Search Strategy Formulation: A Framework For Learning

Dr Andrew MacFarlane
Department of Computer Science
City University London
London EC1V 0HB
+44 (0)20 7040 8386
andym@city.ac.uk

Dr Tony Russell-Rose
UXLabs
3000 Cathedral Hill
Guildford, GU2 7YB
+44 7779 936191
tgr@uxlabs.co.uk

ABSTRACT

Healthcare information professionals perform systematic literature reviews to gather the evidence needed to answer specific research questions and formulate policy. However, performing a systematic review is a resource-intensive and time consuming undertaking, often taking years to complete. Moreover, the output relies heavily on the quality of the initial search strategy in ensuring that the scope is sufficiently exhaustive and not biased by easily accessible studies. In this paper we introduce a structured methodology and a framework for learning which together aim to embody best practices from the community and provide support for many of the common issues in search strategy development.

CCS Concepts

• **Information systems→Information Retrieval→Information Retrieval Query Processing.**

Keywords

Information retrieval; systematic reviews; education; training.

1. INTRODUCTION

Medical knowledge is growing so rapidly that it is difficult for healthcare professionals to keep up. As the volume of published studies increases year by year, the gap between research knowledge and professional practice grows ever wider. *Systematic literature reviews* can play a key role in closing this gap, by synthesizing the complex, incomplete and at times conflicting findings of biomedical research into a form that can readily inform health decision making [1]. A key principle of systematic reviews is that the protocol by which the literature was collected and analyzed should be made transparent and repeatable.

However, undertaking a systematic review is a resource-intensive and time consuming process, sometimes taking years to complete [2]. Even *rapid evidence assessments*, designed to provide quick summaries of what is known about a topic or intervention, can take as long as two to six months [3]. Moreover, new research findings may be published in the interim [4], leading to a lack of currency and potential for inaccuracy. It is therefore vital that the *search strategies* used to identify relevant studies should be published so that the process is seen to be auditable and repeatable.

In this paper we introduce a structured methodology for search strategy development and provide a framework for learning based on this methodology. In section 2 we provide the background to the problem in the context of the domain, and then discuss related pedagogical work in section 3. Section 4 details the structured searching methodology which we then apply in Section 5 to develop a framework for learning. We provide ideas for the future direction of the work in section 6.

CERI'16, June 14–16, 2016, Granada, Spain.

DOI: http://dx.doi.org/10.1145/12345.67890

2. BACKGROUND

At its heart, the process of systematic review relies on painstaking and meticulous searching of multiple literature sources. These include published literature sources such as MEDLINE and other specialist databases and 'grey literature' (i.e. technical reports and other non-peer reviewed sources). The principal way in which such sources are interrogated is through the use of Boolean queries, which utilize a variety of keywords, operators and ontology terms (also referred to as 'subject headings') – see Fig. 1.

```
("etiology"[Subheading] OR
"etiology"[All Fields] OR
"causes"[All Fields] OR
"causality"[MeSH Terms] OR
"causality"[All Fields]) AND
("somnambulism"[MeSH Terms] OR
"somnambulism"[All Fields] OR
("sleep"[All Fields] AND
"walking"[All Fields]) OR "sleep
walking"[All Fields])
```

Fig. 1 – Example of subject headings use

Reviewers incrementally build complex queries line by line, sometimes involving hundreds of terms, which are combined to form an overall search strategy – see Fig. 2.

The choice of search strategy is critical in ensuring that the process is sufficiently exhaustive and that the review is not biased by easily accessible studies [5]. In addition, the strategy needs to be transparent and repeatable, so that others may replicate the methodology. However, there are often mistakes in search strategies reported in the literature that prevent them from being executed in their published form. In one sample of 63 MEDLINE strategies, at least one error was detected in 90% of these, including spelling errors, truncation errors, logical operator error, incorrect query line references, redundancy without rationale, and more [6].

Evidently, despite the dedication and painstaking attention to detail of many individuals (many of whom are trained librarians), creating effective search strategies may be prone to error, often relying on manual processes with limited editorial support. Moreover, once published, strategies are typically stored as free text, and are thus rarely directly executable in their native form. This compromises their ability to be used by others and may unintentionally detract from the rigour of the review.

```
1. randomized controlled trial.pt.
2. controlled clinical trial.pt.
3. randomized.ab.
4. placebo.ab.
5. clinical trials as topic.sh.
6. randomly.ab.
7. trial.ti.
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. (animals not (humans and animals)).sh.
10. 8 not 9
11. exp Child/
12. ADOLESCENT/
13. exp infant/
14. child hospitalized/
15. adolescent hospitalized/
16. (child$ or infant$ or toddler$ or adolescen$ or teenage$).tw.
17. or/11-16
18. Child Nutrition Sciences/
19. exp Dietary Proteins/
20. Dietary Supplements/
21. Dietetics/
22. or/18-21
23. exp Infant, Newborn/
24. exp Overweight/
25. exp Eating Disorders/
26. Athletes/
27. exp Sports/
28. exp Pregnancy/
29. exp Viruses/
30. (newborn$ or obes$ or "eating disorder$" or pregnan$ or childbirth or virus$ or influenza).tw.
31. or/23-30
32. 10 and 17 and 22
33. 32 not 31
```

Fig. 2 – Complex Boolean query example extract

3. RELATED WORK

Students of library and information science may have taken a module on search strategy development in library school, but further education and training is required to deal effectively with the highly complex queries typical of systematic reviews. We provide a broad overview of teaching methods, curricula development, online materials and assessment and feedback below. A much larger review of the area can be found in Fernandez-Luna et al. [7] – here we focus on Level 2 of the taxonomy given in that paper, together with a focus on [B] Educational Goals; [b1] Library and information science. The technical level [A] focuses on the operational aspect of undertaking a query from a given information need.

The literature on teaching methods shows that generic ideas in pedagogy can be used to build frameworks to tackle problems in teaching and learning in information retrieval [7]. There are a couple of tensions which need to be considered. The first of these is that the method can either be process or outcome based. Kuhlthau [8] provides a five stage strategy to assist the process of learning based on prior work [9]. These kinds of ideas are very useful for systemic reviews, where the search process is very complex. McGregor [10] however shows that students tend to be focused on the outcome rather than the process, so any scheme for education must deal with that tension and ensure the student understands the importance of process. The second tension to deal with is whether to focus on theory or practice in teaching [7]: the argument for the former is that understanding the theoretical concepts provides the student with transferable skills (i.e. they can use any system to search). However systems do vary in practice and it is important to give the student practical knowledge of specific systems. In this work, we recommend a balance of the two.

The Cochrane Organization provides guidelines for search strategy development and has developed curricula specifically for search in systematic review [2]. Professional bodies such as CILIP in the U.K and the Medical Library Association in the U.S.A. [11] also provide guidelines for curricula design; of a more general nature in the former case and more specific in the latter case. Other organizations such as the UK Quality Assurance Agency provide subject benchmarks in Library and Information Science which can be used to inform curricula.

After a curriculum has been defined, the instruction method can be derived and there is a clear case for producing materials online through E-learning systems such as Moodle. Much work has been done in the area of developing online resources for IR instruction [7]; for example, the Cochrane Organization materials are online and available to all [2]. E-Learning environments may include interactive elements to help the student e.g. Java Applets [12], or a self-training package which can include pre and post self-assessments for students to measure their own progress [13]. Materials and assessments are designed in conjunction with each other and are integrated to ensure a good learning experience for the student. Face to face methods of delivery can also be used [7].

Assessment and feedback can use a wide variety of methods including those mentioned above in E-learning such as multiple choice questions (MCQs) [14], assignments, projects and tests [15] or even an IR game [16] where the student can assess their performance on a task using an IR system with a standard test collection, allowing them to assess their progress using a graphical tool. Feedback can be automatic via MCQs [14] or via summative or formative textual feedback [15] for more abstract ideas where there is no right and wrong answer (see section 5 below).

4. A STRUCTURED SEARCH METHODOLOGY

When undertaking a search there are a number of stages in the process starting from the realization of an information need (*cognition*) to the creation of a search which is submitted to an information retrieval system (*syntactic*). At City University London a search methodology has been used for many years on various modules which provides a structured approach to the process from beginning to end. The model resembles the framework derived by Taylor in 1968 [17], but is much more elaborate. We outline each of the levels of the structured searching framework in this section.

4.1 Cognitive stage

This stage initiates a search. The user realizes that they do not have sufficient knowledge to undertake a particular task [18], or in the case of systematic review is faced with a complex information need in the form of a specific research question. The search process is often performed by an intermediary (e.g. a librarian) who may have limited knowledge of the subject area. Therefore an ability to understand the needs of the original requestor is key at this stage. Needs can be *visceral* (an unexpressed need) or *conscious* (a within brain description) [17]. For the most part, systematic review needs are *conscious*, as domain knowledge may be required. Clarity of self-reflection is essential here as initial thoughts about the research question will govern the direction of the search process, and further reflection during iterations of searching will assist in understanding the topic and further guiding the direction of the search.

4.2 Linguistic stage

Once the searcher has identified an area or sub-domain in which to undertake a search, some kind of linguistic description is required to identify the underlying concepts. This could take the form of a document which describes the overall need, or it could be a description of facets using schemes such as PICO – Patient, Intervention, Comparison and Outcome [19]. Often it may involve both with the facets defined in the document describing the information need. Other more generic facet analysis schemes include ad-hoc and PMEST – Personality, Matter, Energy, Space and Time [20]. Whichever scheme is used, for each facet a list of terms and synonyms is identified.

4.3 Strategic stage

Assuming Boolean logic is used, the general scheme for taking data from the facet analysis is to apply an OR operator to the terms and synonyms within a facet, and then to apply the AND operator between the facets. There are three well-known search strategies for combining terms and facets: building blocks, successive fractions and citation pearl growing [21]. We outline these schemes here, and assume access to intermediate search sets as in the examples given in Figs. 1 and 2.

4.3.1 Building Blocks

In this method the sets for the facets are formed separately, and once this is done the final set is formed using the AND operator on the facet sets (see Fig. 3).

```
Set1 = etiology OR caus*
Set2 = somnambulism OR sleep
Set3 = walking OR "sleep walking”
Set4 = Set1 AND Set2 AND Set3
```

Fig. 3 – Building Blocks Search Strategy example

The advantage of this approach is that each set (sets 1 to 3 in Fig. 3) can be reused, and the search can continue to develop themes within each facet with no impact on other facets. However the drawback of this approach is that the search may lose focus in terms of its overall direction which may undermine its effectiveness.

4.3.2 Successive Fractions

An alternative to building blocks is successive fractions, where one facet is formed first and subsequent facets are incrementally added to the set to form the final answer (see Fig. 4).

```
Set1 = etiology OR caus*
Set2 = Set1 AND (somnambulism OR sleep)
Set3 = Set2 AND (walking OR "sleep
walking”)
```

Fig. 4 – Successive Fractions Search Strategy example

The searcher can start with the most general facet and refine the query from there. There are fewer steps in this method and it is more holistic, with the searcher having a clearer idea in each iteration of where the search is going. The drawback of the method is that any mistakes in earlier steps (e.g. Set3 in Fig 4), may require the search process to be restarted (e.g. from Set1 in Fig. 4).

4.3.3 Citation Pearl Growing

In this method a known useful item is pre-identified and index terms and or subject headings are extracted from it. The user goes through several iterations of extracting terms from records and testing them out on queries until they are happy with both the terms and their combinations [21]. A final phase is to reuse the building blocks strategy to create the final set of results.

4.4 Tactical stage

Within the strategies outlined above a number of tactics are available to the user. These fall into two broad groups: *choice of terms* and *choice of operators*. Choice of terms relates to the searcher's domain knowledge, whilst choice of operators relates to their knowledge of search techniques. The choice of terms can be augmented by the use of field operators, which depends on the meta-data available (e.g. restricting a search to the AUTHOR field). Operators can be either Boolean (AND, OR and AND-NOT are the only operators available on most systems) or word based. The latter can be either proximity operators (e.g. for phrases: "sleep walking" in Figs. 3/4) or truncation/wildcard operators (e.g. caus* in Figs. 3/4). While word based operators are not strictly Boolean operators, they behave in a similar manner, e.g. by narrowing a query (proximity operators are a special case of AND) or broadening a query (truncation/wildcard operators are a special case of OR).

4.5 Logical stage

A search strategy and its tactics are formed within some kind of logical framework – in the examples used so far we have concentrated on Boolean logic and word based extensions to that logic. This type of search is exact match logic using set retrieval, which is the dominant paradigm in systematic reviews. Another form of logic is best match in which ranked retrieval is addressed, but this is rarely relied upon exclusively in systematic reviews. We address exact match logic only in this paper therefore. Searching requires the users to utilize their knowledge of strategy, tactics and logic all together – which form the *formalized* need [17]. However logic can be considered separately in terms of learning, as we will outline later in section 4.

4.6 Syntactic stage

This is the stage where the user takes their search strategy and executes it on an operational information retrieval system – the *compromised* need (question as presented to the information system) [17]. Each IR system will have its own syntax, and the types of operators available and their range will vary from system to system. In most cases AND and OR are used as is, but most systems use NOT to mean AND-NOT (see set 10 in Fig. 2) which can be confusing since the Boolean operator NOT is actually a unary rather than binary operator.

Many systems used '*' and/or '?' for wildcard characters either for single characters, a given set of characters or any number of characters to an upper limit. Proximity operators often use quotes "" (as per Google), but many offer the ability to choose the number of characters between the terms e.g. WITHIN, NEAR. Proximity operators on blocks of text such as sentences or paragraphs are rare. A further syntactic method used often is to restrict the search to a given field (e.g. `“etiology"[Subheading])` which vary syntactically between systems (e.g. AU(name) in ProQuest Dialog [22]). The meta-data scheme available on the IR system or the source will determine the fields available.

Finally the type of interface will often determine the type of strategy used. The search strategies described in section 4.3 require access to search sets – a typical example of this is ProQuest Dialog [22]. Alternatively, in some cases a form-based user interface is available with columns representing terms, synonyms and phrases and rows representing the facets. In this paper we concentrate on the former, known as command line interfaces.

5. A FRAMEWORK FOR LEARNING

In this section we develop a framework for learning based on the methodology in section 4. Systematic review often requires two main activities [5]: an initial search to identify any existing systematic reviews on the subject, and a full search if/when no prior review is found. The framework can be applied in both activities, but is more critical for the latter due to the complexity of the task. At each stage we outline good practice and identify common sources of error [6], outline learning objectives, curricula and learning materials, teaching methods and assessment and feedback methods.

5.1 Cognitive stage

A key problem at this stage is the growing volume of published medical studies [1],[4] with the number increasing year on year. It is worth stressing the need for the user to reflect on the current state of their search and its ability to identify relevant studies that address the research question. Searchers should also be aware of the importance of sources, i.e. the databases that contain relevant information to fulfill the given information need. This will include peer review literature in prestigious journals, but other sources should be included such as non-English language articles, the 'grey' literature, non-refereed journals, conference proceedings, company reports etc. [3]. This ensures that the searcher understands the comprehensive nature of the requirements of a systematic review, and that the search needs to be exhaustive before any filtering of the literature can take place [3]. However searchers should be aware of information quality given the range and type of material available e.g. potential bias or error in published studies. To this end it would be useful to introduce the searcher to information literacy ideas to think through these issues [23]. Standard checklists are used by search professionals to validate their search strategy [24]. This includes the identification of a gold standard of known relevant records in section B of the standard checklist [24] which can be used to further citation pearl growing search strategies.

5.1.1 Learning Objectives

The learning objectives could include an understanding the following: 1) the importance of sources and potential bias in those sources, assessing information quality; 2) the exhaustive nature of systematic reviews and the process as a whole; 3) the notion of relevance and the use of gold standard records to assist strategy development.

5.1.2 Curricula and learning materials

A key resource would be the *Cochrane Handbook for Systematic Reviews of Interventions* [2], but general schemes on the body of professional knowledge from organizations such as CILIP would also be useful. There is useful work by Bates et al. [26] which surveys LIS curricula in Europe and recommends the use of Wilson's nested model [27] to guide curricula design – this would embed the learning materials in the key academic work on information seeking over many years.

5.1.3 Teaching methods

Clearly the student needs to take a step back and understand the information need in detail before attempting searching as recommended by Cohen [28]. The first author has used this method of a number of years where information needs are taught in conjunction with the linguistic stage, and this allows the student to focus on getting things right at the start. Putting students in groups and getting them to discuss the issues in tutorials has found to be a very successful form of learning [29].

5.1.4 Assessment and Feedback methods

A focus on assessment could involve encouraging students to develop their self-reflection skills through either formative or summative feedback schemes (perhaps even using peer review). Assessment would focus on information literacy tasks, assessing the information quality of sources using knowledge positive and negative examples and their relation to relevance.

5.2 Linguistic stage

A key issue at this stage is the ability to define an appropriate research question [3] given the information need identified at the cognitive stage. It would be useful to practice the writing of a document which describes this question, including its objectives, the subject area, the population concerned, type of evidence for evaluation and outcomes required [3]. Once this is done, support can be provided for the extraction of facets from the healthcare question, using PICO [19] or some other appropriate facet analysis scheme. If appropriate or available, the use of tools for extracting PICO elements or other information using utilities such as ExaCT [5] could be useful. Section A of the standard checklist [24] can be used to collect information about the information need including the authors' stated objective, the focus of the research etc.

5.2.1 Learning Objectives

The learning objectives could include the following: 1) defining and documenting a clear research question; 2) Using facet analysis techniques such as PICO to analyze the research question effectively.

5.2.2 Curricula and learning materials

Reference to good practice provided by Cochrane [2] on how to conduct systematic reviews would be appropriate. This would give the student an overall idea of how to initiate work on a research question and keep it up to date. A collection of example facet analyses for healthcare topics (e.g. with the PICO scheme) would be useful learning materials.

5.2.3 Teaching methods

Since facet analysis is not an exact science, students should be encouraged to develop their own ideas and to refer to case studies and examples illustrating good practice. This can be done individually or in group tutorials, or through online tasks using E-learning materials.

5.2.4 Assessment and Feedback methods

As with the previous level, both formative and summative feedback schemes are appropriate, providing individualized feedback to address specific student issues. The use of MCQs could be considered, but only to address known issues in facet analysis such as placing the terms in the correct PICO element.

5.3 Strategic stage

This stage is concerned with translating the facet analysis to the search strategy. A key issue is understanding the relationships between facets: in the case of Boolean search strategies, OR is typically applied to terms within facets, and the AND operator is applied between facets (see section 4.3). Students can confuse the two and use inappropriate operators e.g. AND within a facet. One author has been teaching this material for 15 years, and it is a common source of error.

A number of key problems at the strategic level are identified by Sampson et al [6]. In some cases the wrong line number is used in a step, either omitting a set or using an incorrect set (this applies to any of the strategies described in section 4.3.1 to 4.3.2). The searcher can avoid this by drawing the relationship between line numbers/sets, to show the relationship between or within facets depending on the focus. MeSH and free text terms used on the same line can compromise reuse. A simple solution to this is to address the granularity of the strategy, and provide examples of when MeSH and free text terms could be decoupled. Terms can be reused leading to redundancy without rationale, which may not harm the search but may slow down run times for large searches and complicate the strategy unnecessarily. A way round this is to check for the use of a given term more than once in a strategy, and ensure that the term is required at that particular stage. In the case where searches are required over a number of databases, training on how to tailor the search strategy to each database should be provided. This should include a clear description of the strategy for the purposes of reproducibility (which is good practice in systematic reviews). Section C of the standard checklist [24] provides examples of issues to think about when forming the search strategy, including adapting an already existing search strategy, using a database thesaurus and thinking about how the final combination of terms were selected (see sections 5.3.1 to 5.3.3 below).

5.3.1 Building blocks

A key issue with this strategy is to get users to understand the drawbacks of the method e.g. a searcher focusing on one facet may lose focus on the whole topic (section 4.3.1). Users should be trained to understand that if they are to use the method, a clear understanding of each facet must be gained. This could include continual review of the information need and any related topics which could be useful for each facet. Links and relations between facets should be identified by the searcher and recorded in a checklist [14].

5.3.2 Successive fractions

In this strategy the sequential order of the facets is crucial, and the user needs to be taught to think about the starting point. Normally this would be the most specific facet first (e.g. the type of patient in PICO), with other more general facets following after (e.g. outcome in PICO). This is particularly important in more ad-hoc methods of analysis where the user has identified their own facets e.g. Object, Activity, Date. In such cases it would be better to start with the Object/Activity facets and finish with the Date facet. As with building blocks links and relations between facets should be identified by the searcher and recorded in a checklist [24].

5.3.3 Citation Pearl Growing

This requires an understanding of the use of gold standard records (the 'pearls') to develop an overall strategy. Section D of the checklist [24] provides useful advice on considering issues such as sensitivity (recall), precision and specificity [25]. The 'pearls' can be used to check each metric and the strategy developed to meet a certain criteria e.g. a preference for a high level of sensitivity (recall) whilst ensuring a threshold of 50% for specificity [25]. This is done by checking to see if the 'pearls' are retrieved by the search strategy, and an interactive process in search strategy development may be needed to in order to ensure that all 'pearls' are retrieved. The balance of the two can be adapted to the given needs of the searcher, but the linkage between the different terms needs to be emphasized. Choice of further strategies such as building blocks can then be addressed.

5.3.4 Learning Objectives

The learning objectives could include the following: 1) effective translation of facet analysis into an appropriate research strategy; 2) understanding the different forms of search strategy, their similarities and differences and when to apply a given strategy for a particular problem.

5.3.5 Curricula and learning materials

The curricula would focus on the different forms of search strategies available, with a clear link made to the facet analysis. The problems identified early in this section should be specifically addressed and built in to the learning materials. Each of the search strategies needs to be clearly explained with appropriate examples, with differences between building blocks, successive fractions and citation pearl growing demonstrated.

5.3.6 Teaching methods

There are a number of different methods for teaching search strategies including Bhavnani et al [30], which uses taxonomies of both tasks and general IR strategies to build a methodology to learn to search by 1) learning specific search strategies for frequent tasks, 2) using strategies for given contexts, 3) learning how to execute a strategy accurately and 4) applying strategies across different applications (in conjunction with the syntactic level below). Use of graphical online tools would also be a useful addition to the learning experience e.g. the relations between intermediate search sets.

5.3.7 Assessment and Feedback methods

The use of MCQs can be used to test understanding of the form of strategy, e.g. MacFarlane [14] specifies an example set of questions (labelled under the group C element of the MATH taxonomy [31]), which would use questions on the different forms to allow the user to assess their own understanding. For example, giving the student a facet analysis and asking them to identify the correct building blocks strategy from a number of distractors. Key problems identified in the Common errors should be built in to the distractors, e.g. using OR between facets instead of AND.

5.4 Tactical stage

The strategic and tactical stages are closely related and often need to be considered simultaneously. This requires thought on the use of terms and operators (section 3.4).

A number of common errors at the tactical level are identified by Sampson et al. [6]. Spelling errors are a significant issue. Applying appropriate thesauri or other knowledge organization schemes (e.g. taxonomies, ontologies) can require further verification of medical terms. Google may be used as a source of verification but has limited value as the terms returned may reflect similar errors made on the web and may not provide relevant terms for the domain. Missed spelling variants can be dealt with by teaching the searcher to think about variations of words and use truncation as a tactic. However, the searcher can inadvertently choose irrelevant MeSH or free text terms, or alternatively miss other useful MeSH terms. A further problem is that MeSH terms can be exploded without any effect if the term is at the bottom of the hierarchy, since no further child terms exist. Encouraging the learner to reflect on the terms used and providing training on the MeSH scheme can help address these issues.

Section C of the standard checklist [24] provides examples of issues to think about when forming tactics, including terms extracted from documents and identifying different types of term checking including terms extracted from gold standard records, terms suggested by experts and from database thesauri etc.

5.4.1 Learning Objectives

The learning objectives could include the following: 1) how to successfully use appropriate tactics within a given strategy, 2) good practice on choosing operators, 3) good practice on choosing terms.

5.4.2 Curricula and learning materials

The learning materials would focus on when to use particular operators in a strategy, e.g. Boolean, proximity or wildcard operators, and best practice on picking terms e.g. those extracted from gold standard records.

5.4.3 Teaching methods

Given the subjective nature of term selection, students can be put into groups and given case studies along with examples of good and bad tactics for those strategies. The use of operators is more objective, and online self-reflection materials can be used.

5.4.4 Assessment and Feedback methods

For term selection tactics, either formative or summative feedback schemes would be appropriate, providing individualized feedback to address specific student issues. MCQs can be used for operator tactics with appropriate use with given distractors, which can be delivered with Group C questions [31] in strategies above [14] but as a separate question set, e.g. the correct use of MeSH terms.

5.5 Logical stage

Closely aligned with the tactical stage is the logical stage of the framework (section 3.4). Two key problems at the tactical stage are identified by Sampson et al [6]. The first of these is confusion between the operators AND, OR with potential serious impact to the overall search strategy (section 3.4). This can occur with users unfamiliar with Boolean logic who are used to thinking in terms of AND as an OR: for example a request such as 'Find me documents about cats and dogs' is linguistically AND, but semantically it implies OR. This contrast can be confusing for students. Clarification on the natural language use of OR and AND needs to be highlighted to the user. The second issue is the inappropriate use of the NOT operator, which must be used with care as relevant documents may be eliminated from results. It should be stressed to the learner that the NOT operator should only be used where a given term or set of terms is known to be harmful to the overall search. Further training could be given on the relationship between the word operators (truncation, proximity) and Boolean operators (OR, AND) ensuring they understand that the former are special cases of the latter (section 4.5).

5.5.1 Learning Objectives

The learning objectives could include the following: 1) correct use of Boolean and extended Boolean operators.

5.5.2 Curricula and learning materials

The material would focus on understanding Boolean logic using methods such as Venn diagrams, together with providing some understanding of the underlying axioms of the mathematics e.g. AND, OR are symmetric, whereas NOT is not symmetric. This material can be drawn from any good textbook on discrete mathematics. The use of word operations e.g. proximity and wildcards can then be further explained from a Boolean logic perspective.

5.5.3 Teaching methods

Online delivery of the material would be appropriate for this level, with examples and self-assessment for each of the operators. The teaching scheme must not assume that the student is familiar with discrete mathematics [29]. Tutorial group tasks have also proved to be successful for face to face students [29].

5.5.4 Assessment and Feedback methods

Group A questions [31] could be used to assess the understanding of Boolean and extended Boolean logic by providing text examples and asking the student to pick which queries would retrieve that text [14].

5.6 Syntactic stage

Implementing the search strategy on an operational information retrieval system is the final stage of the search (section 4.6). The syntax of the different search systems can be very different but there are certain commonalities. In cases where multiple searches are required, training on translation of queries to different systems should be provided. This includes training on unary operators (applied to a single term), binary operators (applied to two terms) and clarification of what operators are symmetric (two different terms can be on either side of the AND, OR operators) and non-symmetric (in Dialog ProQuest [22] the proximity operators ""/PRE impose order on words, whilst NEAR does not).

One particular problem at the syntactic level is identified by Sampson et al. [6]. This is the inappropriate use of truncation e.g. using methods* instead of 'method*' to capture several terms on that concept. Training on truncation operators and their impact needs to be provided and examples given of both appropriate and inappropriate use.

5.6.1 Learning Objectives

The learning objectives could include the following: 1) understanding how to translate a Boolean search strategy with relevant tactics into a form which can be executed by an operational information retrieval system.

5.6.2 Curricula and learning materials

Materials will need to be developed for specific systems e.g. ProQuest Dialog [22], together with a general scheme of how to approach the translation of a generic Boolean query to relevant syntax. This will require a survey of existing systems used in systematic review. The material will need to address problems identified in the literature mentioned above [6].

5.6.3 Teaching methods

At this stage practice on real systems will be required to ensure that the user can truly understand the final stage. This could require the use of PC labs, with specific tasks – perhaps in conjunction with an overall task from all levels of the framework – with work on other levels being done prior to the lab. The teaching method needs to instill some self-reflection, to establish both the process of translation of the Boolean query to the target system, but also to instill confidence in the student in what can be a very complex activity. Online materials and self-assessments on individual elements of the system syntax would also be useful.

5.6.4 Assessment and Feedback methods

Assignments which give the student an opportunity to build their confidence and knowledge in search e.g. providing an example systematic review case study to search for and allowing them to build an operational query to find information for that case study.

In-class tests could also be used, whereby students are provided pre-defined search strategies and given limited time to form real searches using a given system in a lab. Multiple choice questions can be used to tackle Group B questions, focusing on specific issues or known problem with syntax on a given search service [14]. An example would be to give a list of search forms in the given syntax and get the student to choose the number of correct forms [14].

6. SUMMARY AND CONCLUSION

We have introduced here a structured search methodology which is used to inform a framework for learning how to develop search strategies which can be used in systematic reviews. This framework includes a number of discrete but interlinked stages: cognitive, linguistic, strategic, tactical, logical and syntactic. The learning framework applied to each stage is a follows:

Cognitive: In this stage the importance of assessing sources will be stressed, in particular understanding the issue of information quality and potential bias in publications. Ideas and concepts in information literacy can be used to inform this part of the framework.

Linguistic: A key skill here is forming a research question given a clinical need, and using an appropriate facet analysis scheme to identify the complementary concepts of the need. Training in the use of standard facet analysis schemes such as PICO are required, together with training on software which can be used to build the facets.

Strategic: Being able to take the facet analysis and form an appropriate search strategy is the key skill that needs to be developed at this stage. This includes in initial translation from the facet analysis to the strategy (OR is applied within facets, AND between facets), to choosing the type of strategy to be used: building blocks, successive fractions or citation pearl growing.

Tactical: With a strategy, the choice of terms and operators needs to be considered. Choice of terms will depend on domain knowledge and interaction with a subject matter expert, whilst choice of operator requires the appropriate knowledge of Boolean operators and proximity operators that extend Boolean logic in various ways. Training on the use of field operators would also be appropriate.

Logical: An understanding of the operators identified in the tactical stage is required, in particular the differences and relationships between the operators need to be established as well as the appropriate use of operators.

Syntactic: This final stage needs to be carried out with an operational information retrieval system, and an understanding of the systems functionality must be provided. The system's ability to handle intermediate search sets must also be stressed to support the complex search strategies outlined above.

The next stage in this work is to develop learning materials to deliver this learning framework, to engage in outreach activities with users who undertake systematic reviews, and to provide them with a structured learning framework that they can use to improve their knowledge and skills. Guidance on how to develop learning objectives, curricula/learning materials, teaching methods and assessment/feedback for each individual level of the search framework is provided in section 5. It is our plan to develop these concepts further. The proposed outcome of this work is to give users the skills they need to be more effective searchers and to share their knowledge with others who have common interests. A broader outcome is to improve the quality of search strategies used in systematic reviews, thereby improving the quality and accuracy of those reviews.

7. ACKNOWLEDGEMENT

The authors are grateful to Stephen Robertson, Lyn Robinson and David Bawden for their feedback on the original of the structured searching framework, and to Stephen Robertson for recommending the Taylor 1968 reference [7].

8. REFERENCES

[1] Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J. P. T., Mavergames, C., and Gruen, R. L. 2014. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. In: Plos medicine, Vol. 11, No. 2.

[2] Lefebvre, C., Manheimer, E., and Glanville, J. 2011. Searching for Studies. In Higgins, J. P. T., and Green S., Eds. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available on: http://handbook.cochrane.org/.

[3] Hemingway, P. and Brereton, N. 2009. What is a systematic review? 2nd ed. Hayward Medical Communications.

[4] Shojania, K.G., Sampson, M., Ansari, M,T,, Ji, J., Doucette, S. and Moher, D. 2007. How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis. Ann Intern Med.

[5] Tsafnat, G., Glasziou, P., Choong, M.K., Dunn, A, Galgani, F. and Coiera, E. 2014. Systematic review automation technologies. Syst Rev 3,74, 1-15.

[6] Sampson M, McGowan J. 2006. Errors in search strategies were identified by type and frequency. J. Clinical Epidemiology. 59, 10, 1057–63.

[7] Fernandez-Luna J.M., Huete, J.M. MacFarlane, A. and Efthimiadis, E.N. 2009. Teaching and learning in information retrieval. Inf Retrieval, 12, 201-226.

[8] Kuhlthau. C.C. 1997. Learning in digital libraries: an information search process approach. Lib Trends 45,4, 708-724

[9] Kuhlthau, C.C. 1988. Developing a model of the library search process: cognitive and affective aspects. RQ, 232-242.

[10] McGregor, J. 1994. Information seeking and use: students' thinking and their mental models. J. Youth Services Lib, 8, 69-76.

[11] Nicholson, S. 2005. Understanding the foundation: the state of generalist search education in library schools as related to the needs of expert searchers in medical libraries. J. Med. Lib. Asoc. 93, 1, 61-68.

[12] Henrich. A. and Morgenroth, K. 2007. Information retrieval as e-learning course in German – Lessons learned after 5 years of experienced. Proc. 1st international workshop on teaching and learning of information retrieval. Available on: http://tinyurl.com/z2ueueh.

[13] Sacchanand, C. and Jaroenpuntaruk, V. 2006. Development of a web-based self-traning package for information retrieval using the distance education approach. Elec. Lib. 24, 4, 501-516.

[14] MacFarlane, A. 2011. Using multiple choice questions to assist learning for Information Retrieval. In Efthimiadis, E.,

Fernandez-Luna, J.M. Huete, J.F. and MacFarlane, A. Eds. Teaching and Learning in Information Retrieval. Springer Verlag, Berlin, 107-121.

[15] Zhu, L. and Tang, C. 2006. A module-based integration of information retrieval into undergraduate curricula. J. Comp. Sci. Col. 22, 2, 288-294.

[16] Halttunen, E. and Sormunen, E. 2000. Learning information retrieval through an educational game. Is gaming sufficient for learning. Edu. Info, 18, 289-311.

[17] Taylor, R.S. 1968. Question-negotiation and information seeking in libraries. College & Research Libraries, 29, 3, 178-194.

[18] Belkin, N. J., Oddy, R.N. and Brooks, H.M. 1982. ASK for Information retrieval: part 1. background and theory. J.Documentation. 38, 2, 61-71.

[19] Dahlgren Memorial Library. 2016, Evidence-Based Medicine Resource Guide: Types of Clinical Questions. Available on: http://tinyurl.com/zfsa3ob.

[20] Ranganathan, S.R. 2006. Colon Classification (6th Edition). Ess Ess Publications. New Dehli.

[21] Markey, K. and Cochrane, P. 1981. Online training and practice manual for ERIC database searchers (2nd Edition). ERIC Clearinghouse on Information Resources, Syracuse University. Available on: http://tinyurl.com/j5v65wb.

[22] ProQuest Dialog. Available on: http://tinyurl.com/l22vdk9.

[23] Inskip, C. 2014. Information literacy is for life, not just a good degree: a literature review. CILIP. Available on: http://tinyurl.com/kjaujnh.

[24] Glanville, J., Bayliss, S., Booth, A., Dundar, Y., Fleeman, N., Foster, L., Fraser, C., Fry-Smith, A, Golder, S., Lefebvre, C., McNally, R., Miller, C., Paisley, S., Payne, L, Price, A. Shaikh, H., Sutton, A., Welch, K. and Wilkinson, A. 2008. So many filters, so little time: The development of a Search Filter Appraisal Checklist. J.MLA, 96,4, 356-361.

[25] Wilczynski, N.L., Haynes, R.B., Lavis, JN., Ramkissoonsingh, R. and Arnold-Oatley, A. 2004. Optimal search strategies for detecting health services research studies in MEDLINE. CMAJ 171, 10, 1179-1185.

[26] Bates, J,. Bawden, D., Corderio, I. Steinerova, J., Vakkari, P. and Vilar, P. 2005. Information Seeking and information Retrieval. In Kajberg, L. and Lorring, L. Eds. European curriculum reflections on Library and information Science, Denmark. RSLIS.

[27] Wilson, T.D. 1999. Models in information behavior research. J.Doc, 55, 3, 249-270.

[28] Cohen, L.B. 2001. 10 tips for teaching how to search the web. American Libs, Nov, 44-46.

[29] MacFarlane, A. 2009. Teaching mathematics for search using a tutorial style of delivery. Inf Retrieval, 12, 162-178.

[30] Bhavnani, S., Drabenstott, K., and Radev, D. 2001. Towards a united framework of IR tasks and strategies. In Proc. ASSIT annual meeting, 38, 340-354.

[31] Smith, G, Wood, L., Crawford, K., Coupland, M., Ball, G. and Stephenson, B. 1996. Consructing mathematical examinations to assess a range of knowledge and skills. Int J Math Educ Sci Tech, 30, 47-63.