

Comparing Methods for Finding Search Sessions on a Specified Topic: A Double Case Study

Tessel Bogaard¹[0000-0002-5049-9386], Aysenur Bilgin¹[0000-0002-6225-9953], Jan
Wielemaker^{1,2}[0000-0001-5574-5673], Laura Hollink¹[0000-0002-6865-0021], Kees
Ribbens^{3,4}[0000-0003-0266-5877], and Jacco van
Ossenbruggen^{1,2}[0000-0002-7748-4715]

¹ Centrum Wiskunde & Informatica, the Netherlands {`firstname.lastname`}@cwi.nl

² Vrije Universiteit Amsterdam, the Netherlands

³ NIOD Institute for War, Holocaust and Genocide Studies, the Netherlands

⁴ Erasmus School of History, Culture and Communication, the Netherlands

Abstract. Users searching for different topics in a collection may show distinct search patterns. To analyze search behavior of users searching for a specific topic, we need to retrieve the sessions containing this topic. In this paper, we compare different topic representations and approaches to find topic-specific sessions. We conduct our research in a double case study of two topics, World War II and feminism, using search logs of a historical newspaper collection. We evaluate the results using manually created ground truths of over 600 sessions per topic. The two case studies show similar results: The query-based methods yield high precision, at the expense of recall. The document-based methods find more sessions, at the expense of precision. In both approaches, precision improves significantly by manually curating the topic representations. This study demonstrates how different methods to find sessions containing specific topics can be applied by digital humanities scholars and practitioners.

Keywords: Digital libraries · User interests · Log analysis.

1 Introduction

Analysis of search logs is an unobtrusive technique for large-scale investigations into user behavior in digital libraries. Users interested in different topics might display different search behaviors. For example, the work presented in [34] demonstrated different search patterns of users searching for five major religions. In a previous study, we observed a distinct search pattern for users searching for documents related to World War II (WWII) [5]. For these types of studies, we need to be able to retrieve those user interactions from the search logs that relate to a user interest in a specified topic. In this paper, we propose and compare generally applicable methods to find user interactions that relate to a specified topic from a larger set of logged search interactions. We work at the level of sessions (coherent sequences of user interactions with the collection) as they capture the context in which individual user actions occurred and connect search actions to clicks on documents. We address two research questions:

(RQ1) How can we represent a specified topic?

(RQ2) How can we use the topic representation to retrieve relevant sessions?

To answer the first research question, we look into different, consecutive ways to build a term list as a representation of a topic: i) using semantic relations in an explicit knowledge resource, ii) applying local word embeddings trained on the documents in the collection, and iii) in each step, by manual curation of the term lists by domain experts. To answer the second research question, we look into matching the different term lists to user sessions. We match them to either a) the user queries, or b) the contents of the clicked documents. We compare and discuss the combined methods in terms of number of retrieved sessions as well as estimated precision scores. We conduct our research using data from the National Library of the Netherlands, focusing on search in their historical newspaper collection⁵. In previous work [4,5], the search logs of the digital library were already split into user sessions, and we consider this session identification step outside the scope of this paper. We present a double case study in the context of two historical topics with societal relevance: WWII (a pivotal period in Dutch and global history), and feminism (a movement that has had and still has an impact on Dutch society). We evaluate our methods on a ground truth of over 600 manually assessed sessions per topic.

This study contributes insights into how different topic representations and matching approaches perform when retrieving topic-specific sessions. Our results show that when sessions are retrieved based solely on user queries, the precision is high, however, the set of sessions remains small. When the document-based matching approach is used, the set of sessions retrieved increases, but at the expense of precision. Moreover, we find that by manually curating the term lists we improve precision while still preserving a larger set of sessions. The two topics investigated in this paper show similar general patterns in their results, however, we observe a higher overall precision for the more popular topic (WWII). Finally, our study demonstrates how different methods can be applied and combined by digital humanities scholars and practitioners to retrieve topic-specific sessions.

2 Related Work

We discuss work on detection and analysis of user interests; and how knowledge resources and word embeddings have been used to enrich queries and documents.

2.1 Topic-specific Search Log Analysis

Search behavior in digital libraries and archives has been studied frequently, e.g., [7,10,18,21,29,33]. Topics have been detected in search logs for various reasons; for example, to determine user interests [16,19,23,26], to uncover topic-specific search patterns [4,34], or to recognize changes in topic within a session [17].

⁵ The National Library of the Netherlands has granted us access to user logs from their search platform <https://www.delpher.nl>, providing access to collections from the National Library of the Netherlands and other heritage institutions

Other studies observe topic-specific search patterns by analyzing logs from a specific search interface, such as a health portal [8], or a media archive [19,20].

In most cases, topics are detected in search logs by investigating the queries that users entered. Sometimes, in addition to the query, the contents of what was clicked in sessions is also taken into account. For example, query analysis has been combined with mouse-fixation behavior and the metadata of clicked documents [16]. In previous work, we used the metadata of clicked documents, as well as the use of facets to filter search results, to understand search behavior in different parts of a digital library collection [4,5]. In this study, we investigate and compare how query-based and content-based approaches perform.

We represent a topic as a list of terms. This is similar to the work presented in [34], where users searching for five large religions were identified by matching queries to five respective lists of professionally curated terms. The authors of [13] used a list of terms and phrases that signify specific types of questions, and matched these to queries in order to analyze how people learn within sessions.

2.2 External Resources to Enrich Queries or Documents

In previous work, knowledge resources have been used to classify documents in collections, e.g., by finding relevant Wikipedia categories [35]; or by finding relevant concepts [24] for the documents in the collection. In other cases, knowledge resources have provided a semantic enrichment of user queries, e.g., to categorize queries [20,26,36]; or for query expansion during search, e.g., by searching related concepts in Wikipedia [1,2,15]. In the present study, we use Wikipedia as a knowledge resource to expand a single term topic representation. Wikipedia is widely used, publicly available and has a broad coverage, making it applicable to many use cases beyond the ones studied in this paper. This makes Wikipedia an attractive option, even though we are aware of the fact that Wikipedia is biased both with respect to which topics are represented in the articles and the contents of the articles [9,30].

Word embeddings have been used by researchers in several query expansion applications, such as search, text classification, plagiarism detection [3]. In this type of distributed representation, words with similar meanings are more likely to be close together [28]. The semantic associations between words that thus emerge, have been shown to be effective in tackling the query-document vocabulary mismatch problem [14]. We use word embeddings to expand on the terms representing a topic, and as such to be able to increase the number of sessions found. Specifically, we use local embeddings, following [12], where it was demonstrated that corpus-specific embeddings perform better than global embeddings for query expansion.

3 Data

We use collection and log data from the National Library of the Netherlands. As a knowledge resource for the topic representations we use Wikipedia.

3.1 Document Collection and Search Logs

Our research is conducted using a document collection and search logs from the National Library of the Netherlands. The library maintains a number of digitized historic collections, our focus is on the historical newspaper collection spanning almost four hundred years (1618-1995). Within this collection, users can search using full-text search and facets (filters based on the metadata attributes of the collection). The logs used in this study were collected between October 2015 and March 2016 (raw data 200M records). They record the user interactions with the search system. These interactions have previously been grouped in sessions, to be able to study search behavior in context. The log records have been cleaned and processed, and sessions have been identified based on a *clickstream* model as described in [4,5], using the IP address as identifier and connecting sequential HTTP requests to follow a user navigating the search system. For this study, we have retained all sessions which include clicked documents within the newspaper collection, resulting in a total of 204,266 sessions over the six month period. In addition, we received the full text digitization and metadata records of the historical newspaper collection (103M documents at the time).

3.2 Knowledge Sources for Topic Representations

In this double case study, the topics of interest are WWII (“Tweede Wereldoorlog” in Dutch), and feminism (“feminisme” in Dutch). These topics are selected based on their societal relevance, and thus their value to digital humanities scholars. For example, professional historians from the Dutch NIOD Institute for War, Holocaust and Genocide Studies are interested in understanding how people search for topics related to World War II (WWII) in the media, and how this changes over time. We represent the two topics using lists of relevant terms. In the first expansion of the list of relevant terms, we use Wikipedia. As this is a publicly available knowledge resource, with many possible applications in different domains for different topics, it contributes to the general applicability of our methods. Our topics of interest correspond to the existing Wikipedia categories for WWII⁶ and for feminism⁷. We have selected the top-300 Wikipedia articles in these categories, based on the popularity within the same period as the logs (October 2015–March 2016). To collect these Wikipedia articles, we have used the tool Massview Analysis⁸, including the subcategories. The top-300 most popular Wikipedia articles within the WWII category counted to of a total of 4.7 million views, compared to 1.2 million views within the feminism category. The assessed Wikipedia articles for the topics are available online⁹. We use the popularity ranking as an indicator for public interest in the topics described in

⁶ https://nl.wikipedia.org/wiki/Categorie:Tweede_Wereldoorlog

⁷ <https://nl.wikipedia.org/wiki/Categorie:Feminisme>

⁸ <https://pageviews.toolforge.org/massviews/>, by MusikAnimal, Kaldari, and Marcel Ruiz Fornas.

⁹ <https://edu.nl/4arxw> and <https://edu.nl/9qbfr>

the articles as the use of Wikipedia is a strong indicator for how this interest is composed in a country such as the Netherlands.

4 Method

We describe the different methods we compare to find topic-specific sessions. First, we explain the consecutive steps to build term lists representing the topics. Second, we describe how to use the term lists to find topic-specific sessions in a larger set of sessions. Third, we explain how we evaluate the different methods.

4.1 Creating Term Lists

We compare five ways of creating terms lists to represent the topics, where each list builds on the previous list.

List 1. Single term: List 1 contains a single term or phrase to represent the topic, in our case “Tweede Wereldoorlog” (WWII) or “feminisme” (feminism).

List 2. Wikipedia: For this list, we leverage the semantic relations in Wikipedia to find additional terms to represent the topic. First, we match the term in List 1 to their corresponding Wikipedia category and add them to List 2. Then, we take the article titles of pages within that category or any of its sub-categories. To increase the likelihood that these article titles are indeed relevant terms, we select only the top-300 most popular titles based on Wikipedia page view data. Some Wikipedia article titles require preprocessing. Where the title only consists of a named entity, it is used as-is. In the case of a title consisting of a named entity and a class between parentheses, for example, “The Color Purple (film)”, we separate the class from the named entity. In the case of a title consisting of a classifying noun, preposition, named entity title phrase, for example “Bombardement op Rotterdam” (Bombing of Rotterdam), we leave out the preposition when it is not part of a named entity.

List 3. Wikipedia curated: For List 3, we ask domain experts to manually assess the terms in List 2 and remove those that are less relevant, in the assumption that this will improve the quality of the terms on the list and thus improve the precision of the matched sessions. For the WWII terms, experts from the NIOD Institute for War, Holocaust and Genocide Studies were involved in the assessment; for the feminism terms, two of the authors of this paper familiar with the topic. The assessment is based on the question whether it is plausible that someone with a specific interest in WWII or in feminism would consult the subject described in the corresponding Wikipedia article. Articles in which our main topic of interest (WWII or feminism) is only of minor importance – for example, in biographies of politicians, actors and professional sportsmen for whom the WWII period was not pivotal in their lives – were removed from the lists. Similarly, articles referring to a topic occurring outside the time period of the historical newspaper collection (1618-1995) – for example, movies or books published after 1995 – were also removed. We note that in the case of the WWII topics, most of these are topics from the war period itself or from

the period leading to the war, but also included are issues that are part of the post-war remembrance culture and therefore refer to the period after WWII.

List 4. Wikipedia expanded: We expand the terms in List 3 using local word embeddings to create the larger List 4. We describe this process in detail in Section 4.2.

List 5. Wikipedia expanded and curated: To create List 5, we ask domain experts to assess the terms in List 4, using the same process as for List 3.

This results in five term lists for our topics (see Table 1).

Table 1. Number of terms in each term list

	1: single	2: wiki	3: wiki curated	4: wiki expanded	5: wiki exp&cur
WWII	1	300	200	728	364
Feminism	1	300	199	703	327

4.2 Term Expansion Based on Local Word Embeddings

To expand the term lists, we employ a widely used technique based on word embeddings [27], vector representations of words where words that appear close together in the vector space are likely to have a similar meaning. We use local embeddings instead of global embeddings, training on a selected set of topically relevant documents, as we expect term similarity to be highly dependent on the context of the topic, as was shown in [12]. For this purpose, we query the library’s newspaper collection for documents that contain the terms in List 3, and use those as a topically-constrained training corpus. We work with the Indri search engine [32], using default Dirichlet smoothing [31]. The terms are translated to Indri queries, searching for an exact phrase match, or in the case of a title and a class description an exact phrase match and a Boolean AND for the class. We use the gensim library¹⁰ for both preprocessing and to train the embeddings. To preprocess the digitized text in the training corpus, we first identify the combination of symbols and characters that mark the beginning and end of each article, and remove them. Next, we extract the sentences to be broken down into tokens, and lowercase the text. For the configuration of the hyper-parameters of gensim’s word embedding algorithm, we refer to the *set expansion* solution proposed by [25] where the authors suggest setting the word vector size to 100 and the window size to 10¹¹. The reason to use a window size as large as 10, is the empirical evidence that larger window sizes are good at providing more topical similarity [22]. Since we are interested in identifying phrases that can be made up of multiple words (e.g., “Nationaal-Socialistische Beweging”, “Tweede Wereldoorlog”), we instruct the model to learn bigrams and trigrams (phrases

¹⁰ <https://radimrehurek.com/gensim/>

¹¹ https://github.com/NervanaSystems/nlp-architect/tree/master/nlp_architect/solutions/set_expansion

that contain two and three words). With these settings the model is expected to find associations for the single or multi-word target phrase, and suggest related words (made up of phrases consisting of one or more words). Once the model is trained, we query it using the terms in List 3 as seeds. We retain the top-3 most similar words for each term, and add them as expanded terms to List 4.

4.3 Matching Terms to Sessions

We match the terms of the five lists to sessions in two ways: matching the terms to (a) the user queries and to (b) the clicked documents.

In the **query-based** approach, user queries in the sessions are compared to the terms in the lists using exact phrase matching. As there is little context in a user query, we only include the named entity and not any information included in brackets (such as a class or publication year for the terms based on the Wikipedia article titles). Sessions are considered relevant to a topic if they contain at least one query that contains words matching a term from the topical term list.

In the **document-based** approach, we leverage the contents of the documents clicked in the sessions. For the matching of a term with the content of clicked documents, we include – where present – the class or the noun in the set of terms. This results in for example, the terms "*A Bridge Too Far*" *AND film*, or *Bombing AND Rotterdam*. For the WWII matching we include an extra step: we remove all matched clicked documents published before 1920, as WWII is a topic based on a historical period, and any documents from before 1920 are considered not relevant. Thus, we retrieve all sessions in which at least one matching document has been clicked.

4.4 Manually Evaluating Retrieved Sessions

The different methods provide us with sets of sessions for each of the five term lists based on either the query matching, and the document matching, with a total of ten sets of sessions for each topic. To estimate the precision of the resulting ten sets of sessions for each topic, human raters assess samples drawn from these sets. The raters judge whether one of the information needs of the user in that session is to find newspaper documents about a topic that is directly related to the topic of interest. To do this, the rater can inspect the session, using a visualization that includes the search interactions with the queries and selected facets [6], and the clicked documents and their metadata and content. We use inter-rater reliability to check the agreement among the raters.

5 Results

We apply the ten methods described in Section 4 – five ways to represent a topic as a term list, combined with two approaches to match the terms to a session – to the full set of sessions. This results in ten retrieved sets of sessions per topic.

To estimate precision, we draw samples from each set and manually assess a total of 1243 sessions. We compute the inter-rater agreement using Cohen’s κ [11] based on a dual assessment of about 50 sessions per topic. We observe a κ of 0.90 for WWII and 0.84 for feminism, demonstrating good agreement.

Table 2. Size (count) and precision (percentage) for the retrieved session sets

topic representation	query-based matching				document-based matching			
	WWII	Feminism	WWII	Feminism	WWII	Feminism	WWII	Feminism
1: single	89	26	100%	100%	116	37	95%	100%
2: wiki	667	702	83%	44%	10,001	8,260	66%	37%
3: wiki curated	471	222	100%	100%	7,434	5,341	81%	56%
4: wiki expanded	4,977	6,064	52%	20%	103,833	126,656	22%	9%
5: wiki exp&cur	626	339	100%	97%	12,044	10,519	74%	63%

Size and Precision Table 2 shows the number of sessions and the precision of each set. As expected, the use of a single term to represent a topic (List 1) results in almost perfect precision but a low number of retrieved sessions. Precision remains high (97% to 100%) when using the longer, curated lists (Lists 3 and 5) in a query-based matching. This method increases the number of retrieved sessions significantly (5 to 13 times as many, in our case). When these lists (3 and 5) are used for document-based matching, the number of retrieved documents increases even more; however, precision is lower. On the WWII topic, precision of this method may still be acceptable (74% to 81%) but on feminism it is probably not (56% and 63%). For the expanded term lists in their un-curated form (Lists 2 and 4), precision drops depending on matching method and topic. When List 2 is used for query-based matching, precision is 83% on the WWII topic, which may still be acceptable. For document-based matching, and/or when applied to feminism, precision will be too low for most applications (37% to 66%). List 4 results in low precision in all cases (9% to 52%).

Note that Lists 4 and 5 were created by expanding List 3. In theory, the same local embedding-based expansion method could be applied to Lists 1 and 2. However, in practice, this is not promising, as List 1 consists of a single term and List 2 has relatively low precision. For that reason, expansions of List 1 and 2 were not included in our experiment.

Combining Two Matching Methods Table 3 shows the number of sessions that appear in both the query-based and document-based session sets, i.e., the intersection of the two sets. For List 1, the intersection is relatively small: e.g., for WWII, only 23 of the 89 sessions retrieved with the query based method are also in the document-based set. We conclude that when using a single term topic representation, a combination of query-based and document based matching is a good way to increase the number of retrieved sessions. For List 2, 3, 4 and 5, on the other hand, the intersection is relatively large; the majority of

Table 3. Intersection of query- and document-based session sets

topic representation	WWII			Feminism		
	query only	both	doc only	query only	both	doc only
1: single	66	23	93	9	17	20
2: wiki	142	525	9,476	386	316	7,944
3: wiki curated	116	355	7,079	39	183	5,158
4: wiki expanded	107	4,870	98,963	390	5,674	120,982
5: wiki exp&cur	114	512	11,532	45	294	10,225

the sessions retrieved with the query-based methods are also retrieved with the document-based methods. Combining the two methods is less worthwhile here.

Error Analysis Our manual annotation effort gave us insight into the types of errors that occur. Some sessions were incorrectly retrieved because of terms in the term lists that are not unique to the two topics, WWII and feminism. For example “concentration camps” may also occur in documents about the Indonesian Independence War; “emancipation” may occur in documents about slavery or religion; the term “gas chamber” is now almost uniquely associated with WWII, but had a different meaning historically; Anne Frank’s last name is common in the Netherlands and appears in sessions of family historians unrelated to Anne Frank, and in many documents throughout the collection that are not related to WWII. This type of error happens with all methods but is more frequent when using the document-based matching. We hypothesize that users act as a “smart filter”, as they are less likely to use generic or ambiguous query terms without adding meaningful modifying terms. A future direction of research could be to investigate if only selecting terms that users used in their queries might increase precision for a document-based matching.

Another cause for errors in the document-based matching is brought on by mistakes in the digitization process, such as incorrectly set document boundaries, or when the newspaper document contains multiple topics, such as articles summarizing local news or presenting a cultural calendar.

6 Lessons Learned

In general, query-based matching results in higher precision than document-based matching. Document-based matching, on the other hand, results in more retrieved sessions (up to 20 times more, in our experiments) at the loss of precision. We have experimented with a more narrow inclusion of document-based matched sessions (e.g., matching more than one document in a session), but a preliminary inspection did not seem to increase precision. Future work, though, could investigate this further. A combination of query- and document-based matching is useful when a topic is represented as a single term. In this case, the combination retrieves significantly more sessions without losing precision. We hypothesize that the combination is similarly worthwhile when topics are represented as relatively short terms lists.

When a topic representation is expanded to a longer term list, manual curation of the terms is key. This holds for both our expansion methods (using a knowledge resource or local word embeddings). Curation is especially critical for document-based matching. In this work, we leveraged the category structure of Wikipedia. Future work will have to determine how other knowledge resources and other semantic relations perform.

All our methods perform better on the WWII topic than on the feminism topic. This could in part be due to the fact that WWII is a less abstract topic than feminism and as such may be easier to detect. Even so, we hypothesize the prevalence of the topic in our data plays a large part as well: WWII is not only the more popular topic on Wikipedia, the retrieved session sets (both query- and document-based) are larger than the respective sets containing topics related to feminism. It would be interesting to investigate this further using topics at different abstraction and popularity levels. In general, we expect that both knowledge-based and corpus-based expansion methods work better on more popular topics.

7 Conclusion

Understanding search behavior for topics with societal relevance can provide digital humanities scholars insights into the interest in these topics within a collection, and the research presented in this paper supports this objective. We compared different methods on how to retrieve user sessions containing specified topics, using different term lists to represent the topics and applying term matching to user queries and to clicked documents. We observed that when retrieving sessions is based solely on user queries, the precision is high, but the number of sessions retrieved small. Using the document-based matching approach, more sessions are retrieved, at the expense of precision. We found that manual curation is essential, without this step the expanded lists (using a knowledge resource or local word embeddings) perform poorly in terms of precision. This effect was particularly strong for the document-based matching. Furthermore, we observed a higher overall precision for the more popular, WWII topic. In conclusion, we believe this research helps to pave the way for a better understanding and communication of topic-specific user interests within collections for digital humanities scholars as well as collection owners and practitioners.

Acknowledgements We would like to thank the National Library of the Netherlands, and Lynda Hardman (Centrum Wiskunde & Informatica) for their support. The Wikipedia articles related to WWII were assessed by Kees Ribbens with the assistance of Caroline Schoofs and Koen Smilde. This research is partially supported by the VRE4EIC project, a project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 676247.

References

1. Aggarwal, N., Buitelaar, P.: Query Expansion Using Wikipedia and Dbpedia. In: Forner, P., Karlgren, J., Womser-Hacker, C. (eds.) CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012. CEUR Workshop Proceedings, vol. 1178. CEUR-WS.org (2012), <http://ceur-ws.org/Vol-1178/CLEF2012wn-CHiC-AggarwalEt2012.pdf>
2. ALMasri, M., Berrut, C., Chevallet, J.P.: Wikipedia-based Semantic Query Enrichment. In: Proceedings of the Sixth International Workshop on Exploiting Semantic Annotations in Information Retrieval. pp. 5–8. ESAIR '13, ACM, New York, NY, USA (2013). <https://doi.org/10.1145/2513204.2513209>, <http://doi.acm.org/10.1145/2513204.2513209>
3. Azad, H.K., Deepak, A.: Query Expansion Techniques for Information Retrieval: A Survey. *Information Processing and Management* (2019). <https://doi.org/10.1016/j.ipm.2019.05.009>
4. Bogaard, T., Hollink, L., Wielemaker, J., Hardman, L., van Ossenbruggen, J.: Searching for Old News: User Interests and Behavior Within a National Collection. In: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval. pp. 113–121. CHIIR '19, ACM, New York, NY, USA (2019). <https://doi.org/10.1145/3295750.3298925>
5. Bogaard, T., Hollink, L., Wielemaker, J., van Ossenbruggen, J., Hardman, L.: Metadata Categorization for Identifying Search Patterns in a Digital Library. *Journal of Documentation* **75**(2), 270–286 (2019). <https://doi.org/10.1108/JD-06-2018-0087>
6. Bogaard, T., Wielemaker, J., Hollink, L., Hardman, L., van Ossenbruggen, J.: Understanding User Behavior in Digital Libraries Using the MAGUS Session Visualization Tool. In: Hall, M.M., Mercun, T., Risse, T., Duchateau, F. (eds.) *Digital Libraries for Open Knowledge - 24th International Conference on Theory and Practice of Digital Libraries, TPD L 2020, Lyon, France, August 25-27, 2020, Proceedings. Lecture Notes in Computer Science*, vol. 12246, pp. 171–184. Springer (2020). https://doi.org/10.1007/978-3-030-54956-5_13, https://doi.org/10.1007/978-3-030-54956-5_13
7. Borgman, C.L., Smart, L.J., Millwood, K.A., Finley, J.R., Champeny, L., Gilliland, A.J., Leazer, G.H.: Comparing Faculty Information Seeking in Teaching and Research: Implications for the Design of Digital Libraries. *Journal of the American Society for Information Science and Technology* **56**(6), 636–657 (2005). <https://doi.org/10.1002/asi.20154>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20154>
8. Callahan, A., Pernek, I., Stiglic, G., Leskovec, J., Strasberg, H.R., Shah, N.H.: Analyzing Information Seeking and Drug-Safety Alert Response by Health Care Professionals as New Methods for Surveillance. *Journal of medical Internet research* **17**(8), e204 (1 2015). <https://doi.org/10.2196/jmir.4427>, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4642796&tool=pmcentrez&rendertype=abstract>
9. Callahan, E.S., Herring, S.C.: Cultural bias in Wikipedia Content on Famous Persons. *Journal of the American Society for Information Science and Technology* **62**(10), 1899–1915 (2011). <https://doi.org/10.1002/asi.21577>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21577>
10. Clough, P., Hill, T., Paramita, M.L., Goodale, P.: Europeana: What Users Search for and Why. In: Kamps, J., Tsakonias, G., Manolopoulos, Y., Iliadis, L., Karydis,

- I. (eds.) *Research and Advanced Technology for Digital Libraries*. pp. 207–219. Springer International Publishing, Cham (2017)
11. Cohen, J.: A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **20**(1), 37–46 (1960). <https://doi.org/10.1177/001316446002000104>, <https://doi.org/10.1177/001316446002000104>
 12. Diaz, F., Mitra, B., Craswell, N.: Query Expansion with Locally-Trained Word Embeddings. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pp. 367–377. Association for Computational Linguistics, Berlin, Germany (8 2016). <https://doi.org/10.18653/v1/P16-1035>, <https://www.aclweb.org/anthology/P16-1035>
 13. Eickhoff, C., Teevan, J., White, R., Dumais, S.: Lessons from the Journey: A Query Log Analysis of Within-session Learning. In: *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*. pp. 223–232. WSDM '14, ACM, New York, NY, USA (2014). <https://doi.org/10.1145/2556195.2556217>, <http://doi.acm.org/10.1145/2556195.2556217>
 14. Fernández-Reyes, F.C., Hermosillo-Valadez, J., Montes-y Gómez, M.: A Prospect-Guided Global Query Expansion Strategy Using Word Embeddings. *Information Processing and Management* (2018). <https://doi.org/10.1016/j.ipm.2017.09.001>
 15. Guisado-Gómez, J., Prat-Pérez, A., Larriba-Pey, J.L.: Query Expansion via Structural Motifs in Wikipedia Graph. *CoRR* **abs/1602.07217** (2 2016), <http://arxiv.org/abs/1602.07217>
 16. Hienert, D., Kern, D.: Term-Mouse-Fixations As an Additional Indicator for Topical User Interests in Domain-Specific Search. In: *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. pp. 249–252. ICTIR '17, ACM, New York, NY, USA (2017). <https://doi.org/10.1145/3121050.3121088>, <http://doi.acm.org/10.1145/3121050.3121088>
 17. Hienert, D., Kern, D.: Recognizing Topic Change in Search Sessions of Digital Libraries Based on Thesaurus and Classification System. In: Bonn, M., Wu, D., Downie, J.S., Martaus, A. (eds.) *19th ACM/IEEE Joint Conference on Digital Libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019*. pp. 297–300. IEEE (2019). <https://doi.org/10.1109/JCDL.2019.00049>, <https://doi.org/10.1109/JCDL.2019.00049>
 18. Hollink, L., Mika, P., Blanco, R.: Web Usage Mining with Semantic Analysis. In: *Proceedings of the 22Nd International Conference on World Wide Web*. pp. 561–570. WWW '13, ACM, New York, NY, USA (2013). <https://doi.org/10.1145/2488388.2488438>, <http://doi.acm.org/10.1145/2488388.2488438>
 19. Hollink, V., Tsikrika, T., Vries, A.P.d.: Semantic Search Log Analysis: A Method and a Study on Professional Image Search. *Journal of the American Society for Information Science and Technology* **62**(4), 691–713 (6 2011). <https://doi.org/10.1002/asi.21484>, <http://dx.doi.org/10.1002/asi.21484>
 20. Huurnink, B., Hollink, L., Van Heuvel, W.D., De Rijke, M.: Search Behavior of Media Professionals at an Audiovisual Archive: A Transaction Log Analysis. *Journal of the American Society for Information Science and Technology* (2010). <https://doi.org/10.1002/asi.21327>
 21. Kules, B., Capra, R.: Designing Exploratory Search Tasks for User Studies of Information Seeking Support Systems. In: *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*. pp. 419–420. JCDL '09, ACM, New

- York, NY, USA (2009). <https://doi.org/10.1145/1555400.1555492>, <http://doi.acm.org/10.1145/1555400.1555492>
22. Levy, O., Goldberg, Y.: Dependency-Based Word Embeddings. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 302–308. Association for Computational Linguistics, Baltimore, Maryland (6 2014). <https://doi.org/10.3115/v1/P14-2050>
 23. Liu, H., Kešelj, V.: Combined Mining of Web Server Logs and Web Contents for Classifying User Navigation Patterns and Predicting Users’ Future Requests. *Data and Knowledge Engineering* **61**(2), 304–330 (2007). <https://doi.org/10.1016/j.datak.2006.06.001>
 24. Malo, P., Sinha, A., Wallenius, J., Korhonen, P.: Concept-based Document Classification Using Wikipedia and Value Function. *Journal of the American Society for Information Science and Technology* **62**(12), 2496–2511 (12 2011). <https://doi.org/10.1002/asi.21596>
 25. Mamou, J., Pereg, O., Wasserblat, M., Dagan, I., Goldberg, Y., Eirew, A., Green, Y., Guskin, S., Izsak, P., Korat, D.: Term Set Expansion based on Multi-Context Term Embeddings: an End-to-end Workflow. arXiv preprint arXiv:1807.10104 (2018)
 26. Meij, E., Bron, M., Hollink, L., Huurnink, B., de Rijke, M.: Mapping Queries to the Linking Open Data cloud: A Case Study Using DBpedia. *Web Semantics: Science, Services and Agents on the World Wide Web* **9**(4), 418 – 433 (2011). <https://doi.org/http://dx.doi.org/10.1016/j.websem.2011.04.001>, <http://www.sciencedirect.com/science/article/pii/S1570826811000187>
 27. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. In: 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings (1 2013), <http://arxiv.org/abs/1301.3781>
 28. Nematzadeh, A., Meylan, S.C., Griffiths, T.L.: Evaluating Vector-Space Models of Word Representation, or, The Unreasonable Effectiveness of Counting Words Near Other Words. In: *CogSci* (2017)
 29. Niu, X., Hemminger, B.M.: Analyzing the Interaction Patterns in a Faceted Search Interface. *J. Assoc. Inf. Sci. Technol.* **66**(5), 1030–1047 (2015). <https://doi.org/10.1002/asi.23227>, <https://doi.org/10.1002/asi.23227>
 30. Shane Greenstein, B., Zhu, F.: Is Wikipedia Biased? *American Economic Review* **102**(3), 343–48 (2012). <https://doi.org/10.1257/aer.102.3.343>
 31. Smucker, M.D., Allan, J.: An Investigation of Dirichlet Prior Smoothing’s Performance Advantage. Tech. rep., The University of Massachusetts, The Center for Intelligent Information Retrieval (2006)
 32. Strohman, T., Metzler, D., Turtle, H., Croft, W.B.: Indri: A Language Model-based Search Engine for Complex Queries (extended version). CIIR Technical Report (2005)
 33. Walsh, D., Clough, P.D., Hall, M.M., Hopfgartner, F., Foster, J., Kontonatsios, G.: Analysis of Transaction Logs from National Museums Liverpool. In: Doucet, A., Isaac, A., Golub, K., Aalberg, T., Jatowt, A. (eds.) *Digital Libraries for Open Knowledge. Lecture Notes in Computer Science*, vol. 11799, pp. 84–98. Springer (2019). https://doi.org/10.1007/978-3-030-30760-8_7
 34. Wan-Chik, R., Clough, P., Sanderson, M.: Investigating Religious Information Searching Through Analysis of a Search Engine Log. *Journal of the American Society for Information Science and Technology* **64**(12), 2492–2506 (2013). <https://doi.org/10.1002/asi.22945>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22945>

35. Weale, T.: Utilizing Wikipedia Categories for Document Classification (2006)
36. Xu, Y., Jones, G.J.F., Wang, B.: Query Dependent Pseudo-relevance Feedback Based on Wikipedia. In: Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 59–66. SIGIR '09, ACM, New York, NY, USA (2009). <https://doi.org/10.1145/1571941.1571954>, <http://doi.acm.org/10.1145/1571941.1571954>