

# Validating the importance of work tasks as context for professional search

Thomas Schoegje<sup>1,2</sup>  
tmschoegje@gmail.com

Egon L. van den Broek<sup>1</sup>  
vandenbroek@acm.org

Toine Pieters<sup>2</sup>  
t.pieters@uu.nl

<sup>1</sup>Institute of Information and Computing Sciences, Utrecht University

<sup>2</sup>Freudenthal Institute, Utrecht University

## Abstract

In professional search many work tasks share some structure and are likely to recur. We argue that retrieval in the context of such tasks can exploit prior knowledge about these tasks to serve more relevant results. Understanding why someone searches can help to distill and express the user's information desires. In order to validate the value of this approach, we asked users to judge search results with and without a new search filter that narrowed down the results to documents relevant to the work tasks. Initial results suggest the importance of exploiting work tasks as context for professional search, although future work should consider the extent of this effect.

## 1 Introduction

As users seek to fulfill different tasks, they have different information desires guiding their search interactions. In working environments these work tasks are usually different from typical web search, resulting in different desires. Professional search settings can require a high recall of document (paragraphs) over precision, for example [4][5]. Because these tasks may be more structured and more likely to recur, we consider using the user's work tasks as the context for their query. In this way we aim to get a better representation of their information desires. We will consider work tasks as discrete work-related activities, such as handling email or creating a document. Search tasks are a subset of work tasks, where the user aims to find information to support other work tasks. In this paper we report our attempts to identifying work tasks that result in typical search tasks, and tailoring search results to the underlying information desires.

We will consider the user's tasks as the primary source of a user's information desires, which in turn shape the user's search interactions. Although the details of individual work tasks vary, we expect that many professions have generic, characteristic and recurring work tasks. A basic example of the influence of work tasks on information desires is how users can use homonyms to describe different concepts depending on their work domain. Work tasks can lead to typical information desires which are difficult to express in a textual query. Instead we will consider how these desires can be expressed in a way that supplements the textual query. An alternative to textual queries can be as simple as a filter in the search interface, where the users can filter on from their work domain. This approach helps aligning the user's desires with the retrieval engine's target function (i.e. its estimation function of document relevant-ness). This will ideally let the user avoid having to compromise their information need [3]. This is also more directly measuring their intention than implicitly asking them to write a query, which is implicitly the case [5]. This paper will test the following two hypotheses: that work tasks will

---

*Copyright © by the paper's authors. Copying permitted for private and academic purposes.*

In: Joint Proceedings of the First International Workshop on Professional Search (ProfS2018); the Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR); and the International Workshop on Data Search (DATA:SEARCH18). Co-located with SIGIR 2018, Ann Arbor, Michigan, USA – 12 July 2018, published at <http://ceur-ws.org>

often shape typical information desires, and that enabling users to express these desires more accurately will better support them in their tasks.

Section 2 will open by reviewing (authentic) work tasks as concepts. Afterwards, we introduce the work tasks investigated in their context at the municipality Utrecht in Section 3. In Section 4 we describe an experiment to quantify the benefits of filtering documents based on their originating work tasks. Finally we present our conclusions in Section 5 and expand on our plans for future work in Section 6.

## 2 Authentic work tasks

Work tasks are defined as concrete sections of time that include actions towards a goal; the task outcome (e.g. handling email)[6]. A work task may include multiple search tasks, which are sub-tasks that include one or more queries towards an information need. Although the work tasks will rarely be identical, many of them will share characteristic aspects. This is especially the case for professional search, as work tasks are more recurring and structurally defined than the information desires in typical web search.

Some work tasks are shared between users within a team or between users with similar roles in different teams. Others, such as reviewing recent information, are very common. The importance of one such ubiquitous work task, reviewing recent documents, has been quantified [2]. Here it was shown that the number of times any document was accessed in their professional setting followed a logarithmic function.

The actual search process itself also affects the user’s information desires[1][7]. The context of one’s stage in the information seeking process could also be used to refine results.

## 3 Investigating work tasks

The work tasks investigated are performed within the municipality Utrecht in the Netherlands. This is an organization where diverse teams look up information with diverse goals during their work tasks. This diversity is ideal for investigating the types and nature of work tasks, as well as their effects on user interactions with the information systems. Our experiment aims to show that work tasks can often shape typical information desires, and that embedding searches with this context information will better support them in their tasks. First, we will identify typical information desires through exploratory interviews with a focus on the user work tasks and the difficulties they have in completing these. We then provide them with a filter that lets them express their work task context in a search interface. This filter is based on meta-data that can be algorithmically annotated (i.e. the documents can be classified and filtered by class). In order to validate the perceived impact of the new interface we perform an experiment including a dummy filter that does not function as intended but instead randomly filters out documents.

### 3.1 Analysis of work tasks

In our experiment we aim to support policy makers from diverse domains. Although their specific approach may vary per domain and individual, they experienced similar challenges in retrieving information. The first key challenge was retrieving specific documents that were stored without descriptive titles in a semi-structured archive. In this case users tended to resort to systems other than the intended interface to find the documents (in particular by checking email attachments and by asking colleagues). The second key challenge is sorting through a large volume of documents retrieved unrelated to the intended domain. The first challenge was addressed by allowing users to easily access recently viewed documents (as suggested in the literature [2]). The current focus will be on the second challenge; allowing users to more accurately express their information desires in a way that reduces the number of unrelated results.

Based on work tasks, two promising document categorizations were identified through explorative interviews. The first was to consider the information needs at the various (global) stages of policy making. Here, different teams continue building on the same information as the documents produced get less explorative and increasingly specific. The second was to focus on the types of communication between these teams, as a policy maker is required to produce or search for specific types of documents. Whereas the first case groups the work tasks within a team, the second case groups on the communication tasks that policy makers from multiple teams might encounter. These two categorizations will be discussed in order.

The first categorization considers the global policy making process. There are three main steps involving different teams of policy makers: gathering information, forming policy proposals and adapting it. The forming of policy proposals is further separated by domain, with two significant ones shown in Table 1.

Table 1: Global steps in policy making.

Step	Purpose
Information gathering	Exploring within given direction, informing
Debating (domain: city and space)	Forming policy proposals
Debating (domain: man and society)	Forming policy proposals
Adapting	Finalizing policy, guiding direction of future policy

Table 2: Types of policy documents sent between teams.

Name	Purpose	Details
Letter	Informing	
Policy notice	Informing of planned policy	Accompanied by letter
Memo	Technical information	Typically small updates
Proposal	New policy proposal	
Reply	Answer to a question	

The second categorization is based on the policy documents communicated between teams, which can generally be divided into five categories based on their purpose. They are the products of mutually exclusive work tasks, and their purpose is summarized in Table 2.

Our initial solution to reduce the number of irrelevant search results is to allow users to filter on such a categorization based on meta-data. This annotation was approached manually for new documents, where a new interface was introduced to aid employees in selecting one of five appropriate templates (corresponding to these classes). Although this ensures accurate annotation of new documents, users also want to search older documents. This can be approached by exploiting the file location and document title where possible, and using classification for the remaining cases. The algorithmic classification is out of the scope of this paper. This avoids inaccuracies due to classification errors for the remainder of the paper. An experiment will now follow to test the value in this work-task based approach, where classification was verified manually.

## 4 Empirical study

An empirical study was set up to verify whether filters in the search interface help the user to better express the information desire by expressing the context of the work tasks. Users were asked to judge how well a set of results fulfilled a specified information desired, both with and without the filter. In order to test the presence of a placebo effect on the new filter (which was created as they hypothesized it would improve performance), we also introduce a second placebo filter which filters on a random subtype.

### 4.1 Materials

Based on the categorization of the policy making process (see Table 1), there are 4 document types. For each of these 2 search queries were formulated by a user familiar with the system, resulting in 8 queries. They were chosen to represent authentic search tasks. The actual query used is hidden, and the user is instead presented a search question that represents the underlying information desire. Although such verbose question yield poor results when used as the query[5], these authentic search tasks reflect examples where users were interesting in filtering on the categories. Using each of the 8 queries, results were retrieved under the following three conditions:

1. TextSearch: a full-text search using the queries.
2. FilterSearch: the same search, but filtering out all results of the incorrect document types.
3. PlaceboSearch: appears identical to FilterSearch, but instead of filtering on the desired (and indicated) document type it filters on another document type.

This document categorization was chosen over the alternative previously introduced as the placebo filter would be more obvious (the user would recognize that a memo is not actually a letter). The current PlaceboSearch displays documents that were written for a different purpose or domain, but on the same topic.

# Wat zijn de doelstellingen om het verwerken van afval te verbeteren en recycling te promoten?

Filter: off

- Besluit**  
- 28-06-2018  
MV 1 Afval Canal Pride.docx  
Mondelinge vragen **Afval** Canal Pride Eva van Esch, Partij voor de Dieren Op 16 juni jl. vond het feestelijke Canal Pride plaats, maar helaas was het zondag niet zo'n feest voor het milieu.  
Door het late opruimen is veel **afval**, waaronder veel plastic, ballonnen en plastic confetti, van de kade in het water gewaaid.  
Daardoor is het **afval** verder verspreid, met alle gevolgen voor het milieu en dieren van dien.
- Informerend**  
- 28-06-2018  
Raadsbrief Werkgelegenheid van de toekomst.pdf  
'**Afval**' is een grondstof en energie komt van duurzame bronnen.
- Debatresultaat**  
- 21-06-2018  
Coalitieakkoord Utrecht ruimte voor iedereen.pdf  
We stimuleren het hergebruik van grondstoffen in de bouwopgave en verbeteren de gescheiden inzameling van **afval**.
  - We zetten volop in op circulariteit bij het sluiten van (nieuwe) contracten voor de verwerking van **afval** en grondstoffenAls we inwoners willen stimuleren **afval** gescheiden in te zamelen is het belangrijk dat mensen tevreden zijn over de serviceverlening van de gemeente op dit gebied.

Figure 1: The static user interface presented where the user can highlight results. Off-screen are radio buttons for an overall evaluation.

## 4.2 Procedure

5 female and 4 male participants were asked to judge the results for queries. They were presented with a static search interface such as the one shown in Figure 1. Document type is indicated using the color and title. Users were asked to highlight any relevant results by clicking on them (similar to previous work in gathering subjective opinions[8]). Then they were asked to indicate an overall rating of the results on a (Likert) scale of 1 - 5. They could do so at their own tempo, before proceeding to the next set of results. In order to avoid the possibility of users forgetting to answer one of these questions, users were asked to confirm when they found no useful results. The time it took to answer each question (in seconds) was also stored, and the experiment took around 30 minutes. After two practice screens, a total of 24 sets of results were presented. The order was randomized by shuffling in which order the filter types were presented, and then presenting the 8 questions per filter type in a random order per user. This was done so the user could get accustomed to the different filters. The time taken per answer is recorded.

## 4.3 Results

Table 3 shows basic descriptors of how users evaluated the various search engines. Having the proper filter outperform the others in every aspect encourages further study, although further statistical analysis is required. An ANOVA was performed using the mean Likert scores for each set of results shown (8 questions x 3 filters), with the search engine as the dependent variable. The result was not strongly statistically significant ( $F(2,27) = 3.453$ ,  $p = .0505$ ), although the effect size was fairly large ( $\eta^2 = 0.247$ ). In combination with the confidence intervals, this suggests that the  $p$  value would decrease below .05 if more participants were added to the study.

The discrepancy where the TextSearch filter has a large number of results selected but a relatively Likert rating is likely because this version tended to show the same file preview multiple times (from different sources). The reduced decision time might be a related to presenting that are obviously relevant, but also influenced by the TextSearch filter including more documents with lengthy previews.

## 5 Conclusion

We noted the important role of the user's (work) tasks could play in helping users. Users perform queries within the context of a work task, and this context can be used to better understand their information needs. This is especially the case for professional search as many work tasks have a more structured and recurring nature than is often the case in web search. Results suggest the importance of work tasks as the context for professional search, although future work should consider the extent of this effect.

Table 3: Descriptors for user evaluations on the various search engines. Average values are shown, as well as the 95% confidence interval for the overall Likert evaluation of the results. Every participant evaluated every set of question- and results.

Engine	Likert rating	Likert CI95	Number Selected	Decision time (seconds)
TextSearch	2.20	(2.20, 2.79)	2.23	71.26
PlaceboSearch	2.48	(2.18, 2.78)	1.95	61.76
FilterSearch	3.06	(2.76, 3.35)	3.63	60.45

## Acknowledgements

The authors gratefully acknowledge the participants for their time and Arjen de Vries for his comments on the present work.

## References

- [1] Clarkson, K.L.: Supporting the complex dynamics of the information seeking process. PhD thesis, Radboud University, Nijmegen, Netherlands (2018)
- [2] Dumais, S., Cutrell, E., Cadiz, J.J., Jancke, G., Sarin, R., Robbins, D.C.: Stuff I’ve seen: a system for personal information retrieval and re-use. In: ACM SIGIR Forum. Volume 49., ACM (2016) 28–35
- [3] Hoenkamp, E.C.: About the ‘compromised information need’ and optimal interaction as quality measure for search interfaces. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM (2015) 835–838
- [4] Kim, Y., Seo, J., Croft, W.B.: Automatic boolean query suggestion for professional search. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, ACM (2011) 825–834
- [5] Koster, C.H., Seibert, O., Seutter, M.: The phasar search engine. In: International Conference on Application of Natural Language to Information Systems, Springer (2006) 141–152
- [6] Saastamoinen, M., Järvelin, K.: Queries in authentic work tasks: the effects of task type and complexity. *Journal of Documentation* **72**(6) (2016) 1114–1133
- [7] Spink, A.: Multiple search sessions model of end-user behavior: An exploratory study. *Journal of the American Society for Information Science* **47**(8) (1996) 603–609
- [8] Van Den Broek, E.L., Kisters, P.M., Vuurpijl, L.G.: Content-based image retrieval benchmarking: Utilizing color categories and color distributions. *Journal of imaging science and technology* **49**(3) (2005) 293–301