

Proposal Master’s Thesis. “Seeking Research Software. A Qualitative Study of Humanities Scholars’ Information Seeking Behavior.”

Ronny Gey (M24), geyronny@hu-berlin.de

Introduction

Today, software is a central component of science. Throughout the entire research life cycle, researchers use software tools for data collection, transformation, analysis and presentation as well as for generating hypotheses, managing literature and writing scientific papers (Kethers et al 2017, Pan et al 2016, Wolski 2017). Software has changed the way we actually do science. The complexity of the analyses carried out by researchers has increased, as has the amount of data that researchers can process. Software supports the documentation of the research process and enables reproducibility (Dallmeier-Tiessen 2016, Waltemath & Wolkenhauer 2016) and accuracy of results.

Due to the increased importance of research software for research (Katz 2017) and the increase in the sheer number of software, it is all the more important for researchers to identify suitable software and select the one which best fits the research problem, the actual step in the research process, or the research data which has to be processed, and, in consequence, which satisfies the researchers information need. In addition to increased efforts, difficulties in seeking software can also endanger the scientific reproducibility of studies or even lead to multiple developments of software with similar or identical functions (Hucka and Graham 2018).

The information seeking behaviour of researchers is generally of great interest within the field of information behaviour (e.g. Ahmadianyazdi & Chandrashekara 2018, Barrett 2005, Campbell 2017, Catalano 2013, Ellis 1993, Hemminger et al. 2007, Korobili et al. 2011, Liyana & Noorhidawati 2017, Rupp-Serrano & Robbins 2013, Wang et al. 2008). However, seeking software is still rather challenging for researchers (Howison & Bullard 2016). In a recent study, Hucka & Graham (2018, p. 2) surveyed scientists and engineers from several fields to better understand their approaches and selection criteria for seeking software. They found out that “*finding software suitable for a given purpose remains surprisingly difficult*”. Even in the domain of software development, the main challenge for software reuse are difficulties in finding software artifacts as Bauer et al. (2014) revealed in a study on code reuse at Google. Grossmann et al. (2009) identified users unawareness of specific software tools. These results are all the more surprising because the participants in the cited studies come from a group with a greater affinity for software (software developers, engineers).

The lack of awareness of specific software tools among researchers has been addressed by several technical solutions. Code aggregators, specialised search engines, programming language package repositories, code and application repositories, research repositories (e.g. Zenodo or Figshare), and curated web lists and catalogues aid users in discovering software (Struck 2018). Standards and tools for citing software enable researchers to identify, cite and reuse software (e.g. Niemeyer et al. 2016, Smith et al. 2016, Soito & Hwang 2017). Research funding agencies and research organizations (e.g. Haupt et al. 2018, Katerow & Feulner 2018, Scheliga et al. 2019) adopt guidelines for the development and use of research software with the aim of increasing the reusability and quality of the software artifacts developed. In turn, the

technical solutions presented are also aimed more at a technically experienced audience, often even at software developers directly. For researchers with less experience in the use of software, e.g. from the humanities (Rimmer et al. 2006), seeking software remains a difficult undertaking.

The information-seeking behaviour of humanities scholars in general has been addressed widely (see Catalanós (2013) literature review for an overview). In his pioneering work on Grounded Theory in information-seeking, Ellis (1993) identified patterns of information-seeking of social sciences, sciences, and humanities scholars. In 2005, Barrett (2005) analysed information-seeking habits of graduate student researchers in the humanities. Korobili et al. (2011) examined factors influencing information-seeking behaviour at the philosophy faculties. In previous studies, however, the classic research process of humanities scholars has been examined, which is mainly based on literature research. Although the information-seeking in the humanities is also based on software tools, e.g. databases, web-based editions, search engines, or online journals (Barrett 2005, Rimmer 2006), the search for software itself is rarely discussed. One of these rare examples, however a non-humanities one, is Hepworths et al. (2017) study of journalism professors' information-seeking behaviour. While seeking new online tools, journalism professors rely on other journalism professors, followed closely by media-related foundations, media professionals, and conferences.

Problem and Research Questions

The aim of this work is to investigate the information-seeking behaviour of humanities scholars when seeking research software and to compare it with their information-seeking behaviour in regard to literature. Special emphasis will be placed on the respondents' recourse to their own research process and the knowledge and practice structures in their field.

RQ1: What differences and similarities exist in the information-seeking behaviour of humanities scholars between the seeking for literature and the seeking for software?

RQ2: Which factors explain the differences in the information-seeking behaviour?

The work is based on Wilson's (2000) understanding of information-seeking behaviour which he defines as: "*a purposive process of seeking information as a consequence of a need to satisfy some goals*". The empirical work will be guided by either Ellis' (1989), Kuhlthaus (1991), Wilsons (1997), or Niedźwiedzka (2003) information(-seeking) behaviour model. The decision depends on various factors, which have yet to be examined: (1) The effort of working with the model should correspond to the total amount of work available for the master's thesis. (2) A possible sharpening of the research question in the conceptual phase can shift the focus in favour of certain aspects that are represented to varying degrees in the models. (3) Interview data as well as (4) the choice of a certain analysis method (qualitative content analysis) could encourage the use of a particular model

Methods

I chose an exploratory study design (Rinsdorf 2013) where the personal realm of experience of each interviewee lies in the center of the analysis. Interviews are the main data gathering technique which are applied in a semi-structured manner, guided by interview guidelines, and implemented in a face-to-face manner (Bryman 2004) in German language. With the interviews I obtain emotions, thoughts, and intentions of the participants and discover their perspective of the social world (Patton 2002). The interview data will be analyzed with a qualitative content analysis to explore qualified hypotheses (Kohlbacher 2006, Krippendorff 2012, Mayring 2000). It enables the researcher to include textual information and to identify its properties systematically. I will make all data generated during the concept, survey, analysis, and writing phases publicly available on GitHub¹, as long as it meets research

¹https://github.com/geyslein/Masters_Thesis

ethics standards (e.g. interview audio records and unanonymized interview transcripts will be excluded).

Work Plan

The work plan is based on an estimated duration of 17 weeks for writing the master's thesis. As a social scientist I worked 8 years with qualitative data and its analysis.

Week	Task	Methods and Tools	Comments
00-02	conceptual phase, literature research	systematic literature review, discovery systems (Leipzig University Library, BASE,), reference management system (JabRef), notetaking (Markdown), LaTeX text editor (Texmaker)	... already in progress
00-02	scheduling interviewees	phone, email, notetaking	existing contacts to faculty members at Leipzig University
02-03	interview guidelines, consent forms	notetaking	
03-05	conducting and transcribing interviews	notetaking, recording device, transcription tool (easytranscript)	
04-08	data analysis	open coding (Corbin & Strauss 1990), focused coding (Emerson 2011), qualitative content analysis (Kohlbacher 2006, Krippendorff 2012, Mayring 2000), QDA tool (?)	
08-16	Writing	LaTeX text editor, reference management system	
16-17	Proofreading and Submission		

References

Ahmadianyazdi, F. & Chandrashekara, M., Research challenges and delayed gratification in information seeking behavior: a case study of research scholars, *Library Philosophy & Practice*, 2018.

Barrett, A., The Information-Seeking Habits of Graduate Student Researchers in the Humanities, *The Journal of Academic Librarianship*, 2005 , 31 , 324-331.

Bauer, V.; Eckhardt, J.; Hauptmann, B. & Klimek, M. An exploratory study on reuse at google,

Proceedings of the 1 st International Workshop on Software Engineering Research and Industrial Practices - SER&IPs 2014, 2014.

Bryman, A., Social research methods, Oxford Univ. Press, 2004.

Campbell, L., The Information-Seeking Habits of Architecture Faculty College & Research Libraries, American Library Association, 2017 , 78.

Catalano, A. Patterns of graduate students' information seeking behavior: a meta-synthesis of the literature, *Journal of Documentation*, 2013, 69, 243-274.

Corbin, J. & Strauss, A., Grounded theory research: Procedures, canons and evaluative criteria, *Zeitschrift für Soziologie*, 1990, 19, 418-427.

Dallmeier-Tiessen, S., Reproduzierbarkeit Und Open Science: Bestandteile Und Erste Erfahrungswerte - Mit Besonderem Augenmerk Auf Software, Zenodo, 2016.

Ellis, D., Modeling the Information-Seeking Patterns of Academic Researchers: A Grounded Theory Approach, *The Library Quarterly*, 1993, 63, 469-486.

Emerson, R. M.; Fretz, R. I. & Shaw, L. L., Writing ethnographic fieldnotes, University of Chicago Press, 2011.

Greifeneder, E.; Pontis, S.; Blandford, A.; Attalla, H.; Neal, D. & Schiebke, K. Researchers' attitudes towards the use of social networking sites, *Journal of Documentation*, 2018 , 74, 119-136

Grossman, T.; Fitzmaurice, G. & Attar, R., A survey of software learnability, Proceedings of the 27th international conference on human factors in computing systems - CHI 09, 2009.

Haupt, C.; Meinel, M. & Schlauch, T., The software engineering initiative of DLR: overcome the obstacles and develop sustainable software, 2018 IEEE/ACM 13 th International Workshop on Software Engineering for Science (SE4Science), 2018 , 16-19.

Hemminger, B. M.; Lu, D.; Vaughan, K. & Adams, S. J., Information seeking behavior of academic scientists, *Journal of the American Society for Information Science and Technology*, 2007, 58, 2205-2225.

Hepworth, K.; Mensing, D. & Yun, G. W., Journalism Professors' Information-Seeking Behaviors: Finding Online Tools for Teaching, *Journalism & Mass Communication Educator Publications*, 2017 , 73 , 255-270.

Howison, J. & Bullard, J., Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature, *Journal of the Association for Information Science and Technology*, Wiley, 2015 , 67 , 2137-2155.

Hucka, M. & Graham, M. J. Software search is not a science, even among scientists: A survey of how scientists and engineers find software, *Journal of Systems and Software*, 2018, 141, 171-191.

Ingwersen, P. & Järvelin, K., The Turn, Springer-Verlag, 2005.

Katerbow, M. & Feulner, G., Recommendations on the development, use and provision of Research Software, Deutsche Forschungsgemeinschaft, 2018.

Katz, D. S. Software in Research: Underappreciated and Underrewarded, Presented at eResearch Australia, 2017.

Kethers, S.; Treloar, A. & Wu, M., Building Tools to Facilitate Data Reuse, *International Journal of Digital Curation*, 2017 , 11 , 1-12.

Korobili, S.; Malliari, A. & Zapounidou, S., Factors that Influence Information-Seeking Behavior: The Case of Greek Graduate Students, *The Journal of Academic Librarianship*, Elsevier BV, 2011 , 37 , 155-165.

- Kohlbacher, F., The Use of Qualitative Content Analysis in Case Study Research Forum: Qualitative Social Research, 2006, 7, 23.
- S. Kriewel, "Unterstützung beim Finden und Durchführen von Suchstrategien in Digitalen Bibliotheken," Dissertation an der Universität Duisburg-Essen, Fachbereich Ingenieurwissenschaften, 2010.
- Krippendorff, K., Content Analysis: An Introduction to its Methodology, Sage Publications, 2012.
- Liyana, S. & Noorhidawati, A., How graduate students seek for information: Convenience or guaranteed result?, Malaysian Journal of Library & Information Science, 2017 , 19.
- Mayring, P., Qualitative Content Analysis, Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 2000 , 1 , 1-10.
- Murphy-Hill, E.; Lee, D. Y.; Murphy, G. C. & McGrenere, J., How Do Users Discover New Tools in Software Development and Beyond?, Computer Supported Cooperative Work (CSCW), 2015, 24, 389-422.
- Niedźwiedzka, B., A proposed general model of information behaviour Information research, 2003, 9, 9-1.
- Niemeyer, K. E.; Smith, A. M. & Katz, D. S., The Challenge and Promise of Software Citation for Credit, Identification, Discovery, and Reuse, Journal of Data and Information Quality, 2016, 7, 1-5.
- Pan, X.; Yan, E. & Hua, W., Disciplinary differences of software use and impact in scientific literature, Scientometrics, 2016 , 109 , 1593-1610.
- Patton, M. Q., Qualitative Research & Evaluation Methods, SAGE, 2002.
- Rimmer, J.; Warwick, C.; Blandford, A.; Gow, J. & Buchanan, G., Humanities scholars' information-seeking behaviour and use of digital resources, Digital Libraries in the Context of Users' Broader Activities, 2006, 19.
- Rinsdorf, L., Qualitative Methoden, In: Umlauf, K.; Fühles-Ubach, S. & Seadle, M. (Eds.), Handbuch Methoden der Bibliotheks-und Informationswissenschaft, DeGruyter Saur, 2013, 64-79.
- Rowlands, I. & Nicholas, D., Understanding Information Behaviour: How Do Students and Faculty Find Books?, The Journal of Academic Librarianship, Elsevier BV, 2008 , 34 , 3-15.
- Rupp-Serrano, K. & Robbins, S., Information-Seeking Habits of Education Faculty, College & Research Libraries, American Library Association, 2013, 74, 131-142.
- Scheliga, K.; Pampel, H.; Konrad, U.; Fritzsche, B.; Schlauch, T.; Nolden, M.; Zu Castell, W.; Finke, A.; Hammitzsch, M.; Bertuch, O. & Denker, M., Dealing with research software: Recommendations for best practices, Helmholtz Association, Helmholtz Open Science Coordination Office, 2019.
- Smith, A. M.; Katz, D. S. & Niemeyer, K. E., Software citation principles, PeerJ Computer Science, 2016, 2.
- Soito, L. & Hwang, L. J., Citations for Software: Providing Identification, Access and Recognition for Research Software, International Journal of Digital Curation, Edinburgh University Library, 2017, 11, 48-63.
- Struck, A., Research Software Discovery: Challenges, Risks And Opportunities, IEEE eScience 2018, 2018.
- Waltemath, D. & Wolkenhauer, O., How Modeling Standards, Software, and Initiatives Support Reproducibility in Systems Biology and Systems Medicine, IEEE Transactions on Biomedical Engineering, 2016, 63, 1999-2006.

Wang, P.; Dervos, D. A.; Zhang, Y. & Wu, L., Information-seeking behaviors of academic researchers in the internet age: A user study in the United States, China and Greece, *Proceedings of the American Society for Information Science and Technology*, 2008, 44, 1-29.

Wilson, T. Models in information behaviour research, *Journal of Documentation*, 1999, 55, 249-270.

Wilson, T. D., Human information behavior, *Informing science*, Informing Science Institute, 2000, 3, 49-56.

Wolski, M.; Howard, L. & Richardson, J., The importance of tools in the data lifecycle, *Digital Library Perspectives*, 2017, 33, 235-252.