Search and Retrieval Strategies in the Systematic Literature Search

Terminologies and Taxonomies in Information and Library Sciences

Anne-Kathrin Merz, Helge Knüttel
University of Regensburg, Germany

Contact

Anne-Kathrin Merz M.A.
University of Regensburg, Germany, Faculty of Medicine
Mail: anne-kathrin.merz@ukr.de

Dr. Helge Knüttel
University of Regensburg, Germany, University Library
Mail: helge.knuettel@ur.de

INTRODUCTION

Modelling the Systematic Literature Search

The systematic literature search (SLS) is a core element of systematic reviews (SR) in EBM. Developing optimal and reproducible search strategies is a crucial and resource-intensive task[1]. Search strategies aim to find an unbiased, optimal and database-specific representation of the review question in form of a Boolean search query. According to Cochranes MECIR statement this task involves a set of clearly defined sub-tasks[2].

Successful efforts were made to support individual resource-consuming steps of the systematic literature search through software applications but a systematic classification of tools is still missing[1] as well as a complete taxonomy of effective strategies in systematic literature search. Our study focusses these questions.

It is work in progress, so further research has to be done.

Conclusions

Since the 1970's information scientists developed a variety of models to describe search processes and Information Behaviour[3-5]. Information Behaviour describes all interaction with information systems and information units, as well as the variety of methods people employ to discover and gain access to information resources[5]. In general, Information Behaviour can be described by a set of task- and context-dependent tactics[6], which are defined as patterns of moves/activities[7]. Information Behaviour models are core references for the architecture of information systems ranging from classical databases to search engines.

A process-oriented Information Behaviour model concerning the search behaviour in systematic literature searches is still missing. We expect that connecting the views of information science and EBM is a valuable contribution to the understanding of seeking and searching in SR and may help to develop evidence-based, user-friendly software solutions supporting systematic literature searches.

RESULTS

Context and Information Behaviour

In the SLS we identified higher level interaction contexts with specific characteristics

Context	Examples	Characteristics
Information Source (IS)	(bibliographic) databases, clinical trial registers, thesauri, grey literature sources ...	structure, familiarity, availability, reliability
Information Items (II)	citations, references, full-texts, studies, thesauri entries	structure, availability, evidence
Terms/Queries (TQ)	keywords, synonyms, descriptors, artefacts, search query	complexity

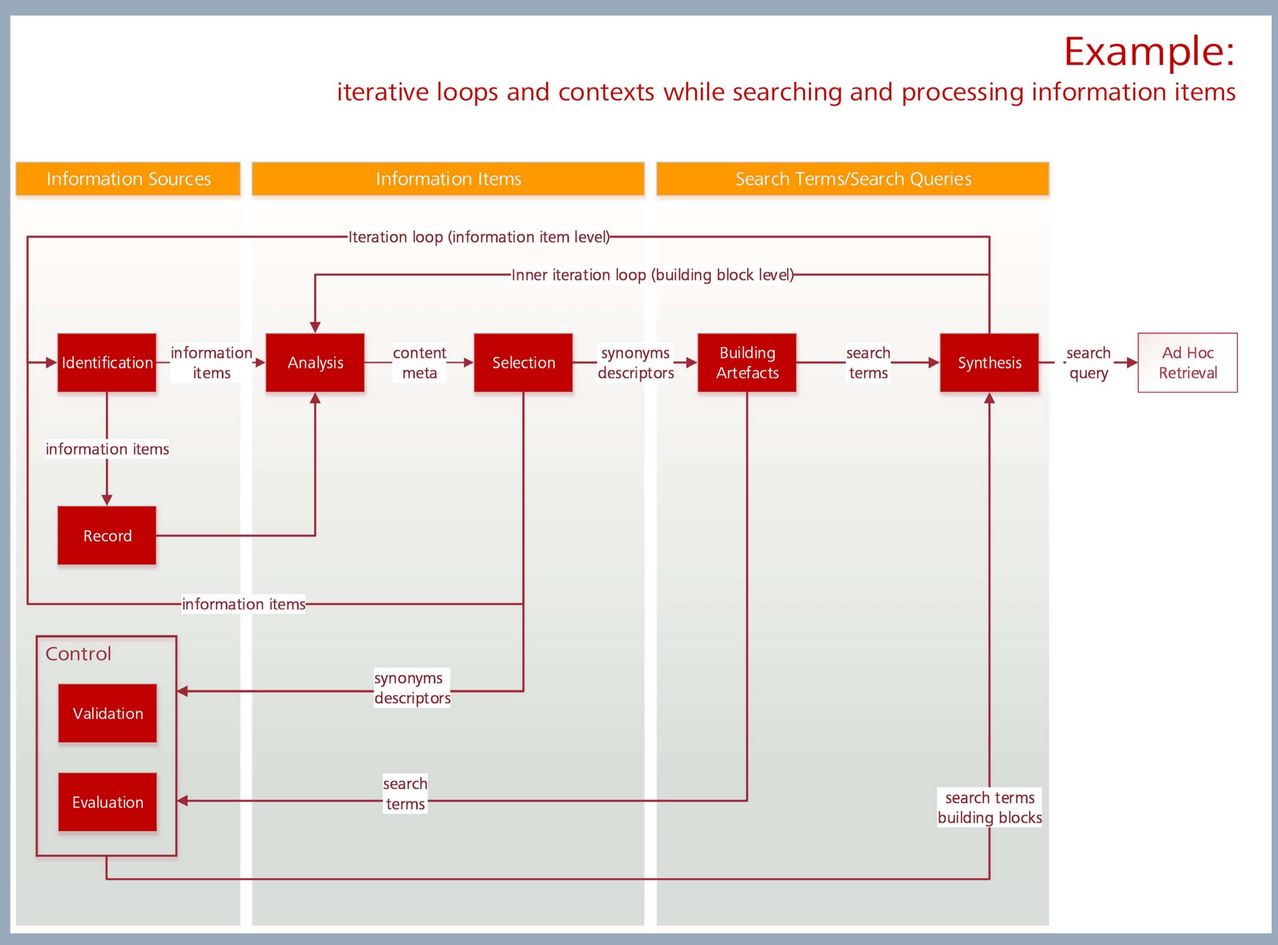
The central SR task "development of a search strategy" contains concrete search tasks and results in different search-related information behaviours depending on different contexts. Searching for relevant information items as a source for high quality search terms are the central search tasks.

Information Behaviour	Context	Examples/Tasks
Identification		
Searching	IS	Ad Hoc Retrieval
Specifying	IS	search for existing systematic reviews, known item search
Chaining	II	bibliographic search (backward), citation search (forward)
Monitoring	IS	E-Mail alerts
Processing		
Analyzing	IS II	thesauri, bibliographic entries, full text, citations
Selecting	IS II	citations, descriptors, synonyms (terms and phrases)
Creating Artefacts*	TQ	generate search terms or new phrases by truncation, operators, wildcard (database specific)
Synthesizing*	TQ	add new search terms to query, re-arrange search queries
Controlling		
Validating**	IS	meaning of free text terms and descriptors
Evaluating**	IS	evaluation of free text terms, phrases, descriptors

* Lower level activities were identified
** These activities should be defined as separate tactics due to their complexity and variants

Patterns and Tactics

Tactics might contain chains or patterns of information behaviours.
We identified several recurring patterns in our empirical data.



Factors affecting the collection of synonyms
Need for further synonyms
Retrieval of known relevant documents
Occurance in specific field (TI-AB-TX)
Missing descriptors
Insufficient indexing
Reasons to conduct test searches
Comparison of broad/narrow descriptors
Unspecific candidate terms
Complexity of phrases
Comparison of different ADJ/NEAR artefacts
Variation in syntax and spelling

We found several factors affecting the collection of synonyms as well as reasons for the need for control loops.

One popular tactic to get high quality search terms is to ensure the retrieval of relevant information items by choosing at least one descriptor and alternatively TI-AB terms per building block.

DISCUSSION/METHODS

Discussion

It is possible to map SR work-tasks like "identify appropriate controlled vocabulary […] and free-text terms[…]"[2] to concrete search-tasks well known to information science.

Citations identified as relevant play an especially important role within the search process. As optimal representative of the information need (clinical question), they are ideal sources for the identification of new search terms as well as new representatives. Therefore we identified and confirmed professional search behaviour described elsewhere[3,4], but also specifics of SLS within systematic reviews: Searching and processing information items to identify new search terms as well as controlling search terms belong to the main search related tasks that were previously not described in the IR literature in detail.

Professionals often use search tactics arising from the early days of online databases[8] and referred to as best practices[9]. Typical database-related tactics, previously found by Bates[7] and others[10], are in common use. During systematic reviews they are not only employed to build ad-hoc search queries but also to test the quality of search queries. Therefore they not only can be interpreted as lower level search tactics but also as lower level control tactics.

Further research needs to be done to confirm our findings and complete our model in more detail.

Issues arise from the diversity of terminology and contexts used in EBM and information sciences: Methods in SR and SLS are closely connected to the structure of Boolean databases, where IB models are often more general covering a wide range of information sources.

Methods

Our work is based on a comprehensive literature review as well as behavioural observations and a free-text questionnaire investigating tactics for search term identification and the development of search strategies.
We observed an information specialists' behaviour during a systematic review in Nov 2014. In protocols from 6 sessions each ranging from 1-3 hours we identified up to 190 single moves per session. Protocols were analysed to classify information behaviour and contexts. Processes were coded into flowcharts. 16 free text questions about tactics to build search queries were answered by as yet 4 information professionals. Answers were analysed, classified and coded into flowcharts, too. In the next step we compared our empirical data with existing theories and models of Information Behaviour during professional searches and mapped our findings to known results.

Literature cited

1. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev 2014; 3, S. 74.
2. Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Methodological standards for the conduct of new Cochrane Intervention Reviews: Cochrane Collaboration; 2012 Dec 17. Available from: URL:http://www.editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR_conduct_standards%202.2%2017122012_0.pdf.
3. Ellis D. A BEHAVIOURAL APPROACH TO INFORMATION RETRIEVAL SYSTEM DESIGN. Journal of Documentation 1989; 45(3), S.171–212.
4. Makri, S; Blandford, A; Cox, A(2008): Investigating the information-seeking behaviour of academic lawyers. From Ellis's model to design. In: Information Processing & Management 44 (2), S. 613–634.
5. Wilson TD. Models in information behaviour research. Journal of Documentation 1999; 55(3):249–70.
6. Xie, I; Joo, S (2012): Factors affecting the selection of search tactics: Tasks, knowledge, process, and systems. Information Processing & Management 48 (2), S. 254–270.
7. Bates, MJ. (1979): Information search tactics. In: J. Am. Soc. Inf. Sci. 30 (4), S. 205–214.
8. Fidel, R (1991): Searchers' selection of search keys: I. The selection routine. In: Journal of the American Society for Information Science 42 (7), S. 490–500.
9. Booth, A (2008): Unpacking your literature search toolbox: on search styles and tactics. In: Health Info Libr J 25 (4), S. 313–317.
10. Markey K, Cochrane PA. Online Training and Practice Manual for ERIC Data Base Searchers. Syracuse, NY: ERIC Clearinghouse on Information Resources; 1981. Available from: URL:http://eric.ed.gov/?id=ED212296.