Supporting Novice Researchers in Domain Familiarization ï A Question-Guided Research Community Map Creation and Reflection Tool

Shruti Hardat[1], Christoph Greven[2] and Ulrik Schroeder[2]

Abstract: The first step to starting research is to familiarize with the research field, which is a difficult task for non-experts. They usually do not know which exact information to gather or which strategy to follow. There exists some literature on supporting them by means of guidelines, visualizations and critical thinking. However, they must be supported such that they can form their own strategies to objectively evaluate the research field. Therefore, a tool was developed to help non-experts identify which information to gather, structure the gathered information by means of a visualization and perform the task in a time-efficient manner. Finally, a user study was conducted to investigate the validity of the proposed approach. Results show that the developed software tool supports non-experts in their literature review process by providing them with an overview of the research field and making the process time-efficient.

Keywords: domain familiarization, literature research, research community, concept map, guided questions, computer science education, e-learning

1 Literature Research for Domain Familiarization

Familiarization with new knowledge domains is a life-long learning process that encompasses various areas. It is important to understand a domain to successfully perform any related task, whether in work or private life. This especially holds true for science, where research work is conducted specific to a certain field. For expert researchers, with extensive research experience, familiarization needs are fulfilled by performing literature reviews, especially in the initial stages of their work. Literature reviews help researchers familiarize themselves with the composition of the field, its researchers, research groups and the state-of-the-art [Ra09]. Additionally, they help researchers in finding spaces of opportunities and improvements [Ra09].

This process, while straightforward for an expert, can be quite complicated for a non-expert, i.e. someone with no prior research experience, due to various reasons. A natural approach to performing a literature review is to firstly acquire and secondly analyze information regarding the research field, including who or what it constitutes of. Experts usually already understand the dynamics of the field and therefore know where to look and what to look for. However, information acquisition can be difficult for non-experts.

[1] RWTH Aachen University, Learning Technologies Research Group, Ahornstraße 55, 52074 Aachen, shruti.hardat@rwth-aachen.de
[2] RWTH Aachen University, Learning Technologies Research Group, Ahornstraße 55, 52074 Aachen, {greven|schroeder}@cs.rwth-aachen.de

Given the large amount of information available, it is critical for researchers to know what exactly to search for to avoid cognitive overload and disorientation [Te05]. Unfortunately, non-experts may simply not possess this knowledge, thereby leading to them adopting a trial-and-error approach instead of a more efficient, strategic one. Furthermore, apart from lacking procedural know-how, non-experts also do not possess domain related knowledge. The domain glossary or widely-used keywords are probably unknown to them. Finally, having acquired the right information, it is also essential for non-experts to be able to critically analyze this information to form an objective overview of the research field. However, the kind of critical thinking required to analyze the information may be beyond their expertise. Therefore, it is vital to provide support to non-expert researchers in their domain familiarization endeavor.

Based on the outlined problems, it becomes clear that non-experts require support tailored to their needs while performing a literature review. The aim is to simplify their familiarization process by helping them form an overview of the research field. Additionally, the aim is to help them in a manner that they learn about concepts and strategies to perform a literature review so that they can replicate expert behavior and perform the task independently in the long run. Since the existing literature does not address these exact needs, a software tool acting like a guidance system was developed. It provides a systematic approach to performing a literature review by supporting the identification of their information needs, so that they realize exactly which information they need to acquire. The end goal is for non-experts to be able to gain an objective understanding of the research field and find relevant literature based on this understanding. The tool allows users to make informed decisions by means of reflective thinking and visualizations. Reflective thinking is a deliberate thinking process that can lead to a deeper understanding of the information [WY96] while visualizations are useful in providing information overviews. The development of the tool is based on literature and user research apart from rigorous usability testing with users. Finally, the software tool is evaluated with the target group to verify whether this approach helps non-expert researchers gain an overview of the research community in a time-efficient manner. Based on statistical and qualitative user evaluation, the outlined approach has positive implications for non-expert researchers.

2 Approaches to Aid Literature Research

Existing literature focuses on various aspects related to research and tools to ease the process. Some research has focused on identifying important information that non-experts must obtain to gain an overview over a research discipline. Additionally, a significant amount of research has looked into visual tools as a means to understanding existing interrelationships in the research field. Finally, research has also investigated the role of concept mapping in establishing critical thinking.

2.1 Useful Information to Gain Insights

Research regarding support for non-expert researchers has largely been bibliography- and paper-oriented [Fu11], [MSL11], [BCB03]. Tools falling under this category generally target non-experts and seek to support their literature review process by somehow automating it. While finding relevant literature is an important aim of a literature review, such systems merely capture the output of the research community. They provide no overview of the community itself and do not support in objectively evaluating the research field. The first step to starting a literature review must be to gather relevant information to objectively evaluate it. Unfortunately, it may not be clear to non-experts which information is relevant. To this end, [RC07] attempted to map a research domain using their knowledge framework. They listed questions which, according to them, must be answered to understand the intellectual structure of the domain. [IK11] conducted expert interviews to identify questions novice researchers need to answer in the early stages of their research. The ideas outlined by both works are in agreement with each other and can serve as guidelines for non-expert researchers when they perform a literature review. According to the literature, non-experts must be able to answer questions like who the core *researchers* in the domain are, their *affiliations*, *research projects* and *interrelationships*. Additionally, they must identify the significant *research topics* and *trends* within the research field, and also important *resources* of literature such as *conferences*. However, the relevance of these questions must firstly be evaluated with target users, as no such evaluation was performed by these authors.

2.2 Adequate Medium to Understand Interrelations

The information collected while familiarizing with the research field consists of various relationships, e.g. those between researchers and their research projects. Visualizations have been extensively used in the literature to depict such links. To understand the participation of researchers and their roles in multiple projects, [Kn11] examined the composition of users and projects of the TeraGrid as a two-mode affiliation network. The work outlines the importance of understanding connections between project groups and research projects to identify how they cooperate or compete with each other. [Eg05] adopted a knowledge mapping framework and performed various analyses using visualizations like author citation and similarity maps, to summarize the medical informatics research domain. These visualizations are effective in identifying connected authors. However, they contain an extremely high amount of information and interlinking. They also do not provide a summary of the entire domain as a whole. [Se11] created Macademia, a website that promotes researcher collaborations based on their interests. The visualization centers either around a researcher or a research interest, and researchers are linked based on the latter, thus easing the process of identifying researchers with similar specialties. Unfortunately, it also suffers from a high amount of interlinking. In general, since most of such research is focused on summarizing the domain, it outlines various criteria that can be interesting when starting a literature review, such as identification of prominent researchers, subtopics within the research etc. These tools also

make use of overview visualizations to efficiently depict various aspects of the research community, although the overview usually does not encompass all possible components. Nonetheless, even if they were to overview the entire domain as a whole, it is important to help non-experts perform the summarization themselves. In this way they can actively form an understanding of the relationships prevalent in the community and gain an overview themselves instead of passively reading about it.

2.3 Establishing Critical Thinking

Visualizations can play a major role in organization and sense-making of knowledge. A special type of visualization called mind tools, computer-supported visualizations to engage learners in thinking critically [Jo96], are especially useful. Mind maps and concept maps are types of mind tools. Applications of computer-supported concept maps as an educational tool have been well-studied. Concept maps consist of graphical representations including concepts, represented by text in circles or boxes, and labelled relationships between these concepts, represented by a connecting line [NC08]. [Fa16] evaluated the understanding of key concepts in undergraduate students enrolled in foundational engineering courses. He found them to perform better when using computer-supported concept maps to illustrate key concepts. [Wa16] developed a personalized and interactive concept mapping environment and applied scaffolding to enhance studentsô understanding of key concepts. Thereby, he was able to target individual areas where a student needed improvement. [KB13] developed Termina, a point-based game-like tool to improve studentsô understanding of a topic while they create a concept map by correctly answering questions asked by the tool. Their user study shows that the tool assists students in learning and forming connections between the information. Therefore, concept mapping is being applied successfully to enhance learning and critical thinking. These ideas could also be transferred to non-experts seeking to familiarize with a research field.

Substantial amount of research has looked at supporting literature reviews and usage of visualizations from various angles. However, it still seems to be lacking in terms of guiding non-expertsôthought processes and decisions while performing a literature review and helping them develop their own strategies through that. This work attempts to bridge this gap.

3 Question-Guided Research Support

This work proposes a guidance system in the form of a software tool to fulfil the following objectives, which are based on problems faced by non-experts, as identified in section 1. Therefore, non-experts must be supported in:

- Identification of information needs during information acquisition,
- Structuring of the collected information to form an overview of it,

- Making their process time-efficient by avoiding need for trial-and-error.

The first objective can be fulfilled by presenting non-experts with guidance questions which they should answer to objectively evaluate the research field. The second objective can be fulfilled by using visualizations and borrowing features from mind tools. The third objective can be fulfilled by providing them with a structured solution, so that they know exactly how they want to proceed with the literature review.

3.1 Elaborated Research-Guiding Questions

The identification of the questions which should be presented to non-experts when they perform a literature review is critical in helping them make informed research decisions. These questions were identified based on literature [Kn11], [Eg05], [RC07], [IK11] and independent user research. For this purpose, a survey in the form of an online questionnaire was conducted. It aimed to identify additional guidance questions based on user needs and to verify the relevance of those questions already identified through literature. The questionnaire received 60 participants, each holding at least a high school degree. The participants were divided into 10 experts and 50 non-experts, where participants involved in research work were categorized as experts. The identified questions were divided into four major categories: *researchers*, *institutions*, *topics* and *publishings*. They were chosen because questions related to them are regarded as highly relevant by experts but are not given the same importance by non-experts, thereby revealing a mismatch in their thought processes. This is an important discovery because the purpose is to help non-expert researchers think about important criteria that they may not think about otherwise. Therefore, it is necessary to understand how experts perform these processes and replicate them for non-experts. Additionally, all the questions identified through literature were chosen as they all received an average relevance score of at least 3, the average of the possible scores. Within the category *researchers*, the important questions relate to the *names* of prominent researchers, their *affiliations*, key *publications*, *research interests* and their *interrelationships*. Within *institutions*, questions about *names* of research institutions involved and also their major *research projects* are included. Under *topics*, questions concerning names of *sub research topics* in the research field and the current *trends* within them are contained. Finally, within *publishings*, questions regarding names of prominent *conferences* and *journals* are included.

3.2 Research Community Map Creation and Reflection Tool

The developed tool had to focus primarily on helping users learn field-related concepts, so that they can make informed decisions and develop individual research strategies. The more specific idea was to present novice researchers with the identified guidance questions and allow them to choose which questions to answer based on their individual needs. Finally, their answers were visualized in a pre-existing visual template that depicts the interrelations present in the data in a manner that allows users to obtain an overview of the research field.

The web-based tool consists of a smaller left panel containing the guidance questions and a larger right panel containing a visual map. To arrive at these panels, the user must first provide her research topic, e.g., *recommender systems*. The questions page consists of the four main categories outlined in section 3.1: *researchers*, *institutions*, *topics* and *publishings*, and they all contain subcategories. The visual map, as seen in fig. 1, acts as a template that depicts the structure of the research field and the relationships between its various components. If a user would like to explore a category, e.g. *researchers*, clicking on it will lead her to the various researcher-related subcategories and questions. Here she can add their names, details, i.e. affiliations, publications, research interest and also associations between those researchers who share specialties, influence or contradict each other, collaborate with each other or have external reasons to agree or disagree. As fig. 2 shows, the provided information is added to both pages simultaneously. Alternatively, she could add the detail directly through the map. Most of the toolôs functions can be performed through either of the pages. If she adds associations between the added researchers, the respective relationship shows up as an icon above individual researcher names, clicking on which highlights the recipients of the relationship. The user can repeat this process in any order for whichever categories she finds relevant. She can also add custom subcategories to further individualize her map and match her own mental model.

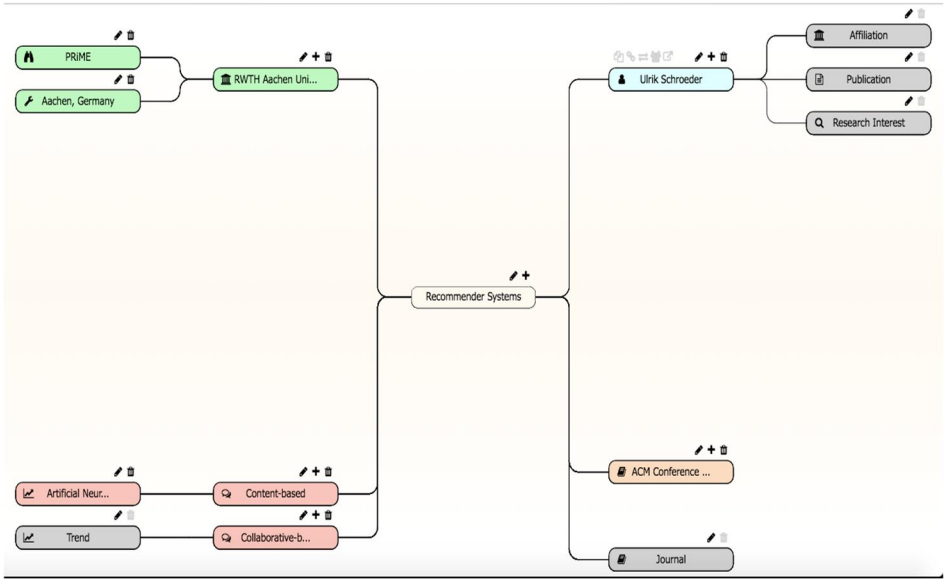

Fig. 1: A partially-filled layout of the visual map. Unfilled parts of the template are gray in color

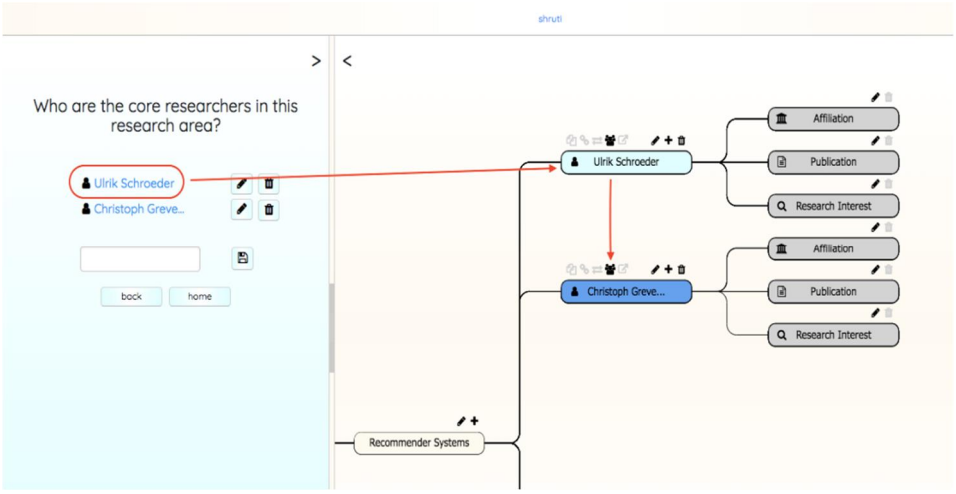

Fig. 2: The researcherôs name is added simultaneously on both sides. Additionally, it can be seen on the map that the researchers share collaborations

The tool provides guidance tips to the users to proffer help on how to proceed. However, the goal is to merely assist them while allowing them the independence to make their own decisions. The aim was to keep the design intuitive and the tool usable so that users can solely focus on their research process. Therefore, rigorous usability testing was conducted throughout the development process to improve the tool based on user feedback. Tasks considered tedious by users, such as typing long names, have been replaced with auto-complete, to allow easy input of information. To aid them while gathering information, all the entered information is clickable and takes users to relevant pages, such as institution websites or conference proceedings. Such features have been added through integration with open-source databases called dblp computer science bibliography[3] and Global Research Identifier Database[4] (GRID).

The tool provides users with a structured approach, where they are guided on which information to look for and how this information is connected. The validity of this approach has been evaluated with target users, as discussed in the following section.

[3] DBLP is a computer science bibliography being constantly updated by the University of Trier. Names of published authors, their publications, names of conferences and journals and their proceedings are being indexed in this project. Link: http://dblp.uni-trier.de/

[4] GRID is a project by Digital Science & Research Solutions Ltd. that is cataloging the worldôs research organizations, along with their webpages when possible, and also seeking to eliminate ambiguities present on the world wide web concerning the names of these institutions. Link: https://grid.ac/

4 User Study

A user study was conducted with target users to validate the approach implemented by the tool. The main aim of this study was to verify whether the objectives outlined in the previous section were achieved.

4.1 Procedure

A between-groups design was adopted to avoid learning effects. The participants were divided into a treatment group that would use the developed tool and search engines and a control group that would only use search engines and was not explicitly offered additional aid. The task was to perform a literature review on a predefined computer science-related topic within a time-frame of 30 minutes. The treatment group received 10 extra minutes initially to familiarize themselves with the tool. A list of the literature they found and the time at which they found it was maintained to perform quantitative analysis. Additionally, participants were informed that an understanding of the research field would also be expected as a part of the output of their literature review. The literature review was followed by a semi-structured interview regarding their experience and gained knowledge. A total of 18 B.Sc. Computer Science students with no knowledge of the predefined topic participated and 9 participants were randomly assigned to each group.

4.2 Hypotheses

The following hypotheses were tested and the subsequent results were obtained through the user study.

1. Hypothesis I: The mean of papers found by the treatment group is higher than the mean of papers found by the control group.
2. Hypothesis II: The control group has a higher mean in the percentage of papers found by unique authors than the treatment group.
3. Hypothesis III: The mean of papers by unique authors found by the treatment group does not equal the mean of papers by unique authors found by the control group.
4. Hypothesis IV: The mean of add-on[5] papers found by the treatment group is higher than the mean of add-on papers found by the control group.
5. Hypothesis V: The difference in means of the timings in which a matched participant found his n-th paper is bigger than 0 (with n being the minimum of papers found among the matched pairing).

[5] Add-on papers are those papers which belong to an author, institution or publishing, i.e. a conference or journal, by whom a paper was already found.

To clarify, hypothesis II compares the relative percentages between the two groups whereas hypotheses III and IV compare only the absolute numbers of a single category, i.e. either papers by unique authors or add-on papers, between the two groups.

4.3 Quantitative Analysis: Results and Discussion

To verify hypotheses I to IV, unpaired two sample t-tests at confidence levels of 95 % were applied. Hypothesis I was validated because the mean of number of papers found by the treatment group was significantly higher than the control group. Hypotheses II to IV were examined to identify where the significant difference came from. In the case of authors, once a paper was counted, all other papers written by that author or any of his co-authors were eliminated from contention. It was observed that the treatment group found significantly more of these add-on papers whereas no significant difference could be found for papers by unique authors. This implies that add-on papers were responsible for the treatment group finding a higher number of papers. Similar results were observed for institutions and publishings. Therefore, it seems plausible that participants in the treatment group were guided by the tool to look deeper into specific authors, and thereby found a higher absolute amount of papers.

Hypothesis V was tested to answer whether the software tool makes the literature review process time-efficient. A one-sided paired sample t-test was performed at a confidence level of 95 %. This showed that the participants in the treatment group found the same amount of papers significantly faster in comparison to their control group counterparts.

In conclusion, the tool helps users find a significantly higher amount of literature in a time-efficient manner. The fact that the number of papers found by the treatment group is higher due to add-on papers indicates that the tool guided their thinking process. The tool suggested that they reflect on entities that they already found interesting. Furthermore, not only did the treatment group participants find a higher amount of literature, they did so faster than the control group. This suggests that utilizing a well-structured strategy, as provided by the tool, made the process time-efficient. However, the user evaluation could have further benefited from a higher number of participants to achieve even more statistically sound results.

4.4 Qualitative Observations and Feedback: Results and Discussion

Apart from quantitative analysis, qualitative observations were also made. During the literature review session, the research methods of both groups were observed. Both started by searching the predefined topic on Google and usually proceeded to read the Wikipedia page on the topic. However, thereafter their strategies seemed to diverge. While the treatment group focused on categories within the software tool and on answering the questions within it, the control group appeared to have no clear strategy.

Both groups were afterwards asked to provide the names of prominent researchers, institutions, research topics and conferences or journals within the research field, to test their knowledge and overview of it. These questions were chosen because they were identified as highly relevant by experts during the user research but were assigned less relevance, relatively, by non-experts. It was observed that participants in the control group were more likely to finish their session early because they believed they had already formed a good overview of the topic. However, when they were asked the aforementioned overview questions, the control group participants did not perform well. Interestingly, even though conferences and journals are a good source of research literature specific to a research field, only 44 % of the control group identified conferences or journals related to the research topic as opposed to 100 % of the participants in the treatment group. This lack of overview may have resulted in the control group finding a lower number of research papers compared to the treatment group. On being asked whether a visualization would have been helpful in better structuring information and being able to answer the overview questions, 8 out of the 9 participants answered "yes". Additionally, the treatment group provided positive feedback about the tool and rated it as 4 out of 5 on average, on a scale of 1 (least) to 5 (most), on its helpfulness for non-expert researchers. This clearly shows that non-expert researchers would appreciate being supported during their review process.

In a real life scenario, non-expert researchers who are usually students are under a time-crunch to perform a literature review and meet deadlines. However, the time-crunch is usually in a range larger than 30 minutes. Therefore, it is necessary to extrapolate the aforementioned results and their implications to real life situations. The control group participants lacked a clear strategy. Additionally, they ended their session early believing they had already formed a good overview, yet did not perform well in overview-related questions. This implies that they already did not make optimal usage of the 30 minutes provided to them. Hence, with more time at hand, it can be assumed that they would still not perform as well as the treatment group. This indicates that the results obtained in the user study for a time limit of 30 minutes can also be applied to real situations.

In summary, while non-expert researchers in general appear to lack a clear strategy while performing a literature review, as depicted by the control group, it appears that the adopted strategic approach supports them in replicating expert behavior. Furthermore, these results bode well for non-expert researchers lacking research expertise, since it appears that these results may also be applicable to real life scenarios.

5 Conclusion and Future Work

The purpose of this research work was to support non-expert researchers in their literature review process. While support for non-experts and their objectives, at large, exists in the current literature, no solution fulfils their exact needs and objectives. Non-experts need support in identifying the information they must acquire, structuring this information and performing the process time-efficiently. The presented software tool, a web-based

application developed to fulfill these objectives, adopts a structured approach by presenting users with guidance questions based on user research. Furthermore, it structures their collected information by means of a visualization to provide an overview. At the same time, the tool allows its users to make decisions based on their user needs and develop individualistic strategies. The goal was to help them learn the literature review process and avoid dependency on the tool in the long run.

A user study, accomplished between a treatment and control group, showed that the former not only found a higher number of papers than the latter but also gained a better overview of the research topic. Since the gained overview was judged based on questions that were deemed highly relevant by experts, it may be stated that the tool aids non-experts in replicating expert behavior. Additionally, the statistical analysis showed that the treatment group participants were able to find the same amount of research papers significantly faster than the control group. Therefore, the software tool helps non-experts perform a literature review in a time-efficient manner. Furthermore, these results seem to be applicable to real life scenarios as well. Hence, the software tool managed to achieve its objectives.

The software tool, however, can be improved. Based on results, users would have preferred a rather topic-centric map layout. This must be further examined. Additionally, the scalability of the map must be addressed. It also needs to be investigated whether the provided guidance helps non-experts develop strategies and become independent researchers. This work also only supports one aspect of the research process. Therefore, it can be further expanded in the future to support the various stages of research apart from reviews. Finally, by means of providing such a structured procedure, the developed approach could even support experts, thereby freeing their minds from making subjective decisions. Hence, they could concentrate on more diverse aspects of a knowledge domain.

References

[BCB03] Börner, K.; Chen, C.; Boyack, K. W.: Visualizing knowledge domains. In: Annual Review of Information Science and Technology, 37(1), pp. 179-255, 2003.

[Eg05] Eggers, S.; Huang, Z.; Chen, H.; Yan, L.; Larson, C.; Rashid, A.; Chau, M.; Lin, C.: Mapping medical informatics research. In: Chen, H., Fuller, S.S., Friedman, C., Hersh, W. (Eds.), Medical Informatics, pp. 36 - 62. Integrated Series in Information Systems, 8, Springer, Boston, MA, 2005.

[Fa16] Fang, N.: Using concept maps to illustrate the evolution of key concepts: Student learning experience in a foundational undergraduate engineering course. In: ASEEôs 123rd Annual Conference and Exposition, New Orleans, LA, 2016.

[Fu11] Fu, J. Y.; Zhang, X.; Zhao, Y. H.; Huang, M. H.; Chen, D. Z.: Bibliometric analysis of complementary and alternative medicine research over three decades. In: Scientometrics, 88(2), pp. 617-626, 2011.

[IK11] Ismail, M. A.; Kareem, S. A.: Identifying how novice researchers search, locate, choose and use web resources at the early stage of research. In: Malaysian Journal of Library & Information Science, 16(3), pp. 67-85, 2011.

[Jo96] Jonassen, D. H.: Computers in the classroom: Mindtools for critical thinking. Prentice-Hall, Inc, 1996.

[KB13] Kneissl, F.; Bry, F.: Fostering Concept Maps Awareness as a Means toLearning. In: Cloud and Green Computing (CGC), pp. 354-357. IEEE, 2013.

[Kn11] Knepper, R.: The shape of the teragrid: Analysis of teragrid users and projects as an affiliation network. In: TeraGrid Conference: Extreme Digital Discovery, Salt Lake City, Utah. ACM, 2011.

[MSL11] Mattsson, P.; Sundberg, C. J.; Laget, P.: Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. In: Scientometrics, 87(1), pp 99-105, 2011.

[NC08] Novak, J. D.; Cañas, A. J.: The theory underlying concept maps and how to construct and use them. 2008.

[Ra09] Randolph, J. J.: A guide to writing the dissertation literature review. In: Practical Assessment, Research & Evaluation, 14(13), pp. 1-13, 2009.

[RC07] Reid, E. F.; Chen, H.: Mapping the contemporary terrorism research domain. In: International Journal of Human-Computer Studies, 65(1), pp. 42-56, 2007.

[Se11] Sen, S.; Charlton, H.; Kerwin, R.; Lim, J.; Maus, B.; Miller, N.; Naminski, M.R.; Schneeman, A.; Tran, A.; Nunes, E.; Sparling, E. I.: Macademia: Semantic visualization of research interests. In: IUI ô11 Proceedings of the 16th International Conference on Intelligent User Interfaces, Palo Alto, CA, pp. 457-458. ACM, 2011.

[Te05] Tergan, S-O.: Digital concept maps for managing knowledge and information. In: Tergan, S.-O. & Keller, T. (Eds.), Knowledge and Information Visualization, Springer, Berlin, Heidelberg, pp. 185-204, 2005.

[Wa16] Wang, S.: Designing Interactive and Personalized Concept Mapping Learning Environments. Proceedings of the 9th International Conference on Educational Data Mining, pp. 678-680, 2016.

[WY96] Wade, R. C.; Yarbrough, D. B.: Portfolios: A tool for reflective thinking in teacher education?. In: Teaching & Teacher Education, 12(1), pp. 63-79, 1996.