaner, Thomas Kerr BMC Public Health. 2016; 16: 668. Published online 2016 Jul 29. doi: 10.1186/s12889-016-3304-6 PMCID: PMC4966816

165: The eClinical Care Pathway Framework: a novel structure for creation of online complex clinical care pathways and its application in the management of sexually transmitted infections Jo Gibbs, Lorna J. Sutcliffe, Voula Gkatzidou, Kate Hone, Richard E. Ashcroft, Emma M. Harding-Esch, Catherine M. Lowndes, S. Tariq Sadiq, Pam Sonnenberg, Claudia S. Estcourt BMC Med Inform Decis Mak. 2016; 16: 98. Published online 2016 Jul 22. doi: 10.1186/s12911-016-0338-8 PMCID: PMC4957844

166: Impacts of casinos on key pathways to health: qualitative findings from American Indian gaming communities in California Stephen R. Kodish, Joel Gittelsohn, Vanessa M. Oddo, Jessica C. Jones-Smith BMC Public Health. 2016; 16: 621. Published online 2016 Jul 22. doi: 10.1186/s12889-016-3279-3 PMCID: PMC4957391

167: Workflow-driven clinical decision support for personalized oncology Anca Bucur, Jasper van Leeuwen, Nikolaos Christodoulou, Kamana Sigdel, Katerina Argyri, Lefteris Koumakis, Norbert Graf, Georgios Stamatakos BMC Med Inform Decis Mak. 2016; 16(Suppl 2): 87. Published online 2016 Jul 21. doi: 10.1186/s12911-016-0314-3 PMCID: PMC4965727

168: A knowledge base for tracking the impact of genomics on population health Wei Yu, Marta Gwinn, W. David Dotson, Ridgely Fisk Green, Mindy Clyne, Anja Wulf, Scott Bowen, Katherine Kolor, Muin J. Khoury Genet Med. Author manuscript; available in PMC 2017 Dec 1.Published in final edited form as: Genet Med. 2016 Dec; 18(12): 1312–1314. Published online 2016 Jun 9. doi: 10.1038/gim.2016.63 PMCID: PMC5133140

169: How to improve vital sign data quality for use in clinical decision support systems? A qualitative study in nine Swedish emergency departments Niclas Skyttberg, Joana Vicente, Rong Chen, Hans Blomqvist, Sabine Koch BMC Med Inform Decis Mak. 2016; 16: 61. Published online 2016 Jun 4. doi: 10.1186/s12911-016-0305-4 PMCID: PMC4893236

170: Reengineering of MeSH thesauri for term selection to optimize literature retrieval and knowledge reconstruction in support of stem cell research Yan Su, James Andrews, Hong Huang, Yue Wang, Liangliang Kong, Peter Cannon, Ping Xu BMC Med Inform Decis Mak. 2016; 16: 54. Published online 2016 May 23. doi: 10.1186/s12911-016-0298-z PMCID: PMC4878086

171: Using qualitative comparative analysis in a systematic review of a complex intervention Leila Kahwati, Sara Jacobs, Heather Kane, Megan Lewis, Meera Viswanathan, Carol E. Golin Syst Rev. 2016; 5: 82. Published online 2016 May 4. doi: 10.1186/s13643-016-0256-y PMCID: PMC4875617

172: Searching for qualitative research for inclusion in systematic reviews: a structured methodological review Andrew Booth Syst Rev. 2016; 5: 74. Published online 2016 May 4. doi: 10.1186/s13643-016-0249-x PMCID: PMC4855695

173: Automated identification of molecular effects of drugs (AIMED) Safa Fathiamini, Amber M Johnson, Jia Zeng, Alejandro Araya, Vijaykumar Holla, Ann M Bailey, Beate C Litzenburger, Nora S Sanchez, Yekaterina Khotskaya, Hua Xu, Funda Meric-Bernstam, Elmer V Bernstam, Trevor

Cohen J Am Med Inform Assoc. 2016 Jul; 23(4): 758–765. Published online 2016 Apr 23. doi: 10.1093/jamia/ocw030 PMCID: PMC4926748

174: Leveraging dialog systems research to assist biomedical researchers' interrogation of Big Clinical Data Julia Hoxha, Chunhua Weng J Biomed Inform. Author manuscript; available in PMC 2018 Mar 29.Published in final edited form as: J Biomed Inform. 2016 Jun; 61: 176–184. Published online 2016 Apr 8. doi: 10.1016/j.jbi.2016.04.003 PMCID: PMC5875984

175: Disparities in a provision of in-hospital post-arrest interventions for out-of-hospital cardiac arrest (OHCA) in the elderly population—protocol for a systematic review Joanna M. Bielecki, Josephine Wong, Nicholas Mitsakakis, Prakesh S. Shah, Murray D. Krahn, Valeria E. Rac Syst Rev. 2016; 5: 55. Published online 2016 Apr 7. doi: 10.1186/s13643-016-0234-4 PMCID: PMC4853855

176: Improving the Utility of MeSH® Terms Using the TopicalMeSH Representation Zhiguo Yu, Elmer Bernstam, Trevor Cohen, Byron C. Wallace, Todd R. Johnson J Biomed Inform. Author manuscript; available in PMC 2017 Jun 1.Published in final edited form as: J Biomed Inform. 2016 Jun; 61: 77–86. Published online 2016 Mar 19. doi: 10.1016/j.jbi.2016.03.013 PMCID: PMC4893983

177: Adapting machine learning techniques to censored time-to-event health record data: a general-purpose approach using inverse probability of censoring weighting David M. Vock, Julian Wolfson, Sunayan Bandyopadhyay, Gediminas Adomavicius, Paul E. Johnson, Gabriela Vazquez-Benitez, Patrick J. O'Connor J Biomed Inform. Author manuscript; available in PMC 2017 Jun 1.Published in final edited form as: J Biomed Inform. 2016 Jun; 61: 119–131. Published online 2016 Mar 16. doi: 10.1016/j.jbi.2016.03.009 PMCID: PMC4893987

178: Outbreak definition by change point analysis: a tool for public health decision? Gaëtan Texier, Magnim Farouh, Liliane Pellegrin, Michael L. Jackson, Jean-Baptiste Meynard, Xavier Deparis, Hervé Chaudet BMC Med Inform Decis Mak. 2016; 16: 33. Published online 2016 Mar 12. doi: 10.1186/s12911-016-0271-x PMCID: PMC4788889

179: Multiplex methods provide effective integration of multi-omic data in genome-scale models Claudio Angione, Max Conway, Pietro Lió BMC Bioinformatics. 2016; 17(Suppl 4): 83. Published online 2016 Mar 2. doi: 10.1186/s12859-016-0912-1 PMCID: PMC4896256

180: How to promote adverse drug reaction reports using information systems – a systematic review and meta-analysis Inês Ribeiro-Vaz, Ana-Marta Silva, Cristina Costa Santos, Ricardo

Cruz-Correia BMC Med Inform Decis Mak. 2016; 16: 27. Published online 2016 Mar 1. doi: 10.1186/s12911-016-0265-8 PMCID: PMC4772685

181: Factors influencing completion of multi-dose vaccine schedules in adolescents: a systematic review K. E. Gallagher, E. Kadokura, L. O. Eckert, S. Miyake, S. Mounier-Jack, M. Aldea, D. A. Ross, D. Watson-Jones BMC Public Health. 2016; 16: 172. Published online 2016 Feb 19. doi: 10.1186/s12889-016-2845-z PMCID: PMC4759915

182: A multiple kernel learning algorithm for drug-target interaction prediction André C. A. Nascimento, Ricardo B. C. Prudêncio, Ivan G. Costa BMC Bioinformatics. 2016; 17: 46. Published online 2016 Jan 22. doi: 10.1186/s12859-016-0890-3 PMCID: PMC4722636

183: Use of network meta-analysis in systematic reviews: a survey of authors Andrew W. Lee Syst Rev. 2016; 5: 8. Published online 2016 Jan 19. doi: 10.1186/s13643-015-0174-4 PMCID: PMC4719661

184: Automated parameter estimation for biological models using Bayesian statistical model checking Faraz Hussain, Christopher J Langmead, Qi Mi, Joyeeta Dutta-Moscato, Yoram Vodovotz, Sumit K Jha BMC Bioinformatics. 2015; 16(Suppl 17): S8. Published online 2015 Dec 7. doi: 10.1186/1471-2105-16-S17-S8 PMCID: PMC4674867

185: Supporting systematic reviews using LDA-based document representations Yuanhan Mo, Georgios Kontonatsios, Sophia Ananiadou Syst Rev. 2015; 4: 172. Published online 2015 Nov 26. doi: 10.1186/s13643-015-0117-0 PMCID: PMC4662004

186: RESTORE: an exploratory trial of a web-based intervention to enhance self-management of cancer-related fatigue: findings from a qualitative process evaluation Michelle Myall, Carl R. May, Chloe Grimmett, Christine M. May, Lynn Calman, Alison Richardson, Claire L. Foster BMC Med Inform Decis Mak. 2015; 15: 94. Published online 2015 Nov 14. doi: 10.1186/s12911-015-0214-y PMCID: PMC4650501

187: Physical Activity through Sustainable Transport Approaches (PASTA): protocol for a multicentre, longitudinal study Evi Dons, Thomas Götschi, Mark Nieuwenhuijsen, Audrey de Nazelle, Esther Anaya, Ione Avila-Palencia, Christian Brand, Tom Cole-Hunter, Mailin Gaupp-Berghausen, Sonja Kahlmeier, Michelle Laeremans, Natalie Mueller, Juan Pablo Orjuela, Elisabeth Raser, David Rojas-Rueda, Arnout Standaert, Erik Stigell, Tina Uhlmann, Regine Gerike, Luc Int Panis BMC Public Health. 2015; 15: 1126. Published online 2015 Nov 14. doi: 10.1186/s12889-015-2453-3 PMCID: PMC4650276

188: How to conduct systematic reviews more expeditiously? Alexander Tsertsvadze, Yen-Fu Chen, David Moher, Paul Sutcliffe, Noel McCarthy Syst Rev. 2015; 4: 160. Published online 2015 Nov 12. doi: 10.1186/s13643-015-0147-7 PMCID: PMC4643500

189: Automatic Classification of Structured Product Labels for Pregnancy Risk Drug Categories, a Machine Learning Approach Laritza M. Rodriguez, Dina Demner Fushman AMIA Annu Symp Proc. 2015; 2015: 1093–1102. Published online 2015 Nov 5. PMCID: PMC4765680

190: Classification of Clinically Useful Sentences in MEDLINE Mohammad Amin Morid, Siddhartha Jonnalagadda, Marcelo Fiszman, Kalpana Raja, Guilherme Del Fiol AMIA Annu Symp Proc. 2015; 2015: 2015–2024. Published online 2015 Nov 5. PMCID: PMC4765649

191: Examining the Distribution, Modularity, and Community Structure in Article Networks for Systematic Reviews Xiaonan Ji, Raghu Machiraju, Alan Ritter, Po-Yin Yen AMIA Annu Symp Proc. 2015; 2015: 1927–1936. Published online 2015 Nov 5. PMCID: PMC4765615

192: Recent Directions in Telemedicine: Review of Trends in Research and Practice Laurence S. Wilson, Anthony J. Maeder Healthc Inform Res. 2015 Oct; 21(4): 213–222. Published online 2015 Oct 31. doi: 10.4258/hir.2015.21.4.213 PMCID: PMC4659877

193: GOTA: GO term annotation of biomedical literature Pietro Di Lena, Giacomo Domeniconi, Luciano Margara, Gianluca Moro BMC Bioinformatics. 2015; 16: 346. Published online 2015 Oct 28. doi: 10.1186/s12859-015-0777-8 PMCID: PMC4625458

194: Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada Katelyn Godin, Jackie Stapleton, Sharon I. Kirkpatrick, Rhona M. Hanning, Scott T. Leatherdale Syst Rev. 2015; 4: 138. Published online 2015 Oct 22. doi: 10.1186/s13643-015-0125-0 PMCID: PMC4619264

195: Rapid Review Summit: an overview and initiation of a research agenda Julie Polisena, Chantelle Garritty, Craig A. Umscheid, Chris Kamel, Kevin Samra, Jeannette Smith, Ann Vosilla Syst Rev. 2015; 4: 111. Published online 2015 Sep 26. doi: 10.1186/s13643-015-0111-6 PMCID: PMC4583747

196: Automatically finding relevant citations for clinical guideline development Duy Duc An Bui, Siddhartha Jonnalagadda, Guilherme Del Fiol J Biomed Inform. Author manuscript; available in PMC 2016 Oct 1.Published in final edited form as: J Biomed Inform. 2015 Oct; 57: 436–445. Published online 2015 Sep 10. doi: 10.1016/j.jbi.2015.09.003 PMCID: PMC4786461

197: Amplitude spectrum distance: measuring the global shape divergence of protein fragments Clovis Galiez, François Coste BMC Bioinformatics. 2015; 16: 256. Published online 2015 Aug 14. doi: 10.1186/s12859-015-0693-y PMCID: PMC4535829

198: Automatic Detection of Protected Health Information from Clinic Narratives Hui Yang, Jonathan M. Garibaldi J Biomed Inform. Author manuscript; available in PMC 2016 Aug 18.Published in final edited form as: J Biomed Inform. 2015 Dec; 58(Suppl): S30–S38. Published online 2015 Jul 29. doi: 10.1016/j.jbi.2015.06.015 PMCID: PMC4989090

199: Details of development of the resource for adults with asthma in the RAISIN (randomized trial of an asthma internet self-management intervention) study Deborah Morrison, Frances S. Mair, Rekha Chaudhuri, Marilyn McGee-Lennon, Mike Thomas, Neil C. Thomson, Lucy Yardley, Sally Wyke BMC Med Inform Decis Mak. 2015; 15: 57. Published online 2015 Jul 28. doi: 10.1186/s12911-015-0177-z PMCID: PMC4517557

200: Examining the influence of a text message-based sleep and physical activity intervention among young adult smokers in the United States A. Jordan Filion, Gerarda Darlington, Jean-Philippe Chaput, Michele Ybarra, Jess Haines BMC Public Health. 2015; 15: 671. Published online 2015 Jul 16. doi: 10.1186/s12889-015-2045-2 PMCID: PMC4502521

201: Preparing a collection of radiology examinations for distribution and retrieval Dina Demner-Fushman, Marc D. Kohli, Marc B. Rosenman, Sonya E. Shooshan, Laritza Rodriguez, Sameer Antani, George R. Thoma, Clement J. McDonald J Am Med Inform Assoc. 2016 Mar; 23(2): 304–310. Published online 2015 Jul 1. doi: 10.1093/jamia/ocv080 PMCID: PMC5009925

202: Using an electronic activity monitor system as an intervention modality: A systematic review Zakkoyya H. Lewis, Elizabeth J. Lyons, Jessica M. Jarvis, Jacques Baillargeon BMC Public Health. 2015; 15: 585. Published online 2015 Jun 24. doi: 10.1186/s12889-015-1947-3 PMCID: PMC4479243

203: RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials Iain J Marshall, Joël Kuiper, Byron C Wallace J Am Med Inform Assoc. 2016 Jan; 23(1): 193–201. Published online 2015 Jun 23. doi: 10.1093/jamia/ocv044 PMCID: PMC4713900

204: Care episode retrieval: distributional semantic models for information retrieval in the clinical domain Hans Moen, Filip Ginter, Erwin Marsi, Laura-Maria Peltonen, Tapio Salakoski, Sanna Salanterä BMC Med Inform Decis Mak. 2015; 15(Suppl 2): S2. Published online 2015 Jun 15. doi: 10.1186/1472-6947-15-S2-S2 PMCID: PMC4474584

205: Louhi 2014: Special issue on health text mining and information analysis Sumithra Velupillai, Martin Duneld, Aron Henriksson, Maria Kvist, Maria Skeppstedt, Hercules Dalianis BMC Med Inform Decis Mak. 2015; 15(Suppl 2): S1. Published online 2015 Jun 15. doi: 10.1186/1472-6947-15-S2-S1 PMCID: PMC4474544

206: Automating data extraction in systematic reviews: a systematic review Siddhartha R. Jonnalagadda, Pawan Goyal, Mark D. Huffman Syst Rev. 2015; 4: 78. Published online 2015 Jun 15. doi: 10.1186/s13643-015-0066-7 PMCID: PMC4514954

207: Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers John Rathbone, Tammy Hoffmann, Paul Glasziou Syst Rev. 2015; 4: 80. Published online 2015 Jun 15. doi: 10.1186/s13643-015-0067-6 PMCID: PMC4472176

208: Innovating to enhance clinical data management using non-commercial and open source solutions across a multi-center network supporting inpatient pediatric care and research in Kenya Timothy Tuti, Michael Bitok, Chris Paton, Boniface Makone, Lucas Malla, Naomi Muinga, David Gathara, Mike English J Am Med Inform Assoc. 2016 Jan; 23(1): 184–192. Published online 2015 Jun 10. doi: 10.1093/jamia/ocv028 PMCID: PMC4681113

209: Effectiveness of a Smartphone application and wearable device for weight loss in overweight or obese primary care patients: protocol for a randomised controlled trial Esther Granado-Font, Gemma Flores-Mateo, Mar Sorlí-Aguilar, Xavier Montaña-Carreras, Carme Ferre-Grau, Maria-Luisa Barrera-Uriarte, Eulàlia Oriol-Colominas, Cristina Rey-Reñones, Iolanda Caules, Eva-María Satué-Gracia, OBSBIT Study Group BMC Public Health. 2015; 15: 531. Published online 2015 Jun 4. doi: 10.1186/s12889-015-1845-8 PMCID: PMC4455326

210: Patient-reported outcomes in a large community-based pain medicine practice: evaluation for use in phenotype modeling David A. Juckett, Fred N. Davis, Mark Gostine, Philip Reed, Rebecca Risko BMC Med Inform Decis Mak. 2015; 15: 41. Published online 2015 May 28. doi: 10.1186/s12911-015-0164-4 PMCID: PMC4446111

211: Weighted protein residue networks based on joint recurrences between residues Wael I. Karain, Nael I. Qaraeen BMC Bioinformatics. 2015; 16: 173. Published online 2015 May 26. doi: 10.1186/s12859-015-0621-1 PMCID: PMC4491895

212: Multiple perspectives on clinical decision support: a qualitative study of fifteen clinical and vendor organizations Joan S Ash, Dean F Sittig, Carmit K McMullen, Adam Wright, Arwen

Bunce, Vishnu Mohan, Deborah J Cohen, Blackford Middleton BMC Med Inform Decis Mak. 2015; 15: 35. Published online 2015 Apr 24. doi: 10.1186/s12911-015-0156-4 PMCID: PMC4447027

213: Conditions potentially sensitive to a Personal Health Record (PHR) intervention, a systematic review Morgan Price, Paule Bellwood, Nicole Kitson, Iryna Davies, Jens Weber, Francis Lau BMC Med Inform Decis Mak. 2015; 15: 32. Published online 2015 Apr 18. doi: 10.1186/s12911-015-0159-1 PMCID: PMC4411701

214: YOC, A new strategy for pairwise alignment of collinear genomes Raluca Uricaru, Célia Michotey, Hélène Chiapello, Eric Rivals BMC Bioinformatics. 2015; 16(1): 111. Published online 2015 Apr 2. doi: 10.1186/s12859-015-0530-3 PMCID: PMC4411659

215: USI: a fast and accurate approach for conceptual document annotation Nicolas Fiorini, Sylvie Ranwez, Jacky Montmain, Vincent Ranwez BMC Bioinformatics. 2015; 16: 83. Published online 2015 Mar 14. doi: 10.1186/s12859-015-0513-4 PMCID: PMC4367850

216: Metabolic syndrome in hypertensive adults from rural Northeast China: an update Shasha Yu, Xiaofan Guo, Hongmei Yang, Liqiang Zheng, Yingxian Sun BMC Public Health. 2015; 15: 247. Published online 2015 Mar 14. doi: 10.1186/s12889-015-1587-7 PMCID: PMC4367840

217: Maternal smoking and the risk of still birth: systematic review and meta-analysis Takawira C Marufu, Anand Ahankari, Tim Coleman, Sarah Lewis BMC Public Health. 2015; 15: 239. Published online 2015 Mar 13. doi: 10.1186/s12889-015-1552-5 PMCID: PMC4372174

218: MeSH ORA framework: R/Bioconductor packages to support MeSH over-representation analysis Koki Tsuyuzaki, Gota Morota, Manabu Ishii, Takeru Nakazato, Satoru Miyazaki, Itoshi Nikaido BMC Bioinformatics. 2015; 16: 45. Published online 2015 Feb 15. doi: 10.1186/s12859-015-0453-z PMCID: PMC4343279

219: Automated confidence ranked classification of randomized controlled trial articles: an aid to evidence-based medicine Aaron M Cohen, Neil R Smalheiser, Marian S McDonagh, Clement Yu, Clive E Adams, John M Davis, Philip S Yu J Am Med Inform Assoc. 2015 May; 22(3): 707–717. Published online 2015 Feb 5. doi: 10.1093/jamia/ocu025 PMCID: PMC4457112

220: A web-based intervention for abused women: the New Zealand isafe randomised controlled trial protocol Jane Koziol-McLain, Alain C Vandal, Shyamala Nada-Raja, Denise Wilson, Nancy E Glass, Karen B Eden, Christine McLean, Terry Dobbs, James Case BMC Public Health. 2015; 15: 56. Published online 2015 Jan 31. doi: 10.1186/s12889-015-1395-0 PMCID: PMC4314812

221: Using text mining for study identification in systematic reviews: a systematic review of current approaches Alison O'Mara-Eves, James Thomas, John McNaught, Makoto Miwa, Sophia Ananiadou Syst Rev. 2015; 4(1): 5. Published online 2015 Jan 14. doi: 10.1186/2046-4053-4-5Correction in: Syst Rev. 2015; 4: 59. PMCID: PMC4320539

222: Identifying tandem Ankyrin repeats in protein structures Broto Chakrabarty, Nita Parekh BMC Bioinformatics. 2014; 15(1): 6599. Published online 2014 Dec 30. doi: 10.1186/s12859-014-0440-9 PMCID: PMC4307672

223: Specific genotypes of human papillomavirus in 125 high-grade squamous lesions and invasive cervical cancer cases from Congolese women Luc Magloire Anicet Boumba, Lahoucine Hilali, Mustapha Mouallif, Donatien Moukassa, Moulay Mustapha Ennaji BMC Public Health. 2014; 14: 1320. Published online 2014 Dec 23. doi: 10.1186/1471-2458-14-1320 PMCID: PMC4391118

224: Design of a multi-site multi-state clinical trial of home monitoring of chronic disease in the community in Australia Branko G Celler, Ross Sparks, Surya Nepal, Leila Alem, Marlien Varnfield, Jane Li, Julian Jang-Jaccard, Simon J McBride, Rajiv Jayasena BMC Public Health. 2014; 14: 1270. Published online 2014 Dec 15. doi: 10.1186/1471-2458-14-1270 PMCID: PMC4320478

225: The remote exercise monitoring trial for exercise-based cardiac rehabilitation (REMOTE-CR): a randomised controlled trial protocol Ralph Maddison, Jonathan C Rawstorn, Anna Rolleston, Robyn Whittaker, Ralph Stewart, Jocelyne Benatar, Ian Warren, Yannan Jiang, Nicholas Gant BMC Public Health. 2014; 14: 1236. Published online 2014 Nov 28. doi: 10.1186/1471-2458-14-1236 PMCID: PMC4289048

226: Perceptions of sexual risk behavior among Palestinian youth in the West Bank: a qualitative investigation Salwa G Massad, Rita Karam, Ryan Brown, Peter Glick, Mohammed Shaheen, Sebastian Linnemayr, Umaiyeh Khammash BMC Public Health. 2014; 14: 1213. Published online 2014 Nov 24. doi: 10.1186/1471-2458-14-1213 PMCID: PMC4247558

227: Coverage of Rare Disease Names in Standard Terminologies and Implications for Patients, Providers, and Research Kin Wah Fung, Rachel Richesson, Olivier Bodenreider AMIA Annu Symp Proc. 2014; 2014: 564–572. Published online 2014 Nov 14. PMCID: PMC4419993

228: Extracting Concepts Related to Homelessness from the Free Text of VA Electronic Medical Records Adi V. Gundlapalli, Marjorie E. Carter, Guy Divita, Shuying Shen, Miland Palmer, Brett South, B.S. Begum Durgahee, Andrew Redd, Matthew Samore AMIA Annu Symp Proc. 2014; 2014: 589–598. Published online 2014 Nov 14. PMCID: PMC4419940

229: G-Bean: an ontology-graph based web tool for biomedical literature retrieval James Z Wang, Yuanyuan Zhang, Liang Dong, Lin Li, Pradip K Srimani, Philip S Yu BMC Bioinformatics. 2014; 15(Suppl 12): S1. Published online 2014 Nov 6. doi: 10.1186/1471-2105-15-S12-S1 PMCID: PMC4243180

230: Impact of long-term lifestyle programmes on weight loss and cardiovascular risk factors in overweight/obese participants: a systematic review and network meta-analysis Lukas Schwingshackl, Sofia Dias, Georg Hoffmann Syst Rev. 2014; 3: 130. Published online 2014 Oct 30. doi: 10.1186/2046-4053-3-130 PMCID: PMC4227972

231: Is planned adaptation to heat reducing heat-related mortality and illness? A systematic review Melanie Boeckmann, Ines Rohn BMC Public Health. 2014; 14: 1112. Published online 2014 Oct 28. doi: 10.1186/1471-2458-14-1112 PMCID: PMC4219109

232: Considering methodological options for reviews of theory: illustrated by a review of theories linking income and health Mhairi Campbell, Matt Egan, Theo Lorenc, Lyndal Bond, Frank Popham, Candida Fenton, Michaela Benzeval Syst Rev. 2014; 3: 114. Published online 2014 Oct 13. doi: 10.1186/2046-4053-3-114 PMCID: PMC4208031

233: An update on overweight and obesity in rural Northeast China: from lifestyle risk factors to cardiometabolic comorbidities Xiaofan Guo, Zhao Li, Liang Guo, Liqiang Zheng, Shasha Yu, Hongmei Yang, Lu Zou, Ying Zhou, Yaowen Zhang, Luoning Zhu, Yonghong Zhang, Yingxian Sun BMC Public Health. 2014; 14: 1046. Published online 2014 Oct 8. doi: 10.1186/1471-2458-14-1046 PMCID: PMC4198624

234: Reducing systematic review workload through certainty-based screening Makoto Miwa, James Thomas, Alison O'Mara-Eves, Sophia Ananiadou J Biomed Inform. 2014 Oct; 51: 242–253. doi: 10.1016/j.jbi.2014.06.005 PMCID: PMC4199186

235: A Microsoft-Excel-based tool for running and critically appraising network meta-analyses—an overview and application of NetMetaXL Stephen Brown, Brian Hutton, Tammy Clifford, Doug Coyle, Daniel Grima, George Wells, Chris Cameron Syst Rev. 2014; 3: 110. Published online 2014 Sep 29. doi: 10.1186/2046-4053-3-110 PMCID: PMC4195340

236: Rural Environments and Community Health (REACH): a randomised controlled trial protocol for an online walking intervention in rural adults Braden L Mitchell, Nicole R Lewis, Ashleigh E Smith, Alex V Rowlands, Gaynor Parfitt, James Dollman BMC Public Health. 2014; 14: 969. Published online 2014 Sep 18. doi: 10.1186/1471-2458-14-969 PMCID: PMC4177164

237: Assembly-free genome comparison based on next-generation sequencing reads and variable length patterns Matteo Comin, Michele Schimd BMC Bioinformatics. 2014; 15(Suppl 9): S1. Published online 2014 Sep 10. doi: 10.1186/1471-2105-15-S9-S1 PMCID: PMC4168702

238: An update on the prevalence of metabolic syndrome and its associated factors in rural northeast China Shasha Yu, Xiaofan Guo, Hongmei Yang, Liqiang Zheng, Yingxian Sun BMC Public Health. 2014; 14: 877. Published online 2014 Aug 26. doi: 10.1186/1471-2458-14-877 PMCID: PMC4153886

239: The Effects of Data Sources, Cohort Selection, and Outcome Definition on a Predictive Model of Risk of Thirty-Day Hospital Readmissions Colin Walsh, George Hripcsak J Biomed Inform. Author manuscript; available in PMC 2015 Dec 1.Published in final edited form as: J Biomed Inform. 2014 Dec; 52: 418–426. Published online 2014 Aug 23. doi: 10.1016/j.jbi.2014.08.006 PMCID: PMC4261028

240: Comparing a paper based monitoring and evaluation system to a mHealth system to support the national community health worker programme, South Africa: an evaluation Sunisha Neupane, Willem Odendaal, Irwin Friedman, Waasila Jassat, Helen Schneider, Tanya Doherty BMC Med Inform Decis Mak. 2014; 14: 69. Published online 2014 Aug 9. doi: 10.1186/1472-6947-14-69 PMCID: PMC4150556

241: Routine piloting in systematic reviews—a modified approach? Linda Long Syst Rev. 2014; 3: 77. Published online 2014 Jul 18. doi: 10.1186/2046-4053-3-77 PMCID: PMC4108964

242: Limestone: High-throughput candidate phenotype generation via tensor factorization Joyce C. Ho, Joydeep Ghosh, Steve R. Steinhubl, Walter F. Stewart, Joshua C. Denny, Bradley A. Malin, Jimeng Sun J Biomed Inform. 2014 Dec; 52: 199–211. Published online 2014 Jul 16. doi: 10.1016/j.jbi.2014.07.001 PMCID: PMC6563906

243: Systematic review automation technologies Guy Tsafnat, Paul Glasziou, Miew Keen Choong, Adam Dunn, Filippo Galgani, Enrico Coiera Syst Rev. 2014; 3: 74. Published online 2014 Jul 9. doi: 10.1186/2046-4053-3-74 PMCID: PMC4100748

244: How much salt do adults consume in climate vulnerable coastal Bangladesh? Sabrina Rasheed, Shamshad Jahan, Tamanna Sharmin, Shahidul Hoque, Masuma Akter Khanam, Mary Anne Land, Mohammad Iqbal, Syed Manzoor Ahmed Hanifi, Fatema Khatun, Abul Kasem Siddique, Abbas Bhuiya BMC Public Health. 2014; 14: 584. Published online 2014 Jun 11. doi: 10.1186/1471-2458-14-584 PMCID: PMC4059094

245: MiningABs: mining associated biomarkers across multi-connected gene expression datasets Chun-Pei Cheng, Christopher DeBoever, Kelly A Frazer, Yu-Cheng Liu, Vincent S Tseng BMC Bioinformatics. 2014; 15: 173. Published online 2014 Jun 8. doi: 10.1186/1471-2105-15-173 PMCID: PMC4068973

246: Proteomics, lipidomics, metabolomics: a mass spectrometry tutorial from a computer scientist's point of view Rob Smith, Andrew D Mathis, Dan Ventura, John T Prince BMC Bioinformatics. 2014; 15(Suppl 7): S9. Published online 2014 May 28. doi: 10.1186/1471-2105-15-S7-S9 PMCID: PMC4110734

247: MEIGO: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics Jose A Egea, David Henriques, Thomas Cokelaer, Alejandro F Villaverde, Aidan MacNamara, Diana-Patricia Danciu, Julio R Banga, Julio Saez-Rodriguez BMC Bioinformatics. 2014; 15: 136. Published online 2014 May 10. doi: 10.1186/1471-2105-15-136 PMCID: PMC4025564

248: A review of machine learning methods to predict the solubility of overexpressed recombinant proteins in Escherichia coli Narjeskhatoon Habibi, Siti Z Mohd Hashim, Alireza Norouzi, Mohammed Razip Samian BMC Bioinformatics. 2014; 15: 134. Published online 2014 May 8. doi: 10.1186/1471-2105-15-134 PMCID: PMC4098780

249: Direct-to-patient disclosure of results of mismatch repair (MMR) screening for Lynch syndrome via electronic personal health record (ePHR): A feasibility study Michael J Hall, Meagan M Herda, Elizabeth A Handorf, Christina C Rybak, Cindy A Keleher, Mark Siemon, Mary B Daly Genet Med. 2014 Nov; 16(11): 854–861. Published online 2014 May 1. doi: 10.1038/gim.2014.42 PMCID: PMC4216634

250: Probabilistic drug connectivity mapping Juuso A Parkkinen, Samuel Kaski BMC Bioinformatics. 2014; 15: 113. Published online 2014 Apr 17. doi: 10.1186/1471-2105-15-113 PMCID: PMC4011783

251: Clinical research data warehouse governance for distributed research networks in the USA: a systematic review of the literature John H Holmes, Thomas E Elliott, Jeffrey S Brown, Marsha A Raebel, Arthur Davidson, Andrew F Nelson, Annie Chung, Pierre La Chance, John F Steiner J Am Med Inform Assoc. 2014 Jul; 21(4): 730–736. Published online 2014 Mar 28. doi: 10.1136/amiajnl-2013-002370 PMCID: PMC4078282

252: Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties Feixiong Cheng, Zhongming Zhao J Am Med Inform Assoc. 2014 Oct; 21(e2): e278–e286. Published online 2014 Mar 18. doi: 10.1136/amiajnl-2013-002512 PMCID: PMC4173180

253: Combining techniques for screening and evaluating interaction terms on high-dimensional time-to-event data Murat Sariyar, Isabell Hoffmann, Harald Binder BMC Bioinformatics. 2014; 15: 58. Published online 2014 Feb 26. doi: 10.1186/1471-2105-15-58 PMCID: PMC3945780

254: Peak picking NMR spectral data using non-negative matrix factorization Suhas Tikole, Victor Jaravine, Vladimir Rogov, Volker Dötsch, Peter Güntert BMC Bioinformatics. 2014; 15: 46. Published online 2014 Feb 11. doi: 10.1186/1471-2105-15-46 PMCID: PMC3931316

255: Development and tuning of an original search engine for patent libraries in medicinal chemistry Emilie Pasche, Julien Gobeill, Olivier Kreim, Fatma Oezdemir-Zaech, Therese Vachon, Christian Lovis, Patrick Ruch BMC Bioinformatics. 2014; 15(Suppl 1): S15. Published online 2014 Jan 10. doi: 10.1186/1471-2105-15-S1-S15 PMCID: PMC4015144

256: Altering micro-environments to change population health behaviour: towards an evidence base for choice architecture interventions Gareth J Hollands, Ian Shemilt, Theresa M Marteau, Susan A Jebb, Michael P Kelly, Ryota Nakamura, Marc Suhrcke, David Ogilvie BMC Public Health. 2013; 13: 1218. Published online 2013 Dec 21. doi: 10.1186/1471-2458-13-1218 PMCID: PMC3881502

257: Is retirement good for your health? A systematic review of longitudinal studies Iris van der Heide, Rogier M van Rijn, Suzan JW Robroek, Alex Burdorf, Karin I Proper BMC Public Health. 2013; 13: 1180. Published online 2013 Dec 13. doi: 10.1186/1471-2458-13-1180 PMCID: PMC4029767

258: Risks to patient safety associated with implementation of electronic applications for medication management in ambulatory care - a systematic review Cheryl LL Carling, Ingvild Kirkehei, Therese Kristine Dalsbø, Elizabeth Paulsen BMC Med Inform Decis Mak. 2013; 13: 133. Published online 2013 Dec 5. doi: 10.1186/1472-6947-13-133 PMCID: PMC3913838

259: Biobanking across the phenome - at the center of chronic disease research Medea Imboden, Nicole M Probst-Hensch BMC Public Health. 2013; 13: 1094. Published online 2013 Nov 25. doi: 10.1186/1471-2458-13-1094 PMCID: PMC4222669

260: Comparison and combination of several MeSH indexing approaches Antonio Jose Jimeno Yepes, James G. Mork, Dina Demner-Fushman, Alan R. Aronson AMIA Annu Symp Proc. 2013; 2013: 709–718. Published online 2013 Nov 16. PMCID: PMC3900212

261: A Large-Scale Analysis of the Reasons Given for Excluding Articles that are Retrieved by Literature Search During Systematic Review Tracy Edinger, Aaron M. Cohen AMIA Annu Symp Proc. 2013; 2013: 379–387. Published online 2013 Nov 16. PMCID: PMC3900186

262: The Ontology of Clinical Research (OCRe): An Informatics Foundation for the Science of Clinical Research Ida Sim, Samson W. Tu, Simona Carini, Harold P. Lehmann, Brad H. Pollock, Mor Peleg, Knut M. Wittkowski J Biomed Inform. Author manuscript; available in PMC 2015 Dec 1.Published in final edited form as: J Biomed Inform. 2014 Dec; 52: 78–91. Published online 2013 Nov 13. doi: 10.1016/j.jbi.2013.11.002 PMCID: PMC4019723

263: enRoute: dynamic path extraction from biological pathway maps for exploring heterogeneous experimental datasets Christian Partl, Alexander Lex, Marc Streit, Denis Kalkofen, Karl Kashofer, Dieter Schmalstieg BMC Bioinformatics. 2013; 14(Suppl 19): S3. Published online 2013 Nov 12. doi: 10.1186/1471-2105-14-S19-S3 PMCID: PMC3980897

264: A lifestyle intervention supported by mobile health technologies to improve the cardiometabolic risk profile of individuals at risk for cardiovascular disease and type 2 diabetes: study rationale and protocol Melanie I Stuckey, Sheree Shapiro, Dawn P Gill, Robert J Petrella BMC Public Health. 2013; 13: 1051. Published online 2013 Nov 7. doi: 10.1186/1471-2458-13-1051 PMCID: PMC3922899

265: Towards generating a patient's timeline: Extracting temporal relationships from clinical notes Azadeh Nikfarjam, Ehsan Emadzadeh, Graciela Gonzalez J Biomed Inform. Author manuscript; available in PMC 2014 Dec 1.Published in final edited form as: J Biomed Inform. 2013 Dec; 46(0): S40–S47. Published online 2013 Nov 7. doi: 10.1016/j.jbi.2013.11.001 PMCID: PMC3974721

266: Jimena: efficient computing and system state identification for genetic regulatory networks Stefan Karl, Thomas Dandekar BMC Bioinformatics. 2013; 14: 306. Published online 2013 Oct 11. doi: 10.1186/1471-2105-14-306 PMCID: PMC3853020

267: Cochrane diagnostic test accuracy reviews Mariska MG Leeflang, Jonathan J Deeks, Yemisi Takwoingi, Petra Macaskill Syst Rev. 2013; 2: 82. Published online 2013 Oct 7. doi: 10.1186/2046-4053-2-82 PMCID: PMC3851548

268: Methodological developments in searching for studies for systematic reviews: past, present and future? Carol Lefebvre, Julie Glanville, L Susan Wieland, Bernadette Coles, Alison L Weightman Syst Rev. 2013; 2: 78. Published online 2013 Sep 25. doi: 10.1186/2046-4053-2-78 PMCID: PMC4015986

269: Cochrane methods - twenty years experience in developing systematic review methods Jackie Chandler, Sally Hopewell Syst Rev. 2013; 2: 76. Published online 2013 Sep 20. doi: 10.1186/2046-4053-2-76 PMCID: PMC3849105

270: Helping people make well-informed decisions about health care: old and new challenges to achieving the aim of the Cochrane Collaboration Andrew D Oxman Syst Rev. 2013; 2: 77. Published online 2013 Sep 20. doi: 10.1186/2046-4053-2-77 PMCID: PMC3848654

271: Electronic Health Record Design and Implementation for Pharmacogenomics: a Local Perspective Josh F. Peterson, Erica Bowton, Julie R. Field, Marc Beller, Jennifer Mitchell, Jonathan Schildcrout, William Gregg, Kevin Johnson, Jim N Jirjis, Dan M. Roden, Jill M. Pulley, Josh C. Denny Genet Med. Author manuscript; available in PMC 2014 Apr 1.Published in final edited form as: Genet Med. 2013 Oct; 15(10): 833–841. Published online 2013 Sep 5. doi: 10.1038/gim.2013.109 PMCID: PMC3925979

272: MeSH indexing based on automatically generated summaries Antonio J Jimeno-Yepes, Laura Plaza, James G Mork, Alan R Aronson, Alberto Díaz BMC Bioinformatics. 2013; 14: 208. Published online 2013 Jun 26. doi: 10.1186/1471-2105-14-208 PMCID: PMC3706357

273: ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing Robert C. Green, Jonathan S. Berg, Wayne W. Grody, Sarah S. Kalia, Bruce R. Korf, Christa L. Martin, Amy McGuire, Robert L. Nussbaum, Julianne M. O'Daniel, Kelly E. Ormond, Heidi L. Rehm, Michael S. Watson, Marc S. Williams, Leslie G. Biesecker Genet Med. Author manuscript; available in PMC 2014 Jan 1.Published in final edited form as: Genet Med. 2013 Jul; 15(7): 565–574. Published online 2013 Jun 20. doi: 10.1038/gim.2013.73Correction in: Genet Med. 2017 May; 19(5): 606. PMCID: PMC3727274

274: Clustering cliques for graph-based summarization of the biomedical research literature Han Zhang, Marcelo Fiszman, Dongwook Shin, Bartlomiej Wilkowski, Thomas C Rindflesch BMC Bioinformatics. 2013; 14: 182. Published online 2013 Jun 7. doi: 10.1186/1471-2105-14-182 PMCID: PMC3682874

275: Imaging informatics for consumer health: towards a radiology patient portal Corey W Arnold, Mary McNamara, Suzie El-Saden, Shawn Chen, Ricky K Taira, Alex A T Bui J Am Med Inform Assoc. 2013 Nov; 20(6): 1028–1036. Published online 2013 Jun 5. doi: 10.1136/amiajnl-2012-001457 PMCID: PMC3822110

276: A systematic review of the literature on the evaluation of handoff tools: implications for research and practice Joanna Abraham, Thomas Kannampallil, Vimla L Patel J Am Med Inform Assoc. 2014 Jan; 21(1): 154–162. Published online 2013 May 23. doi: 10.1136/amiajnl-2012-001351 PMCID: PMC3912721

277: Leveraging Concept-based Approaches to Identify Potential Phyto-therapies Vivekanand Sharma, Indra Neil Sarkar J Biomed Inform. Author manuscript; available in PMC 2014 Aug 1.Published in final edited form as: J Biomed Inform. 2013 Aug; 46(4): 602–614. Published online 2013 May 9. doi: 10.1016/j.jbi.2013.04.008 PMCID: PMC3723125

278: Greedy feature selection for glycan chromatography data with the generalized Dirichlet distribution Marie C Galligan, Radka Saldova, Matthew P Campbell, Pauline M Rudd, Thomas B Murphy BMC Bioinformatics. 2013; 14: 155. Published online 2013 May 7. doi: 10.1186/1471-2105-14-155 PMCID: PMC3703279

279: Implementation of routine screening for Lynch syndrome in university and safety-net health system settings: successes and challenges Evelyn Marquez, Zhuo Geng, Sarah Pass, Pia Summerour, Linda Robinson, Venetia Sarode, Samir Gupta Genet Med. Author manuscript; available in PMC 2017 Jan 5.Published in final edited form as: Genet Med. 2013 Dec; 15(12): 925–932. Published online 2013 Apr 18. doi: 10.1038/gim.2013.45 PMCID: PMC5215584

280: “Reducing unnecessary testing in a CPOE system through implementation of a targeted CDS intervention” Donald L Levick, Glenn Stern, Chad D Meyerhoefer, Aaron Levick, David Pucklavage BMC Med Inform Decis Mak. 2013; 13: 43. Published online 2013 Apr 8. doi: 10.1186/1472-6947-13-43 PMCID: PMC3629995

281: Predictors of Medication Adherence in Elderly Patients with Chronic Diseases Using Support Vector Machine Models Soo Kyoung Lee, Bo-Yeong Kang, Hong-Gee Kim, Youn-Jung Son Healthc Inform Res. 2013 Mar; 19(1): 33–41. Published online 2013 Mar 31. doi: 10.4258/hir.2013.19.1.33 PMCID: PMC3633170

282: Use of a support vector machine for categorizing free-text notes: assessment of accuracy across two institutions Adam Wright, Allison B McCoy, Stanislav Henkin, Abhivyakti Kale, Dean

F Sittig J Am Med Inform Assoc. 2013 Sep; 20(5): 887–890. Published online 2013 Mar 30. doi: 10.1136/amiajnl-2012-001576 PMCID: PMC3756266

283: Managing protected health information in distributed research network environments: automated review to facilitate collaboration Christine E Bredfeldt, Amy Butani, Sandhyasree Padmanabhan, Paul Hitz, Roy Pardee BMC Med Inform Decis Mak. 2013; 13: 39. Published online 2013 Mar 22. doi: 10.1186/1472-6947-13-39 PMCID: PMC3617086

284: Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence Felix Köpcke, Benjamin Trinczek, Raphael W Majeed, Björn Schreiweis, Joachim Wenk, Thomas Leusch, Thomas Ganslandt, Christian Ohmann, Björn Bergh, Rainer Röhrig, Martin Dugas, Hans-Ulrich Prokosch BMC Med Inform Decis Mak. 2013; 13: 37. Published online 2013 Mar 21. doi: 10.1186/1472-6947-13-37 PMCID: PMC3606452

285: A method for estimating from thermometer sales the incidence of diseases that are symptomatically similar to influenza Ricardo Villamarín, Gregory Cooper, Michael Wagner, Fu-Chiang Tsui, Jeremy U. Espino J Biomed Inform. Author manuscript; available in PMC 2015 Oct 18.Published in final edited form as: J Biomed Inform. 2013 Jun; 46(3): 444–457. Published online 2013 Mar 14. doi: 10.1016/j.jbi.2013.02.003 PMCID: PMC4609543

286: Towards public health decision support: a systematic review of bidirectional communication approaches Brian E Dixon, Roland E Gamache, Shaun J Grannis J Am Med Inform Assoc. 2013 May-Jun; 20(3): 577–583. Published online 2013 Mar 6. doi: 10.1136/amiajnl-2012-001514 PMCID: PMC3628068

287: Cost effectiveness of a computer-delivered intervention to improve HIV medication adherence Raymond L Ownby, Drenna Waldrop-Valverde, Robin J Jacobs, Amarilis Acevedo, Joshua Caballero BMC Med Inform Decis Mak. 2013; 13: 29. Published online 2013 Feb 28. doi: 10.1186/1472-6947-13-29 PMCID: PMC3599639

288: Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data Carlos A Alvarez, Christopher A Clark, Song Zhang, Ethan A Halm, John J Shannon, Carlos E Girod, Lauren Cooper, Ruben Amarasingham BMC Med Inform Decis Mak. 2013; 13: 28. Published online 2013 Feb 27. doi: 10.1186/1472-6947-13-28 PMCID: PMC3599266

289: Prevalence and comorbidity of diabetes mellitus among non-institutionalized older adults in Germany - results of the national telephone health interview survey 'German Health

Update (GEDA)' 2009 Yong Du, Christin Heidemann, Antje Gößwald, Patrick Schmich, Christa Scheidt-Nave BMC Public Health. 2013; 13: 166. Published online 2013 Feb 23. doi: 10.1186/1471-2458-13-166 PMCID: PMC3599814

290: A New Iterative Method to Reduce Workload in the Systematic Review Process Siddhartha Jonnalagadda, Diana Petitti Int J Comput Biol Drug Des. Author manuscript; available in PMC 2014 Feb 21.Published in final edited form as: Int J Comput Biol Drug Des. 2013; 6(0): 5–17. Published online 2013 Feb 21. doi: 10.1504/IJCBDD.2013.052198 PMCID: PMC3787693

291: Development of a web-based intervention for the indicated prevention of depression Saskia M Kelders, Wendy TM Pots, Maarten Jan Oskam, Ernst T Bohlmeijer, Julia EWC van Gemert-Pijnen BMC Med Inform Decis Mak. 2013; 13: 26. Published online 2013 Feb 20. doi: 10.1186/1472-6947-13-26 PMCID: PMC3598782

292: Visualization of protein interaction networks: problems and solutions Giuseppe Agapito, Pietro Hiram Guzzi, Mario Cannataro BMC Bioinformatics. 2013; 14(Suppl 1): S1. Published online 2013 Jan 14. doi: 10.1186/1471-2105-14-S1-S1 PMCID: PMC3548679

293: A Study on Pubmed Search Tag Usage Pattern: Association Rule Mining of a Full-day Pubmed Query Log Abu Saleh Mohammad Mosa, Illhoi Yoo BMC Med Inform Decis Mak. 2013; 13: 8. Published online 2013 Jan 9. doi: 10.1186/1472-6947-13-8 PMCID: PMC3552776

294: Nutrition and physical activity programs for obesity treatment (PRONAF study): methodological approach of the project Augusto G Zapico, Pedro J Benito, Marcela González-Gross, Ana B Peinado, Esther Morencos, Blanca Romero, Miguel A Rojo-Tirado, Rocio Cupeiro, Barbara Szendrei, Javier Butragueño, Maite Bermejo, María Alvarez-Sánchez, Miguel García-Fuentes, Carmen Gómez-Candela, Laura M Bermejo, Ceila Fernandez-Fernandez, Francisco J Calderón BMC Public Health. 2012; 12: 1100. Published online 2012 Dec 21. doi: 10.1186/1471-2458-12-1100 PMCID: PMC3577471

295: Window Classification of Brain CT Images in Biomedical Articles Zhiyun Xue, Sameer Antani, L. Rodney Long, Dina Demner-Fushman, George R. Thoma AMIA Annu Symp Proc. 2012; 2012: 1023–1029. Published online 2012 Nov 3. PMCID: PMC3540547

296: Towards the Creation of a Visual Ontology of Biomedical Imaging Entities Matthew S. Simpson, Daekeun You, Md Mahmudur Rahman, Sameer K. Antani, George R. Thoma, Dina Demner-Fushman AMIA Annu Symp Proc. 2012; 2012: 866–875. Published online 2012 Nov 3. PMCID: PMC3540530

297: Barriers to Retrieving Patient Information from Electronic Health Record Data: Failure Analysis from the TREC Medical Records Track Tracy Edinger, Aaron M. Cohen, Steven Bedrick, Kyle Ambert, William Hersh AMIA Annu Symp Proc. 2012; 2012: 180–188. Published online 2012 Nov 3. PMCID: PMC3540501

298: Logical Differential Prediction Bayes Net, improving breast cancer diagnosis for older women Houssam Nassif, Yirong Wu, David Page, Elizabeth Burnside AMIA Annu Symp Proc. 2012; 2012: 1330–1339. Published online 2012 Nov 3. PMCID: PMC3540455

299: Survival Prediction and Treatment Recommendation with Bayesian Techniques in Lung Cancer M. Berkan Sesen, Timor Kadir, Rene-Banares Alcantara, John Fox, Michael Brady, Sir AMIA Annu Symp Proc. 2012; 2012: 838–847. Published online 2012 Nov 3. PMCID: PMC3540451

300: A systematic review of the use of financial incentives and penalties to encourage uptake of healthy behaviors: protocol Jean Adams, Emma L Giles, Shannon Robalino, Elaine McColl, Falko F Sniehotta Syst Rev. 2012; 1: 51. Published online 2012 Oct 31. doi: 10.1186/2046-4053-1-51 PMCID: PMC3499145

301: MEDLINE clinical queries are robust when searching in recent publishing years Nancy L Wilczynski, K Ann McKibbon, Stephen D Walter, Amit X Garg, R Brian Haynes J Am Med Inform Assoc. 2013 Mar-Apr; 20(2): 363–368. Published online 2012 Sep 27. doi: 10.1136/amiajnl-2012-001075 PMCID: PMC3638187

302: Drug–drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records Shobha Phansalkar, Heleen van der Sijs, Alisha D Tucker, Amrita A Desai, Douglas S Bell, Jonathan M Teich, Blackford Middleton, David W Bates J Am Med Inform Assoc. 2013 May-Jun; 20(3): 489–493. Published online 2012 Sep 25. doi: 10.1136/amiajnl-2012-001089 PMCID: PMC3628052

303: Improving the efficiency and relevance of evidence-based recommendations in the era of whole-genome sequencing: an EGAPP methods update David L. Veenstra, Margaret Piper, James E. Haddow, Stephen G. Pauker, Roger Klein, Carolyn Sue Richards, Sean R. Tunis, Benjamin Djulbegovic, Michael Marrone, Jennifer S. Lin, Alfred O. Berg, Ned Calonge Genet Med. Author manuscript; available in PMC 2014 Feb 23.Published in final edited form as: Genet Med. 2013 Jan; 15(1): 14–24. Published online 2012 Sep 6. doi: 10.1038/gim.2012.106 PMCID: PMC3932295

304: Improving image retrieval effectiveness via query expansion using MeSH hierarchical structure Mariano Crespo Azcárate, Jacinto Mata Vázquez, Manuel Maña López J Am Med Inform

Assoc. 2013 Nov; 20(6): 1014–1020. Published online 2012 Sep 5. doi: 10.1136/amiajnl-2012-000943 PMCID: PMC3822102

305: Deficiencies in the transfer and availability of clinical trials evidence: a review of existing systems and standards Gert van Valkenhoef, Tommi Tervonen, Bert de Brock, Hans Hillege BMC Med Inform Decis Mak. 2012; 12: 95. Published online 2012 Sep 4. doi: 10.1186/1472-6947-12-95 PMCID: PMC3534489

306: A conceptual framework and protocol for defining clinical decision support objectives applicable to medical specialties Justin W Timbie, Cheryl L Damberg, Eric C Schneider, Douglas S Bell BMC Med Inform Decis Mak. 2012; 12: 93. Published online 2012 Sep 3. doi: 10.1186/1472-6947-12-93 PMCID: PMC3536635

307: A classification of errors in lay comprehension of medical documents Alla Keselman, Catherine Arnott Smith J Biomed Inform. Author manuscript; available in PMC 2013 Dec 1.Published in final edited form as: J Biomed Inform. 2012 Dec; 45(6): 1151–1163. Published online 2012 Aug 20. doi: 10.1016/j.jbi.2012.07.012 PMCID: PMC3504163

308: Advancing clinical decision support using lessons from outside of healthcare: an interdisciplinary systematic review Helen W Wu, Paul K Davis, Douglas S Bell BMC Med Inform Decis Mak. 2012; 12: 90. Published online 2012 Aug 17. doi: 10.1186/1472-6947-12-90 PMCID: PMC3524755

309: Optimal experiment selection for parameter estimation in biological differential equation models Mark K Transtrum, Peng Qiu BMC Bioinformatics. 2012; 13: 181. Published online 2012 Jul 27. doi: 10.1186/1471-2105-13-181 PMCID: PMC3536579

310: SNP interaction detection with Random Forests in high-dimensional genetic data Stacey J Winham, Colin L Colby, Robert R Freimuth, Xin Wang, Mariza de Andrade, Marianne Huebner, Joanna M Biernacka BMC Bioinformatics. 2012; 13: 164. Published online 2012 Jul 15. doi: 10.1186/1471-2105-13-164 PMCID: PMC3463421

311: Identification of methicillin-resistant Staphylococcus aureus within the Nation's Veterans Affairs Medical Centers using natural language processing Makoto Jones, Scott L DuVall, Joshua Spuhl, Matthew H Samore, Christopher Nielson, Michael Rubin BMC Med Inform Decis Mak. 2012; 12: 34. Published online 2012 Jul 11. doi: 10.1186/1472-6947-12-34 PMCID: PMC3394221

312: Knowledge integration at the center of genomic medicine Muin J. Khoury, Marta Gwinn, W. David Dotson, Sheri D. Schully Genet Med. Author manuscript; available in PMC 2015 Dec

16.Published in final edited form as: Genet Med. 2012 Jul; 14(7): 643–647. doi: 10.1038/gim.2012.43 PMCID: PMC4681509

313: Modeling and mining term association for improving biomedical information retrieval performance Qinmin Hu, Jimmy Xiangji Huang, Xiaohua Hu BMC Bioinformatics. 2012; 13(Suppl 9): S2. Published online 2012 Jun 11. doi: 10.1186/1471-2105-13-S9-S2 PMCID: PMC3372456

314: Clarifying differences between review designs and methods David Gough, James Thomas, Sandy Oliver Syst Rev. 2012; 1: 28. Published online 2012 Jun 9. doi: 10.1186/2046-4053-1-28 PMCID: PMC3533815

315: Immigrant women's experiences of maternity-care services in Canada: a protocol for systematic review using a narrative synthesis Gina M A Higginbottom, Myfanwy Morgan, Jayantha Dassanayake, Helgi Eyford, Mirande Alexandre, Yvonne Chiu, Joan Forgeron, Deb Kocay Syst Rev. 2012; 1: 27. Published online 2012 May 31. doi: 10.1186/2046-4053-1-27 PMCID: PMC3433387

316: Blood pressure and particulate air pollution in schoolchildren of Lahore, Pakistan Muhammad Sughis, Tim S Nawrot, Syed Ihsan-ul-Haque, Asad Amjad, Benoit Nemery BMC Public Health. 2012; 12: 378. Published online 2012 May 25. doi: 10.1186/1471-2458-12-378 PMCID: PMC3403904

317: Learning Sparse Representations for Fruit-Fly Gene Expression Pattern Image Annotation and Retrieval Lei Yuan, Alexander Woodard, Shuiwang Ji, Yuan Jiang, Zhi-Hua Zhou, Sudhir Kumar, Jieping Ye BMC Bioinformatics. 2012; 13: 107. Published online 2012 May 23. doi: 10.1186/1471-2105-13-107 PMCID: PMC3434040

318: Text summarization as a decision support aid T Elizabeth Workman, Marcelo Fiszman, John F Hurdle BMC Med Inform Decis Mak. 2012; 12: 41. Published online 2012 May 23. doi: 10.1186/1472-6947-12-41 PMCID: PMC3461485

319: Feature engineering combined with machine learning and rule-based methods for structured information extraction from narrative clinical discharge summaries Yan Xu, Kai Hong, Junichi Tsujii, Eric I-Chao Chang J Am Med Inform Assoc. 2012 Sep-Oct; 19(5): 824–832. Published online 2012 May 14. doi: 10.1136/amiajnl-2011-000776 PMCID: PMC3422834

320: ProDis-ContSHC: learning protein dissimilarity measures and hierarchical context coherently for protein-protein comparison in protein database retrieval Jingyan Wang, Xin Gao, Quanquan Wang, Yongping Li BMC Bioinformatics. 2012; 13(Suppl 7): S2. Published online 2012 May 8. doi: 10.1186/1471-2105-13-S7-S2 PMCID: PMC3348016

321: USING PharmGKB TO TRAIN TEXT MINING APPROACHES FOR IDENTIFYING POTENTIAL GENE TARGETS FOR PHARMACOGENOMIC STUDIES S. PAKHOMOV, B.T. MCINNES, J. LAMBA, Y. LIU, G.B. MELTON, Y. GHODKE, N. BHISE, V. LAMBA, A.K. BIRNBAUM J Biomed Inform. Author manuscript; available in PMC 2013 Oct 1.Published in final edited form as: J Biomed Inform. 2012 Oct; 45(5): 862–869. Published online 2012 May 4. doi: 10.1016/j.jbi.2012.04.007 PMCID: PMC3438361

322: Using EHRs to integrate research with patient care: promises and challenges Chunhua Weng, Paul Appelbaum, George Hripcsak, Ian Kronish, Linda Busacca, Karina W Davidson, J Thomas Bigger J Am Med Inform Assoc. 2012 Sep-Oct; 19(5): 684–687. Published online 2012 Apr 29. doi: 10.1136/amiajnl-2012-000878 PMCID: PMC3422845

323: The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature Marcia Vervloet, Annemiek J Linn, Julia C M van Weert, Dinny H de Bakker, Marcel L Bouvy, Liset van Dijk J Am Med Inform Assoc. 2012 Sep-Oct; 19(5): 696–704. Published online 2012 Apr 25. doi: 10.1136/amiajnl-2011-000748 PMCID: PMC3422829

324: Clinical research informatics: a conceptual perspective Michael G Kahn, Chunhua Weng J Am Med Inform Assoc. 2012 Jun; 19(e1): e36–e42. Published online 2012 Apr 20. doi: 10.1136/amiajnl-2012-000968 PMCID: PMC3392857

325: Studying the potential impact of automated document classification on scheduling a systematic review update Aaron M Cohen, Kyle Ambert, Marian McDonagh BMC Med Inform Decis Mak. 2012; 12: 33. Published online 2012 Apr 19. doi: 10.1186/1472-6947-12-33 PMCID: PMC3420236

326: CDAPubMed: a browser extension to retrieve EHR-based biomedical literature David Perez-Rey, Ana Jimenez-Castellanos, Miguel Garcia-Remesal, Jose Crespo, Victor Maojo BMC Med Inform Decis Mak. 2012; 12: 29. Published online 2012 Apr 5. doi: 10.1186/1472-6947-12-29 PMCID: PMC3366875

327: Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, Christopher H. Schmid, Lars Bertram, Christina M. Lill, Joshua T. Cohen, Thomas A. Trikalinos Genet Med. 2012 Jul; 14(7): 663–669. Published online 2012 Apr 5. doi: 10.1038/gim.2012.7 PMCID: PMC3908550

328: Improving the Performance of Text Categorization Models used for the Selection of High Quality Articles Seunghee Kim, Jinwook Choi Healthc Inform Res. 2012 Mar; 18(1): 18–28. Published online 2012 Mar 31. doi: 10.4258/hir.2012.18.1.18 PMCID: PMC3324751

329: Leveraging medical thesauri and physician feedback for improving medical literature retrieval for case queries Parikshit Sondhi, Jimeng Sun, ChengXiang Zhai, Robert Sorrentino, Martin S Kohn J Am Med Inform Assoc. 2012 Sep-Oct; 19(5): 851–858. Published online 2012 Mar 21. doi: 10.1136/amiajnl-2011-000293 PMCID: PMC3422816

330: Patient, physician, encounter, and billing characteristics predict the accuracy of syndromic surveillance case definitions Geneviève Cadieux, David L Buckeridge, André Jacques, Michael Libman, Nandini Dendukuri, Robyn Tamblyn BMC Public Health. 2012; 12: 166. Published online 2012 Mar 8. doi: 10.1186/1471-2458-12-166 PMCID: PMC3378465

331: Impact of electronic medical record on physician practice in office settings: a systematic review Francis Lau, Morgan Price, Jeanette Boyd, Colin Partridge, Heidi Bell, Rebecca Raworth BMC Med Inform Decis Mak. 2012; 12: 10. Published online 2012 Feb 24. doi: 10.1186/1472-6947-12-10 PMCID: PMC3315440

332: Evaluating the state of the art in coreference resolution for electronic medical records Ozlem Uzuner, Andreea Bodnari, Shuying Shen, Tyler Forbush, John Pestian, Brett R South J Am Med Inform Assoc. 2012 Sep-Oct; 19(5): 786–791. Published online 2012 Feb 24. doi: 10.1136/amiajnl-2011-000784 PMCID: PMC3422835

333: Automatic classification of mammography reports by BI-RADS breast tissue composition class Bethany Percha, Houssam Nassif, Jafi Lipson, Elizabeth Burnside, Daniel Rubin J Am Med Inform Assoc. 2012 Sep-Oct; 19(5): 913–916. Published online 2012 Jan 29. doi: 10.1136/amiajnl-2011-000607 PMCID: PMC3422822

334: Enabling international adoption of LOINC through translation Daniel J. Vreeman, Maria Teresa Chiaravalloti, John Hook, Clement J. McDonald J Biomed Inform. Author manuscript; available in PMC 2013 Aug 1.Published in final edited form as: J Biomed Inform. 2012 Aug; 45(4): 667–673. Published online 2012 Jan 21. doi: 10.1016/j.jbi.2012.01.005 PMCID: PMC3376691

335: Evaluation of computer-based medical histories taken by patients at home Warner V Slack, Hollis B Kowaloff, Roger B Davis, Tom Delbanco, Steven E Locke, Charles Safran, Howard L Bleich J Am Med Inform Assoc. 2012 Jul-Aug; 19(4): 545–548. Published online 2012 Jan 11. doi: 10.1136/amiajnl-2011-000580 PMCID: PMC3384115

336: Systematic review and evaluation of web-accessible tools for management of diabetes and related cardiovascular risk factors by patients and healthcare providers Catherine H Yu, Robinder Bahniwal, Andreas Laupacis, Eman Leung, Michael S Orr, Sharon E Straus J Am Med Inform Assoc. 2012 Jul-Aug; 19(4): 514–522. Published online 2012 Jan 3. doi: 10.1136/amiajnl-2011-000307 PMCID: PMC3384097

337: Morphometic analysis of TCGA glioblastoma multiforme Hang Chang, Gerald V Fontenay, Ju Han, Ge Cong, Frederick L Baehner, Joe W Gray, Paul T Spellman, Bahram Parvin BMC Bioinformatics. 2011; 12: 484. Published online 2011 Dec 20. doi: 10.1186/1471-2105-12-484 PMCID: PMC3271112

338: Evaluation of an automated safety surveillance system using risk adjusted sequential probability ratio testing Michael E Matheny, Sharon-Lise T Normand, Thomas P Gross, Danica Marinac-Dabic, Nilsa Loyo-Berrios, Venkatesan D Vidi, Sharon Donnelly, Frederic S Resnic BMC Med Inform Decis Mak. 2011; 11: 75. Published online 2011 Dec 14. doi: 10.1186/1472-6947-11-75 PMCID: PMC3262755

339: Trends in biomedical informatics: most cited topics from recent years Hyeon-Eui Kim, Xiaoqian Jiang, Jihoon Kim, Lucila Ohno-Machado J Am Med Inform Assoc. 2011 Dec; 18(Suppl 1): i166–i170. doi: 10.1136/amiajnl-2011-000706 PMCID: PMC3241182

340: Discovery of error-tolerant biclusters from noisy gene expression data Rohit Gupta, Navneet Rao, Vipin Kumar BMC Bioinformatics. 2011; 12(Suppl 12): S1. Published online 2011 Nov 24. doi: 10.1186/1471-2105-12-S12-S1 PMCID: PMC3247082

341: Enhancing clinical concept extraction with distributional semantics Siddhartha Jonnalagadda, Trevor Cohen, Stephen Wu, Graciela Gonzalez J Biomed Inform. Author manuscript; available in PMC 2013 Feb 1.Published in final edited form as: J Biomed Inform. 2012 Feb; 45(1): 129–140. Published online 2011 Nov 7. doi: 10.1016/j.jbi.2011.10.007 PMCID: PMC3272090

342: A multi-layered framework for disseminating knowledge for computer-based decision support Aziz A Boxwala, Beatriz H Rocha, Saverio Maviglia, Vipul Kashyap, Seth Meltzer, Jihoon Kim, Ruslana Tsurikova, Adam Wright, Marilyn D Paterno, Amanda Fairbanks, Blackford Middleton J Am Med Inform Assoc. 2011 Dec; 18(Suppl 1): i132–i139. Published online 2011 Nov 3. doi: 10.1136/amiajnl-2011-000334 PMCID: PMC3241169

343: Towards case-based medical learning in radiological decision making using content-based image retrieval Petra Welter, Thomas M Deserno, Benedikt Fischer, Rolf W Günther, Cord

Spreckelsen BMC Med Inform Decis Mak. 2011; 11: 68. Published online 2011 Oct 27. doi: 10.1186/1472-6947-11-68 PMCID: PMC3217894

344: Anomaly and Signature Filtering Improve Classifier Performance For Detection Of Suspicious Access To EHRs Jihoon Kim, Janice M Grillo, Aziz A Boxwala, Xiaoqian Jiang, Rose B Mandelbaum, Bhakti A Patel, Debra Mikels, Staal A Vinterbo, Lucila Ohno-Machado AMIA Annu Symp Proc. 2011; 2011: 723–731. Published online 2011 Oct 22. PMCID: PMC3243249

345: The Role of the Electronic Medical Record in the Assessment of Health Related Quality of Life Serguei V.S. Pakhomov, Nilay D. Shah, Holly K. Van Houten, Penny L. Hanson, Steven A. Smith AMIA Annu Symp Proc. 2011; 2011: 1080–1088. Published online 2011 Oct 22. PMCID: PMC3243229

346: A bottom-up approach to MEDLINE indexing recommendations Antonio Jimeno-Yepes, Bartłomiej Wilkowski, James G. Mork, Elizabeth Van Lenten, Dina Demner Fushman, Alan R. Aronson AMIA Annu Symp Proc. 2011; 2011: 1583–1592. Published online 2011 Oct 22. PMCID: PMC3243198

347: Reconciling Pairs of Concurrently Used Clinical Practice Guidelines Using Constraint Logic Programming Szymon Wilk, Martin Michalowski, Wojtek Michalowski, Marisela Mainegra Hing, Ken Farion AMIA Annu Symp Proc. 2011; 2011: 944–953. Published online 2011 Oct 22. PMCID: PMC3243171

348: Leveraging Terminologies for Retrieval of Radiology Reports with Critical Imaging Findings Graham I. Warden, Ronilda Lacson, Ramin Khorasani AMIA Annu Symp Proc. 2011; 2011: 1481–1488. Published online 2011 Oct 22. PMCID: PMC3243125

349: SCOWLP update: 3D classification of protein-protein, -peptide, -saccharide and -nucleic acid interactions, and structure-based binding inferences across folds Joan Teyra, Sergey A Samsonov, Sven Schreiber, M Teresa Pisabarro BMC Bioinformatics. 2011; 12: 398. Published online 2011 Oct 13. doi: 10.1186/1471-2105-12-398 PMCID: PMC3210135

350: BioCreative III interactive task: an overview Cecilia N Arighi, Phoebe M Roberts, Shashank Agarwal, Sanmitra Bhattacharya, Gianni Cesareni, Andrew Chatr-aryamontri, Simon Clematide, Pascale Gaudet, Michelle Gwinn Giglio, Ian Harrow, Eva Huala, Martin Krallinger, Ulf Leser, Donghui Li, Feifan Liu, Zhiyong Lu, Lois J Maltais, Naoaki Okazaki, Livia Perfetto, Fabio Rinaldi, Rune Sætre, David Salgado, Padmini Srinivasan, Philippe E Thomas, Luca Toldo, Lynette

Hirschman, Cathy H Wu BMC Bioinformatics. 2011; 12(Suppl 8): S4. Published online 2011 Oct 3. doi: 10.1186/1471-2105-12-S8-S4 PMCID: PMC3269939

351: The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text Martin Krallinger, Miguel Vazquez, Florian Leitner, David Salgado, Andrew Chatr-aryamontri, Andrew Winter, Livia Perfetto, Leonardo Briganti, Luana Licata, Marta Iannuccelli, Luisa Castagnoli, Gianni Cesareni, Mike Tyers, Gerold Schneider, Fabio Rinaldi, Robert Leaman, Graciela Gonzalez, Sergio Matos, Sun Kim, W John Wilbur, Luis Rocha, Hagit Shatkay, Ashish V Tendulkar, Shashank Agarwal, Feifan Liu, Xinglong Wang, Rafal Rak, Keith Noto, Charles Elkan, Zhiyong Lu, Rezarta Islamaj Dogan, Jean-Fred Fontaine, Miguel A Andrade-Navarro, Alfonso Valencia BMC Bioinformatics. 2011; 12(Suppl 8): S3. Published online 2011 Oct 3. doi: 10.1186/1471-2105-12-S8-S3 PMCID: PMC3269938

352: Simple and efficient machine learning frameworks for identifying protein-protein interaction relevant articles and experimental methods used to study the interactions Shashank Agarwal, Feifan Liu, Hong Yu BMC Bioinformatics. 2011; 12(Suppl 8): S10. Published online 2011 Oct 3. doi: 10.1186/1471-2105-12-S8-S10 PMCID: PMC3269933

353: Overview of the BioCreative III Workshop Cecilia N Arighi, Zhiyong Lu, Martin Krallinger, Kevin B Cohen, W John Wilbur, Alfonso Valencia, Lynette Hirschman, Cathy H Wu BMC Bioinformatics. 2011; 12(Suppl 8): S1. Published online 2011 Oct 3. doi: 10.1186/1471-2105-12-S8-S1 PMCID: PMC3269932

354: Recognizing Temporal Information in Korean Clinical Narratives through Text Normalization Youngho Kim, Jinwook Choi Healthc Inform Res. 2011 Sep; 17(3): 150–155. Published online 2011 Sep 30. doi: 10.4258/hir.2011.17.3.150 PMCID: PMC3212741

355: Search filters to identify geriatric medicine in Medline Esther M M van de Glind, Barbara C van Munster, René Spijker, Rob J P M Scholten, Lotty Hooft J Am Med Inform Assoc. 2012 May-Jun; 19(3): 468–472. Published online 2011 Sep 23. doi: 10.1136/amiajnl-2011-000319 PMCID: PMC3341784

356: Healthcare in the Pocket: Mapping the Space of Mobile-Phone Health Interventions Predrag Klasnja, Wanda Pratt J Biomed Inform. Author manuscript; available in PMC 2013 Feb 1.Published in final edited form as: J Biomed Inform. 2012 Feb; 45(1): 184–198. Published online 2011 Sep 9. doi: 10.1016/j.jbi.2011.08.017 PMCID: PMC3272165

357: Auditing Complex Concepts of SNOMED using a Refined Hierarchical Abstraction Network Yue Wang, Michael Halper, Duo Wei, Huanying Gu, Yehoshua Perl, Junchuan Xu, Gai Elhanan, Yan Chen, Kent A. Spackman, James T. Case, George Hripcsak J Biomed Inform. Author manuscript; available in PMC 2013 Feb 1.Published in final edited form as: J Biomed Inform. 2012 Feb; 45(1): 1–14. Published online 2011 Sep 1. doi: 10.1016/j.jbi.2011.08.016 PMCID: PMC3313651

358: The effectiveness of integrated health information technologies across the phases of medication management: a systematic review of randomized controlled trials K Ann McKibbon, Cynthia Lokker, Steven M Handler, Lisa R Dolovich, Anne M Holbrook, Daria O'Reilly, Robyn Tamblyn, Brian J Hemens, Runki Basu, Sue Troyan, Pavel S Roshanov J Am Med Inform Assoc. 2012 Jan-Feb; 19(1): 22–30. Published online 2011 Aug 18. doi: 10.1136/amiajnl-2011-000304 PMCID: PMC3240758

359: A framework for evaluating the appropriateness of clinical decision support alerts and responses Allison B McCoy, Lemuel R Waitman, Julia B Lewis, Julie A Wright, David P Choma, Randolph A Miller, Josh F Peterson J Am Med Inform Assoc. 2012 May-Jun; 19(3): 346–352. Published online 2011 Aug 17. doi: 10.1136/amiajnl-2011-000185 PMCID: PMC3341775

360: Lessons from the Canadian national health information technology plan for the United States: opinions of key Canadian experts Eyal Zimlichman, Ronen Rozenblum, Claudia A Salzberg, Yeona Jang, Melissa Tamblyn, Robyn Tamblyn, David W Bates J Am Med Inform Assoc. 2012 May-Jun; 19(3): 453–459. Published online 2011 Jul 15. doi: 10.1136/amiajnl-2011-000127 PMCID: PMC3341773

361: Using statistical and machine learning to help institutions detect suspicious access to electronic health records Aziz A Boxwala, Jihoon Kim, Janice M Grillo, Lucila Ohno-Machado J Am Med Inform Assoc. 2011 Jul-Aug; 18(4): 498–505. doi: 10.1136/amiajnl-2011-000217 PMCID: PMC3128412

362: Machine learning with naturally labeled data for identifying abbreviation definitions Lana Yeganova, Donald C Comeau, W John Wilbur BMC Bioinformatics. 2011; 12(Suppl 3): S6. Published online 2011 Jun 9. doi: 10.1186/1471-2105-12-S3-S6 PMCID: PMC3111592

363: Improving a gold standard: treating human relevance judgments of MEDLINE document pairs W John Wilbur, Won Kim BMC Bioinformatics. 2011; 12(Suppl 3): S5. Published online 2011 Jun 9. doi: 10.1186/1471-2105-12-S3-S5 PMCID: PMC3111591

364: Exploiting MeSH indexing in MEDLINE to generate a data set for word sense disambiguation Antonio J Jimeno-Yepes, Bridget T McInnes, Alan R Aronson BMC Bioinformatics. 2011; 12: 223. Published online 2011 Jun 2. doi: 10.1186/1471-2105-12-223 PMCID: PMC3123611

365: RAG: An update to the RNA-As-Graphs resource Joseph A Izzo, Namhee Kim, Shereef Elmetwaly, Tamar Schlick BMC Bioinformatics. 2011; 12: 219. Published online 2011 May 31. doi: 10.1186/1471-2105-12-219 PMCID: PMC3123240

366: Evaluating the utility of syndromic surveillance algorithms for screening to detect potentially clonal hospital infection outbreaks Randy J Carnevale, Thomas R Talbot, William Schaffner, Karen C Bloch, Titus L Daniels, Randolph A Miller J Am Med Inform Assoc. 2011 Jul-Aug; 18(4): 466–472. Published online 2011 May 23. doi: 10.1136/amiajnl-2011-000216 PMCID: PMC3128411

367: Mechanistic Indicators of Childhood Asthma (MICA) Study: piloting an integrative design for evaluating environmental health Jane Gallagher, Edward Hudgens, Ann Williams, Jefferson Inmon, Scott Rhoney, Gina Andrews, David Reif, Brooke Heidenfelder, Lucas Neas, Ronald Williams, Markey Johnson, Haluk Özkaynak, Stephen Edwards, Elaine Cohen Hubal BMC Public Health. 2011; 11: 344. Published online 2011 May 19. doi: 10.1186/1471-2458-11-344 PMCID: PMC3112137

368: Discrimination of approved drugs from experimental drugs by learning methods Kailin Tang, Ruixin Zhu, Yixue Li, Zhiwei Cao BMC Bioinformatics. 2011; 12: 157. Published online 2011 May 14. doi: 10.1186/1471-2105-12-157 PMCID: PMC3120701

369: Degree centrality for semantic abstraction summarization of therapeutic studies Han Zhang, Marcelo Fiszman, Dongwook Shin, Christopher M. Miller, Graciela Rosemblat, Thomas C. Rindflesch J Biomed Inform. Author manuscript; available in PMC 2012 Oct 1.Published in final edited form as: J Biomed Inform. 2011 Oct; 44(5): 830–838. Published online 2011 May 8. doi: 10.1016/j.jbi.2011.05.001 PMCID: PMC3170688

370: Using information mining of the medical literature to improve drug safety Kanaka D Shetty, Siddhartha R Dalal J Am Med Inform Assoc. 2011 Sep-Oct; 18(5): 668–674. Published online 2011 May 5. doi: 10.1136/amiajnl-2011-000096 PMCID: PMC3168306

371: Using the time and motion method to study clinical work processes and workflow: methodological inconsistencies and a call for standardized research Kai Zheng, Michael H Guo, David A Hanauer J Am Med Inform Assoc. 2011 Sep-Oct; 18(5): 704–710. Published online 2011 Apr 27. doi: 10.1136/amiajnl-2011-000083 PMCID: PMC3168304

372: BICEPP: an example-based statistical text mining method for predicting the binary characteristics of drugs Frank PY Lin, Stephen Anthony, Thomas M Polasek, Guy Tsafnat, Matthew P Doogue BMC Bioinformatics. 2011; 12: 112. Published online 2011 Apr 21. doi: 10.1186/1471-2105-12-112 PMCID: PMC3110144

373: Collaborative search in electronic health records Kai Zheng, Qiaozhu Mei, David A Hanauer J Am Med Inform Assoc. 2011 May-Jun; 18(3): 282–291. Published online 2011 Apr 12. doi: 10.1136/amiajnl-2011-000009 PMCID: PMC3078661

374: MRCQuant- an accurate LC-MS relative isotopic quantification algorithm on TOF instruments William E Haskins, Konstantinos Petritis, Jianqiu Zhang BMC Bioinformatics. 2011; 12: 74. Published online 2011 Mar 15. doi: 10.1186/1471-2105-12-74 PMCID: PMC3072341

375: A novel approach to the clustering of microarray data via nonparametric density estimation Riccardo De Bin, Davide Risso BMC Bioinformatics. 2011; 12: 49. Published online 2011 Feb 8. doi: 10.1186/1471-2105-12-49 PMCID: PMC3042915

376: A genome alignment algorithm based on compression Minh Duc Cao, Trevor I Dix, Lloyd Allison BMC Bioinformatics. 2010; 11: 599. Published online 2010 Dec 16. doi: 10.1186/1471-2105-11-599 PMCID: PMC3022628

377: Pash 3.0: A versatile software package for read mapping and integrative analysis of genomic and epigenomic variation using massively parallel DNA sequencing Cristian Coarfa, Fuli Yu, Christopher A Miller, Zuozhou Chen, R Alan Harris, Aleksandar Milosavljevic BMC Bioinformatics. 2010; 11: 572. Published online 2010 Nov 23. doi: 10.1186/1471-2105-11-572 PMCID: PMC3001746

378: Detecting Salient Aspects in Online Reviews of Health Providers Samuel Brody, Noémie Elhadad AMIA Annu Symp Proc. 2010; 2010: 202–206. Published online 2010 Nov 13. PMCID: PMC3041395

379: Evaluating the Importance of Image-related Text for Ad-hoc and Case-based Biomedical Article Retrieval Matthew S. Simpson, Dina Demner-Fushman, George R. Thoma AMIA Annu Symp Proc. 2010; 2010: 752–756. Published online 2010 Nov 13. PMCID: PMC3041350

380: A Prospective Evaluation of an Automated Classification System to Support Evidence-based Medicine and Systematic Review Aaron M. Cohen, Kyle Ambert, Marian McDonagh AMIA Annu Symp Proc. 2010; 2010: 121–125. Published online 2010 Nov 13. PMCID: PMC3041348

381: A High Throughput Semantic Concept Frequency Based Approach for Patient Identification: A Case Study Using Type 2 Diabetes Mellitus Clinical Notes Wei-Qi Wei, Cui Tao, Guoqian

Jiang, Christopher G. Chute AMIA Annu Symp Proc. 2010; 2010: 857–861. Published online 2010 Nov 13. PMCID: PMC3041302

382: A review on systematic reviews of health information system studies Francis Lau, Craig Kuziemsky, Morgan Price, Jesse Gardner J Am Med Inform Assoc. 2010 Nov-Dec; 17(6): 637–645. doi: 10.1136/jamia.2010.004838 PMCID: PMC3000756

383: An integrative method for scoring candidate genes from association studies: application to warfarin dosing Nicholas P Tatonetti, Joel T Dudley, Hersh Sagreiya, Atul J Butte, Russ B Altman BMC Bioinformatics. 2010; 11(Suppl 9): S9. Published online 2010 Oct 28. doi: 10.1186/1471-2105-11-S9-S9 PMCID: PMC2967750

384: Feasibility of incorporating genomic knowledge into electronic medical records for pharmacogenomic clinical decision support Casey Lynnette Overby, Peter Tarczy-Hornoch, James I Hoath, Ira J Kalet, David L Veenstra BMC Bioinformatics. 2010; 11(Suppl 9): S10. Published online 2010 Oct 28. doi: 10.1186/1471-2105-11-S9-S10 PMCID: PMC2967740

385: Data-driven approach for creating synthetic electronic medical records Anna L Buczak, Steven Babin, Linda Moniz BMC Med Inform Decis Mak. 2010; 10: 59. Published online 2010 Oct 14. doi: 10.1186/1472-6947-10-59 PMCID: PMC2972239

386: Medie and Info-pubmed: 2010 update Tomoko Ohta, Takuya Matsuzaki, Naoaki Okazaki, Makoto Miwa, Rune Sætre, Sampo Pyysalo, Jun'ichi Tsujii BMC Bioinformatics. 2010; 11(Suppl 5): P7. Published online 2010 Oct 6. doi: 10.1186/1471-2105-11-S5-P7 PMCID: PMC2956400

387: ExaCT: automatic extraction of clinical trial characteristics from journal publications Svetlana Kiritchenko, Berry de Bruijn, Simona Carini, Joel Martin, Ida Sim BMC Med Inform Decis Mak. 2010; 10: 56. Published online 2010 Sep 28. doi: 10.1186/1472-6947-10-56 PMCID: PMC2954855

388: Development of a universal psycho-educational intervention to prevent common postpartum mental disorders in primiparous women: a multiple method approach Heather J Rowe, Jane RW Fisher BMC Public Health. 2010; 10: 499. Published online 2010 Aug 18. doi: 10.1186/1471-2458-10-499 PMCID: PMC2931475

389: The Eat Smart Study: A randomised controlled trial of a reduced carbohydrate versus a low fat diet for weight loss in obese adolescents Helen Truby, Kimberley A Baxter, Paula Barrett, Robert S Ware, John C Cardinal, Peter SW Davies, Lynne A Daniels, Jennifer A Batch BMC Public Health. 2010; 10: 464. Published online 2010 Aug 9. doi: 10.1186/1471-2458-10-464 PMCID: PMC2925340

390: Survival dimensionality reduction (SDR): development and clinical application of an innovative approach to detect epistasis in presence of right-censored data Lorenzo Beretta, Alessandro Santaniello, Piet LCM van Riel, Marieke JH Coenen, Raffaella Scorza BMC Bioinformatics. 2010; 11: 416. Published online 2010 Aug 6. doi: 10.1186/1471-2105-11-416 PMCID: PMC2928804

391: Automatically Extracting Information Needs from Complex Clinical Questions Yong-gang Cao, James J Cimino, John Ely, Hong Yu J Biomed Inform. Author manuscript; available in PMC 2011 Dec 1.Published in final edited form as: J Biomed Inform. 2010 Dec; 43(6): 962–971. Published online 2010 Jul 27. doi: 10.1016/j.jbi.2010.07.007 PMCID: PMC2991382

392: Supporting Retrieval of Diverse Biomedical Data Using Evidence-aware Queries Eithon Cadag, Peter Tarczy-Hornoch J Biomed Inform. Author manuscript; available in PMC 2011 Dec 1.Published in final edited form as: J Biomed Inform. 2010 Dec; 43(6): 873–882. Published online 2010 Jul 17. doi: 10.1016/j.jbi.2010.07.005 PMCID: PMC3059407

393: A flexible R package for nonnegative matrix factorization Renaud Gaujoux, Cathal Seoighe BMC Bioinformatics. 2010; 11: 367. Published online 2010 Jul 2. doi: 10.1186/1471-2105-11-367 PMCID: PMC2912887

394: A new algorithm for reducing the workload of experts in performing systematic reviews Stan Matwin, Alexandre Kouznetsov, Diana Inkpen, Oana Frunza, Peter O'Blenis J Am Med Inform Assoc. 2010 Jul-Aug; 17(4): 446–453. doi: 10.1136/jamia.2010.004325 PMCID: PMC2995653

395: UFFizi: a generic platform for ranking informative features Assaf Gottlieb, Roy Varshavsky, Michal Linial, David Horn BMC Bioinformatics. 2010; 11: 300. Published online 2010 Jun 3. doi: 10.1186/1471-2105-11-300 PMCID: PMC2893168

396: Motivational interviewing for screening and feedback and encouraging lifestyle changes to reduce relative weight in 4-8 year old children: design of the MInT study Rachael W Taylor, Deirdre Brown, Anna M Dawson, Jill Haszard, Adell Cox, Elaine A Rose, Barry J Taylor, Kim Meredith-Jones, Lee Treacy, Jim Ross, Sheila M William BMC Public Health. 2010; 10: 271. Published online 2010 May 24. doi: 10.1186/1471-2458-10-271 PMCID: PMC2888742

397: An overview of MetaMap: historical perspective and recent advances Alan R Aronson, François-Michel Lang J Am Med Inform Assoc. 2010 May-Jun; 17(3): 229–236. doi: 10.1136/jamia.2009.002733 PMCID: PMC2995713

398: Concept-based query expansion for retrieving gene related publications from MEDLINE Sérgio Matos, Joel P Arrais, João Maia-Rodrigues, José Luis Oliveira BMC Bioinformatics. 2010; 11: 212. Published online 2010 Apr 28. doi: 10.1186/1471-2105-11-212 PMCID: PMC2873540

399: TECNOB: study design of a randomized controlled trial of a multidisciplinary telecare intervention for obese patients with type-2 diabetes Gianluca Castelnuovo, Gian Mauro Manzoni, Paola Cuzziol, Gian Luca Cesa, Cristina Tuzzi, Valentina Villa, Antonio Liuzzi, Maria Letizia Petroni, Enrico Molinari BMC Public Health. 2010; 10: 204. Published online 2010 Apr 23. doi: 10.1186/1471-2458-10-204 PMCID: PMC2873580

400: Predicting protein-protein interactions in unbalanced data using the primary structure of proteins Chi-Yuan Yu, Lih-Ching Chou, Darby Tien-Hao Chang BMC Bioinformatics. 2010; 11: 167. Published online 2010 Apr 2. doi: 10.1186/1471-2105-11-167 PMCID: PMC2868006

401: Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes Wei Yu, Tiebin Liu, Rodolfo Valdez, Marta Gwinn, Muin J Khoury BMC Med Inform Decis Mak. 2010; 10: 16. Published online 2010 Mar 22. doi: 10.1186/1472-6947-10-16 PMCID: PMC2850872

402: Semi-automated screening of biomedical citations for systematic reviews Byron C Wallace, Thomas A Trikalinos, Joseph Lau, Carla Brodley, Christopher H Schmid BMC Bioinformatics. 2010; 11: 55. Published online 2010 Jan 26. doi: 10.1186/1471-2105-11-55 PMCID: PMC2824679

403: Classification of protein sequences by means of irredundant patterns Matteo Comin, Davide Verzotto BMC Bioinformatics. 2010; 11(Suppl 1): S16. Published online 2010 Jan 18. doi: 10.1186/1471-2105-11-S1-S16 PMCID: PMC3009487

404: Use of statistical analysis in the biomedical informatics literature Matthew Scotch, Mona Duggal, Cynthia Brandt, Zhenqui Lin, Richard Shiffman J Am Med Inform Assoc. 2010 Jan-Feb; 17(1): 3–5. doi: 10.1197/jamia.M2853 PMCID: PMC2995622

405: Formal Representations of Eligibility Criteria: A Literature Review Chunhua Weng, Samson W. Tu, Ida Sim, Rachel Richesson J Biomed Inform. Author manuscript; available in PMC 2011 Jun 1.Published in final edited form as: J Biomed Inform. 2010 Jun; 43(3): 451–467. Published online 2009 Dec 23. doi: 10.1016/j.jbi.2009.12.004 PMCID: PMC2878905

406: Evaluation of Probabilistic and Logical Inference for a SNP Annotation System Terry H. Shen, Peter Tarczy-Hornoch, Landon T. Detwiler, Eithon Cadag, Christopher S. Carlson J Biomed Inform. Author manuscript; available in PMC 2011 Jun 1.Published in final edited form as: J Biomed

Inform. 2010 Jun; 43(3): 407–418. Published online 2009 Dec 14. doi: 10.1016/j.jbi.2009.12.002 PMCID: PMC2878960

407: Computer Surveillance of Hospital-Acquired Infections: A 25 year Update R. Scott Evans, Rouett H. Abouzelof, Caroline W. Taylor, Vickie Anderson, Sharon Sumner, Sharon Soutter, Ruth Kleckner, James F. Lloyd AMIA Annu Symp Proc. 2009; 2009: 178–182. Published online 2009 Nov 14. PMCID: PMC2815388

408: Enhancing navigation in biomedical databases by community voting and database-driven text classification Timo Duchrow, Timur Shtatland, Daniel Guettler, Misha Pivovarov, Stefan Kramer, Ralph Weissleder BMC Bioinformatics. 2009; 10: 317. Published online 2009 Oct 3. doi: 10.1186/1471-2105-10-317 PMCID: PMC2768718

409: Challenges for automatically extracting molecular interactions from full-text articles Tara McIntosh, James R Curran BMC Bioinformatics. 2009; 10: 311. Published online 2009 Sep 24. doi: 10.1186/1471-2105-10-311 PMCID: PMC2761905

410: The first step in the development of text mining technology for cancer risk assessment: identifying and organizing scientific evidence in risk assessment literature Anna Korhonen, Ilona Silins, Lin Sun, Ulla Stenius BMC Bioinformatics. 2009; 10: 303. Published online 2009 Sep 22. doi: 10.1186/1471-2105-10-303 PMCID: PMC2759963

411: Cross-Topic Learning for Work Prioritization in Systematic Review Creation and Update Aaron M. Cohen, Kyle Ambert, Marian McDonagh J Am Med Inform Assoc. 2009 Sep-Oct; 16(5): 690–704. doi: 10.1197/jamia.M3162Correction in: J Am Med Inform Assoc. 2009 Nov-Dec; 16(6): 898. PMCID: PMC2744720

412: Improved mutation tagging with gene identifiers applied to membrane protein stability prediction Rainer Winnenburg, Conrad Plake, Michael Schroeder BMC Bioinformatics. 2009; 10(Suppl 8): S3. Published online 2009 Aug 27. doi: 10.1186/1471-2105-10-S8-S3 PMCID: PMC2745585

413: What can Natural Language Processing do for Clinical Decision Support? Dina Demner-Fushman, Wendy W. Chapman, Clement J. McDonald J Biomed Inform. Author manuscript; available in PMC 2010 Oct 1.Published in final edited form as: J Biomed Inform. 2009 Oct; 42(5): 760–772. Published online 2009 Aug 13. doi: 10.1016/j.jbi.2009.08.007 PMCID: PMC2757540

414: EasyCluster: a fast and efficient gene-oriented clustering tool for large-scale transcriptome data Ernesto Picardi, Flavio Mignone, Graziano Pesole BMC Bioinformatics. 2009; 10(Suppl 6): S10. Published online 2009 Jun 16. doi: 10.1186/1471-2105-10-S6-S10 PMCID: PMC2697633

415: Can differences in medical drug compliance between European countries be explained by social factors: analyses based on data from the European Social Survey, round 2 John Larsen, Henrik Stovring, Jakob Kragstrup, Dorte G Hansen BMC Public Health. 2009; 9: 145. Published online 2009 May 16. doi: 10.1186/1471-2458-9-145 PMCID: PMC2687449

416: Automated Semantic Indexing of Figure Captions to Improve Radiology Image Retrieval Charles E. Kahn, Jr., Daniel L. Rubin J Am Med Inform Assoc. 2009 May-Jun; 16(3): 380–386. doi: 10.1197/jamia.M2945 PMCID: PMC2732225

417: Sequential Result Refinement for Searching the Biomedical Literature L. Y. Tanaka, J. R. Herskovic, M. S. Iyengar, E. V. Bernstam J Biomed Inform. Author manuscript; available in PMC 2010 Aug 1.Published in final edited form as: J Biomed Inform. 2009 Aug; 42(4): 678–684. Published online 2009 Mar 9. doi: 10.1016/j.jbi.2009.02.009 PMCID: PMC2722929

418: How to successfully select and implement electronic health records (EHR) in small ambulatory practice settings Nancy M Lorenzi, Angelina Kouroubali, Don E Detmer, Meryl Bloomrosen BMC Med Inform Decis Mak. 2009; 9: 15. Published online 2009 Feb 23. doi: 10.1186/1472-6947-9-15 PMCID: PMC2662829

419: Empirical Distributional Semantics: Methods and Biomedical Applications Trevor Cohen, Dominic Widdows J Biomed Inform. Author manuscript; available in PMC 2010 Apr 1.Published in final edited form as: J Biomed Inform. 2009 Apr; 42(2): 390–405. Published online 2009 Feb 14. doi: 10.1016/j.jbi.2009.02.002 PMCID: PMC2750802

420: BioProspecting: novel marker discovery obtained by mining the bibleome Peter L Elkin, Mark S Tuttle, Brett E Trusko, Steven H Brown BMC Bioinformatics. 2009; 10(Suppl 2): S9. Published online 2009 Feb 5. doi: 10.1186/1471-2105-10-S2-S9 PMCID: PMC2646243

421: Is searching full text more effective than searching abstracts? Jimmy Lin BMC Bioinformatics. 2009; 10: 46. Published online 2009 Feb 3. doi: 10.1186/1471-2105-10-46 PMCID: PMC2695361

422: Towards Automatic Recognition of Scientifically Rigorous Clinical Research Evidence Halil Kilicoglu, Dina Demner-Fushman, Thomas C. Rindflesch, Nancy L. Wilczynski, R. Brian

Haynes J Am Med Inform Assoc. 2009 Jan-Feb; 16(1): 25–31. doi: 10.1197/jamia.M2996 PMCID: PMC2605595

423: Outstanding Submissions to the AMIA Annual Symposium Now Featured in JAMIA Lucila Ohno-Machado, Randolph A. Miller J Am Med Inform Assoc. 2009 Jan-Feb; 16(1): 143–144. doi: 10.1197/jamia.M3021 PMCID: PMC2605591

424: Mining clinical relationships from patient narratives Angus Roberts, Robert Gaizauskas, Mark Hepple, Yikun Guo BMC Bioinformatics. 2008; 9(Suppl 11): S3. Published online 2008 Nov 19. doi: 10.1186/1471-2105-9-S11-S3 PMCID: PMC2586752

425: New Standards and Enhanced Utility for Family Health History Information in the Electronic Health Record: An Update from the American Health Information Community's Family Health History Multi-Stakeholder Workgroup W. Gregory Feero, Mary Beth Bigley, Kristin M. Brinner, The Family Health History Multi-Stakeholder Workgroup of the American Health Information Community J Am Med Inform Assoc. 2008 Nov-Dec; 15(6): 723–728. doi: 10.1197/jamia.M2793 PMCID: PMC2585527

426: PredGPI: a GPI-anchor predictor Andrea Pierleoni, Pier Luigi Martelli, Rita Casadio BMC Bioinformatics. 2008; 9: 392. Published online 2008 Sep 23. doi: 10.1186/1471-2105-9-392 PMCID: PMC2571997

427: The Informatics Opportunities at the Intersection of Patient Safety and Clinical Informatics Peter M. Kilbridge, David C. Classen J Am Med Inform Assoc. 2008 Jul-Aug; 15(4): 397–407. doi: 10.1197/jamia.M2735 PMCID: PMC2442268

428: Advancing Personalized Health Care through Health Information Technology: An Update from the American Health Information Community's Personalized Health Care Workgroup John Glaser, Douglas E. Henley, Gregory Downing, Kristin M. Brinner, Personalized Health Care Workgroup of the American Health Information Community J Am Med Inform Assoc. 2008 Jul-Aug; 15(4): 391–396. doi: 10.1197/jamia.M2718 PMCID: PMC2442266

429: GAPscreener: An automatic tool for screening human genetic association literature in PubMed using the support vector machine technique Wei Yu, Melinda Clyne, Siobhan M Dolan, Ajay Yesupriya, Anja Wulf, Tiebin Liu, Muin J Khoury, Marta Gwinn BMC Bioinformatics. 2008; 9: 205. Published online 2008 Apr 22. doi: 10.1186/1471-2105-9-205 PMCID: PMC2387176

430: Network motif-based identification of transcription factor-target gene relationships by integrating multi-source biological data Yuji Zhang, Jianhua Xuan, Benildo G de los Reyes, Robert

Clarke, Habtom W Ressom BMC Bioinformatics. 2008; 9: 203. Published online 2008 Apr 21. doi: 10.1186/1471-2105-9-203 PMCID: PMC2386822

431: Modeling the Distribution of Nursing Effort Using Structured Labor and Delivery Documentation Eric S. Hall, Mollie R. Poynton, Scott P. Narus, Sidney N. Thornton J Biomed Inform. Author manuscript; available in PMC 2009 Dec 1.Published in final edited form as: J Biomed Inform. 2008 Dec; 41(6): 1001–1008. Published online 2008 Apr 20. doi: 10.1016/j.jbi.2008.04.005 PMCID: PMC2649767

432: Gene Ontology density estimation and discourse analysis for automatic GeneRiF extraction Julien Gobeill, Imad Tbahriti, Frédéric Ehrler, Anaïs Mottaz, Anne-Lise Veuthey, Patrick Ruch BMC Bioinformatics. 2008; 9(Suppl 3): S9. Published online 2008 Apr 11. doi: 10.1186/1471-2105-9-S3-S9 PMCID: PMC2352866

433: Exploiting and integrating rich features for biological literature classification Hongning Wang, Minlie Huang, Shilin Ding, Xiaoyan Zhu BMC Bioinformatics. 2008; 9(Suppl 3): S4. Published online 2008 Apr 11. doi: 10.1186/1471-2105-9-S3-S4 PMCID: PMC2349297

434: Is plant mitochondrial RNA editing a source of phylogenetic incongruence? An answer from in silico and in vivo data sets Ernesto Picardi, Carla Quagliariello BMC Bioinformatics. 2008; 9(Suppl 2): S14. Published online 2008 Mar 26. doi: 10.1186/1471-2105-9-S2-S14 PMCID: PMC2323663

435: MScanner: a classifier for retrieving Medline citations Graham L Poulter, Daniel L Rubin, Russ B Altman, Cathal Seoighe BMC Bioinformatics. 2008; 9: 108. Published online 2008 Feb 19. doi: 10.1186/1471-2105-9-108 PMCID: PMC2263023

436: Bioinformatics research in the Asia Pacific: a 2007 update Shoba Ranganathan, Michael Gribskov, Tin Wee Tan BMC Bioinformatics. 2008; 9(Suppl 1): S1. Published online 2008 Feb 13. doi: 10.1186/1471-2105-9-S1-S1 PMCID: PMC2259402

437: A genetic approach for building different alphabets for peptide and protein classification Loris Nanni, Alessandra Lumini BMC Bioinformatics. 2008; 9: 45. Published online 2008 Jan 24. doi: 10.1186/1471-2105-9-45 PMCID: PMC2246158

438: SYRIAC: The SYstematic Review Information Automated Collection System A Data Warehouse for Facilitating Automated Biomedical Text Classification Jianji J. Yang, Aaron M. Cohen, Marian S. McDonagh AMIA Annu Symp Proc. 2008; 2008: 825–829. Published online 2008. PMCID: PMC2656099

439: Optimizing Feature Representation for Automated Systematic Review Work Prioritization Aaron M. Cohen AMIA Annu Symp Proc. 2008; 2008: 121–125. Published online 2008. PMCID: PMC2656096

440: An Integrated Computerized Triage System in the Emergency Department Dominik Aronsky, Ian Jones, Bill Raines, Robin Hemphill, Scott R Mayberry, Melissa A Luther, Ted Slusser AMIA Annu Symp Proc. 2008; 2008: 16–20. Published online 2008. PMCID: PMC2656061

441: Automatic Quality of Life Prediction Using Electronic Medical Records Serguei Pakhomov, Nilay Shah, Penny Hanson, Saranya Balasubramaniam, Steven A. Smith AMIA Annu Symp Proc. 2008; 2008: 545–549. Published online 2008. PMCID: PMC2656045

442: Electronic Medical Record (EMR) Utilization for Public Health Surveillance Zaruhi R. Mnatsakanyan, Daniel J. Mollura, John R. Ticehurst, Mohammad R. Hashemian, Lang M. Hung AMIA Annu Symp Proc. 2008; 2008: 480–484. Published online 2008. PMCID: PMC2656004

443: Physician Use of Outpatient Electronic Health Records to Improve Care Adam Wilcox, Watson A. Bowes, III, Sidney N. Thornton, Scott P. Narus AMIA Annu Symp Proc. 2008; 2008: 809–813. Published online 2008. PMCID: PMC2655996

444: Integrating an automatic classification method into the medical image retrieval process Epaphrodite Uwimana, Miguel E. Ruiz AMIA Annu Symp Proc. 2008; 2008: 747–751. Published online 2008. PMCID: PMC2655992

445: Evaluation of a Document Search Engine in a Clinical Department System Stefan Schulz, Philipp Daumke, Pascal Fischer, Marcel Müller AMIA Annu Symp Proc. 2008; 2008: 647–651. Published online 2008. PMCID: PMC2655987

446: The Value of Personal Health Record (PHR) Systems David Kaelber, Eric C Pan AMIA Annu Symp Proc. 2008; 2008: 343–347. Published online 2008. PMCID: PMC2655982

447: Identification of Inactive Medications in Narrative Medical Text Eugene M. Breydo, Julia T. Chu, Alexander Turchin AMIA Annu Symp Proc. 2008; 2008: 66–70. Published online 2008. PMCID: PMC2655977

448: An Electronic Health Record Based on Structured Narrative Stephen B. Johnson, Suzanne Bakken, Daniel Dine, Sookyung Hyun, Eneida Mendonça, Frances Morrison, Tiffani Bright, Tielman Van Vleck, Jesse Wrenn, Peter Stetson J Am Med Inform Assoc. 2008 Jan-Feb; 15(1): 54–64. doi: 10.1197/jamia.M2131 PMCID: PMC2274868

449: A coherent graph-based semantic clustering and summarization approach for biomedical literature and a new summarization evaluation method Illhoi Yoo, Xiaohua Hu, Il-Yeol Song BMC Bioinformatics. 2007; 8(Suppl 9): S4. Published online 2007 Nov 27. doi: 10.1186/1471-2105-8-S9-S4 PMCID: PMC2217662

450: Identification of DNA-binding proteins using support vector machines and evolutionary profiles Manish Kumar, Michael M Gromiha, Gajendra PS Raghava BMC Bioinformatics. 2007; 8: 463. Published online 2007 Nov 27. doi: 10.1186/1471-2105-8-463 PMCID: PMC2216048

451: Computational analyses of eukaryotic promoters Michael Q Zhang BMC Bioinformatics. 2007; 8(Suppl 6): S3. Published online 2007 Sep 27. doi: 10.1186/1471-2105-8-S6-S3 PMCID: PMC1995544

452: Dissecting complex transcriptional responses using pathway-level scores based on prior information Harmen J Bussemaker, Lucas D Ward, Andre Boorsma BMC Bioinformatics. 2007; 8(Suppl 6): S6. Published online 2007 Sep 27. doi: 10.1186/1471-2105-8-S6-S6 PMCID: PMC1995543

453: A machine learning approach for the identification of odorant binding proteins from sequence-derived properties Ganesan Pugalenthi, Ke Tang, PN Suganthan, G Archunan, R Sowdhamini BMC Bioinformatics. 2007; 8: 351. Published online 2007 Sep 19. doi: 10.1186/1471-2105-8-351 PMCID: PMC2216042

454: Protein subcellular localization prediction based on compartment-specific features and structure conservation Emily Chia-Yu Su, Hua-Sheng Chiu, Allan Lo, Jenn-Kang Hwang, Ting-Yi Sung, Wen-Lian Hsu BMC Bioinformatics. 2007; 8: 330. Published online 2007 Sep 8. doi: 10.1186/1471-2105-8-330 PMCID: PMC2040162

455: Automating document classification for the Immune Epitope Database Peng Wang, Alexander A Morgan, Qing Zhang, Alessandro Sette, Bjoern Peters BMC Bioinformatics. 2007; 8: 269. Published online 2007 Jul 26. doi: 10.1186/1471-2105-8-269 PMCID: PMC1965490

456: Using contextual and lexical features to restructure and validate the classification of biomedical concepts Jung-Wei Fan, Hua Xu, Carol Friedman BMC Bioinformatics. 2007; 8: 264. Published online 2007 Jul 24. doi: 10.1186/1471-2105-8-264 PMCID: PMC2014782

457: Communication Outcomes of Critical Imaging Results in a Computerized Notification System Hardeep Singh, Harvinder S. Arora, Meena S. Vij, Raghuram Rao, Myrna M. Khan, Laura A.

Petersen J Am Med Inform Assoc. 2007 Jul-Aug; 14(4): 459–466. doi: 10.1197/jamia.M2280 PMCID: PMC2244901

458: Essie: A Concept-based Search Engine for Structured Biomedical Text Nicholas C. Ide, Russell F. Loane, Dina Demner-Fushman J Am Med Inform Assoc. 2007 May-Jun; 14(3): 253–263. doi: 10.1197/jamia.M2233 PMCID: PMC2244877

459: Conceptual Knowledge Acquisition in Biomedicine: A Methodological Review Philip R.O. Payne, Eneida A. Mendonça, Stephen B. Johnson, Justin B. Starren J Biomed Inform. Author manuscript; available in PMC 2008 Oct 1.Published in final edited form as: J Biomed Inform. 2007 Oct; 40(5): 582–602. Published online 2007 Mar 27. doi: 10.1016/j.jbi.2007.03.005 PMCID: PMC2082059

460: The BioPrompt-box: an ontology-based clustering tool for searching in biological databases Claudio Corsi, Paolo Ferragina, Roberto Marangoni BMC Bioinformatics. 2007; 8(Suppl 1): S8. Published online 2007 Mar 8. doi: 10.1186/1471-2105-8-S1-S8 PMCID: PMC1885860

461: Biowep: a workflow enactment portal for bioinformatics applications Paolo Romano, Ezio Bartocci, Guglielmo Bertolini, Flavio De Paoli, Domenico Marra, Giancarlo Mauri, Emanuela Merelli, Luciano Milanesi BMC Bioinformatics. 2007; 8(Suppl 1): S19. Published online 2007 Mar 8. doi: 10.1186/1471-2105-8-S1-S19 PMCID: PMC1885848

462: A simulation study comparing aberration detection algorithms for syndromic surveillance Michael L Jackson, Atar Baer, Ian Painter, Jeff Duchin BMC Med Inform Decis Mak. 2007; 7: 6. Published online 2007 Mar 1. doi: 10.1186/1472-6947-7-6 PMCID: PMC1821319

463: Bioinformatics analysis of the early inflammatory response in a rat thermal injury model Eric Yang, Timothy Maguire, Martin L Yarmush, Francois Berthiaume, Ioannis P Androulakis BMC Bioinformatics. 2007; 8: 10. Published online 2007 Jan 10. doi: 10.1186/1471-2105-8-10 PMCID: PMC1797813

464: Identification of Misspelled Words without a Comprehensive Dictionary Using Prevalence Analysis Alexander Turchin, Julia T. Chu, Maria Shubina, Jonathan S. Einbinder AMIA Annu Symp Proc. 2007; 2007: 751–755. Published online 2007. PMCID: PMC2813663

465: The CLEF Corpus: Semantic Annotation of Clinical Text Angus Roberts, Robert Gaizauskas, Mark Hepple, Neil Davis, George Demetriou, Yikun Guo, Jay (Subbarao) Kola, Ian Roberts, Andrea Setzer, Archana Tapuria, Bill Wheeldin AMIA Annu Symp Proc. 2007; 2007: 625–629. Published online 2007. PMCID: PMC2655900

466: A comparative analysis of retrieval features used in the TREC 2006 Genomics Track passage retrieval task Hari Krishna Rekapalli, Aaron M. Cohen, William R. Hersh AMIA Annu Symp Proc. 2007; 2007: 620–624. Published online 2007. PMCID: PMC2655837

467: Investigating CBIR Techniques for Cervicographic Images Zhiyun Xue, Sameer Antani, L. Rodney Long, Jose Jeronimo, George R. Thoma AMIA Annu Symp Proc. 2007; 2007: 826–830. Published online 2007. PMCID: PMC2655825

468: Prevalence and predictors of home and automobile smoking bans and child environmental tobacco smoke exposure: a cross-sectional study of U.S.- and Mexico-born Hispanic women with young children Melissa Gonzales, Lorraine Halinka Malcoe, Michelle C Kegler, Judith Espinoza BMC Public Health. 2006; 6: 265. Published online 2006 Oct 27. doi: 10.1186/1471-2458-6-265 PMCID: PMC1636637

469: Approaches to the evaluation of outbreak detection methods Rochelle E Watkins, Serryn Eagleson, Robert G Hall, Lynne Dailey, Aileen J Plant BMC Public Health. 2006; 6: 263. Published online 2006 Oct 24. doi: 10.1186/1471-2458-6-263 PMCID: PMC1626088

470: SVRMHC prediction server for MHC-binding peptides Ji Wan, Wen Liu, Qiqi Xu, Yongliang Ren, Darren R Flower, Tongbin Li BMC Bioinformatics. 2006; 7: 463. Published online 2006 Oct 23. doi: 10.1186/1471-2105-7-463 PMCID: PMC1626489

471: Creating a medical English-Swedish dictionary using interactive word alignment Mikael Nyström, Magnus Merkel, Lars Ahrenberg, Pierre Zweigenbaum, Håkan Petersson, Hans Åhlfeldt BMC Med Inform Decis Mak. 2006; 6: 35. Published online 2006 Oct 12. doi: 10.1186/1472-6947-6-35 PMCID: PMC1624822

472: Advancing Biomedical Image Retrieval: Development and Analysis of a Test Collection William R. Hersh, Henning Müller, Jeffery R. Jensen, Jianji Yang, Paul N. Gorman, Patrick Ruch J Am Med Inform Assoc. 2006 Sep-Oct; 13(5): 488–496. doi: 10.1197/jamia.M2082 PMCID: PMC1561788

473: A Comparison of Citation Metrics to Machine Learning Filters for the Identification of High Quality MEDLINE Documents Yindalon Aphinyanaphongs, Alexander Statnikov, Constantin F. Aliferis J Am Med Inform Assoc. 2006 Jul-Aug; 13(4): 446–455. doi: 10.1197/jamia.M2031 PMCID: PMC1513679

474: Retrieval with gene queries Aditya K Sehgal, Padmini Srinivasan BMC Bioinformatics. 2006; 7: 220. Published online 2006 Apr 21. doi: 10.1186/1471-2105-7-220 PMCID: PMC1482725

475: Individual characteristics, area social participation, and primary non-concordance with medication: a multilevel analysis Kristina Johnell, Martin Lindström, Jan Sundquist, Charli Eriksson, Juan Merlo BMC Public Health. 2006; 6: 52. Published online 2006 Mar 2. doi: 10.1186/1471-2458-6-52 PMCID: PMC1409782

476: A multivariate prediction model for microarray cross-hybridization Yian A Chen, Cheng-Chung Chou, Xinghua Lu, Elizabeth H Slate, Konan Peck, Wenying Xu, Eberhard O Voit, Jonas S Almeida BMC Bioinformatics. 2006; 7: 101. Published online 2006 Mar 1. doi: 10.1186/1471-2105-7-101 PMCID: PMC1409802

477: MachineProse: an Ontological Framework for Scientific Assertions Deendayal Dinakarpandian, Yugyung Lee, Kartik Vishwanath, Rohini Lingambhotla J Am Med Inform Assoc. 2006 Mar-Apr; 13(2): 220–232. doi: 10.1197/jamia.M1910 PMCID: PMC1447552

478: Reducing Workload in Systematic Review Preparation Using Automated Citation Classification A.M. Cohen, W.R. Hersh, K. Peterson, Po-Yin Yen J Am Med Inform Assoc. 2006 Mar-Apr; 13(2): 206–219. doi: 10.1197/jamia.M1929 PMCID: PMC1447545

479: Prediction of protein structural class with Rough Sets Youfang Cao, Shi Liu, Lida Zhang, Jie Qin, Jiang Wang, Kexuan Tang BMC Bioinformatics. 2006; 7: 20. Published online 2006 Jan 14. doi: 10.1186/1471-2105-7-20 PMCID: PMC1363362

480: Prospective Validation of Text Categorization Filters for Identifying High-Quality, Content-Specific Articles in MEDLINE. Y. Aphinyanaphongs, C.F. Aliferis AMIA Annu Symp Proc. 2006; 2006: 6–10. PMCID: PMC1839419

481: An Effective General Purpose Approach for Automated Biomedical Document Classification Aaron M. Cohen AMIA Annu Symp Proc. 2006; 2006: 161–165. PMCID: PMC1839342

482: Assisting Consumer Health Information Retrieval with Query Recommendations Qing T. Zeng, Jonathan Crowell, Robert M. Plovnick, Eunjung Kim, Long Ngo, Emily Dibble J Am Med Inform Assoc. 2006 Jan-Feb; 13(1): 80–90. doi: 10.1197/jamia.M1820 PMCID: PMC1380203

483: Using Citation Data to Improve Retrieval from MEDLINE Elmer V. Bernstam, Jorge R. Herskovic, Yindalon Aphinyanaphongs, Constantin F. Aliferis, Madurai G. Sriram, William R. Hersh J Am Med Inform Assoc. 2006 Jan-Feb; 13(1): 96–105. doi: 10.1197/jamia.M1909 PMCID: PMC1380202

484: KnowledgeLink: Impact of Context-Sensitive Information Retrieval on Clinicians' Information Needs Saverio M. Maviglia, Catherine S. Yoon, David W. Bates, Gilad Kuperman J Am Med Inform Assoc. 2006 Jan-Feb; 13(1): 67–73. doi: 10.1197/jamia.M1861 PMCID: PMC1380199

485: Automatically Identifying Health Outcome Information in MEDLINE Records Dina Demner-Fushman, Barbara Few, Susan E. Hauser, George Thoma J Am Med Inform Assoc. 2006 Jan-Feb; 13(1): 52–60. doi: 10.1197/jamia.M1911 PMCID: PMC1380197

486: Exploring and Developing Consumer Health Vocabularies Qing T. Zeng, Tony Tse J Am Med Inform Assoc. 2006 Jan-Feb; 13(1): 24–29. doi: 10.1197/jamia.M1761 PMCID: PMC1380193

487: Relationship Structures and Semantic Type Assignments of the UMLS Enriched Semantic Network Li Zhang, Michael Halper, Yehoshua Perl, James Geller, James J. Cimino J Am Med Inform Assoc. 2005 Nov-Dec; 12(6): 657–666. doi: 10.1197/jamia.M1605 PMCID: PMC1294037

488: PHSkb: A knowledgebase to support notifiable disease surveillance Timothy J Doyle, Haobo Ma, Samuel L Groseclose, Richard S Hopkins BMC Med Inform Decis Mak. 2005; 5: 27. Published online 2005 Aug 16. doi: 10.1186/1472-6947-5-27 PMCID: PMC1201144

489: Data-poor categorization and passage retrieval for Gene Ontology Annotation in Swiss-Prot Frédéric Ehrler, Antoine Geissbühler, Antonio Jimeno, Patrick Ruch BMC Bioinformatics. 2005; 6(Suppl 1): S23. Published online 2005 May 24. doi: 10.1186/1471-2105-6-S1-S23 PMCID: PMC1869016

490: Evaluation of BioCreAtIvE assessment of task 2 Christian Blaschke, Eduardo Andres Leon, Martin Krallinger, Alfonso Valencia BMC Bioinformatics. 2005; 6(Suppl 1): S16. Published online 2005 May 24. doi: 10.1186/1471-2105-6-S1-S16 PMCID: PMC1869008

491: Overview of BioCreAtIvE: critical assessment of information extraction for biology Lynette Hirschman, Alexander Yeh, Christian Blaschke, Alfonso Valencia BMC Bioinformatics. 2005; 6(Suppl 1): S1. Published online 2005 May 24. doi: 10.1186/1471-2105-6-S1-S1 PMCID: PMC1869002

492: Personalized online information search and visualization Dongquan Chen, Helmuth F Orthner, Susan M Sell BMC Med Inform Decis Mak. 2005; 5: 6. Published online 2005 Mar 14. doi: 10.1186/1472-6947-5-6 PMCID: PMC1079857

493: A Statistical Approach to Scanning the Biomedical Literature for Pharmacogenetics Knowledge Daniel L. Rubin, Caroline F. Thorn, Teri E. Klein, Russ B. Altman J Am Med Inform Assoc. 2005 Mar-Apr; 12(2): 121–129. doi: 10.1197/jamia.M1640Correction in: J Am Med Inform Assoc. 2005 May-Jun; 12(3): 364. PMCID: PMC551544

494: Extracting Drug-Drug Interaction Articles from MEDLINE to Improve the Content of Drug Databases Stephany Duda, Constantin Aliferis, Randolph Miller, Alexander Statnikov, Kevin Johnson AMIA Annu Symp Proc. 2005; 2005: 216–220. PMCID: PMC1560879

495: Physician use of electronic medical records: Issues and successes with direct data entry and physician productivity Paul D. Clayton, Scott P. Narus, Watson A. Bowes, III, Tammy S. Madsen, Adam B. Wilcox, Garth Orsmond, Beatriz Rocha, Sidney N. Thornton, Spencer Jones, Craig A. Jacobsen, Marc R. Udall, Michael L. Rhodes, Brent E. Wallace, Wayne Cannon, Jerry Gardner, Stan M. Huff, Linda Leckman AMIA Annu Symp Proc. 2005; 2005: 141–145. PMCID: PMC1560588

496: VA QUERI Informatics Paper: Information Technology for Clinical Guideline Implementation: Perceptions of Multidisciplinary Stakeholders Stacie Salsbury Lyons, Toni Tripp-Reimer, Bernard A. Sorofman, Jane E. DeWitt, Bonnie J. BootsMiller, Thomas E. Vaughn, Bradley N. Doebbeling J Am Med Inform Assoc. 2005 Jan-Feb; 12(1): 64–71. doi: 10.1197/jamia.M1495 PMCID: PMC543828

497: Wang Y, Zhao Y, Therneau TM, Atkinson EJ, Tafti AP, Zhang N, Amin S, Limper AH, Khosla S, Liu H. Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records. J Biomed Inform. 2020 Feb;102:103364. doi: 10.1016/j.jbi.2019.103364. Epub 2019 Dec 28. PMID: 31891765; PMCID: PMC7028517.

498: Masud MT, Mamun MA, Thapa K, Lee DH, Griffiths MD, Yang SH. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. J Biomed Inform. 2020 Mar;103:103371. doi: 10.1016/j.jbi.2019.103371. Epub 2020 Jan 11. PMID: 31935462.

499: De Silva K, Jönsson D, Demmer RT. A combined strategy of feature selection and machine learning to identify predictors of prediabetes. J Am Med Inform Assoc. 2020 Mar 1;27(3):396-406. doi: 10.1093/jamia/ocz204. PMID: 31889178.

500: Birgmeier J, Deisseroth CA, Hayward LE, Galhardo LMT, Tierno AP, Jagadeesh KA, Stenson PD, Cooper DN, Bernstein JA, Haeussler M, Bejerano G. AVADA: toward automated pathogenic variant evidence retrieval directly from the full-text literature. Genet Med. 2020 Feb;22(2):362-370. doi: 10.1038/s41436-019-0643-6. Epub 2019 Aug 30. PMID: 31467448; PMCID: PMC7301356.

501: Kim Y, Meystre SM. Ensemble method-based extraction of medication and related information from clinical texts. J Am Med Inform Assoc. 2020 Jan 1;27(1):31-38. doi: 10.1093/jamia/ocz100. PMID: 31282932; PMCID: PMC7489099.

502: Dezső Z, Ceccarelli M. Machine learning prediction of oncology drug targets based on protein and network properties. BMC Bioinformatics. 2020 Mar 14;21(1):104. doi: 10.1186/s12859-020-3442-9. PMID: 32171238; PMCID: PMC7071582.

503: Prahladh S, van Wyk J. Protocol for a scoping review of the current data practices in forensic medicine. Syst Rev. 2020 Apr 8;9(1):76. doi: 10.1186/s13643-020-01308-7. PMID: 32268922; PMCID: PMC7140479.

504: Yao X, Tsang T, Sun Q, Quinney S, Zhang P, Ning X, Li L, Shen L. Mining and visualizing high-order directional drug interaction effects using the FAERS database. BMC Med Inform Decis Mak. 2020 Mar 18;20(Suppl 2):50. doi: 10.1186/s12911-020-1053-z. PMID: 32183790; PMCID: PMC7079342.

505: Mantelakis A, Khajuria A. The applications of machine learning in plastic and reconstructive surgery: protocol of a systematic review. Syst Rev. 2020 Feb 28;9(1):44. doi: 10.1186/s13643-020-01304-x. PMID: 32111260; PMCID: PMC7047352.

506: Petti U, Baker S, Korhonen A. A systematic literature review of automatic Alzheimer's disease detection from speech and language. J Am Med Inform Assoc. 2020 Nov 1;27(11):1784-1797. doi: 10.1093/jamia/ocaa174. PMID: 32929494.

507: Ibrahim ZM, Wu H, Hamoud A, Stappen L, Dobson RJB, Agarossi A. On classifying sepsis heterogeneity in the ICU: insight using machine learning. J Am Med Inform Assoc. 2020 Mar 1;27(3):437-443. doi: 10.1093/jamia/ocz211. PMID: 31951005; PMCID: PMC7025363.

508: Omenka OI, Watson DP, Hendrie HC. Understanding the healthcare experiences and needs of African immigrants in the United States: a scoping review. BMC Public Health. 2020 Jan 8;20(1):27. doi: 10.1186/s12889-019-8127-9. PMID: 31914960; PMCID: PMC6950921.

509: Zheng L, He Z, Wei D, Keloth V, Fan JW, Lindemann L, Zhu X, Cimino JJ, Perl Y. A review of auditing techniques for the Unified Medical Language System. J Am Med Inform Assoc. 2020 Oct 1;27(10):1625-1638. doi: 10.1093/jamia/ocaa108. PMID: 32766692; PMCID: PMC7566540.

510: Yang X, Bian J, Fang R, Bjarnadottir RI, Hogan WR, Wu Y. Identifying relations of medications with adverse drug events using recurrent convolutional neural networks and gradient boosting. J Am Med Inform Assoc. 2020 Jan 1;27(1):65-72. doi: 10.1093/jamia/ocz144. PMID: 31504605; PMCID: PMC7489076.

511: Alfattni G, Peek N, Nenadic G. Extraction of temporal relations from clinical free text: A systematic review of current approaches. J Biomed Inform. 2020 Aug;108:103488. doi: 10.1016/j.jbi.2020.103488. Epub 2020 Jul 13. PMID: 32673788.

512: Moradi M, Dashti M, Samwald M. Summarization of biomedical articles using domain-specific word embeddings and graph ranking. J Biomed Inform. 2020 Jul;107:103452. doi: 10.1016/j.jbi.2020.103452. Epub 2020 May 19. PMID: 32439479.

513: Rios P, Radhakrishnan A, Williams C, Ramkissoon N, Pham B, Cormack GV, Grossman MR, Muller MP, Straus SE, Tricco AC. Preventing the transmission of COVID-19 and other coronaviruses in older adults aged 60years and above living in long-term care: a rapid review. Syst Rev. 2020 Sep 25;9(1):218. doi: 10.1186/s13643-020-01486-4. PMID: 32977848; PMCID: PMC7517751.

514: Vasilakes J, Bompelli A, Bishop JR, Adam TJ, Bodenreider O, Zhang R. Assessing the enrichment of dietary supplement coverage in the Unified Medical Language System. J Am Med Inform Assoc. 2020 Oct 1;27(10):1547-1555. doi: 10.1093/jamia/ocaa128. PMID: 32940692; PMCID: PMC7566420.

515: Massi MC, Ieva F, Lettieri E. Data mining application to healthcare fraud detection: a two-step unsupervised clustering method for outlier detection with administrative databases. BMC Med Inform Decis Mak. 2020 Jul 14;20(1):160. doi: 10.1186/s12911-020-01143-9. PMID: 32664923; PMCID: PMC7362640.

516: Yu G, Yang Y, Wang X, Zhen H, He G, Li Z, Zhao Y, Shu Q, Shu L. Adversarial active learning for the identification of medical concepts and annotation inconsistency. J Biomed Inform. 2020 Aug;108:103481. doi: 10.1016/j.jbi.2020.103481. Epub 2020 Jul 18. PMID: 32687985.

517: Pereira MF, Prahm C, Kolbenschlag J, Oliveira E, Rodrigues NF. Application of AR and VR in hand rehabilitation: A systematic review. J Biomed Inform. 2020 Nov;111:103584. doi: 10.1016/j.jbi.2020.103584. Epub 2020 Oct 2. PMID: 33011296.

518: Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. Syst Rev. 2020 Apr 2;9(1):73. doi: 10.1186/s13643-020-01324-7. PMID: 32241297; PMCID: PMC7118839.

519: Alimova I, Tutubalina E. Multiple features for clinical relation extraction: A machine learning approach. J Biomed Inform. 2020 Mar;103:103382. doi: 10.1016/j.jbi.2020.103382. Epub 2020 Feb 3. PMID: 32028051.

520: Lucini FR, Krewulak KD, Fiest KM, Bagshaw SM, Zuege DJ, Lee J, Stelfox HT. Natural language processing to measure the frequency and mode of communication between healthcare professionals and family members of critically ill patients. J Am Med Inform Assoc. 2020 Nov 17:ocaa263. doi: 10.1093/jamia/ocaa263. Epub ahead of print. PMID: 33201981.

521: Corny J, Rajkumar A, Martin O, Dode X, Lajonchère JP, Billuart O, Bézie Y, Buronfosse A. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. J Am Med Inform Assoc. 2020 Nov 1;27(11):1688-1694. doi: 10.1093/jamia/ocaa154. PMID: 32984901.

522: Olakotan OO, Yusof MM. Evaluating the alert appropriateness of clinical decision support systems in supporting clinical workflow. J Biomed Inform. 2020 Jun;106:103453. doi: 10.1016/j.jbi.2020.103453. Epub 2020 May 14. PMID: 32417444.

523: DeLozier S, Speltz P, Brito J, Tang LA, Wang J, Smith JC, Giuse D, Phillips E, Williams K, Strickland T, Davogustto G, Roden D, Denny JC. Real-time clinical note monitoring to detect conditions for rapid follow-up: A case study of clinical trial enrollment in drug-induced torsades de pointes and Stevens- Johnson syndrome. J Am Med Inform Assoc. 2020 Oct 29:ocaa213. doi: 10.1093/jamia/ocaa213. Epub ahead of print. PMID: 33120413.

524: Kronzer VL, Wang L, Liu H, Davis JM, Sparks JA, Crowson CS. Investigating the impact of disease and health record duration on the eMERGE algorithm for rheumatoid arthritis. J Am Med Inform Assoc. 2020 Apr 1;27(4):601-605. doi: 10.1093/jamia/ocaa014. PMID: 32134444.

525: Henry S, Wang Y, Shen F, Uzuner O. The 2019 National Natural language processing (NLP) Clinical Challenges (n2c2)/Open Health NLP (OHNLP) shared task on clinical concept normalization for clinical records. J Am Med Inform Assoc. 2020 Oct 1;27(10):1529-1537. doi: 10.1093/jamia/ocaa106. PMID: 32968800.

526: Thabtah F, Peebles D, Retzler J, Hathurusingha C. Dementia medical screening using mobile applications: A systematic review with a new mapping model. J Biomed Inform. 2020 Nov;111:103573. doi: 10.1016/j.jbi.2020.103573. Epub 2020 Sep 20. PMID: 32961306.

527: Prieto-González D, Castilla-Rodríguez I, González E, Couce ML. Automated generation of decision-tree models for the economic assessment of interventions for rare diseases using the RaDiOS ontology. J Biomed Inform. 2020 Oct;110:103563. doi: 10.1016/j.jbi.2020.103563. Epub 2020 Sep 12. PMID: 32931923.

528: Payrovnaziri SN, Chen Z, Rengifo-Moreno P, Miller T, Bian J, Chen JH, Liu X, He Z. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. J Am Med Inform Assoc. 2020 Jul 1;27(7):1173-1185. doi: 10.1093/jamia/ocaa053. PMID: 32417928.

529: Wang Y, Coiera E, Magrabi F. Can Unified Medical Language System-based semantic representation improve automated identification of patient safety incident reports by type and severity? J Am Med Inform Assoc. 2020 Oct 1;27(10):1502-1509. doi: 10.1093/jamia/ocaa082. PMID: 32574362; PMCID: PMC7566533.

530: Ji J, Hu L, Liu B, Li Y. Identifying and assessing the impact of key neighborhood-level determinants on geographic variation in stroke: a machine learning and multilevel modeling approach. BMC Public Health. 2020 Nov 7;20(1):1666. doi: 10.1186/s12889-020-09766-3. PMID: 33160324; PMCID: PMC7648288.

531: Memmel S, Sisario D, Zimmermann H, Sauer M, Sukhorukov VL, Djuzenova CS, Flentje M. FocAn: automated 3D analysis of DNA repair foci in image stacks acquired by confocal fluorescence microscopy. BMC Bioinformatics. 2020 Jan 28;21(1):27. doi: 10.1186/s12859-020-3370-8. PMID: 31992200; PMCID: PMC6986076.

532: Oliwa T, Furner B, Schmitt J, Schneider J, Ridgway JP. Development of a predictive model for retention in HIV care using natural language processing of clinical notes. J Am Med Inform Assoc. 2020 Nov 5:ocaa220. doi: 10.1093/jamia/ocaa220. Epub ahead of print. PMID: 33150369.

533: Alharbi A, Stevenson M. Refining Boolean queries to identify relevant studies for systematic review updates. J Am Med Inform Assoc. 2020 Nov 1;27(11):1658-1666. doi: 10.1093/jamia/ocaa148. PMID: 33067630.

534: Rybinski M, Xu J, Karimi S. Clinical trial search: Using biomedical language understanding models for re-ranking. J Biomed Inform. 2020 Sep;109:103530. doi: 10.1016/j.jbi.2020.103530. Epub 2020 Aug 18. PMID: 32818666.

535: Shen L, Wright A, Lee LS, Jajoo K, Nayor J, Landman A. Clinical decision support system, using expert consensus-derived logic and natural language processing, decreased sedation-type order errors for patients undergoing endoscopy. J Am Med Inform Assoc. 2020 Nov 11:ocaa250. doi: 10.1093/jamia/ocaa250. Epub ahead of print. PMID: 33175157.

536: You Y, Ru X, Lei W, Li T, Xiao M, Zheng H, Chen Y, Zhang L. Developing the novel bioinformatics algorithms to systematically investigate the connections among survival time, key genes and proteins for Glioblastoma multiforme. BMC Bioinformatics. 2020 Sep 17;21(Suppl 13):383. doi: 10.1186/s12859-020-03674-4. PMID: 32938364; PMCID: PMC7646399.

537: Irvin JA, Kondrich AA, Ko M, Rajpurkar P, Haghgoo B, Landon BE, Phillips RL, Petterson S, Ng AY, Basu S. Incorporating machine learning and social determinants of health indicators into prospective risk adjustment for health plan payments. BMC Public Health. 2020 May 1;20(1):608. doi: 10.1186/s12889-020-08735-0. PMID: 32357871; PMCID: PMC7195714.

538: Zhao Y, Zhou Y, Liu Y, Hao Y, Li M, Pu X, Li C, Wen Z. Uncovering the prognostic gene signatures for the improvement of risk stratification in cancers by using deep learning algorithm coupled with wavelet transform. BMC Bioinformatics. 2020 May 19;21(1):195. doi: 10.1186/s12859-020-03544-z. PMID: 32429941; PMCID: PMC7236453.

539: Ahmadzadeh M, Christie GJ, Cosco TD, Moreno S. Neuroimaging and analytical methods for studying the pathways from mild cognitive impairment to Alzheimer's disease: protocol for a rapid systematic review. Syst Rev. 2020 Apr 2;9(1):71. doi: 10.1186/s13643-020-01332-7. PMID: 32241302; PMCID: PMC7118884.

540: Yu P, Jiang T, Hailey D, Ma J, Qian S. The contribution of electronic health records to risk management through accreditation of residential aged care homes in Australia. BMC Med Inform Decis Mak. 2020 Mar 20;20(1):58. doi: 10.1186/s12911-020-1070-y. PMID: 32192492; PMCID: PMC7082951.

541: Kumar P, Nestsiarovich A, Nelson SJ, Kerner B, Perkins DJ, Lambert CG. Imputation and characterization of uncoded self-harm in major mental illness using machine learning. J Am Med Inform Assoc. 2020 Jan 1;27(1):136-146. doi: 10.1093/jamia/ocz173. PMID: 31651956.

542: Christopoulou F, Tran TT, Sahu SK, Miwa M, Ananiadou S. Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods. J Am Med Inform Assoc. 2020 Jan 1;27(1):39-46. doi: 10.1093/jamia/ocz101. PMID: 31390003; PMCID: PMC6913215.

543: Lee DH, Yetisgen M, Vanderwende L, Horvitz E. Predicting severe clinical events by learning about life-saving actions and outcomes using distant supervision. J Biomed Inform. 2020 Jul;107:103425. doi: 10.1016/j.jbi.2020.103425. Epub 2020 Apr 26. PMID: 32348850.

544: Manduchi E, Fu W, Romano JD, Ruberto S, Moore JH. Embedding covariate adjustments in tree-based automated machine learning for biomedical big data analyses. BMC Bioinformatics. 2020 Oct 1;21(1):430. doi: 10.1186/s12859-020-03755-4. PMID: 32998684; PMCID: PMC7528347.

545: Feder A, Vainstein D, Rosenfeld R, Hartman T, Hassidim A, Matias Y. Active deep learning to detect demographic traits in free-form clinical notes. J Biomed Inform. 2020 Jul;107:103436. doi: 10.1016/j.jbi.2020.103436. Epub 2020 May 16. PMID: 32428572.

546: Zhang A, Teng L, Alterovitz G. An explainable machine learning platform for pyrazinamide resistance prediction and genetic feature identification of Mycobacterium tuberculosis. J Am Med Inform Assoc. 2020 Nov 20:ocaa233. doi: 10.1093/jamia/ocaa233. Epub ahead of print. PMID: 33215194.

547: Hamel C, Corace K, Hersi M, Rice D, Willows M, Macpherson P, Sproule B, Flores-Aranda J, Garber G, Esmaeilisaraji L, Skidmore B, Porath A, Ortiz Nunez R, Hutton B. Psychosocial and pharmacologic interventions for methamphetamine addiction: protocol for a scoping review of the literature. Syst Rev. 2020 Oct 24;9(1):245. doi: 10.1186/s13643-020-01499-z. PMID: 33099314; PMCID: PMC7585172.

548: Jauk S, Kramer D, Großauer B, Rienmüller S, Avian A, Berghold A, Leodolter W, Schulz S. Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study. J Am Med Inform Assoc. 2020 Jul 1;27(9):1383-1392. doi: 10.1093/jamia/ocaa113. PMID: 32968811.

549: Reddy SM, Patel S, Weyrich M, Fenton J, Viswanathan M. Comparison of a traditional systematic review approach with review-of-reviews and semi- automation as strategies to update the evidence. Syst Rev. 2020 Oct 19;9(1):243. doi: 10.1186/s13643-020-01450-2. PMID: 33076975; PMCID: PMC7574591.

550: Zheng H, Fei J, Zhang L, Zhou W, Ding Z, Hu W. Risk factor analysis of insufficient fluid intake among urban adults in Wuxi, China: a classification and regression tree analysis. BMC Public Health. 2020 Mar 4;20(1):286. doi: 10.1186/s12889-020-8380-y. PMID: 32131783; PMCID: PMC7057576.

551: Sharma B, Dligach D, Swope K, Salisbury-Afshar E, Karnik NS, Joyce C, Afshar M. Publicly available machine learning models for identifying opioid misuse from the clinical notes of hospitalized patients. BMC Med Inform Decis Mak. 2020 Apr 29;20(1):79. doi: 10.1186/s12911-020-1099-y. PMID: 32349766; PMCID: PMC7191715.

552: Zhao Z, Cristian A, Rosen G. Keeping up with the genomes: efficient learning of our increasing knowledge of the tree of life. BMC Bioinformatics. 2020 Sep 21;21(1):412. doi: 10.1186/s12859-020-03744-7. PMID: 32957925; PMCID: PMC7507296.

553: Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med Inform Decis Mak. 2020 Feb 3;20(1):16. doi: 10.1186/s12911-020-1023-5. PMID: 32013925; PMCID: PMC6998201.

554: Li Y, Du G, Xiang Y, Li S, Ma L, Shao D, Wang X, Chen H. Towards Chinese clinical named entity recognition by dynamic embedding using domain-specific knowledge. J Biomed Inform. 2020 Jun;106:103435. doi: 10.1016/j.jbi.2020.103435. Epub 2020 Apr 29. PMID: 32360988.

556: Metsker O, Magoev K, Yakovlev A, Yanishevskiy S, Kopanitsa G, Kovalchuk S, Krzhizhanovskaya VV. Identification of risk factors for patients with diabetes: diabetic polyneuropathy case study. BMC Med Inform Decis Mak. 2020 Aug 24;20(1):201. doi: 10.1186/s12911-020-01215-w. PMID: 32831065; PMCID: PMC7444272.

557: Nam JH, Couch D, da Silveira WA, Yu Z, Chung D. PALMER: improving pathway annotation based on the biomedical literature mining with a constrained latent block model. BMC Bioinformatics. 2020 Oct 2;21(1):432. doi: 10.1186/s12859-020-03756-3. PMID: 33008309; PMCID: PMC7532116.

558: Zhu J, Gallego B. Targeted estimation of heterogeneous treatment effect in observational survival analysis. J Biomed Inform. 2020 Jul;107:103474. doi: 10.1016/j.jbi.2020.103474. Epub 2020 Jun 18. PMID: 32562899.

559: Fu JT, Sholle E, Krichevsky S, Scandura J, Campion TR. Extracting and classifying diagnosis dates from clinical notes: A case study. J Biomed Inform. 2020 Oct;110:103569. doi: 10.1016/j.jbi.2020.103569. Epub 2020 Sep 16. PMID: 32949781.

560: Bari V, Hirsch JS, Narvaez J, Sardinia R, Bock KR, Oppenheim MI, Meytlis M. An approach to predicting patient experience through machine learning and social network analysis. J Am Med Inform Assoc. 2020 Oct 26:ocaa194. doi: 10.1093/jamia/ocaa194. Epub ahead of print. PMID: 33104210.

561: Narayan V, Hoong PB, Chuin S. Innovative Use of Health Informatics to Augment Contact Tracing during the COVID19 Pandemic in an Acute Hospital. J Am Med Inform Assoc. 2020 Aug 24:ocaa184. doi: 10.1093/jamia/ocaa184. Epub ahead of print. PMID: 32835358; PMCID: PMC7499570.

562: Carrell DS, Malin BA, Cronkite DJ, Aberdeen JS, Clark C, Li MR, Bastakoty D, Nyemba S, Hirschman L. Resilience of clinical text de-identified with "hiding in plain sight" to hostile reidentification attacks by human readers. J Am Med Inform Assoc. 2020 Jul 1;27(9):1374-1382. doi: 10.1093/jamia/ocaa095. PMID: 32930712.

563: Chen Q, Du J, Kim S, Wilbur WJ, Lu Z. Deep learning with sentence embeddings pre-trained on biomedical corpora improves the performance of finding similar sentences in electronic medical records. BMC Med Inform Decis Mak. 2020 Apr 30;20(Suppl 1):73. doi: 10.1186/s12911-020-1044-0. PMID: 32349758; PMCID: PMC7191680.

564: Chang YC, Wu JT, Hong MY, Tung YA, Hsieh PH, Yee SW, Giacomini KM, Oyang YJ, Chen CY; Alzheimer's Disease Neuroimaging Initiative. GenEpi: gene-based epistasis discovery using machine learning. BMC Bioinformatics. 2020 Feb 24;21(1):68. doi: 10.1186/s12859-020-3368-2. PMID: 32093643; PMCID: PMC7041299.

565: Mull HJ, Stolzmann K, Kalver E, Shin MH, Schweizer ML, Asundi A, Mehta P, Stanislawski M, Branch-Elliman W. Novel methodology to measure pre-procedure antimicrobial prophylaxis: integrating text searches with structured data from the Veterans Health Administration's electronic medical record. BMC Med Inform Decis Mak. 2020 Jan 30;20(1):15. doi: 10.1186/s12911-020-1031-5. PMID: 32000780; PMCID: PMC6993312.

566: Johnson SG, Pruinelli L, Westra BL. Machine Learned Mapping of Local EHR Flowsheet Data to Standard Information Models using Topic Model Filtering. AMIA Annu Symp Proc. 2020 Mar 4;2019:504-513. PMID: 32308844; PMCID: PMC7153147.

567: Liu H, Perl Y, Geller J. Concept placement using BERT trained by transforming and summarizing biomedical ontology structure. J Biomed Inform. 2020 Oct 22;112:103607. doi: 10.1016/j.jbi.2020.103607. Epub ahead of print. PMID: 33098987.

568: Slikboer R, Muir SD, Silva SSM, Meyer D. A systematic review of statistical models and outcomes of predicting fatal and serious injury crashes from driver crash and offense history data. Syst Rev. 2020 Sep 28;9(1):220. doi: 10.1186/s13643-020-01475-7. PMID: 32988410; PMCID: PMC7523043.

569: Rosenthal S, Das S, Hsueh PS, Barker K, Chen CH. Efficient goal attainment and engagement in a care manager system using unstructured notes. J Am Med Inform Assoc. 2020 Mar 6;3(1):62–9. doi: 10.1093/jamiaopen/ooaa001. Epub ahead of print. PMID: 32142137; PMCID: PMC7309242.

570: Kate RJ. Automatic full conversion of clinical terms into SNOMED CT concepts. J Biomed Inform. 2020 Nov;111:103585. doi: 10.1016/j.jbi.2020.103585. Epub 2020 Oct 2. PMID: 33011295.

571: Marshall IJ, Nye B, Kuiper J, Noel-Storr A, Marshall R, Maclean R, Soboczenski F, Nenkova A, Thomas J, Wallace BC. Trialstreamer: A living, automatically updated database of clinical trial reports. J Am Med Inform Assoc. 2020 Sep 17:ocaa163. doi: 10.1093/jamia/ocaa163. Epub ahead of print. PMID: 32940710.

572: O'Connor AM, Glasziou P, Taylor M, Thomas J, Spijker R, Wolfe MS. A focus on cross-purpose tools, automated recognition of study design in multiple disciplines, and evaluation of automation tools: a summary of significant discussions at the fourth meeting of the International Collaboration for Automation of Systematic Reviews (ICASR). Syst Rev. 2020 May 4;9(1):100. doi: 10.1186/s13643-020-01351-4. PMID: 32366302; PMCID: PMC7199360.

573: Eickelberg G, Sanchez-Pinto LN, Luo Y. Predictive modeling of bacterial infections and antibiotic therapy needs in critically ill adults. J Biomed Inform. 2020 Sep;109:103540. doi: 10.1016/j.jbi.2020.103540. Epub 2020 Aug 16. PMID: 32814200; PMCID: PMC7530142.

574: Li X, Zhang H, Zhou XH. Chinese clinical named entity recognition with variant neural structures based on BERT methods. J Biomed Inform. 2020 Jul;107:103422. doi: 10.1016/j.jbi.2020.103422. Epub 2020 Apr 28. PMID: 32353595.

575: Yuan C, Wang Y, Shang N, Li Z, Zhao R, Weng C. A graph-based method for reconstructing entities from coordination ellipsis in medical text. J Am Med Inform Assoc. 2020 Jul 1;27(9):1364-1373. doi: 10.1093/jamia/ocaa109. PMID: 32719840.

575: Bitton Y, Cohen R, Schifter T, Bachmat E, Elhadad M, Elhadad N. Cross- lingual Unified Medical Language System entity linking in online health communities. J Am Med Inform Assoc. 2020 Oct 1;27(10):1585-1592. doi: 10.1093/jamia/ocaa150. PMID: 32910823; PMCID: PMC7566404.

576: McCradden MD, Joshi S, Anderson JA, Mazwi M, Goldenberg A, Zlotnik Shaul R. Patient safety and quality improvement: Ethical principles for a regulatory approach to bias in healthcare machine learning. J Am Med Inform Assoc. 2020 Jun 25:ocaa085. doi: 10.1093/jamia/ocaa085. Epub ahead of print. PMID: 32585698.

577: Weeks HL, Beck C, McNeer E, Williams ML, Bejan CA, Denny JC, Choi L. medExtractR: A targeted, customizable approach to medication extraction from electronic health records. J Am Med Inform Assoc. 2020 Mar 1;27(3):407-418. doi: 10.1093/jamia/ocz207. PMID: 31943012; PMCID: PMC7025369.

578: Shaham A, Chodick G, Shalev V, Yamin D. Personal and social patterns predict influenza vaccination decision. BMC Public Health. 2020 Feb 12;20(1):222. doi: 10.1186/s12889-020-8327-3. PMID: 32050948; PMCID: PMC7017468.

579: Dai HJ, Wang FD, Chen CW, Su CH, Wu CS, Jonnagaddala J. Cohort selection for clinical trials using multiple instance learning. J Biomed Inform. 2020 Jul;107:103438. doi: 10.1016/j.jbi.2020.103438. Epub 2020 May 1. PMID: 32360937.

580: Kunneman F, Lambooij M, Wong A, Bosch AVD, Mollema L. Monitoring stance towards vaccination in twitter messages. BMC Med Inform Decis Mak. 2020 Feb 18;20(1):33. doi: 10.1186/s12911-020-1046-y. PMID: 32070334; PMCID: PMC7029499.

581: Hartman T, Howell MD, Dean J, Hoory S, Slyper R, Laish I, Gilon O, Vainstein D, Corrado G, Chou K, Po MJ, Williams J, Ellis S, Bee G, Hassidim A, Amira R, Beryozkin G, Szpektor I, Matias Y. Customization scenarios for de-identification of clinical notes. BMC Med Inform Decis Mak. 2020 Jan 30;20(1):14. doi: 10.1186/s12911-020-1026-2. PMID: 32000770; PMCID: PMC6993314.

582: Sinha S, Lynn AM, Desai DK. Implementation of homology based and non- homology based computational methods for the identification and annotation of orphan enzymes: using Mycobacterium tuberculosis H37Rv as a case study. BMC Bioinformatics. 2020 Oct 19;21(1):466. doi: 10.1186/s12859-020-03794-x. PMID: 33076816; PMCID: PMC7574302.

583: Yu KH, Wang F, Berry GJ, Ré C, Altman RB, Snyder M, Kohane IS. Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks. J Am Med Inform Assoc. 2020 May 1;27(5):757-769. doi: 10.1093/jamia/ocz230. PMID: 32364237; PMCID: PMC7309263.

584: Kim SM, Peña MI, Moll M, Bennett GN, Kavraki LE. Improving the organization and interactivity of metabolic pathfinding with precomputed pathways. BMC Bioinformatics. 2020 Jan 10;21(1):13. doi: 10.1186/s12859-019-3328-x. PMID: 31924164; PMCID: PMC6954563.

585: Koslovsky MD, Vannucci M. MicroBVS: Dirichlet-tree multinomial regression models with Bayesian variable selection - an R package. BMC Bioinformatics. 2020 Jul 13;21(1):301. doi: 10.1186/s12859-020-03640-0. PMID: 32660471; PMCID: PMC7359232.

586: Wang CCN, Jin J, Chang JG, Hayakawa M, Kitazawa A, Tsai JJP, Sheu PC. Identification of most influential co-occurring gene suites for gastrointestinal cancer using biomedical literature mining and graph-based influence maximization. BMC Med Inform Decis Mak. 2020 Sep 3;20(1):208. doi: 10.1186/s12911-020-01227-6. PMID: 32883271; PMCID: PMC7469322.

587: Cherian RP, Thomas N, Venkitachalam S. Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm. J Biomed Inform. 2020 Oct;110:103543. doi: 10.1016/j.jbi.2020.103543. Epub 2020 Aug 26. PMID: 32858167.

588: Kauchak D, Leroy G, Pei M, Colina S. Predicting Transition Words Between Sentence for English and Spanish Medical Text. AMIA Annu Symp Proc. 2020 Mar 4;2019:523-531. PMID: 32308846; PMCID: PMC7153060.

589: Campbell EA, Bass EJ, Masino AJ. Temporal condition pattern mining in large, sparse electronic health record data: A case study in characterizing pediatric asthma. J Am Med Inform Assoc. 2020 Apr 1;27(4):558-566. doi: 10.1093/jamia/ocaa005. PMID: 32049282; PMCID: PMC7075539.

590: Fang AHS, Lim WT, Balakrishnan T. Early warning score validation methodologies and performance metrics: a systematic review. BMC Med Inform Decis Mak. 2020 Jun 18;20(1):111. doi: 10.1186/s12911-020-01144-8. PMID: 32552702; PMCID: PMC7301346.

591: Kumar A, Sinha N, Bhardwaj A. A novel fitness function in genetic programming for medical data classification. J Biomed Inform. 2020 Nov 14;112:103623. doi: 10.1016/j.jbi.2020.103623. Epub ahead of print. PMID: 33197613.

592: Datta S, Si Y, Rodriguez L, Shooshan SE, Demner-Fushman D, Roberts K. Understanding spatial language in radiology: Representation framework, annotation, and spatial relation extraction from chest X-ray reports using deep learning. J Biomed Inform. 2020 Aug;108:103473. doi: 10.1016/j.jbi.2020.103473. Epub 2020 Jun 18. PMID: 32562898.

593: Jiang S, Wu W, Tomita N, Ganoe C, Hassanpour S. Multi-Ontology Refined Embeddings (MORE): A hybrid multi-ontology and corpus-based semantic representation model for biomedical concepts. J Biomed Inform. 2020 Nov;111:103581. doi: 10.1016/j.jbi.2020.103581. Epub 2020 Oct 1. PMID: 33010425; PMCID: PMC7665985.

594: Yokobori Y, Matsuura J, Sugiura Y, Mutemba C, Nyahoda M, Mwango C, Kazhumbula L, Yuasa M, Chiluba C. Analysis of causes of death among brought-in- dead cases in a third-level Hospital in Lusaka, Republic of Zambia, using the tariff method 2.0 for verbal autopsy: a cross-sectional study. BMC Public Health. 2020 Apr 10;20(1):473. doi: 10.1186/s12889-020-08575-y. PMID: 32272924; PMCID: PMC7147005.

595: Rafique O, Mir AH. Weighted dimensionality reduction and robust Gaussian mixture model based cancer patient subtyping from gene expression data. J Biomed Inform. 2020 Nov 11;112:103620. doi: 10.1016/j.jbi.2020.103620. Epub ahead of print. PMID: 33188907.

596: Eslami Manoochehri H, Nourani M. Drug-target interaction prediction using semi-bipartite graph model and deep learning. BMC Bioinformatics. 2020 Jul 6;21(Suppl 4):248. doi: 10.1186/s12859-020-3518-6. PMID: 32631230; PMCID: PMC7336396.

597: Golob JL, Minot SS. In silico benchmarking of metagenomic tools for coding sequence detection reveals the limits of sensitivity and precision. BMC Bioinformatics. 2020 Oct 15;21(1):459. doi: 10.1186/s12859-020-03802-0. PMID: 33059593; PMCID: PMC7559173.

598: Wang Y, Zhang X, Wang T, Xing J, Wu Z, Li W, Wang J. A machine learning framework for accurately recognizing circular RNAs for clinical decision- supporting. BMC Med Inform Decis Mak. 2020 Jul 9;20(Suppl 3):137. doi: 10.1186/s12911-020-1117-0. PMID: 32646420; PMCID: PMC7346313.

599: Wright CF, Eberhardt RY, Constantinou P, Hurles ME, FitzPatrick DR, Firth HV; DDD Study. Evaluating variants classified as pathogenic in ClinVar in the DDD Study. Genet Med. 2020 Nov 5. doi: 10.1038/s41436-020-01021-9. Epub ahead of print. PMID: 33149276.

600: Hu Y, Xi X, Yang Q, Zhang X. SCeQTL: an R package for identifying eQTL from single-cell parallel sequencing data. BMC Bioinformatics. 2020 May 11;21(1):184. doi: 10.1186/s12859-020-3534-6. PMID: 32393315; PMCID: PMC7216638.

601: Chen S, Ghandikota S, Gautam Y, Mersha TB. MI-MAAP: marker informativeness for multi-ancestry admixed populations. BMC Bioinformatics. 2020 Apr 3;21(1):131. doi: 10.1186/s12859-020-3462-5. PMID: 32245404; PMCID: PMC7119171.

602: Hamzeh O, Alkhateeb A, Zheng J, Kandalam S, Rueda L. Prediction of tumor location in prostate cancer tissue using a machine learning system on gene expression data. BMC Bioinformatics. 2020 Mar 11;21(Suppl 2):78. doi: 10.1186/s12859-020-3345-9. PMID: 32164523; PMCID: PMC7068980.

603: Melissa H, Christopher C, Kenneth G. The National COVID Cohort Collaborative (N3C): Rationale, Design, Infrastructure, and Deployment. J Am Med Inform Assoc. 2020 Aug 17:ocaa196. doi: 10.1093/jamia/ocaa196. Epub ahead of print. PMID: 32805036; PMCID: PMC7454687.

604: Lou S, Li T, Spakowicz D, Yan X, Chupp GL, Gerstein M. Latent-space embedding of expression data identifies gene signatures from sputum samples of asthmatic patients. BMC Bioinformatics. 2020 Oct 15;21(1):457. doi: 10.1186/s12859-020-03785-y. PMID: 33059594; PMCID: PMC7560063.

605: Giang TT, Nguyen TP, Tran DH. Stratifying patients using fast multiple kernel learning framework: case studies of Alzheimer's disease and cancers. BMC Med Inform Decis Mak. 2020 Jun 16;20(1):108. doi: 10.1186/s12911-020-01140-y. PMID: 32546157; PMCID: PMC7296686.

606: Lemsara A, Ouadfel S, Fröhlich H. PathME: pathway based multi-modal sparse autoencoders for clustering of patient-level multi-omics data. BMC Bioinformatics. 2020 Apr 16;21(1):146. doi: 10.1186/s12859-020-3465-2. PMID: 32299344; PMCID: PMC7161108.

607: Wang L, Wampfler J, Dispenzieri A, Xu H, Yang P, Liu H. Achievability to Extract Specific Date Information for Cancer Research. AMIA Annu Symp Proc. 2020 Mar 4;2019:893-902. PMID: 32308886; PMCID: PMC7153063.

608: Hallgrímsson B, Aponte JD, Katz DC, Bannister JJ, Riccardi SL, Mahasuwan N, McInnes BL, Ferrara TM, Lipman DM, Neves AB, Spitzmacher JAJ, Larson JR, Bellus GA, Pham AM, Aboujaoude E, Benke TA, Chatfield KC, Davis SM, Elias ER, Enzenauer RW, French BM, Pickler LL, Shieh JTC, Slavotinek A, Harrop AR, Innes AM, McCandless SE, McCourt EA, Meeks NJL, Tartaglia NR, Tsai AC, Wyse JPH, Bernstein JA, Sanchez-Lara PA, Forkert ND, Bernier FP, Spritz RA, Klein OD. Automated syndrome diagnosis by three-dimensional facial imaging. Genet Med. 2020 Oct;22(10):1682-1693. doi: 10.1038/s41436-020-0845-y. Epub 2020 Jun 1. PMID: 32475986; PMCID: PMC7521994.

609: Shi W, Kelsey T, Sullivan F. Efficient identification of patients eligible for clinical studies using case-based reasoning on Scottish Health Research register (SHARE). BMC Med Inform Decis Mak. 2020 Apr 19;20(1):70. doi: 10.1186/s12911-020-1091-6. PMID: 32306964; PMCID: PMC7169032.

610: Kinkead L, Allam A, Krauthammer M. AutoDiscern: rating the quality of online health information with hierarchical encoder attention-based neural networks. BMC Med Inform Decis Mak. 2020 Jun 9;20(1):104. doi: 10.1186/s12911-020-01131-z. PMID: 32517759; PMCID: PMC7285491.

611: Yang X, Gong Y, Waheed N, March K, Bian J, Hogan WR, Wu Y. Identifying Cancer Patients at Risk for Heart Failure Using Machine Learning Methods. AMIA Annu Symp Proc. 2020 Mar 4;2019:933-941. PMID: 32308890; PMCID: PMC7153053.

612: Gruenstaeudl M, Jenke N. PACVr: plastome assembly coverage visualization in R. BMC Bioinformatics. 2020 May 24;21(1):207. doi: 10.1186/s12859-020-3475-0. PMID: 32448146; PMCID: PMC7245912.

613: Blumenberg L, Ruggles KV. Hypercluster: a flexible tool for parallelized unsupervised clustering optimization. BMC Bioinformatics. 2020 Sep 29;21(1):428. doi: 10.1186/s12859-020-03774-1. PMID: 32993491; PMCID: PMC7525959.

614: Liu T, Wang Z. MASS: predict the global qualities of individual protein models using random forests and novel statistical potentials. BMC Bioinformatics. 2020 Jul 6;21(Suppl 4):246. doi: 10.1186/s12859-020-3383-3. PMID: 32631256; PMCID: PMC7336608.

615: Thomas LSV, Gehrig J. Multi-template matching: a versatile tool for object- localization in microscopy images. BMC Bioinformatics. 2020 Feb 5;21(1):44. doi: 10.1186/s12859-020-3363-7. PMID: 32024462; PMCID: PMC7003318.

616: Yazdani A, Mendez-Giraldez R, Yazdani A, Kosorok MR, Roussos P. Differential gene regulatory pattern in the human brain from schizophrenia using transcriptomic-causal network. BMC Bioinformatics. 2020 Oct 21;21(1):469. doi: 10.1186/s12859-020-03753-6. PMID: 33087039; PMCID: PMC7579819.

617: Ewing E, Planell-Picola N, Jagodic M, Gomez-Cabrero D. GeneSetCluster: a tool for summarizing and integrating gene-set analysis results. BMC Bioinformatics. 2020 Oct 7;21(1):443. doi: 10.1186/s12859-020-03784-z. PMID: 33028195; PMCID: PMC7542881.

618: Buetti-Dinh A, Herold M, Christel S, El Hajjami M, Delogu F, Ilie O, Bellenberg S, Wilmes P, Poetsch A, Sand W, Vera M, Pivkin IV, Friedman R, Dopson M. Reverse engineering directed gene regulatory networks from transcriptomics and proteomics data of biomining bacterial communities with approximate Bayesian computation and steady-state signalling simulations. BMC Bioinformatics. 2020 Jan 21;21(1):23. doi: 10.1186/s12859-019-3337-9. PMID: 31964336; PMCID: PMC6975020.

619: Chang WH, Mashouri P, Lozano AX, Johnstone B, Husić M, Olry A, Maiella S, Balci TB, Sawyer SL, Robinson PN, Rath A, Brudno M. Phenotate: crowdsourcing phenotype annotations as exercises in undergraduate classes. Genet Med. 2020 Aug;22(8):1391-1400. doi: 10.1038/s41436-020-0812-7. Epub 2020 May 5. Erratum in: Genet Med. 2020 Jun 18;: PMID: 32366968.

620: Lin Y, Li Y, Lu K, Ma C, Zhao P, Gao D, Fan Z, Cheng Z, Wang Z, Yu S. Long- distance disorder-disorder relation extraction with bootstrapped noisy data. J Biomed Inform. 2020 Sep;109:103529. doi: 10.1016/j.jbi.2020.103529. Epub 2020 Aug 7. PMID: 32771539.

621: Rosenbaum JE, Jennings J, Ellen JM, Borkovic LM, Scott JA, Wylie C, Rompalo A. Giving syphilis and gonorrhea to friends: using in-person friendship networks to find additional cases of gonorrhea and syphilis. BMC Public Health. 2020 Oct 21;20(1):1526. doi: 10.1186/s12889-020-09589-2. PMID: 33081743; PMCID: PMC7575333.

622: Rachid Zaim S, Kenost C, Berghout J, Chiu W, Wilson L, Zhang HH, Lussier YA. binomialRF: interpretable combinatoric efficiency of random forests to identify biomarker interactions. BMC Bioinformatics. 2020 Aug 28;21(1):374. doi: 10.1186/s12859-020-03718-9. Erratum in: BMC Bioinformatics. 2020 Nov 2;21(1):495. PMID: 32859146; PMCID: PMC7456085.

623: Qin L, Xu X, Ding L, Li Z, Li J. Identifying diagnosis evidence of cardiogenic stroke from Chinese echocardiograph reports. BMC Med Inform Decis Mak. 2020 Jul 9;20(Suppl 3):126. doi: 10.1186/s12911-020-1106-3. PMID: 32646410; PMCID: PMC7346320.

627: Liu J, Tan G, Lan W, Wang J. Identification of early mild cognitive impairment using multi-modal data and graph convolutional networks. BMC Bioinformatics. 2020 Nov 18;21(Suppl 6):123. doi: 10.1186/s12859-020-3437-6. PMID: 33203351; PMCID: PMC7672960.

628: Palombo V, Milanesi M, Sferra G, Capomaccio S, Sgorlon S, D'Andrea M. PANEV: an R package for a pathway-based network visualization. BMC Bioinformatics. 2020 Feb 6;21(1):46. doi: 10.1186/s12859-020-3371-7. PMID: 32028885; PMCID: PMC7006390.

629: Adnan N, Lei C, Ruan J. Robust edge-based biomarker discovery improves prediction of breast cancer metastasis. BMC Bioinformatics. 2020 Sep 30;21(Suppl 14):359. doi: 10.1186/s12859-020-03692-2. PMID: 32998692; PMCID: PMC7526355.

630: Pan Y, Zhou S, Guan J. Computationally identifying hot spots in protein-DNA binding interfaces using an ensemble approach. BMC Bioinformatics. 2020 Sep 17;21(Suppl 13):384. doi: 10.1186/s12859-020-03675-3. PMID: 32938375; PMCID: PMC7495898.

631: Adu-Gyamfi D, Zhang F, Kwansah Ansah AK. EDDAMAP: efficient data-dependent approach for monitoring asymptomatic patient. BMC Med Inform Decis Mak. 2020 Sep 29;20(1):245. doi: 10.1186/s12911-020-01258-z. PMID: 32993640; PMCID: PMC7523348.

632: Nath A, Leier A. Improved cytokine-receptor interaction prediction by exploiting the negative sample space. BMC Bioinformatics. 2020 Oct 31;21(1):493. doi: 10.1186/s12859-020-03835-5. PMID: 33129275; PMCID: PMC7603689.

633: Zhu X, Liu L, He J, Fang T, Xiong Y, Mitchell JC. iPNHOT: a knowledge-based approach for identifying protein-nucleic acid interaction hot spots. BMC Bioinformatics. 2020 Jul 6;21(1):289. doi: 10.1186/s12859-020-03636-w. PMID: 32631222; PMCID: PMC7336410.

634: do Nascimento PM, Medeiros IG, Falcão RM, Stransky B, de Souza JES. A decision tree to improve identification of pathogenic mutations in clinical practice. BMC Med Inform Decis Mak. 2020 Mar 10;20(1):52. doi: 10.1186/s12911-020-1060-0. PMID: 32151256; PMCID: PMC7063785.

635: Hung J, Goodman A, Ravel D, Lopes SCP, Rangel GW, Nery OA, Malleret B, Nosten F, Lacerda MVG, Ferreira MU, Rénia L, Duraisingh MT, Costa FTM, Marti M, Carpenter AE. Keras R-CNN: library for cell detection in biological images using deep neural networks. BMC Bioinformatics. 2020 Jul 11;21(1):300. doi: 10.1186/s12859-020-03635-x. PMID: 32652926; PMCID: PMC7353739.

636: Thutkawkorapin J, Eisfeldt J, Tham E, Nilsson D. pyCancerSig: subclassifying human cancer with comprehensive single nucleotide, structural and microsatellite mutational signature deconstruction from whole genome sequencing. BMC Bioinformatics. 2020 Apr 3;21(1):128. doi: 10.1186/s12859-020-3451-8. PMID: 32245405; PMCID: PMC7118897.

637: Lublóy Á. Medical crowdfunding in a healthcare system with universal coverage: an exploratory study. BMC Public Health. 2020 Nov 9;20(1):1672. doi: 10.1186/s12889-020-09693-3. PMID: 33167927; PMCID: PMC7653851.

638: He Y, Zhou X, Chen Z, Deng X, Gehring A, Ou H, Zhang L, Shi X. PRAP: Pan Resistome analysis pipeline. BMC Bioinformatics. 2020 Jan 15;21(1):20. doi: 10.1186/s12859-019-3335-y. PMID: 31941435; PMCID: PMC6964052.

639: Chen X, Xiong Y, Liu Y, Chen Y, Bi S, Zhu X. m5CPred-SVM: a novel method for predicting m5C sites of RNA. BMC Bioinformatics. 2020 Oct 30;21(1):489. doi: 10.1186/s12859-020-03828-4. PMID: 33126851; PMCID: PMC7602301.

640: Rybiński M, Möller S, Sunnåker M, Lormeau C, Stelling J. TopoFilter: a MATLAB package for mechanistic model identification in systems biology. BMC Bioinformatics. 2020 Jan 29;21(1):34. doi: 10.1186/s12859-020-3343-y. PMID: 31996136; PMCID: PMC6990465.

641: Cernea A, Fernández-Martínez JL, deAndrés-Galiana EJ, Fernández-Ovies FJ, Alvarez-Machancoses O, Fernández-Muñiz Z, Saligan LN, Sonis ST. Robust pathway sampling in phenotype prediction. Application to triple negative breast cancer. BMC Bioinformatics. 2020 Mar 11;21(Suppl 2):89. doi: 10.1186/s12859-020-3356-6. PMID: 32164540; PMCID: PMC7068866.

642: Huang LC, Yeung W, Wang Y, Cheng H, Venkat A, Li S, Ma P, Rasheed K, Kannan N. Quantitative Structure-Mutation-Activity Relationship Tests (QSMART) model for protein kinase inhibitor response prediction. BMC Bioinformatics. 2020 Nov 12;21(1):520. doi: 10.1186/s12859-020-03842-6. PMID: 33183223; PMCID: PMC7664030.

643: Wang P, Huang X, Qiu W, Xiao X. Identifying GPCR-drug interaction based on wordbook learning from sequences. BMC Bioinformatics. 2020 Apr 20;21(1):150. doi: 10.1186/s12859-020-3488-8. PMID: 32312232; PMCID: PMC7171867.

644: Karabayir I, Goldman SM, Pappu S, Akbilgic O. Gradient boosting for Parkinson's disease diagnosis from voice recordings. BMC Med Inform Decis Mak. 2020 Sep 15;20(1):228. doi: 10.1186/s12911-020-01250-7. PMID: 32933493; PMCID: PMC7493334.

645: Yang F, Fan K, Song D, Lin H. Graph-based prediction of Protein-protein interactions with attributed signed graph embedding. BMC Bioinformatics. 2020 Jul 21;21(1):323. doi: 10.1186/s12859-020-03646-8. PMID: 32693790; PMCID: PMC7372763.

646: Paul George AA, Lacerda M, Syllwasschy BF, Hopp MT, Wißbrock A, Imhof D. HeMoQuest: a webserver for qualitative prediction of transient heme binding to protein motifs. BMC Bioinformatics. 2020 Mar 27;21(1):124. doi: 10.1186/s12859-020-3420-2. PMID: 32216745; PMCID: PMC7099796.

647: Du L, Meng Q, Chen Y, Wu P. Subcellular location prediction of apoptosis proteins using two novel feature extraction methods based on evolutionary information and LDA. BMC Bioinformatics. 2020 May 24;21(1):212. doi: 10.1186/s12859-020-3539-1. PMID: 32448129; PMCID: PMC7245797.

648: Zheng X, Fu X, Wang K, Wang M. Deep neural networks for human microRNA precursor detection. BMC Bioinformatics. 2020 Jan 13;21(1):17. doi: 10.1186/s12859-020-3339-7. PMID: 31931701; PMCID: PMC6958766.

649: Coff L, Chan J, Ramsland PA, Guy AJ. Identifying glycan motifs using a novel subtree mining approach. BMC Bioinformatics. 2020 Feb 4;21(1):42. doi: 10.1186/s12859-020-3374-4. PMID: 32019496; PMCID: PMC7001330.

650: Zhang Y, Long Y, Kwoh CK. Deep learning based DNA:RNA triplex forming potential prediction. BMC Bioinformatics. 2020 Nov 12;21(1):522. doi: 10.1186/s12859-020-03864-0. PMID: 33183242; PMCID: PMC7663897.

651: Vargo AHS, Gilbert AC. A rank-based marker selection method for high throughput scRNA-seq data. BMC Bioinformatics. 2020 Oct 23;21(1):477. doi: 10.1186/s12859-020-03641-z. PMID: 33097004; PMCID: PMC7585212.

652: Durai P, Ko YJ, Pan CH, Park K. Evolutionary chemical binding similarity approach integrated with 3D-QSAR method for effective virtual screening. BMC Bioinformatics. 2020 Jul 14;21(1):309. doi: 10.1186/s12859-020-03643-x. PMID: 32664863; PMCID: PMC7362480.

653: Du W, Sun Y, Li G, Cao H, Pang R, Li Y. CapsNet-SSP: multilane capsule network for predicting human saliva-secretory proteins. BMC Bioinformatics. 2020 Jun 9;21(1):237. doi: 10.1186/s12859-020-03579-2. PMID: 32517646; PMCID: PMC7285745.

Appendix B

CLEF Datasets Characteristics

Table B.1: CLEF2019 DTA test dataset characteristics.

Review	No. of Studies	No. of Relevant Abstracts
CD008874	2,382	130 (5.5%)
CD009044	3,169	47 (1.5%)
CD011686	9,729	74 (0.8%)
CD012080	6,643	85 (1.3%)
CD012233	472	54 (11.4 %)
CD012567	6,735	12 (0.2%)
CD012669	1,260	82 (6.50%)
CD012768	131	100 (76.3%)
Total	**30,521**	**584 (1.91%)**

Table B.2: CLEF2019 Interventions training dataset characteristics.

Review	No. of Studies	No. of Relevant Abstracts
CD005139	5,392	112 (2.08%)
CD005253	2,014	4 (0.20%)
CD006715	149	13 (8.72%)
CD007868	300	5 (1.67%)
CD008018	739	17 (2.30%)
CD008170	12,320	88 (0.71%)
CD008201	3,574	11 (0.31%)
CD010019	728	1 (0.14%)
CD010355	43	9 (20.93%)
CD010526	652	21 (3.22%)
CD010778	339	26 (7.67%)
CD011380	66	8 (12.12%)
CD011436	290	25 (8.62%)
CD011571	146	15 (10.27%)
CD012120	169	7 (4.14%)
CD012164	55	6 (10.91%)
CD012223	2,456	12 (0.49%)
CD012347	1,098	16 (1.46%)
CD012521	375	2 (0.53%)
CD012930	740	50 (6.76%)
Total	**26,253**	**336 (1.28%)**

Table B.3: CLEF2019 Interventions test dataset characteristics.

Review	**No. of Studies**	**No. of Relevant Abstracts**
CD000996	281	10 (3.60%)
CD001261	571	85 (14.90%)
CD004414	336	32 (9.50%)
CD006468	3,874	91 (2.30%)
CD007867	943	31 (3.30%)
CD009069	1,757	94 (5.40%)
CD009642	1,922	90 (4.70%)
CD010038	8,867	36 (0.40%)
CD010239	224	23 (10.30%)
CD010558	2,815	75 (2.70%)
CD010753	2,539	35 (1.40%)
CD011140	289	4 (1.40%)
CD011571	146	21 (14.40%)
CD011768	9,160	81 (0.90%)
CD011977	195	65 (33.30%)
CD012069	3,479	425 (12.20%)
CD012164	61	10 (16.40%)
CD012342	2,353	9 (0.40%)
CD012455	1,593	12 (0.80%)
CD012551	591	86 (14.60%)
Total	**41,715**	**1,305 (3.13%)**

Table B.4: CLEF2019 Prognosis test dataset characteristics.

Review	**No. of Studies**	**No. of Relevant Abstracts**
CD012661	3,367	527 (15.65%)

Table B.5: CLEF2019 Qualitative test dataset characteristics.

Review	**No. of Studies**	**No. of Relevant Abstracts**
CD011558	2,168	51 (2.35%)
CD011787	4,369	125 (2.86%)
Total	**6,537**	**176 (2.69%)**

References

Ahmed Abdoaziz, Ahmed Abdulla, Hongfei Lin, Bo Xu, and Santosh Kumar Banbhrani. Improving biomedical information retrieval by linear combinations of different query expansion techniques. *BMC Bioinformatics*, 17(7):10–11, 2016.

About Cochrane Reviews. About Cochrane Reviews | Cochrane Library. [Online] Available at: https://www.cochranelibrary.com/about/about-cochrane-reviews, 2019. [Accessed: 2019-05-1].

Basant Agarwal and Namita Mittal. Text classification using machine learning methods-a survey. *Advances in Intelligent Systems and Computing*, 236(1):701–709, 2014.

Amal Alharbi and Mark Stevenson. Ranking abstracts to identify relevant evidence for systematic reviews: The University of Sheffield's approach to CLEF eHealth 2017 Task 2 . In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Amal Alharbi and Mark Stevenson. Improving Ranking for Systematic Reviews Using Query Adaptation. In Fabio Crestani, Martin Braschler, Jacques Savoy, Andreas Rauber, Henning Müller, David E Losada, Gundula Heinatz Bürki, Linda Cappellato, and Nicola Ferro, editors, *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 141–148, Cham, 2019a. Springer International Publishing.

Amal Alharbi and Mark Stevenson. Ranking studies for systematic reviews using query adaptation: University of sheffield's approach to clef ehealth 2019 task 2. In *Working*

Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, Lugano, Switzerland, September 2019b.

Amal Alharbi and Mark Stevenson. A dataset of systematic review updates. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, page 1257–1260, New York, NY, USA, 2019c. Association for Computing Machinery.

Amal Alharbi and Mark Stevenson. Refining Boolean queries to identify relevant studies for systematic review updates. *Journal of the American Medical Informatics Association*, 27(11):1658–1666, nov 2020. doi: 10.1093/jamia/ocaa148.

Amal Alharbi, William Briggs, and Mark Stevenson. Retrieving and ranking studies for systematic reviews: University of sheffield's approach to clef ehealth 2018 task 2. In *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Avignon, France, September 2018.

Kate Alldred, Yemisi Takwoingi, Boliang Guo, Mary Pennant, Jonathan Deeks, James Neilson, and Zarko Alfirevic. First trimester serum tests for down's syndrome screening. *Cochrane Database of Systematic Reviews*, (11), 2015.

Allard Altena and Silvia Olabarriaga. Predicting publication inclusion for diagnostic accuracy test reviews using random forests and topic modelling. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Lagopoulos Athanasios Tsoumakas Grigorios Anagnostou, Antonios and Ioannis Vlahavas. Combining inter-review learning-to-rank and intra-review incremental training for title and abstract screening in systematic reviews. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Elliott Antman, Joseph Lau, Bruce Kupelnick, Frederick Mostelle, and Thomas Chalmers. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA*, 268(2):240–8, 1992.

Yindalon Aphinyanaphongs, Ioannis Tsamardinos, Alexander Statnikov, Douglas Hardin, and Constantin F Aliferis. Text categorization models for high-quality article retrieval in internal medicine. *Journal of the American Medical Informatics Association: JAMIA*, 12 (2):207–216, 2005.

Professor Tony Avery, Mrs Gill Gookey, Dr Rachel Spencer, Dr Richard Knox, Ms Kate Marsden, and Dr Ndeshi Salema. Selecting the right drug. *InnovAiT*, 6(8):478–487, 2013.

Hiteshwar Kumar Azad and Akshay Deepak. Query expansion techniques for information retrieval: A survey. *Information Processing & Management*, 56(5):1698–1735, 2019.

Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Publishing Company, Boston, MA, USA, 2nd edition, 2011.

Jing Bai and Jian Yun Nie. Adapting information retrieval to query contexts. *Information Processing and Management*, 44(6):1901–1922, 2008.

Shariq Bashir, Akmal Saeed Khattak, and Mohammed Ali Alshara. Automatically transforming full length biomedical articles into search queries for retrieving related articles. *Egyptian Informatics Journal*, 2020.

Hilda Bastian, Paul Glasziou, and Iain Chalmers. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS medicine*, 7(9):e1000326, 2010.

Michel Beauverd, John Wokke, and Gian Borasio. Recombinant human insulin-like growth factor I (rhIGF-I) for the treatment of amyotrophic lateral sclerosis/motor neuron disease. *Cochrane Database of Systematic Reviews*, (11), 2012.

Yves Bestgen. Getting rid of the chi-square and log-likelihood tests for analysing vocabulary differences between corpora. Quaderns de Filologia: Estudis Lingüístics, 2018.

Jiantao Bian, Mohammad Amin Morid, Siddhartha Jonnalagadda, Gang Luo, and Guilherme Del Fiol. Automatic identification of high impact articles in PubMed to support clinical decision making. *Journal of Biomedical Informatics*, 73:95–103, 2017.

Angela Boland, Gemma Cherry, and Rumona Dickson. *Doing a systematic liteature review: a student's guide*. SAGE, 2014.

Andrew Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recogn.*, 30(7):1145–1159, 1997.

Eric Brill. A simple rule-based part of speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing*, ANLC '92, page 152–155, USA, 1992. Association for Computational Linguistics.

Eric Brill. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging. *Comput. Linguist.*, 21(4):543–565, dec 1995.

Duy Duc An Bui, Siddhartha Jonnalagadda, and Guilherme Del Fiol. Automatically finding relevant citations for clinical guideline development. *Journal of Biomedical Informatics*, 57:436–445, 2015.

Mudge David Craig Jonathan Johnson David Tong Allison Campbell, Denise and Giovanni Strippoli. Antimicrobial agents for preventing peritonitis in peritoneal dialysis patients. *Cochrane Database of Systematic Reviews*, (4), 2017.

Lefebvre Carol, Julie Glanville, Simon Briscoe, Christopher Marshall, Maria-Inti Metzendorf, Anna Noel-Storr, Rader Tamara, Farhad Shokraneh, James Thomas, and Lisa Wieland. Chapter 4: Searching for and selecting studies. In *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, 6.1 edition, 2020.

Claudio Carpineto and Giovanni Romano. A Survey of Automatic Query Expansion in Information Retrieval. *ACM Computing Surveys*, 44(1):1–50, 2012.

Peter Castaldi, Michael Cho, Matthew Cohn, Fawn Langerman, et al. The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. *Human molecular genetics*, 19(3):526–534, 2009.

Iain Chalmers. Helping Physicians To Keep Abreast of the Medical Literature: Medical and Philosophical Commentaries, 1773–1795. *Annals of Internal Medicine*, 133(3):238–243, 2000.

Jacqueline Chandler and Miranda Cumpston. Chapter IV: Updating a review. In *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, 6 edition, 2019.

Jacqueline Chandler, Miranda Cumpston, James Thomas, Julian Higgins, et al. Chapter I: Introduction | Cochrane Training. In *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, 6 edition, 2019.

Jiayi Chen, Su Chen, Yang Song, Hongyu Liu, et al. Ecnu at 2017 ehealth task 2: Technologically assisted reviews in empirical medicine. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Kiyomi Chujo and Masao Utiyama. Selecting level-specific specialized vocabulary using statistical measures. *System*, 34(2):255–269, 2006.

Archie Cochrane. *1931–1971: a critical review with particular reference to the medical profession*. Medicines for the year 2000, London: Office of Health Economics, 1979.

Aaron Cohen. An effective general purpose approach for automated biomedical document classification. *Annual Symposium proceedings.*, (1):161–165, 2006.

Aaron Cohen. Optimizing feature representation for automated systematic review work prioritization. *Annual Symposium Proceedings*, (1):121–125, 2008.

Aaron Cohen and Neil Smalheiser. OHSU CLEF 2018 task 2 diagnostic test accuracy ranking using publication type cluster similarity measures. In *Working Notes of CLEF*

2018 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, Avignon, France, September 2018.

Aaron Cohen, William Hersh, Kim Peterson, and Po-Yin Yen. Reducing workload in systematic review preparation using automated citation classification. *Journal of the American Medical Informatics Association: JAMIA*, 13(2):206–19, 2006.

Aaron Cohen, Kyle Ambert, and Marian Mcdonagh. Cross-topic learning for work prioritization in systematic review creation and update introduction and background. *Journal of the American Medical Informatics Association: JAMIA*, 16:690–704, 2009.

Aaron Cohen, Kyle Ambert, and Marian McDonagh. A prospective evaluation of an automated classification system to support evidence-based medicine and systematic review. *AMIA Annual Symposium proceedings*, 2010:121–125, 2010.

Aaron Cohen, Kyle Ambert, and Marian McDonagh. Studying the potential impact of automated document classification on scheduling a systematic review update. *BMC Medical Informatics and Decision Making*, 12(1):33, 2012.

Aaron Cohen, Neil Smalheiser, Marian McDonagh, Clement Yu, et al. Automated confidence ranked classification of randomized controlled trial articles: An aid to evidence-based medicine. *Journal of the American Medical Informatics Association: JAMIA*, 22(3): 707–717, 2015.

Gordon Cormack and Maura Grossman. Navigating imprecision in relevance assessments on the road to total recall. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '17*, New York, USA, 2017a. ACM Press.

Gordon Cormack and Maura Grossman. Technology-assisted review in empirical medicine: Waterloo participation in CLEF eHealth 2018. In *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Avignon, France, September 2018.

Gordon V. Cormack and Maura R. Grossman. Technology-Assisted Review in Empirical Medicine: Waterloo Participation in CLEF eHealth 2017. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017b.

Simon Corston-Oliver and Michael Gamon. Combining Decision Trees And Transformation-Based Learning To Correct Transferred Linguistic Representations. Association for Machine Translation in the Americas, 2003.

Jonathan Culpeper. Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, 14(1):29–59, 2009.

Siddhartha Dalal, Paul Shekelle, Susanne Hempel, Sydne Newberry, et al. *A Pilot Study Using Machine Learning and Domain Knowledge To Facilitate Comparative Effectiveness Review Updating*. Agency for Healthcare Research and Quality (US), 2012.

Jonathan Deeks, Julian Higgins, and Douglas Altman. Chapter 10: Analysing data and undertaking meta-analyses. In *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, 6 edition, 2019.

Ish Kumar Dhammi and Sudhir Kumar. Medical subject headings (mesh) terms. *Indian journal of orthopaedics*, 48(5):443–4, 2014.

Manuel Díaz-Galiano, Maria Martín-Valdivia, and Alfonso Ureña-López. Query expansion with a medical ontology to improve a multimodal information retrieval system. *Computers in Biology and Medicine*, 39(4):396–403, 2009.

Soo Downe, Kenneth Finlayson, Özge Tunçalp, and Ahmet Gülmezoglu. Provision and uptake of routine antenatal services: a qualitative evidence synthesis. *Cochrane Database of Systematic Reviews*, (6), 2019.

Ted Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1):61–74, 1993.

Mark Elkins. Updating systematic reviews. *Journal of Physiotherapy*, 64(1):1–3, 2018.

Julian Elliott, Anneliese Synnot, Tari Turner, et al. Living systematic review: 1. Introduction - the why, what, when, and how. *Journal of Clinical Epidemiology*, 91:23–30, 2017.

Oana Frunza, Diana Inkpen, and Stan Matwin. Building systematic reviews using automatic text classification techniques. In *Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference*, 2010.

Paul Garner, Sally Hopewell, Jackie Chandler, Harriet Maclehose, et al. When and how to update systematic reviews: consensus and checklist. *the BMJ Holger J Schünemann*, 354354(45), 2016.

Rayid Ghani, Rosie Jones, and Dunja Mladenic. Building Minority Language Corpora by Learning to Generate Web Search Queries. *Knowledge and Information Systems*, 7(1): 56–83, 2005.

David Gough, Sandy Oliver, and James Thomas. *An introduction to systematic reviews*. SAGE, London, first edition, 2012a.

David Gough, James Thomas, and Sandy Oliver. Clarifying differences between review designs and methods. *Systematic Reviews*, 1:28, 2012b.

Erik Graf, Leif Azzopardi, and Keith van Rijsbergen. Automatically generating queries for prior art search. In *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

Sally Green, Julian Higgins, Philip Alderson, Mike Clarke, et al. *Cochrane handbook for systematic reviews of interventions*, volume 4. John Wiley & Sons, 2011.

Helen Handoll and Paddy Pearce. Interventions for treating isolated diaphyseal fractures of the ulna in adults. *Cochrane Database of Systematic Reviews*, (6), 2012.

Harriet MacLehose. Withdrawing published Cochrane Reviews - Cochrane Editorial and Publishing Policy Resource - Confluence, 2018. [Online] Available at:

https://documentation.cochrane.org/display/EPPR/Withdrawing+published+Cochrane+Reviews [Accessed 7-June-2018].

Greg Harris, Anand Panangadan, and Viktor Prasanna. Interactive query refinement for boolean search. 2014.

Kazuma Hashimoto, Georgios Kontonatsios, Makoto Miwa, and Sophia Ananiadou. Topic detection using paragraph vectors to support active learning in systematic reviews. *Journal of Biomedical Informatics*, 62:59–65, 2016.

Brian Haynes and Nancy Wilczynski. Optimal search strategies for retrieving scientifically strong studies of diagnosis from medline: analytical survey. *BMJ*, 328(7447):1040, 2004.

Susanne Heiwe, , and Stefan Jacobson. Exercise training for adults with chronic kidney disease. *Cochrane Database of Systematic Reviews*, (10), 2011.

Harri Hemila and Elizabeth Chalker. Vitamin C for preventing and treating the common cold. *Cochrane Database of Systematic Reviews*, (1), 2013.

Knut Hofland and Johansson Stig. Word frequencies in British and American English. *Bergen, Norway: The Norwegian Computing Centre for the Humanities*, 1982.

Noah Hollmann and Carsten Eickhoff. Relevance-based stopping for recall-centric medical document retrieval. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017a.

Noah Hollmann and Carsten Eickhoff. Ranking and feedback-based stopping for recall-centric document retrieval. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017b.

Sally Hopewell, Mike Clarke, Carol Lefebvre, and Roberta Scherer. Handsearching versus electronic searching to identify reports of randomized trials. *Cochrane Database of Systematic Reviews*, (2), 2007.

Edward Hughes, Julie Brown, John Collins, Cindy Farquhar, et al. Ovulation suppression for endometriosis for women with subfertility. *Cochrane Database of Systematic Reviews*, (3), 2007.

Osman Ibrahim and Dario Landa-Silva. ES-Rank: Evolution Strategy learning to rank approach. In *Proceedings of the Symposium on Applied Computing*, SAC '17, New York, USA, 2017.

S. Jabri, A. Dahbi, T. Gadi, and A. Bassir. Ranking of text documents using tf-idf weighting and association rules mining. In *2018 4th International Conference on Optimization and Applications (ICOA)*, pages 1–6, 2018.

Nusrat Jahan, Sadiq Naveed, Muhammad Zeshan, and Muhammad A Tahir. How to Conduct a Systematic Review: A Narrative Literature Review. *Cureus*, 8(11), 2016.

Johns Hopkins University. Coronavirus covid-19 global cases by johns hopkins CSSE, 2020. [Online] Available at: https://www.arcgis.com/apps/opsdashboard [Accessed 20-May-2020].

Siddhartha Jonnalagadda and Diana Petitti. A new iterative method to reduce workload in the systematic review process. *International journal of computational biology and drug design*, 6(0):5–17, 2013.

Georgiadis Georgios Kalphov, Vassil and Leif Azzopardi. SiS at CLEF 2017 eHealth TAR Task. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Evangelos Kanoulas, Dan Li, Leif Azzopardi, and Rene Spijker. CLEF 2017 technologically assisted reviews in empirical medicine overview. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation forum, Dublin, Ireland*, September 2017.

Evangelos Kanoulas, Rene Spijker, Dan Li, and Leif Azzopardi. CLEF 2018 Technology Assisted Reviews in Empirical Medicine Overview. In *CLEF 2018 Evaluation Labs and Workshop: Online Working Notes*, Avignon, France, September 2018.

Evangelos Kanoulas, Dan Li, Leif Azzopardi, and Rene Spijker. CLEF 2019 Technology Assisted Reviews in Empirical Medicine Overview. In *CLEF 2019 Evaluation Labs and Workshop: Online Working Notes*, Lugano, Switzerland, September 2019.

Sarvnaz Karimi, Stefan Pohl, Falk Scholer, Lawrence Cavedon, et al. Boolean versus ranked querying for biomedical systematic reviews. *BMC medical informatics and decision making*, 10(1):1–20, 2010.

Madian Khabsa, Ahmed Elmagarmid, Ihab Ilyas, Hossam Hammady, et al. Learning to identify relevant studies for systematic reviews using random forest and external information. *Machine Learning*, 102(3):465–482, 2016.

Khalid Khan, Regina Kunz, Jos Kleijnen, and Gerd Antes. Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 96(3):118–121, 2003.

Halil Kilicoglu, Dina Demner-Fushman, Thomas Rindflesch, Nancy Wilczynski, and Brian Haynes. Towards automatic recognition of scientifically rigorous clinical research evidence. *Journal of the American Medical Informatics Association: JAMIA*, 16:25–31, 2009.

Seunghee Kim and Jinwook Choi. Improving the performance of text categorization models used for the selection of high quality articles. *Healthcare Informatics Research*, 18(1):1–18, 2012.

Seunghee Kim and Jinwook Choi. An SVM-based high-quality article classifier for systematic reviews. *Journal of Biomedical Informatics*, 47:153–159, 2014.

Youngho Kim, Jangwon Seo, and W. Bruce Croft. Automatic Boolean query suggestion for professional search. In *SIGIR'11 - Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 825–834. Association for Computing Machinery, 2011.

Georgios Kontonatsios, Austin J Brockmeier, Piotr Przybyła, John McNaught, et al. A semi-supervised approach using label propagation to support citation screening. *Journal of biomedical informatics*, 72:67–76, 2017.

Zoe Kopsaftis, Richard Wood-Baker, and Phillippa Poole. Influenza vaccine for chronic obstructive pulmonary disease (COPD). *Cochrane Database of Systematic Reviews*, (6), 2018.

Grace Lee. A Study of Convolutional Neural Networks for Clinical Document Classification in Systematic Reviews: SysReview at CLEF eHealth 2017. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Grace Lee and Aixin Sun. Seed-driven document ranking for systematic reviews in evidence-based medicine. In *41st International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2018*. Association for Computing Machinery, Inc, 2018.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, feb 2020.

Ivan Lerner, Perrine Créquit, Philippe Ravaud, and Ignacio Atal. Automatic screening using word embeddings achieved high sensitivity and workload reduction for updating living network meta-analyses. *Journal of Clinical Epidemiology*, 108:86–94, 2019.

Stephanie Lewis and Martin Clarke. Forest plots: trying to see the wood and the trees. *BMJ*, 322 7300:1479–80, 2001.

Dan Li and Evangelos Kanoulas. Automatic Thresholding by Sampling Documents and Estimating Recall. In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Lugano, Switzerland, September 2019.

Ersilia Lucenteforte, Alessandra Bettiol, Salvatore De Masi, and Gianni Virgili. *Updating Diagnostic Test Accuracy Systematic Reviews: Which, When, and How Should They Be Updated?* Springer International Publishing, Cham, 2018.

Heneghan Carl Glasziou Paul Mahtani, Kamal and Rafael Perera. Reminder packaging for improving adherence to self-administered long-term medications. *Cochrane Database of Systematic Reviews*, (9), 2011.

Christopher Manning and Hinrich Schütze. *Foundations of statistical natural language processing*. The MIT Press, Cambridge, MA, 1999.

Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze. *Relevance feedback and query expansion*. Cambridge University Press, 2008a.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Relevance feedback and query expansion*, page 162–177. Cambridge University Press, 2008b.

David Martinez, Sarvnaz Karimi, Lawrence Cavedon, and Timothy Baldwin. Facilitating biomedical systematic reviews using ranked text retrieval and classification. In *13th Australasian Document Computing Symposium (ADCS)*, Hobart Tasmania, 2008.

Izet Masic, Milan Miokovic, and Belma Muhamedagic. Evidence based medicine - new approaches and challenges. *Acta informatica medica : AIM : journal of the Society for Medical Informatics of Bosnia & Herzegovina : casopis Drustva za medicinsku informatiku BiH*, 16(4):219–225, 2008.

Stan Matwin, Alexandre Kouznetsov, Diana Inkpen, Oana Frunza, et al. A new algorithm for reducing the workload of experts in performing systematic reviews. *Journal of the American Medical Informatics Association: JAMIA*, 17(4):446–453, 2010.

Jessie McGowan and Margaret Sampson. Systematic reviews need systematic searchers. *Journal of the Medical Library Association*, 93(1):74, 2005.

Medical Subject Headings (MeSH®) in MEDLINE®/PubMed®: A Tutorial. [Online] Available at: https://www.nlm.nih.gov/bsd/disted/meshtutorial/introduction/, 2012. [Accessed 28-November-2017].

Manuele Michelessi, Ersilia Lucenteforte, Francesco Oddone, Miriam Brazzelli, et al. Optic nerve head and fibre layer imaging for diagnosing glaucoma. *Cochrane Database of Systematic Reviews*, (11), 2015.

Adamantios Minas, Athanasios Lagopoulos, and Grigorios Tsoumakas. Aristotle university's approach to the technologically assisted reviews in empirical medicine task of the 2018 clef eHealth lab. In *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Avignon, France, September 2018.

Makoto Miwa, James Thomas, Alison O'Mara-Eves, and Sophia Ananiadou. Reducing systematic review workload through certainty-based screening. *Journal of Biomedical Informatics*, 51:242–253, 2014.

David Moher and Alexander Tsertsvadze. Systematic reviews: When is an update an update? *Lancet*, 367:881–883, 2006.

Shoichi Nakagawa, Isao Makino, Takashi Ishizaki, and Ichiro Dohi. Dissolution of cholesterol gallstones by ursodeoxycholic acid. *The Lancet*, 310(8034):367–369, 1977.

Gonzalo Navarro. A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31–88, 2001.

Karen New, Vicki Flenady, and Mark Davies. Transfer of preterm infants from incubator to open cot at lower versus higher body weight. *Cochrane Database of Systematic Reviews*, (9), 2011.

David Newman, Sarvnaz Karimi, and Lawrence Cavedon. Using topic models to interpret MEDLINE's medical subject headings. Springer, Berlin, Heidelberg, 2009.

Grace Ngai and Radu Florian. Transformation based learning in the fast lane. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*, 2001.

Vicki Nisenblat, Patrick MM Bossuyt, Cindy Farquhar, Neil Johnson, et al. Imaging modalities for the non-invasive diagnosis of endometriosis. *Cochrane Database of Systematic Reviews 2016*, 2, 2016.

Marcello Nisio, Alessandro Squizzato, Anne Rutjes, Harry Büller, et al. Diagnostic accuracy of D-dimer test for exclusion of venous thromboembolism: a systematic review. *Journal of Thrombosis and Haemostasis*, 5(2):296–304, 2007.

Christopher Norman, Leeflang Mariska, and Aurélie Névéol. Limsi@CLEF eHealth 2017 task 2: Logistic regression for automatic article ranking. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Christopher Norman, Leeflang Mariska, and Aurélie Névéol. Limsi@CLEF eHealth 2018 task 2: Technology assisted reviews by stacking active and static learning. In *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Avignon, France, September 2018.

Jane Noyes, Andrew Booth, Margaret Cargo, Kate Flemming, Angela Harden, Janet Harris, Ruth Garside, Karin Hannes, Tomás Pantoja, and James Thomas. Chapter 21: Qualitative evidence. In *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, 6 edition, 2019.

Giorgio Nunzio, Federica Beghini, Federica Vezzani, and Genevi'eve Henrot. An Interactive Two-Dimensional Approach to Query Aspects Rewriting in Systematic Reviews. IMS Unipd At CLEF eHealth Task 2. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Giorgio Nunzio, Giacomo Ciuffreda, and Federica Vezzani. Interactive sampling for systematic reviews. IMS unipd at CLEF 2018 eHealth task 2. In *Working Notes of CLEF 2018*

- *Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Avignon, France, September 2018.

Giorgio Maria Di Nunzio. A Distributed Effort Approach for Systematic Reviews. IMS Unipd At CLEF 2019 eHealth Task 2. In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Lugano, Switzerland, September 2019.

Hanna Olofsson, Agneta Brolund, Christel Hellberg, Rebecca Silverstein, Karin Stenström, Marie Österberg, and Jessica Dagerhamn. Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. *Research Synthesis Methods*, 8 (3):275–280, 2017. doi: 10.1002/jrsm.1237.

Alison O'Mara-Eves, James Thomas, John McNaught, Makoto Miwa, and Sophia Ananiadou. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews*, 4(1):5, 2015.

Andrew Oxman. The science of reviewing research. *Annals of the New York Academy of Sciences*, 703:125–134, 1993.

Matthew Page, Larissa Shamseer, Douglas Altman, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. *PLOS Medicine*, 13(5):1–30, 2016.

Elisabeth Pain. How to keep up with the scientific literature. *Science*, 354(6316), 2016.

PMC Overview. [Online] Available at: https://www.ncbi.nlm.nih.gov/pmc/about/intro/, 2020. [Accessed 28-May-2020].

Punjaporn Pojanapunya and Richard Watson Todd. Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Ling. Theory*, 14(1):133–167, 2018.

PubMed Tutorial. PubMed Tutorial - Glossary. [Online] Available at: https://www.nlm.nih.gov/bsd/disted/pubmedtutorial/glossary.html, 2017. [Accessed 03-November-2017].

Pia Bastholm Rahmner, Birgit Eiermann, Seher Korkmaz, Lars L Gustafsson, Magnus Gruvén, Simon Maxwell, Hans-Georg Eichle, and Anikó Vég. Physicians' reported needs of drug information at point of care in Sweden. *British journal of clinical pharmacology*, 73(1):115–125, 2012.

John Rathbone, Tammy Hoffmann, and Paul Glasziou. Faster title and abstract screening? evaluating abstrackr, a semi-automated online screening program for systematic reviewers. *Systematic reviews*, 4(1):80, 2015.

Paul Rayson. From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4):519–549, 2008.

Paul Rayson. *Corpus Analysis of Key Words*. American Cancer Society, 2012.

Paul Rayson. Corpus analysis of key words. In *Chapelle, Carol (Ed) The Concise Encyclopedia of Applied Linguistics*, pages 320–326. Wiley, 2019.

Ian Ruthven and Mounia Lalmas. A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review*, 18(2):95–145, 2003.

Margaret Sampson, Jennifer Tetzlaff, and Christine Urquhart. Precision of healthcare systematic review searches in a cross-sectional sample. *Research Synthesis Methods*, 2 (2):119–125, 2011.

Harrisen Scells and Guido Zuccon. Generating better queries for systematic reviews. In *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '18, New York, USA, 2018. ACM.

Harrisen Scells, Guido Zuccon, Anthony Deacon, and Bevan Koopman. QUT ielab at CLEF 2017 technology assisted reviews track: Initial experiments with learning to rank. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017a.

Harrisen Scells, Guido Zuccon, Bevan Koopman, Leif Azzopardi, et al. A test collection for evaluating retrieval of studies for inclusion in systematic reviews. In *40th International*

ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, August 2017b.

Harrisen Scells, Guido Zuccon, and Bevan Koopman. Automatic boolean query refinement for systematic review literature search. In *The World Wide Web Conference*, WWW '19, New York, USA, May 2019. ACM.

Harrisen Scells, Guido Zuccon, and Bevan Koopman. You Can Teach an Old Dog New Tricks: Rank Fusion applied to Coordination Level Matching for Ranking in Systematic Reviews. In Joemon M Jose, Emine Yilmaz, João Magalhães, Pablo Castells, Nicola Ferro, Mário J Silva, and Flávio Martins, editors, *Advances in Information Retrieval*, pages 399–414, Cham, 2020. Springer International Publishing.

Mike Scott and Christopher Tribble. *Textual Patterns: Key Words and Corpus Analysis In Language Education*. John Benjamins Publishing, 2006.

Burr Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52(55-66): 11, 2010.

Burr Settles. *Active Learning*. Morgan & Claypool Publishers, 2012.

Walid Shalaby and Wlodek Zadrozny. Patent retrieval: a literature review. *Knowledge and Information Systems*, 61(2):631–660, 2019.

Dou Shen, Rong Pan, Jian Tao Sun, Jeffrey Junfeng Pan, et al. Query enrichment for web-query classification. *ACM Transactions on Information Systems*, 24(3):320–352, 2006.

Gangadhar Shobha and Shanta Rangaswamy. *Computational analysis and understanding of natural languages: Principles, methods and applications Chapter 8 - Machine learning*. Elsevier B.V., 2018.

Kaveh Shojania, Margaret Sampson, Mohammed Ansari, Jun Ji, et al. How quickly do systematic reviews go out of date? a survival analysis. *Annals of Internal Medicine*, 147: 224–233, 2007.

Gaurav Singh, Iain Marshall, James Thomas, and Byron Wallace. Identifying diagnostic test accuracy publications using a deep model. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Jaspreet Singh and Lini Thomas. IIIT-H at CLEF eHealth 2017 task 2: technologically assisted reviews in empirical medicine. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

David C Slawson and Sean W Reed. Finding high-quality review articles. *American family physician*, 79(10):875–7, 2009.

Justin JongSu Song, Wookey Lee, and Jafar Afshar. An effective High Recall Retrieval method. *Data & Knowledge Engineering*, 123:101603, 2019.

Hanna Suominen, Liadh Kelly, Lorraine Goeuriot, Evangelos Kanoulas, et al. Overview of the CLEF eHealth Evaluation Lab 2018. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, Cham, 2018. Springer International Publishing.

Hanna Suominen, Liadh Kelly, Lorraine Goeuriot, and Martin Krallinger. CLEF eHealth Evaluation Lab 2020. In Joemon M Jose, Emine Yilmaz, João Magalhães, Pablo Castells, Nicola Ferro, Mário J Silva, and Flávio Martins, editors, *Advances in Information Retrieval*, pages 587–594, Cham, 2020. Springer International Publishing.

Didi Surian, Adam Dunn, Liat Orenstein, Rabia Bashir, et al. A shared latent space matrix factorisation method for recommending new trial evidence for systematic review updates. *Journal of Biomedical Informatics*, 79:32–40, 2018.

Susan Sutherland. An introduction to systematic reviews. *Journal of Evidence-Based Dental Practice*, 4(1):47–51, 2004.

Teruhiko Terasawa, Tomas Dvorak, Stanley Ip, Gowri Raman, et al. Systematic review: charged-particle radiation therapy for cancer. *Annals of internal medicine*, 151(8): 556–65, 2009.

Grant Theron, Jonny Peter, Marty Richardson, Rob Warren, et al. GenoType® MTBDRsl assay for resistance to second-line anti-tuberculosis drugs. *Cochrane Database of Systematic Reviews*, (9), 2016.

James Thom and Falk Scholer. A comparison of evaluation measures given how users perform on search tasks. In *Proceedings of the 12th Australasian Document Computing Symposium*, Melbourne, Australia, December 2007.

Yves Tillé. Sampling Algorithms. In *Sampling Algorithms*. Springer New York, 2006.

Federico Tomassetti, Giuseppe Rizzo, Antonio Vetro, Luca Ardito, et al. Linked data approach for selection process automation in systematic reviews. In *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, Durham, UK, April 2011.

Andrew Trotman and Kat Lilly. Jassjr: The minimalistic bm25 search engine for teaching and learning information retrieval. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 2185–2188, New York, NY, USA, 2020. Association for Computing Machinery.

Guy Tsafnat, Paul Glasziou, Miew Keen Choong, Adam Dunn, et al. Systematic review automation technologies. *Systematic Reviews*, 3(1):74, 2014.

Roelien Van de Vrie, Marianne Rutten, Joyce Asseler, Mariska Leeflang, et al. Laparoscopy for diagnosing resectability of disease in women with advanced ovarian cancer. *Cochrane Database of Systematic Reviews*, (3), 2019.

Siw Waffenschmidt, Marco Knelangen, Wiebke Sieben, Stefanie Bühn, and Dawid Pieper. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. *BMC Medical Research Methodology*, 19 (1):132, 2019.

Byron Wallace, Thomas Trikalinos, Joseph Lau, Carla Brodley, et al. Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics*, 11(1), 2010.

Byron Wallace, Kevin Small, Carla Brodley, Joseph Lau, et al. Deploying an interactive machine learning system in an evidence-based practice center. In *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*, New York, USA, 2012a. ACM Press.

Byron Wallace, Kevin Small, Carla Brodley, Joseph Lau, et al. Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining. *Genetics in Medicine*, 14:663, 2012b.

Maggie Westby, Jo Dumville, Nikki Stubbs, Gill Norman, et al. Protease activity as a prognostic factor for wound healing in venous leg ulcers. *Cochrane Database of Systematic Reviews*, (9), 2018.

Valeire White, Julie Glanville, Carol Lefebvre, and Trevor Sheldon. A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy. *Journal of Information Science*, 27(6):357–370, 2001.

Christopher Williams, Nicholas Henschke, Christopher Maher, Maurits van Tulder, et al. Red flags to screen for vertebral fracture in patients presenting with low-back pain. *Cochrane Database of Systematic Reviews 2013*, 1, 2013.

Huaying Wu, Tingting Wang, Jiayi Chen, Su Chen, et al. Ecnu at 2018 eHealth task 2: Technologically assisted reviews in empirical medicine. In *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Avignon, France, September 2018.

Wei Yu, Melinda Clyne, Siobhan Dolan, Ajay Yesupriya, et al. GAPscreener: an automatic tool for screening human genetic association literature in PubMed using the support vector machine technique. *BMC bioinformatics*, 9:205, 2008.

Zhe Yu and Tim Menzies. Data balancing for technologically assisted reviews: Undersampling or reweighting. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Dublin, Ireland, September 2017.

Zhaohao Zeng and Tetsuya Sakai. BM25 Pseudo Relevance Feedback Using Anserini at Waseda University. In *Proceedings of the Open-Source IR Replicability Challenge co-located with 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, OSIRRC@SIGIR 2019, Paris, France, July 25, 2019*, volume 2409, pages 62–63. CEUR-WS.org, 2019.

Haotian Zhang, Jimmy Lin, Gordon Cormack, and Mark Smucker. Sampling strategies and active learning for volume estimation. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval - SIGIR '16*, New York, USA, July 2016. ACM Press.

Guido Zuccon, Harrisen Scells, Mohamed A. Sharaf, and Bevan Koopman. Sampling query variations for learning to rank to improve automatic boolean query generation in systematic reviews. In *Proceedings of The Web Conference 2020*, WWW '20, page 3041–3048, New York, NY, USA, 2020. Association for Computing Machinery.