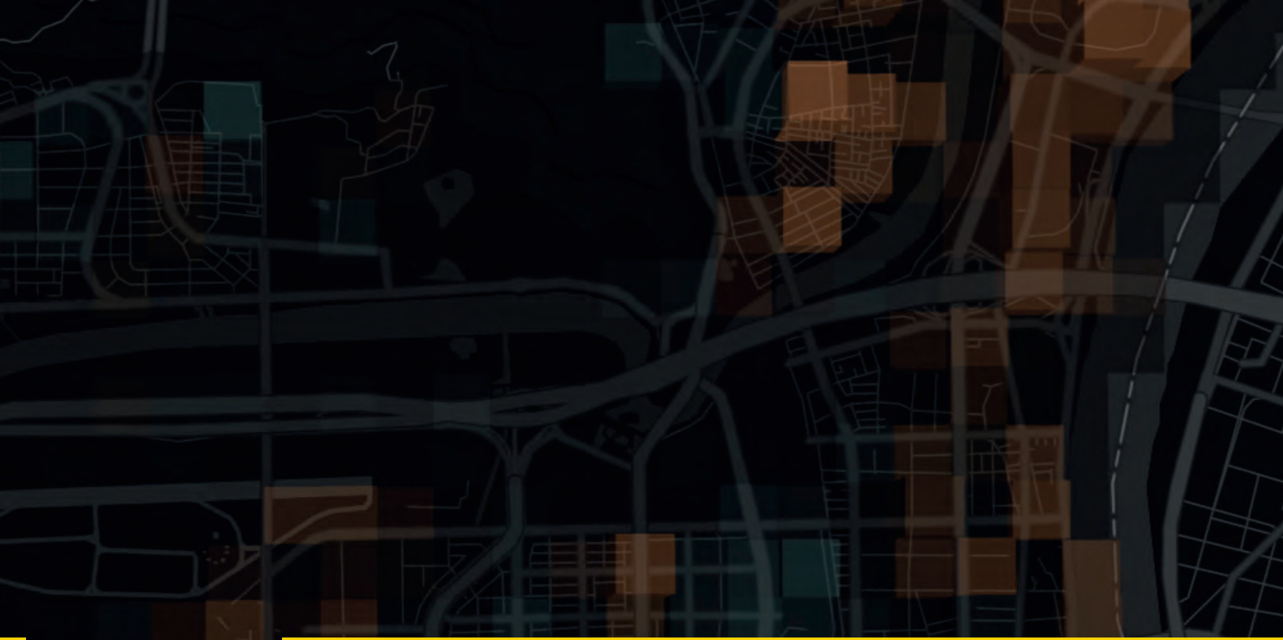
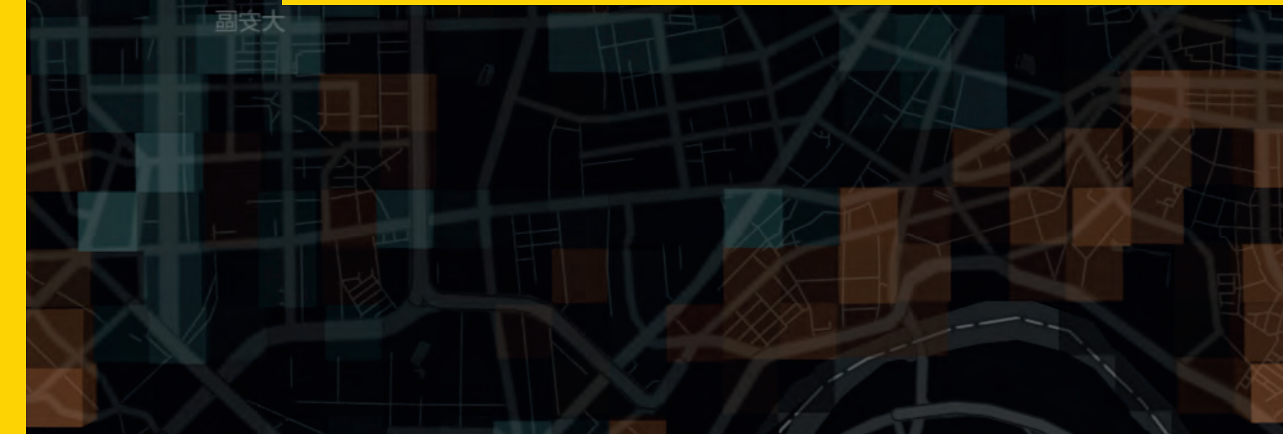


Task-based search for municipal knowledge workers

Designing and implementing specialised search functionality

THOMAS SCHOEGJE



TASK-BASED SEARCH FOR MUNICIPAL KNOWLEDGE WORKERS

Designing and implementing specialised search functionality

Thomas Schoegje

Task-based search for municipal knowledge workers - Designing and implementing specialised search functionality; Thomas Schoegje – Utrecht: Freudenthal Institute, Faculty of Science, Utrecht University / FI Scientific Library (formerly published as CD-β Scientific Library), no. 123, 2024.

Dissertation Utrecht University. With references. Met een samenvatting in het Nederlands.

ISBN: 978-90-73346-81-9
Cover design: Thomas Schoegje
Printed by: Canon

TASK-BASED SEARCH FOR MUNICIPAL KNOWLEDGE WORKERS

Designing and implementing
specialised search functionality

TAAK-GEBASEERD ZOEKEN VOOR GEMEENTELIJKE KENNISWERKERS

Het ontwerpen en implementeren van
gespecialiseerde zoekfunctionaliteit

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 3 december 2024 des ochtends te 10.15 uur

door

Thomas Schoegje

geboren op 30 december 1990
te Zaanstad

Promotoren:
Prof. dr. Toine Pieters
Prof. dr. ir. Arjen P. de Vries
Prof. dr. Lynda Hardman

Beoordelingscommissie:
Prof. dr. G.T. Barkema
Prof. dr. K. Bystrom
Prof. dr. J.T. Jeuring
Prof. dr. A.J. Meijer
Dr. M.T. Schaefer

Dit proefschrift werd mogelijk gemaakt met financiële steun van de Gemeente Utrecht.

Table of contents

Chapter 1 Introduction

1. Introduction

In this thesis we investigate how organisations with overwhelming amounts of information can support the diverse information needs of users by developing internal search engines tailored to the users' tasks. The empirical work focuses on developing task-based search systems to meet the information needs of the municipality of Utrecht.

This setting allowed us to compare general-purpose search engines to search engines designed to address the varied needs of different municipal knowledge workers, such as council members and policy workers.

Our research examines how to characterise and develop support for the information needs of user groups with complex and (domain-)specific tasks. We first introduce why the previous search systems at the municipality were insufficient. These general-purpose search engines were developed when the municipality and its information storages grew (section 1.1), but offered little support for the specific information needs of council members and policy workers. We designed search functionality to address the needs of specific user groups in the organisation (section 1.2). We considered how to identify and characterise the user tasks that require information (section 1.3) and how to design support for these tasks (section 1.4). This leads us to our research questions (section 1.5), thesis contributions (section 1.6), and thesis outline (section 1.7).

1.1 One search for all

The municipality of Utrecht has over 5000 employees[1], including knowledge workers, public administrators, and municipal workers who increasingly work with information. In the early days of the municipality, there was no need for information systems. As council members, policy workers and others worked with a growing amount of information, they started housing information in domain-specific storage systems (see Figure 1 phase 1). These systems allowed users to search for information by retrieving documents relevant to their search terms. By 2018, the municipality maintained numerous isolated information systems, including a content-management system, a system for archiving documents, a system for searching for internal colleagues, and a system for planning council meetings and archiving council information. In this fragmented information infrastructure, it became challenging to know what information existed in the organisation, or even where to start searching. Organisations typically solve this by combining all information in a single generic search system [47] (see Figure 1 phase 2).

Pooling all documents in one system introduces new challenges however, as different user groups engage with information for diverse purposes. For example, users may require different answers to the same search input. A

1 https://www.utrecht.nl/bestuur-en-organisatie/werken-bij-de-gemeente/personeel-in-cijfers/, accessed 05-12-2023

council member searching for the 'Uithoflijn' might be learning about a sensitive political issue around a long-delayed tram line. Meanwhile, an employee who lives at the Uithof might search for the 'Uithoflijn' to find out when the tram line will finish construction. Satisfying the needs of different user groups within the same system requires a means to characterise how their needs differ [47], and appropriate functionality to support those needs (see Figure 1 phase 3). The goal of this thesis is to investigate how we can specify the requirements for targeted search applications, by developing such search engines for different user groups at the municipality of Utrecht.

Figure 1. A search maturity model illustrating how organisations typically develop their search applications over time. Typically, separate search applications are developed for different domains at first (phase 1: searching in silos). Eventually, organisations tend to seek a central access point for their information (phase 2: one search for all). However, the requirements of different user groups can differ significantly, necessitating tailored entrances towards the data (phase 3: targeted search applications). It remains unclear how the requirements of these targeted search applications can be specified. The model and accompanying image were created by Findwise based on their survey data [47].

1.2 Targeted search applications

As different user groups have distinct requirements, there is a benefit to tailoring the search engine to these requirements. This is reflected in the search engine industry, as vendors shifted from enterprise search engines towards insight engines. Whereas enterprise search engines are typically marketed around searching through all of a company's data sources (see Figure 1 phase 2), insight engines are positioned to leverage artificial intelligence to personalise the search experience towards the needs of different users and user groups (see Figure 1 phase 3).

The primary way that systems personalise results is using Artificial Intelligence (AI) techniques such as learning to rank which search results are relevant given a search context [100]. Increasingly, large language models are also offering the option to generate information that is relevant to a textual prompt. Search engines have made significant advancements in retrieving (and generating) relevant information, particularly in domains with ample training data such as web search. However, it remains unclear if similar degrees of personalisation can be achieved through learning to rank within small to medium organisations, which possess substantially smaller datasets. There is limited scientific research measuring whether such personalising systems yield significant improvements in work performance [169]. Moreover, generic search functionality and generic AI techniques often fall short for users who have complex and (domain-)specific information needs, which necessitate specialised functionalities. For instance, a council member seeking to comprehend a council information dossier would ideally have a visualisation of all pertinent information from an overview perspective, rather than a set of relevant documents.

There is a need to establish a method for specifying the requirements of user groups, facilitating the design of targeted search applications. To establish such a method, we first consider scenarios where users may require specialised support, and subsequently consider the nature of this support.

1.3 Characterising search tasks

The end goal of users is typically not merely to find information, but rather to utilise that information for some specific purpose. We start from the premise that the best search engine is the one that is most useful for this broader task that the searcher is trying to perform [14]. For example, a council member's task might involve understanding how the previous city council decided to construct the tram line 'the Uithoflijn', which requires them to find information about the decision-making process.

Understanding the users' tasks allows us to discern when specialised search functionality is necessary. While generic search functionality may suffice for simple search tasks, it may not be enough to support complex tasks. (Municipal) knowledge workers in particular may require support, as they perform intellectual tasks which are creative, relatively unstructured and build upon the organisation's knowledge in unexpected ways [133]. In their roles as domain specialists and professional searchers, municipal knowledge workers face complex [128] and specific [165] search tasks.

Specialised support for a task can reduce its complexity. In systems lacking adequate support, users may find it unclear how to perform tasks in that system, or users may be forced to perform sub-tasks (e.g. finding precise keywords). Both of these factors increase task complexity [94, 26, 23], and can lead to the search process being slower [83], more uncertain and more frustrating [87]. Consequently, users may abandon their search for information [138], or make decisions based on incomplete information [17]. By designing specialised search functionalities, we aim to mitigate the overall task complexity, thereby facilitating a more informed (and hopefully better) decision-making process. Complex and (domain-)specific tasks in particular frequently necessitate specialised functionalities. However, it is unclear how to characterise those tasks for the purpose of designing supporting functionality.

In this thesis, we identify and characterise the search tasks of municipal knowledge workers. This allows us contextualise and support their tasks. We develop a task characterisation for the purpose of designing specialised search functionality. Previous work has predominantly focused on domain-agnostic task characterisations, such as Li and Belkin's faceted task framework [94], alongside others (e.g. [25, 130, 125]). These frameworks are useful to investigate fundamental research questions on, for example, how tasks affect information interaction (e.g. [95]). However, these models divorce tasks from their specific domain and therefore they do not provide details on the information that is sought in any particular use-case. To enable the development of targeted search applications, we adapt their faceted task framework [94] to encompass domain-specific task requirements.

Once we can characterise user tasks, we need a way to design specialised search functionality around those tasks.

1.4 Specialised search functionality

Arguably, current search engines tend to support a small set of basic tasks, leaving much of the work of complex tasks to the user [32]. Enabling systems to perform more complex tasks requires a better understanding of the users' tasks and domains.

Search functionality can be specialised in at least two dimensions: it can be domain-specific (or domain-agnostic) search, and task-specific (or general-purpose). As a result, we can distinguish between four types of search systems:

Domain-agnostic general-purpose functionality supports basic search tasks. Such generic search engines are designed to work across different domains. Examples of such generic search engines include most out-of-the-box solutions, such as Elastic search [116].

Domain-agnostic task-specific functionality supports search tasks that occur across multiple domains. For example, assisting council members in (re) formulating their search query. This type of system can provide situational and adaptive support. For example, adapting query suggestions [6] and the ranking of search results [137, 101] to the task. The search interface itself can also be adaptive and change which functionalities are presented during different task stages [70]. In addition to adapting to information-needs, task-based tools can allow users to organise information within a search task more effectively [32, 42]. An example of meta-task support is recommending tasks to users [53], which may also help them reduce the complexity of tasks by clarifying their constituent sub-tasks [30]. For the municipality, an example of generic task functionality is when the search engine offers query suggestions to help users disambiguate what aspect of a dossier they are interested in (e.g. ’the Uithoflijn audit’ or ’the Uithoflijn completion’).

Domain-specific general-purpose functionality supports information needs that are common and/or characteristic to a domain, without specifying the exact tasks being supported. For example, a domain-specific search engine that searches council information.

Such systems provide ways to view and interact with information to satisfy needs that are unique to a domain. Specialised interactions with information include navigating [39] or filtering [90] information based on domain-specific concepts. Interfaces can present tailored perspectives, such as displaying relationships between domain-specific concepts [80] or presenting search results in domain-specific perspectives [112]. An example for the municipality is iBabs. iBabs is a meeting planning system for council information which includes a search functionality to filter search results by the policy area and responsible alderperson.

Domain-specific task-specific functionality supports domain-specific tasks. For example, helping council members to understand the decision-making process around a council document, by displaying search results within their political context. This class of system combines the strengths of both task-specific (adaptive, situational) and domain-specific (specialised) systems. For instance, users’ information needs can be supported throughout different task stages by helping users formulate, refine and search for their questions [102]. Another example is to only display document genres (e.g. council letters or Powerpoint presentations) during specific tasks where they are known to be useful [50]. Designing support around both the user’s domain and tasks allows us to design support that is unique to the domain. This level of specificity is not possible when focusing on solely the domain or the task. Previous studies have identified tasks of knowledge workers in different domains and created tailored solutions (e.g. [174, 117, 101, 153]). For instance, by supporting marine ecology experts in the monitoring fish populations through computer vision and classification [10]; or by enabling cultural heritage experts to navigate artworks based on semantic similarity [66]. An example of a situational and specialised system for the municipality could be to support council members when their search task involves understanding the political context of a topic, by providing an overview perspective of political dossiers.

Studies on search functionality for municipal or other policy-related domains are primarily domain-specific and technology driven. For example, some studies introduce a semantic web for parliamentary data [11, 71] or focus on data mining political events [85]. A more detailed understanding of the information needs of municipal knowledge workers is still lacking. A more thorough understanding of the user would enable us to design functionalities that let users perform tasks more effectively, efficiently and with a higher user satisfaction.

As the specificity of the domain and task support increases, we can provide support by providing information that is more relevant (e.g. [174, 100]), presenting information from a perspective that is meaningful for the task (e.g. [153, 61]) and enabling users to interact with information in ways that are useful to the task (e.g. [101]. There is a need to characterise the tasks of knowledge workers more rigorously. This would enable us to identify when tasks in different domains have similar requirements, and therefore similar solutions.

1.5 Research questions

We focus our research on developing specialised search functionalities for user groups at the municipality of Utrecht. We focus on knowledge workers, who engage in a higher proportion of intellectual and complex tasks [133] and thus stand to benefit the most from improved search functionality. This leads us to our main thesis research question:

> **TRQ** *How can we design search functionality to support the search tasks of municipal knowledge workers?*

We investigate task-based support for council members and policy workers, which first requires us to understand the tasks these users perform:

> **TRQ1** *How can we characterise the tasks of municipal knowledge workers for the purpose of developing (domain-)specific search functionality (Chapter 2)?*

Every characterisation is made for a purpose (see e.g. [22]), and inherent to our purpose is ensuring that the task-description can represent domain-specific requirements of the search tasks. We explored this question while identifying the tasks of council members, as we developed a domain-specific search engine for council information. Over the course of this project, we found that not every task warrants specialised functionality. For example, generic search engines may suffice for tasks such as looking up a known document. The benefit of task-specific support needs to outweigh the cost of building that support. Hence we turn to the second question:

> **TRQ2** *How can we determine which search tasks of municipal knowledge workers warrant the development of specialised search functionality (Chapter 3)?*

We investigated this question while developing a search engine for the web-based search tasks of policy workers. After identifying the tasks of policy workers we considered which ones required additional support, by comparing the users' search tasks with the functionality available. An important finding was that complex web-based search tasks tend to transition into expert search tasks, and that these expert search tasks were not sufficiently supported. By considering what that support entails, we arrive at the third thesis research question:

> **TRQ3** *How can we design specialised search functionality for municipal knowledge workers around a faceted description of their search tasks (Chapter 4)?*

As a case-study, we developed a search engine tailored to the expert search tasks identified in the previous study. This expert search system was designed around the task facets used to characterise the task, with particular emphasis on whether the retrieval unit should be a person or a document. We compared variations of this system using simulated search tasks, to assess whether the system based on the faceted task descriptions was more effective, efficient and satisfactory.

So far, we have focused on search tasks centered around retrieving individual documents, such as when a user needs to 'retrieve the council proposal about the Uithoflijn'. However, other search tasks can require different ways to interact with information. For instance, the search task 'create an overview of the council's decision-making process when they created the proposal for the Uithoflijn' necessitates that users understand how various search results are interrelated. We now investigate how to support this type of task, as a use-case for designing functionality that facilitates alternative methods of interacting with information. Hence the research question:

> **TRQ4** *How can we support council members in their task of obtaining an overview perspective of how individual (council) documents are interrelated?*

We present a study where we provided council members with an overview of council information dossiers that the city council has voted on. First, decision histories of policy proposals were reconstructed, by identifying links between council documents. Then, we explored how to display such overviews in the search interface.

These research questions represent incremental steps towards a broader research goal: a method to design search applications that support the tasks of their target users. It remains a challenge to formalise an approach for designing useful search functionality around user tasks. In this thesis, we detail the application of one such method within a specific organisation. The results of these studies demonstrate how the complex tasks of knowledge workers at the municipality of Utrecht can be supported.

1.6 Contributions

Our studies contribute to the understanding of domain-specific systems and the experts in those domains, as well as to the task-based design of search applications. To present an overview of our main contributions, we organise them around the CHI contribution types [172]:

1.6.1 Empirical contributions

Identifying and characterising the work tasks and search tasks of Dutch knowledge workers:

- Identified task facets to characterise the tasks of municipal knowledge workers (Chapter 2, validated in Chapter 3)
- Provided an overview of tasks of council members at Utrecht and Hollands Kroon (Chapter 2)
- Provided an overview of tasks of policy workers at Utrecht (Chapter 3)

Insight into the information seeking of municipal knowledge workers:

- Found that council members and policy workers exhibit similarities to professional searchers, as they engage in specific and complex search tasks. However, they differ in that they often lack advanced search literacy and are typically constrained by time, which limits their ability to search thoroughly (Chapters 2 and 3)
- Found that policy workers perform complex web-based search tasks by finding and asking colleagues with expertise for help (Chapter 4)
- Found that, to help council members understand intricate political dossiers, policyworkers manually compile decision histories. This is a 'best effort' approach with subjective inclusion criteria (Chapter 5)

Improved search functionality for specific tasks, as tested in studies where users perform simulated search tasks:

- Showed that a search index with only authoritative sources helps policy workers perform simple web-based search tasks more effectively and efficiently (Chapter 3)
- Showed that a candidate-based expert search interface helps policy workers perform expert search tasks more effectively (Chapter 4)
- Proposed an automatic approach to construct decision histories of policy proposals (Chapter 5)

1.6.2 Artifact contributions

Search engines and functionality:

- Developed a prototype search engine for council information. An enhanced version of this system is currently in production and can be accessed at ureka.utrecht.nl, last accessed 05/05/2024 (Chapter 2)
- Developed a prototype search engine that supports the web-based tasks of policy workers (Chapter 3)
- Developed a prototype search engine for people search. The system finds internal colleagues who have expertise, as evidenced by the documents they have authored (Chapter 4)
- Developed functionality to construct decision-histories of policy proposals of the city council, and integrated this in the council information search system (Chapter 5)

1.6.3 Methodological contributions

Steps taken towards task-based design:

- Task facets were used to characterise domain-specific tasks (Chapters 2 and 3)
- The mismatch between task description and available systems was used to assess which tasks require improved search functionality (Chapters 3, 4 and 5)
- Task descriptions were used to improve the retrieval and presentation of search results (Chapters 3, 4 and 5)
- Task descriptions were used to model and present information around the user's needs (Chapters 4 and 5)

Altogether, these contributions enable us to better diagnose the problems that users encounter during complex search tasks, rather than optimising functionalities for ill-defined problems. This allows the municipality to move from generic search functionality towards targeted search applications (see the transition from phase 2 to phase 3 in Figure 1). Furthermore, our work represents early steps towards adapting information seeking models from scientific literature towards the purpose of developing domain-specific search engines.

1.7 Thesis outline

We initially investigate the tasks of council members and how to characterise them (chapter 2). We then consider the tasks of policy workers, which ones need support and how to support them (chapter 3). We then consider how to support one type of (expert search) task in more detail (chapter 4). Finally, we move beyond typical retrieval tasks by considering how to support a task where council members need an overview of political information (chapter 5). We conclude the thesis by revisiting our research questions and discussing the implications of our findings (chapter 6) and how our work could be continued in future work (chapter 7).

In **chapter 2** we investigate how to identify and characterise user tasks. We extend a pre-existing faceted task model to include domain-specific task facets. We identify these facets by investigating how one target user group (council members) described their own tasks, and analysing which aspects of tasks they used. We find that five task facets are most useful in representing domain-specific information needs: the task objective, topic aspect, information source, retrieval unit, and task specificity. The material in this chapter has been published as:

> Schoegje, T., de Vries, A., & Pieters, T. (2022, August). Adapting a Faceted Search Task Model for the Development of a Domain-Specific Council Information Search Engine. In *International Conference on Electronic Government* (pp. 402-418). Cham: Springer International Publishing. https://doi.org/10.1145/3657054.3657116

The first author was responsible for theoretical and practical contributions, as well as the manuscript. The co-authors provided incidental theoretical suggestions and significant improvements to the manuscript. This is the case for all publications, except where stated otherwise.

We validate the task model created in chapter 2 in **chapter 3** by identifying how an independent user group describes their own tasks. We then compare their tasks to the search functionality that is already available to them, to identify where specialised search functionality is required. We introduce a search system to better support policy workers during their simple web-based search tasks by only indexing authoritative sources and verify in a quantitative user study that users are more effective and efficient. We also find that policy workers solve more complex web-based search tasks by finding a colleague with expertise. The material in this chapter has been published as:

> Schoegje, T., de Vries, A., Hardman, L., & Pieters, T. (2023). Improving the Effectiveness and Efficiency of Web-Based Search Tasks for Policy Workers. In *Information*, 14(7), 371. https://doi.org/10.3390/info14070371

Chapter 4 explores how to design the ranking method and retrieval units for expert search tasks. We compare document-centric ranking (ranking the text of each individual document) with candidate-centric ranking (ranking all of the text written by an author appended to a single string), as well as a document-centric interface (showing a single document) with a candidate-centric interface (showing the most relevant documents of an author). Some key findings are that the document-based ranking leads to faster task completion times, and that the candidate-based interface leads to more effective task completion (finding a higher rate of relevant experts). Additionally, we find that users evaluate search results more thoroughly when they are presented more information, which may be preferable if we want to avoid approaching people without the correct expertise. The material in this chapter has been published as:

> Schoegje, T., Hardman, L., De Vries, A., & Pieters, T. (2024, March). Improving expert search effectiveness: Comparing ways to rank and present search results. In Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval (pp. 56-65). https://doi.org/10.1145/3627508.3638296

Although search engines excel at presenting the most relevant fragments of information, some search tasks require users to piece those fragments together. In **chapter 5** we present a search task where users require an overview of the city council's decision making process. To support users we introduce a method for reconstructing an overview of council decisions based on how council documents are co-referenced during multiple meetings, and how documents reference older documents in their text. In addition we define guidelines for how to design the user interface for such overviews based on expert interviews. The material in Chapter 5 is published as:

> Schoegje, T., Hardman, L., De Vries, A., & Pieters, T. (2024, June). Reconstructing the decision-making processes of a city council based on references between documents. In Proceedings of the 25th Annual International Conference on Digital Government Research (pp. 525-533). https://doi.org/10.1145/3657054.3657116

The thesis has also indirectly benefited from an exploratory study on whether a user's search task is a useful signal to improve search rankings. In this work we used a domain-specific task taxonomy and classified user queries into one of these tasks. We then performed term expansion based on the task of the query. As this work did not yield a peer reviewed publication, we instead briefly discuss this work and our findings in Appendix A.

> Schoegje, T., Kamphuis, C., Dercksen, K., Hiemstra, D., Pieters, T., & de Vries, A. P. (2020). Exploring term expansion for task-based retrieval at the TREC-COVID track. arXiv preprint https://doi.org/10.48550/arXiv.2010.12674

The first author was responsible for the main theoretical and practical contributions, as well as the manuscript. All co-authors contributed conceptual suggestions.

Chapter 2 Adapting a faceted search task model for the development of a domain-specific council information search engine

Domain specialists, such as council members may benefit from specialised search functionality. However, formalising their search requirements for developing a search system is challenging. To address this, we adapt a faceted task model to characterise the tasks of our target user group.

Initially, we identified which task facets council members use to describe their tasks. We then characterised their tasks based on those facets. Finally, we discuss design implications of these tasks for the development of a search engine. Through two studies conducted at the same municipality, we identified a set of task facets and used these to define the tasks of council members. By coding the task descriptions of council members, we identified five key task facets: the task objective, topic aspect, information source, retrieval unit, and task specificity. A third study at a different municipality confirmed the consistency of our findings.

We then explored design implications of these tasks, which influence 1) how information should be modelled, and 2) how information can be presented in context, and it provides implicit suggestions for 3) how users want to interact with information.

Our work represents a step towards better understanding the search requirements of target user groups within an organisation. A task model enables organisations to prioritise their investment in new technology when developing search systems.

Chapter contribution

The following chapter investigates how to adapt the faceted task classification to model (domain-)specific tasks (TRQ1), for the purpose of creating a domain-specific search engine based on those tasks.

The chapter is published as:

Schoegje, T., de Vries, A., & Pieters, T. (2022, August). Adapting a Faceted Search Task Model for the Development of a Domain-Specific Council Information Search 21 Engine. In *International Conference on Electronic Government* (pp. 402-418). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-15086-9_ 26

2.1 Introduction

Users and information needs at the municipality of Utrecht are diverse. Some of these user groups have complex information needs [134], which may require specialised search functionality before search satisfaction can be achieved [48]. If organisations do not identify these specialised requirements they risk investing in search solutions that are unsatisfactory, which must then be replaced within a few years. We investigated the needs of council members as a target user group because they perform complex intellectual tasks [134] and are therefore likely to benefit from specialised search functionality. Our main research goal was to extend an existing (search) task model from the information seeking and retrieval literature to adapt it for the purpose of identifying council member search requirements. We applied this design approach to develop an e-government search application that supports council member search tasks.

We found that existing task models are not designed for the purpose of representing domain-specific information needs. To address this, we adapted an existing task model based on users' descriptions of their tasks, which includes a domain-specific topic facet that describes how tasks relate to information subsets in that domain. Below we discuss how the adapted task model can inform the design of a search engine, as the interface should clearly reflect where users can perform each of their tasks; what information is necessary for each task; and how users want to interact with this information.

We first contextualise the tasks by describing the information seeking of council members, which we do by describing the users, information sources and channels involved. We then identify council member tasks and relate these to the available information systems. This reveals tasks that are not adequately supported by those systems. We then discuss the design implications of the task model for better support of those tasks, by extending the information model and creating a better interface. We end by discussing how the proposed system is more suitable to council tasks than the pre-existing system.

We performed our analysis at two Dutch municipalities with consistent findings, suggesting that the search functionality for council members in similar contexts can be standardised.

2.2 Related Work

Developing useful search tools for target user groups requires an understanding of how and why they search. Users start searching for information as a part of performing some broader (work) task [73]. In professional settings work tasks often follow from the professional's responsibilities [94, 78, 27], which in turn follow from their role within the organisation [89]. If we can model the situational context of information tasks, we can assess whether a search engine is able to aid the user in their goals [60] and therefore whether useful for the user in task completion [12, 161].

We introduce several domain-independent task models from the information seeking and retrieval literature as complementary perspectives at the basis of the current work. We start with the broadest model and ending with the narrowest, and then introduce increasingly domain-specific models.

2.2.1 Domain Independent Task Models

Different research purposes have resulted in different kinds of conceptual models, with characteristics such as 1) how much is modelled (broad or narrow), 2) whether a process or static situation is modelled, 3) whether something directly observable or something more theoretical is modelled (concrete or abstract) and 4) how much the model generalises beyond domains (general or specific) [76, 73].

Ingwersen and Järvelin argue that all information seeking and searching is performed by actors, who interpret information and their tasks through their own perspectives. Therefore, they propose a model of information seeking from the cognitive perspective [73]. This broad and abstract framework contains five main components, where the interrelation between these components (over time) provide important context for each individual component. The five components are 1) the information seeker/user, 2) the interface, 3) the IT systems underlying the interface, 4) the information objects in those systems and 5) their social/organisational/cultural context. They note that work tasks may originate from the organisational context, but they may also originate from e.g. newly found information. In this model, work tasks arise as a result of interactions between the actor and the other components. Work tasks are directed from components towards the actor, and search tasks are directed from the actor towards other components.

Taylor proposed that all information acquisition and use in a professional setting is performed in a contextual environment [154], and defined such contextual factors. Byström et al. extended this work [25] to represent the context of Workplace Information Environments. These describe important variables grouped in four categories: 1) sets of people, 2) work tasks, 3) settings and 4) task resolutions. They note that one's profession leads to work activities, which in turn necessitate (work) tasks.

Byström and Hansen present a narrower conceptual model focusing on what work tasks, information seeking tasks and search tasks are, and how they are interrelated [24]. In their definition, an information seeking task includes all steps a user takes to gather information, including interpersonal ones. Information seeking can consist of several information search tasks, which are search episodes with their own search goal. A search task performed within a search engine or database is also known as a retrieval task. Rather than focusing on tasks from an individual's perspective, Byström considers how to characterise the tasks of a user groups, and the social practices that will affect how tasks are performed [27].

Li and Belkin reviewed the literature on information tasks and proposed a narrower model [94]. They characterised tasks with independent facets that affect the information behaviour and the subtasks that a given task produces. This framework has been used to describe tasks and create simulated tasks [64, 149]. An extension included the information level of results as a facet [35].

Byström and Hansen show that we can describe an information need in three levels of increasing context [24]. At the lowest level there is a topical description and query, similar to traditional laboratory experiments [73]. The second level includes a situational description, which is the context of the task at hand. The faceted task framework is a step towards gathering research on this area [94]. The third level is all the contextual factors that affect the search, such as the four categories of the workplace information environment [25] and the five components of the cognitive framework [73].

These models have some overlap, but also different strengths. Ingwersen and Järvelin's model is useful for a cognitive user centered perspective. The model by Byström 24 et al. model for the workplace information environment is valuable for investigating work practices in the workplace. Byström and Hansen's task level model is good at describing tasks from a process perspective. Li and Belkin's model is a powerful tool for describing information behaviour. These models all generalise beyond domains, which gives them explanatory power in a wide range of settings. We find however, that there is a trade-off between a model's ability to generalise beyond domains, and its expressive power within a narrower domain. Li and Belkin were unable to add task topic as a facet because without a domain a list of topics would be unlimited [94]. Byström and Hansen noted the importance of finding a better way to characterise context- and situation-aware descriptions based on real-life data [24]. Ingwersen and Järvelin noted that our increased understanding of tasks had not yet translated to better design criteria for information (retrieval) systems. To address these concerns we now turn to increasingly domain-specific task models.

2.2.2 Domain Specific Task Models

The above models are applicable during information seeking and retrieval in most domains. We now focus on smaller domains, which allows us to introduce task descriptions that capture increasingly more situational and contextual context.

The classic example to model the user's goal is the web search taxonomy of query intents [20, 125]. Within enterprise search there may exist a similar taxonomy of search tasks [132], and between search tasks there seem to be recurring sequences of task types [132]. User interfaces can better support the search experience when they are designed to reflect the typical flow of the users' tasks. When we consider tasks within professional search domains we find complex tasks with similar characteristics [165, 164]. When we focus on the tasks within a single organisation we can introduces an element of the organisations' objectives [78], but these are still far removed from the users' immediate goals.

There are also examples of search task studies outside of a professional/ work setting, which often analyse all users of a system collectively [150, 159], or focus on supporting specific tasks [156, 42].

Taylor provided a contextual description of the information use environment of American legislators [154]. This information use environment describes the people, problems, setting and problem resolutions. Legislative culture is described as verbal, nonhierarchical, time-constrained and as having the political party as a major centralising force. An observational study of knowledge workers at a municipal administration identified four main types of work tasks that involved search tasks [134]. Working on legislative processes includes complex information seeking tasks [134, 106], possibly because knowledge creation (such as the creation of legislation) is cognitively complex [84]. Complex tasks are more likely to require specialised search functionality.

The present study approach focuses more specifically on one target user group, and adapting an existing task model for the purpose of search design. This approach contrasts with previous work for similar user groups (e.g. [82, 44], http://zoek. openraadsinformatie.nl), which typically focus on the data- or technology-driven innovations.

2.3 Council Member Information Seeking

Council members in the Netherlands have three main responsibilities:

1. Prescribing guidelines for new legislation
2. Verifying whether the municipal workers have adequately translated the council's decisions to concrete policy
3. Representing the citizens' interests while forming legislation guidelines

These are the same responsibilities that Taylor identified for American legislators [154]. Work responsibilities form the highest level of motivation for work tasks and subsequent search tasks [27]. The first and second responsibilities lead to active tasks, whereas the third occurs passively during the other tasks. In performing the first responsibility the council is informed by experts, debates solutions, and decides on new policy over a series of meetings. Members aim to create solutions and arguments that extend or modify existing policies. Members then try to persuade others to support their solutions during meetings. Most of these solutions and knowledge are created during the preparation for domain-specific commission meetings. When the council is not unanimous on a solution, members refine and prepare their proposals before a final discussion in a council meeting. In this article we investigate the sub-tasks of preparing for council meetings, because all council work is oriented around these meetings. Users complained that the previous system was not satisfactory for performing the search tasks within this work task, and hence specialised search functionality may be valuable. A critical challenge is that council members must often extend existing policies, that were created before the members joined the council. It is thus crucial that they understand the existing policy and how it was formed.

2.3.1 Council Members and Supporting User Groups

Multiple user groups support council members during their tasks. Because of space limitations we only briefly and informally introduce these in Table 1. Here we characterise them using their knowledge types [73]. Domain knowledge can be declarative (what is it about) and procedural (how to do it). Search knowledge can also be declarative (where will I be successful) procedural (how do I search effectively). Professional searchers typically have a high domain and search knowledge, and are thorough when searching [4, 89, 129]. Table 1 is based on observations of search behaviour and interviews at two municipalities.

Table 1. Characterising the knowledge of council members and supporting groups

	Domain Knowledge		*Search Knowledge*		
User	**Decl.**	**Proc.**	**Decl.**	**Proc.**	**Thoroughness**
Council member	Experience dependent		Diverse	Diverse	Time-limited
Faction Staff	Experience dependent		Diverse	Diverse	Thorough
Adviser	High	High	Unknown	Unknown	Unknown
Search expert	Low	Low	High	High	Thorough
Public servants	High	High	Unknown	Unknown	Thorough

Council members become professional searchers with domain expertise over time, but the election cycle leads to the replacement of experts with inexperienced members. A notable difference from typical professional search domains is that council members are not trained to search effectively, unlike other experts [127, 4]. Many are unfamiliar with Boolean operators and strategies for effective query formulation. Council members may therefore benefit from search training and/or a search interface that supports them in expressing complex queries.

2.3.2 Information Sources

Due to space limitations, we will only briefly contextualise the information systems we observed during interviews and interactive search sessions at two municipalities. The two primary information sources were web search engines (mainly Google) and an internal system called iBabs. iBabs is an app used for planning meetings and archiving the official policy documents used during these meetings. A copy of the public data can be accessed at http://api.openraadsinformatie.nl/ (accessed May 2022).

Information seeking on a new topic typically began with performing a web search to find general background information (from indexed news outlets or information published on the municipality's homepage for example). This was followed up by searching in iBabs. This archival system provides an internal search engine that allows users to (re)find known documents by filtering facets such as the date and title of the meeting. This type of functionality is less useful for non-specific needs. This is consistent with findings at a Finnish municipality, which showed that the organisation's internal systems tended to perform well for specific tasks (such as re-finding a known document), but less well for more amorphous tasks (such as exploring a topic) [135].

As we may expect from previous literature [135], the other prominent internal information sources and channels included e-mail (personal or collective faction inboxes) and human sources (colleagues, party-neutral advisers and a temporarily appointed clerk whose main responsibility was to search for information). Some larger political parties created internal solutions to share information (e.g. documenting plans in the cloud), although their main information advantage appears to be in having council members with more experience. Experienced council members remember older events and documents, which is a significant benefit given the difficulties in exploring archived information.

2.4 Methodology

In the first of two analyses we used a codebook to analyse interview data to identify what tasks facets council members used to describe their tasks. In the second analysis we summarised the council member tasks we identified in the interview data and an observational study and characterised them based on the facets we previously identified.

The two analyses were based on interview data from three studies. Both analyses were initially performed on the interview data that resulted from two studies performed at the municipality of Utrecht, and then a third study was conducted at the municipality Hollands Kroon to test if the results could be reproduced with a similar user group in a different organisation.

The municipality of Utrecht is one of the largest and oldest municipalities in the country, whereas municipality Hollands Kroon is of average size and was formed less than 10 years ago as a fusion of smaller municipalities. Selecting municipalities of such different sizes and histories allows us to determine whether the work tasks we identify are organisation-dependent, or whether council member tasks are similar across organisations.

2.4.1 Participants

Each study included a sample of council members that were diverse in terms of experience (years in council), demography (gender, age) and the political parties that they represent (size, ideology). This sample was selected by council clerks.

2.4.2 Data Collection

The first study was a series of six one-hour interviews performed to construct a customer journey for preparing a council meeting. These semi-structured interviews were not limited to search-related questions, but aimed to identify all work tasks. The study aimed to map out relevant information channels and sources; relevant user groups; communication channels; the triggers that move users to actions; and noted which steps went well and which did not. We only report the aspects relevant to the present research scope.

During the second study these same participants (except for one replacement) performed simulated search tasks in an interactive session. These simulated tasks were recreations of real council tasks in a laboratory setting where we asked users to search for pre-defined topics. Each participant in the session had an observer who asked them unstructured questions about their information seeking. This setup allowed us to observe more instrumental search tasks (i.e. tasks that were not explicit user goals, but necessary sub-steps).

The third study at municipality Hollands Kroon consisted of five semi-structured interviews designed to first identify the work tasks performed in preparation of a council meeting. For each work task we focused on the search tasks involved, and we concluded by asking for (recent) examples of each search task.

2.4.3 Analysis 1: Identifying Task Facets

We first identified which task facets characterised tasks in the domain. We analysed the task facets that users used to describe their tasks by developing a codebook based on the interview data. Coding is a qualitative method used to analyse interview data by annotating (potentially overlapping) fragments of interviews with codes by theme. It allows the researcher to identify concepts and relations between concepts [40]. The development of a codebook is an iterative process that occurs over multiple studies. With every study analysed, one tries to improve the codebook until it explains all new data from new studies. New codes are found in two main ways. Data-driven codes emerge to represent themes and recurring concepts in the data. Theory-driven codes are added when the data reflects themes from the relevant literature. In our case the theory driven codes include the task facets. We focused on analysing the task related themes and generate the codebook on data from of Utrecht. We then apply the codebook to the data from Hollands Kroon to test whether its completeness for describing tasks performed in this new context.

2.4.4 Analysis 2: Characterising Council Member Tasks

We characterised the work tasks that users described and showed us during the studies at Utrecht using the task facets we identified using the codebook. We then compared the tasks identified to those we found at Hollands Kroon to identify whether our list of tasks is exhaustive.

2.5 Results 1: Identifying Task Facets

By applying the codebook to the data from the first study we found four task-related codes: the task objective, information sources, topic aspect and task specificity. We applied the codebook to the second study at Utrecht to find further evidence for the previous codes and the retrieval unit as a new code.

The task objective is a description from the users' perspective. The information sources are the systems they mentioned, implying which underlying datasets are necessary for the task. The topic aspect represents different types of declarative domain knowledge. Consider the example topic 'the sound leak in Tivoli'. Over time users may be interested in different aspects of this, such as the background of the issue; how the council has dealt with this topic in the past; and what aldermen have previously promised to do about the issue. We found a limited set of topic aspects that are important for many topics. These aspects are reflected in the interview data when users implicitly switch their definition of what a topic's 'context' is. These topic aspects are closely related to the four kinds of information that Taylor identified among American legislators [154]. The only difference is that we found a distinction between background information and policy information, which Taylor grouped as one information type.

The retrieval unit code reflects that users do not always seek documents, but can instead seek, for example, a fact or (the contact details of) a person [123]. It is related to the information level facet but captures more of the user's goal. The final facet is the task specificity, which indicates how specific the information is that users are looking for in a search task.

Table 2. The work tasks identified at Utrecht, described using the task facets council members use to describe their tasks

Task ID	Task Objective	Topic Aspect	Information Sources	Retrieval Unit
WT1	*Understand the agenda item*	Background Information	Google	Facts
WT2	*Evaluate existing or proposed policy*	Policy	iBabs, Google	Document
WT3	*Analyse previous council decisions*	Decision History	iBabs	Document(s), timeline
WT4	*Understand political positions*	Political Context	iBabs, Google	Statement
WT5	*Create an argument*	Mixed	iBabs, Google	Mixed
WT6	*Create or defend a perspective*	Mixed	iBabs, Google	Mixed
WT7	*Evaluate progress on policy execution*	Administrative Context	BMT	Statement

Table 3. The search tasks identified at Utrecht, and the work tasks during which they occur

Task Identifier	Search Task	Associated WTs
ST1	Find news articles, municipality publications and other substantive public documents	WT1, WT2, WT6
ST2	Find reports and other (internally generated) substantive documents	WT1, WT2
ST3	Find the aldermens' commitments (formal agreements to the council)	WT7
ST4	Find agenda items and corresponding transcripts where topic was previously discussed	WT3
ST5	Find documents that were key in the previous discussion of this topic	WT3
ST6	Find documents containing the current policy	WT4
ST7	Find previous statements from aldermen or colleagues	WT4, WT2, WT6
ST8	Find public articles containing political standpoints	WT4
ST9	Find sources supporting an argument	WT5
ST10	Find the alderman responsible for this topic	WT3, WT4, WT2

Table 4. Characterising the search tasks identified in Table 3 using the task facets. Many tasks can be either amorphous (am) or specific (spec). These are joined in the Table for formatting.

Task ID	Topic Aspect	Info Sources	Retrieval unit	Task Specificity
ST1	Topic Background	Google	Document	Am or Spec
ST2	Topic Background	iBabs	Document	Am or Spec
ST3	Decision History	BMT	Statement	Am or Spec
ST4	Decision History	iBabs	TImeline	Am or Spec
ST5	Decision History	iBabs	Document(s)	Am or Spec
ST6	Policy	iBabs/Web	Document	Am or Spec
ST7	Political Context	iBabs	Statement	Am or Spec
ST8	Political Context	Google	Mixed	Am or Spec
ST9	Mixed	Mixed	Mixed	Mixed
ST10	Admin. Context	Google	Mixed	Specific

2.5.1 Generalisation of Codebook

We applied the codebook developed at Utrecht to the data from Hollands Kroon. The codebook was able to explain all task-related themes. This suggested that the codes we used for task facets were stable (also known as theoretical saturation) and can properly represent tasks in this domain.

Five task codes were found based on how council members characterised their tasks: the task objective, the information sources, the topic aspect, the retrieval unit and the task specificity. We adopt these five codes as the task facets to describe the domain-specific task context.

2.6 Results 2: Council Member Tasks

Table 2 introduces the work tasks found at Utrecht. Tables 3 and 4 respectively describe and characterise the search tasks identified.

2.6.1 Generalisation of Tasks

To test whether the list of council member tasks is exhaustive we performed a study at Hollands Kroon and compared the findings to those at Utrecht. At Hollands Kroon we found work tasks WT1-6 from Table 2, but not WT7: evaluating the progress on alderman's commitments.

This may be because the municipality is smaller, making it easier to keep track of such commitments. At Hollands Kroon we found all search tasks except ST3 and ST8. ST3 is less significant in this municipality because WT7 is less significant. It is unclear why users here search for fewer public articles containing political standpoints (ST8). Perhaps the municipality has a smaller profile in the news because it is smaller. There is a slight difference in how users search for previously discussed topics (ST4), as the municipality Hollands Kroon does not maintain transcripts of each meeting. Their council is only provided with the video recordings of meetings. Because these are not searchable, this search task is not well supported. This is because Hollands Kroon has less resources. 33 The council tasks identified at municipality of Utrecht are a superset of those found at municipality Hollands Kroon. We expect that our list of council member tasks within the Netherlands is fairly exhaustive.

2.7 Supporting Specialised Council Task Functionality

When comparing tasks identified with the existing systems (see Table 5) we found that 1) filter-based search functionality is insufficient for non-specific tasks and 2) there is no support for investigating different topic aspects. We discuss how to design a more suitable domain-specific search engine based on the task model. We specifically consider how the interface should enable each of these tasks, what information is necessary for each of these tasks and how users want to interact it. We introduce the domainspecific search engine we developed in cooperation with Spinque, publicly available at https://ureka.utrecht.nl/app/.

2.7.1 Linking Tasks to Information Subsets

The task topic aspect indicates which datasets and document genres are relevant for a given task, informing how information in the domain should be modelled. Within council information we found that tasks related to the topic background aspect should search within public web sources. Tasks involving the political context aspect involve searching the political statements made during meetings (i.e. segments of the council meeting minutes). A search engine that supports the policy aspect should enable searching all council documents. Tasks involving the decision history aspect involve the specific document genre council proposals, and finding the meetings that discuss these proposals.

The retrieval unit facet indicates how these document genres should be indexed: users search for the official council proposal documents in some tasks, but only look for segments of the meeting minutes in other tasks. Identifying the relevant document genres and retrieval units can indicate how the information model that the search engine is based on should be extended (e.g. by extracting political statements from meeting minutes).

2.7.2 Interface Design Implications

Work tasks reflect user goals and inform how the user approaches the system. Hence it should be clear to the user where he should go for any given work task. Information is ideally presented in a useful context, which depends on the topic aspects we identified. The format and presentation of individual results depends on the retrieval units we identified. Based on these guidelines we designed a different view (page) in the interface for each topic aspect, as shown in Figures 1-4. We developed search verticals for existing policy, political context and administrative context. The decision history of council proposals was added as a contextual view when clicking a search result. We did not include functionality for the background information topic aspect, as interviews indicated that web search is satisfactory for this.

The search tasks reflect how users want to interact with the information within these views. The current model does not capture these requirements explicitly, but is a step in that direction. The task specificity facet indicates whether users will need filtering functionality (with high precision) or explorative functionality (with high recall). For example, when users search for statements by specific people (ST7) there is an implied requirement for filtering statements by the speaker. We could investigate the concrete requirements (e.g. on what information features does the user want to filter) by asking users about example tasks or by observing users perform the tasks in questions. Future work may include a search task facet that captures which filters should be included for specific tasks, and which dimensions are of interest in amorphous tasks.

2.7.3 Comparing the Proposed Improvements

In this paper we focus on the design process that resulted in a new search system, rather than individual improvements for specific tasks. As a result, the new system introduces many changes (e.g. the interface, result ranking, the datasets included) and it is both unfeasible and not our goal to evaluate the impact of each variable we changed. To show the value of our design approach we instead compare the proposed system to the existing system. We compare systems based on their ability to facilitate user tasks, because the best search system is the one that is most useful for the user’s goals [13, 161].

Figure 2. Vertical for the Policy topic aspect.

Figure 3. Vertical for the Political Context topic aspect.

Figure 4. Vertical for the Administrative Context topic aspect.

Figure 5. Search result view for political documents (Decision History topic aspect).

Table 5 summarizes the tasks that users want to perform (based on our previous results), and how both systems support these tasks. Because our design approach led to a better understanding of the target user group's requirements, we were able to develop more useful functionalities. This can aid developers and organisations in prioritising the importance of different functionalities. Informally, we report an enthusiastic adoption by its users, and the interest of other local municipalities. This suggests that our approach was successful at specifying the user search requirements.

Table 5. A comparison of the existing and proposed search systems. We summarise the tasks by their facets, because tasks with the same facets require the same functionality.

Requirement	iBabs	Proposed
TA: Background Info	Web search	Web search
TA: Existing Policy	Filtering	Vertical in Figure 1
TA: Political Context	None	Vertical in Figure 2
TA: Administrative Context	None	Vertical in Figure 3
TA: Decision History	None	Result page in Figure 4
Retrieval Unit	Document/meeting	TA dependent (Figures)
Specific Search Tasks	Filters	Filters on the same features
Amorphous Search Tasks	None	Timeline of Decision History

2.8 Conclusion

A target user group may require specialised search functionality to perform their work effectively. In this paper we investigate how to model the search requirements by extending the faceted task model with facets that capture domain-specific information. Comparing these tasks to the existing systems allows us to find initial design implications for improving the search experience, because it illustrates 1) how each task relates to subsets of information in the domain and 2) how users want to interface with this information.

We characterised council members as professional searchers who have not had time to specialise in their domain, and have not had any search literacy training. We found that council members information seeking usually begins with a web search to identify background information, using news sites and municipal websites. They then access internal council information systems to inform themselves about different topic aspects.

We found that existing task classifications were generic by design, and unable to represent domain-specific aspects of tasks. We extended this work by identifying the five task facets that council members used to characterise their tasks, and discussed how these can be used to represent domain-specific tasks. We found the task objective, the topic aspect, the information sources, the retrieval unit and the task specificity. We discussed how tasks have implications for how the information should be modelled, and how the interface should facilitate them.

We used the topic aspect to determine which datasets and document genres are important for which tasks (similar to search verticals). We used the retrieval unit to determine how to index (segments) of documents. For the interface design we used the task aspect of work tasks to present information in a useful context, resulting in a different interface views for different topic aspects. The retrieval unit informed how individual search results should be presented. The search task specificity is a first step towards understanding how users want to interact with the information. Once we identify task specificity, we can investigate what type of filters are beneficial for a (high precision) specific task, or what dimensions users want to explore in (high recall) explorative tasks.

We found the same task facets and the same tasks at two municipalities. If the task model generalises to municipalities in similar contexts, then the search functionality we developed could be standardised across Dutch municipalities.

2.9 Chapter discussion

This chapter identified five task facets that can be used to describe the domain-specific tasks of council members (TRQ1): the task objective, topic aspect, information source, retrieval unit, and task specificity. By characterising council member work and search tasks using these facets, we designed a search engine for council information that was quickly adopted over the pre-existing system. We highlighted tasks of council members such as the need to find expertise and to get an overview perspective on political dossiers.

The task model presented represents an initial step towards adapting the existing information seeking models towards the purpose of designing domain-specific search engines. Since the task facets were identified through the coding of interviews, the findings should be validated by replication with other user groups. Our functionality is designed to support the tasks, rather than the specific organisation or users, suggesting 38 that other users performing the same tasks will benefit from the same functionality. We expect the proposed functionality could be standardised for the search engines of council members in other municipalities in the Netherlands. Additionally, we expect the functionality can be adapted to tasks in other domains that have the same task facets.

Future work could consider how to make the task facets more specific, enabling more concrete design implications based on a given task description.

2.10 Chapter outcomes

We applied these findings for developing search engines at the municipality. We used the task facets to characterise the search tasks of council members at the municipality of Utrecht and Hollands Kroon, as well as policy workers at the municipality of Utrecht. This enabled us to help develop a domain-specific search engine for council information[2], where different search pages were designed to support different search tasks.

2 Available at ureka.utrecht.nl, accessed 15-05-2024

Chapter 3 Improving the Effectiveness and Efficiency of Web-Based Search Tasks for Policy Workers

We build upon existing literature on search tasks to develop a domain-specific search engine tailored to support the search tasks of policy workers. To characterise the search tasks we conducted two rounds of interviews with policy workers at the municipality of Utrecht. Our findings revealed that the challenges they face vary with the complexity of the task. For simple tasks, policy workers struggle with information overload and time pressures, especially during web-based searches. For complex tasks, policy workers prefer to locate domain experts within their organisation to obtain the necessary information, necessitating a different type of search functionality.

To support simple tasks, we developed a web search engine that indexes web pages exclusively from authoritative sources. We confirmed the hypothesis that users prefer expert search over web search for complex tasks and found that supporting complex tasks requires integrating functionality that enables finding internal experts within the broader web search engine. We constructed representative tasks to evaluate the proposed system's effectiveness and efficiency, and found that it improved user performance.

The search functionality developed could be standardised for use by policy workers in various municipalities within the Netherlands.

Chapter contribution

This chapter uses the faceted task model from the previous chapter to analyse the tasks of policy workers instead of council members. This approach allows us to both validate the task model and to characterise the tasks of this new target user group. By comparing the tasks of policy workers with the functionalities provided by available search engines, we can identify which tasks lack adequate support (TRQ2). We then develop and support for these tasks, and test it by having users perform simulated search tasks.

The chapter is published as:

Schoegje, T., de Vries, A., Hardman, L., & Pieters, T. (2023). Improving the Effectiveness and Efficiency of Web-Based Search Tasks for Policy Workers. In *Information*, 14(7), 371. https://doi.org/10.3390/info14070371

3.1 Introduction

We investigate how to apply the theoretical understanding of (search) tasks and information seeking on the practical development of search engines. A growing body of scientific work is investigating how a user's search tasks affect their information seeking (e.g., [30, 94, 165]). Simultaneously, a growing number of professional search applications are being developed for specific domains, such as patent search [167] and real estate search [62], as such search engines allow for higher search satisfaction [47, 113]. Less investigated is how to connect these two worlds: how to develop a domain-specific search engine based on a task analysis. In previous work, a task analysis was proposed to facilitate the design of a search engine for council members [140]. The current work extends and refines this approach while using it to design a search engine for policy workers at a municipality.

The municipality of Utrecht is one of the largest and oldest in the Netherlands, with over 4000 employees. Policy workers (PWs) are responsible for translating the city council's vision into concrete policies (see Figure 6), which involves a large amount of information seeking [135]. Ineffective search systems cost PWs time and lead to incomplete search results, which can negatively impact the quality of their work. As our initial research question, we investigate the challenges during PW search by investigating their goals and the tools they have available to achieve them.

> **RQ1** *What work and search tasks do PWs at the municipality of Utrecht perform?*

The findings enable us to design support for those tasks during the second research question, focusing on web-based tasks, as we find that these present the biggest challenges.

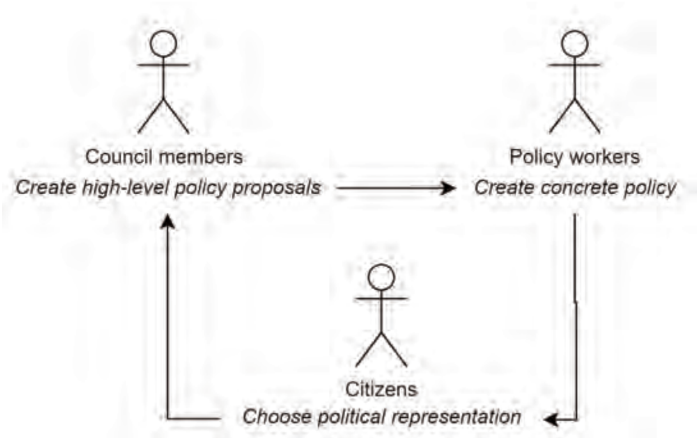


Figure 6. The role of policy workers and other important groups.

RQ2 *How can we design functionality for the more effective and efficient completion of the web-based search tasks of PWs?*

The paper is structured in two parts that each discuss one of these questions. To answer the first question, we performed semi-structured interviews to characterise the information seeking of policy workers, and identified the tasks that are not well supported. The method of interviewing and the results are described in Sections 3.3 and 3.4, and the search tasks are compared to the available functionality. Based on the findings, two hypotheses are formulated on how to improve the search experience. The first hypothesis (H1) is that a search engine that only includes authoritative web sources would be more effective than how PWs at Utrecht currently search. The second hypothesis (H2) is that PWs prefer to address complex web-based search tasks by finding a domain expert within the organisation. This suggests that supporting complex web-based tasks may be achieved through integrating expert search into the web search engine. These findings contribute to our understanding of information seeking in this domain, which might generalise to other knowledge organisations. Additionally, the approach to designing a search engine helps bring the theoretical understanding of search tasks closer to the practical development of search engines (through the modelling of information, the ranking and the interface).

In the second part of the paper, we develop a system to better support PW tasks, based on hypotheses H1 and H2. To test H1, a web search engine is developed that only indexes authoritative web domains in Section 3.5. We then test whether this system is more useful for simulated web-based search tasks than existing search tools (H1) in an experiment described in Section 3.6. We then investigate when policy workers would turn to seeking an expert colleague in their organisation (H2), instead of completing the task themselves. The results are presented in Section 3.7. Finally, we discuss the value of a task-based approach to designing search systems for target user groups.

3.2 Related Work

Our first research question regards identifying and characterising the search tasks of policy workers (PWs). We discuss how existing task models are typically generic by design, and how they can be extended to describe domain-specific tasks. Finally, we examine previous literature on the search tasks of PWs specifically.

3.2.1 Generic Task Models

An influential conceptual model to understand work tasks and search tasks is the conceptual model proposed by Byström and Hansen [24]. This model describes information needs in three levels of increasing context. The first level is a topical description and query. The second level is a situational description, which is the context of the specific task at hand. The third level is a description of broader contextual factors that affect how tasks are performed, such as characteristics of the individual. There are task models that describe tasks at each of these levels.

The first level is the topical description and query. Broder proposed a taxonomy of web-based search intentions that characterise queries into three types: informational, navigational, and transactional [20]. Later, Rose and Levinson extended this taxonomy, replacing the transactional category with a broader "resource" category, where users intend to access or interact with a resource on a page [125]. They also subcategorised the informational and resource categories. Search intents can be classified based on queries [2], which allows developing support for specific types of tasks (e.g., procedural tasks [31]).

The faceted task framework is the most comprehensive model for the second level of the situational context [94]. This framework characterises tasks based on several independent facets that affect the user's information behaviour and subsequent subtasks.

The third level consists of contextual factors. Notable works to model these are the cognitive framework of information seeking and search [74] and the workplace information environment descriptions [25]. The former focuses on the perspective of a single cognitive actor and explores how individuals interact with information. The latter focuses on the practices of groups of actors and is more relevant when describing the needs of a specific user group.

A more extensive overview of how search can be contextualised within the users' tasks is provided by Shah et al. [143]. For the purpose of supporting user tasks, the situational context is most important (as opposed to the generic characteristics of the user or the overly specifics of a momentary search intent). Hence, we focus on the situational context of PWs.

3.2.2 Applying Task Models to a Domain

While the previous task models are generalisable and domain independent, they are less suitable for expressing domain-specific information needs. When designing a search system, it is preferable to use a model that is concrete enough to link user tasks to specific subsets of the information in the domain (e.g., linking tasks to specific document genres).

The faceted task model is domain independent and does not include a facet for the task topic aspect, as the list of potential topics is unlimited in the absence of such context [94]. In a specific domain, such as the PW domain, more specific tasks can be identified, which allows for more context to be added to user goals. A more specific domain also enables a more specific information model that can link user tasks to subsets of information in this model.

The current work uses a domain-specific task model made by analysing council member tasks. The faceted task model was extended with domain-specific facets. This was used to identify search requirements and develop a search system [140] as shown in Figure 7. The faceted task model was expanded with domain-specific facets based on how council members described their own tasks. This allowed for a more comprehensive representation of the situational context. The task model describes tasks using several facets, and was developed by coding interview data. An inventory of tasks was created based on the interview data and an observational study. The authors found that the task model provides insight for (1) how to model information, (2) how to present information in the interface in context, (3) and how to allow users to interact with the information. In contrast to previous technology-centred projects at the municipality, which were temporary fixes, their user-centred approach addresses specific underlying problems that hinder the completion of certain search tasks.

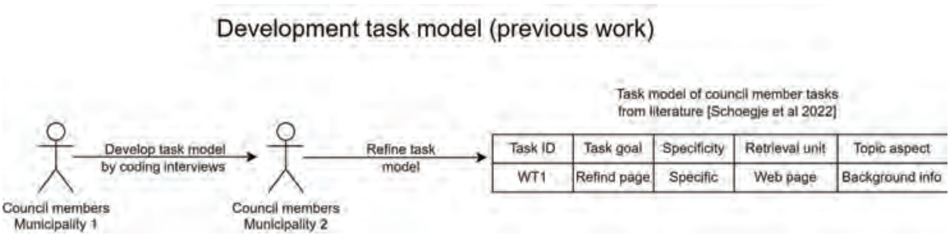


Figure 7. Development of the task model based on how council members describe their tasks [140].

Although the present study builds upon work from some years ago, it should be noted that the development of domain-specific search engines is an active research topic. More recent works investigated the tasks in domains such as mobile web search [3], music retrieval [176], question answering [29] and intelligent assistants [160]. Other studies investigated more generic domains, such as professional search [165], tasks that occur during teamwork [171] or extracting task context from activities [15]. Notably, the DoSSIER project has 15 PhD students working on topics such as characterising the tasks coming from knowledge work [142].

3.2.3 Tasks of Policy Workers

Although no task model has been applied to PWs, there is prior work describing PW tasks. These findings help inform the task model we develop in the present work. The information-use environment of American legislators was described in terms of its people, problems, setting, and problem resolutions [154]. Within this context, legislative aides, such as PWs in the municipality, help filter information and play a role as decision formulators who present options, alternatives, and recommendations 45 [155]. PWs work with ill-structured problems, shifting goals, time stress, and action– feedback loops. It is also unclear to them when they have found all useful information [17]. PWs are experts with less search literacy than expected of professional searchers [37, 41, 140], and their thoroughness can be limited due to time pressures [17]. When characterised based on their types of knowledge, PWs are considered experts with high declarative (what is it) and procedural (how to do it) domain knowledge, moderate declarative search knowledge (where to find it), and various levels of procedural search knowledge (successful ways to find it) [74, 140].

Although previous work identified these challenges, previous systems for PWs were typically not designed for their tasks. Instead, these focused on, for example, the structure of information [82, 44] or technology-driven innovations.

3.3 Method to Identifying Policy Worker Tasks

To identify the tasks of policy workers, we applied a task analysis method that was previously developed with council members. The approach is shown in Figure 8. An overview of the tasks was obtained through interviews with policy workers, as search tasks can be an abstract concept for users, and an open format allows the researcher to ask clarifying questions. Additionally, users might report the goals they focus on during work and forget to report sub-tasks necessary to achieve that goal.

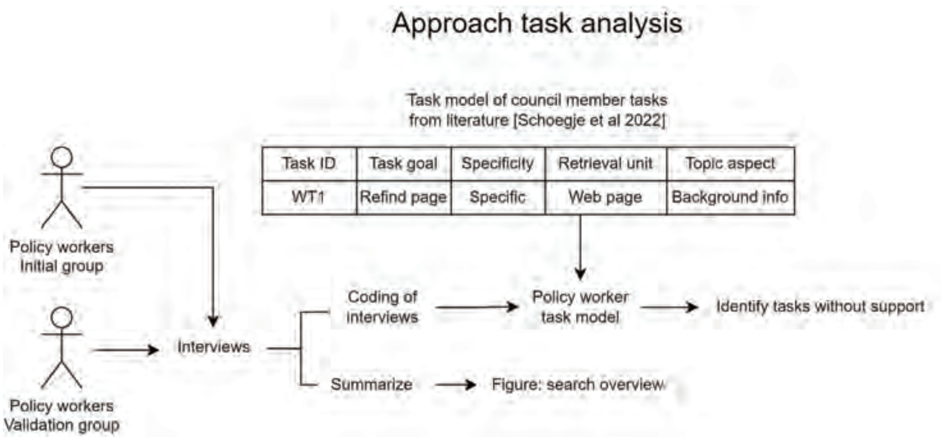


Figure 8. Figure 8: The tasks of policy workers were analysed as shown, by applying the task model that was previously developed with council members [140].

Two rounds of interviews were conducted to identify the tasks. In the first round, the work tasks and search tasks of policy workers were identified. A second round of interviews was performed to validate the findings from the first round and ask questions about how PWs approach search in general (independent of their current task).

Tasks were identified and characterised (RQ1) by performing a task analysis. The task analysis was performed by coding speech segments from the interviews, starting with a codebook developed in a study with council members [140]. The codes in this codebook are the task facets that were previously identified with council members. Another code was added for web search intention, which we introduced in the related work [125], because it is a specific descriptor of the task goal for the domain of web search.

3.3.1 Round 1 Interviews

Participants were recruited by inviting employees who work on municipal policy. Additional participants with other roles were recruited through a convenience sample. All participants had between 1.5 and 2.5 years of experience working at the municipality.

Six one-hour interviews and two half-hour interviews were conducted with eight public servants (five female, three male), including five PWs, two managers and a coordinator for municipal activities in a neighbourhood. The participants had over a year of experience in their current roles, and consisted of six early-career professionals and two late-career professionals.

Semi-structured interviews consisted primarily of identifying the participants' different work tasks and then the search tasks these involve (while eliciting information sources and frequent difficulties). The interviewer used the task facets from the task model as a structure to characterise each task. At the end of the interview, the broader topic of finding data on the web was discussed. The interviews concluded by asking the participants what would help them in a new search system. The task perspective during this first round of interviews gave an overview of user goals and how they achieve them.

Codebook analysis of the interview data was performed by the first author. The known task facets in the codebook were applied to the speech fragments. It was checked whether the codebook was stable (saturated) for describing PW tasks, or whether further unexplained themes were present in the data.

3.3.2 Round 2 interviews

A second round of interviews was performed to see if the findings from round 1 were consistent and to determine whether the codebook needed to be extended.

Participants in this round were five other PWs from the municipality (three female, two male), working in diverse domains of expertise. These users had between 0.5 and 1.5 years of experience.

Interviews were conducted in a similar manner to the first round, but an additional list of topics was prepared to investigate.

3.4 Task Analysis Results

The work and search tasks of PWs are presented based on the interviews conducted. An overview of the PW information seeking process is also presented. Finally, consideration is given to how to better support PW tasks.

3.4.1 Work Tasks

The work tasks of PWs were characterised by their facets in Table 6 (the facets corresponded to the codes from the codebook). During the second round of interviews, evidence for all work tasks was found, except for WT3 (answering council member questions) and WT6 (performing internal services). No additional work tasks were found, indicating that the most relevant work tasks had been identified. The results are discussed in Table 6 per facet.

Table 6. Descriptions of work tasks that involve web search tasks.

ID	**Task Description**	**Topic Aspect**	**Information Sources**	Retr. Unit	Complexity
WT1	Monitor my domain	Topic background	Domain-dependent	Various	Hard
WT2	Learn a new domain	Topic background	Domain-dep., Experts	Various	Hard
WT3	Answer CM questions	Topic background, policy	web	Various	Easy
WT4	Give advice on a project	Topic background, policy	web, domain-dep.	Various	Medium
WT5	Research complex problem	Topic background, policy	web, domain-dep.	Various	Hard
WT6	Perform internal service	Resource	Intranet, colleagues	Action	Easy
WT7	Maintain/ update info	Resource	Utrecht's webpages	Pages, Documents	Easy

We now discuss the results for each column. The task IDs are shown in the first column, and an informal task description is provided in the second column. The topic aspect facet encompasses various types of declarative domain knowledge in this area, which is the third column. Our findings suggest that tasks mainly focus on the detailed substance of a topic and the current policy related to it.

Information sources for each task are listed in the fourth column. Multiple tasks identified webpages as an important information source. The main webpages mentioned were those belonging to government entities (local, provincial, and national) and webpages containing public sector research, such as those from national statistics bureaus. The specific set of important webpages differed between PWs, possibly due to their different domains of work, and may even vary between individuals.

There were insufficient data to investigate if there exists a finite and authoritative list of important webpages for each domain. However, it was observed that the web domains owned by the municipality are the most authoritative sources of information, as these are maintained by the colleagues of the users.

A large diversity of retrieval units was found as shown in the fifth column. Information is sought in various forms, such as facts, documents, contact information, or datasets. This diversity is likely a consequence of the study not focusing on specific domains but rather on all PWs. This broad scope means that the results have aggregated tasks from different domains, resulting in some loss of detail on specific tasks.

The estimated task complexity is in the final column, which was validated by a PW in a clarification session. The authors found task complexity and frequency useful during system development to prioritise what features should be developed first and to decide when it is worth investing in specialised search functionality.

3.4.2 Search Tasks

Supporting PW tasks first requires an understanding of the search tasks that occur. A high-level overview of the search tasks we found is shown in Table 7, and described with more facets in Table 8. During the second round of interviews, we found more evidence for all search tasks except for ST3 (exploring colleagues' tasks), and found no additional tasks.

Table 7. High-level descriptions of search tasks.

ID	Task Objective	Task Motivation	(Web) Search Intention	Found PWs	n Times Non-PWs
ST1	Find a domain expert	Ask for advice	Informational/ Advice	5	3
ST2	Find out who works on x	Ask a request	Resource/Obtain	7	2
ST3	Explore colleagues' tasks	Avoid overlap in work	Informational/List	5	3
ST4	Find a data coach	Find all relevant data	Informational/ Advice	4	1
ST5	Re-find most recent policy	Check compliance	Navigational	4	2
ST6	Find structured data	Decision-making or giving advice	Informational/ Download	4	1
ST7	Quickly (re)find a fact	Answer council's question	Informational/ Closed directed	6	0
ST8	Find intranet page	Find info or perform action	Resource/Interact	0	2

The mapping between the work and search tasks is not obvious, unlike the findings of a similar study that was performed with council members [140]. The work tasks of PWs are more domain-dependent than those of council members. Aggregating the tasks from multiple domains mandates adopting a more abstract and generic view level than in the previous work. Because of this, we cannot use the user goals in the work tasks to add context to why a search task is performed. Instead, we add a search task motivation facet to describe the context. It paraphrases the broader purpose that users state for a given search task. We find no mapping between tasks, but note that more complex work tasks seem to include more search tasks as was also found previously [93].

The results in Table 7 are discussed per column. It starts with the facets of task ID, task objective and task motivation. The web search intention [125] is in the fourth column. We find a substantial number of references to navigational tasks, such as '(re)finding the most current version of a policy document'. Informational tasks also occur, for which users often report asking for help from a colleague. Most complex tasks are informational tasks. Only users who are the expert on a given topic seem to be willing to invest the time and effort to thoroughly search for information, rather than finding an appropriate internal expert and asking them.

There is a difference between the tasks that were mentioned by PWs and the tasks mentioned by users with other functions as shown in the final columns. Which tasks users perform depends on (at least) their role in the organisation and their experience [57]. PWs have more tasks related to finding data and facts online, whereas others have more tasks that involve finding experts (such as PWs). This reflects the role of PWs as the domain experts that eventually answer specialised questions.

The rest of the task facets is used to characterise the same search tasks in more depth in Table 8. The table starts with the task ID and an informal description of the task objective. The information channels used to complete the task are shown in the third column. The primary channels for finding experts are either networking through colleagues or consulting the Who-Is-Who (WIW) system. This is an internal social media system that presents HR data along with the user's self-described expertise. There is a large vocabulary gap between colleagues, however (even within the same domain), and not all users enter exhaustive information. It is frequently used to find pictures of colleagues that users are about to meet, but the system does not represent the users' tasks and responsibilities in a findable way.

Other PW search tasks are primarily performed using search on the web or in personal information systems (usually in email or on the Desktop). The iBabs system is used to archive policy information, although this information is also frequently accessed by users through a web search engine.

The topic aspect reflects what type of declarative domain knowledge is applicable to the task and is recorded in the fourth column. PWs may be interested in different aspects of the same topic, and when they are interested in a particular topic aspect, this changes what they understand as the relevant context. The topic aspect of interest indicates which document genres are relevant (e.g., the document subset containing background information vs. the document subset containing political discussions) and how to present search results in the interface (e.g., using speech fragments as the retrieval unit as opposed to whole documents).

Table 8. Descriptions of (web) search tasks using a faceted task classification. The dash line separates which tasks involve searching for people, and which involve searching for documents.

Task	Task Objective	Channels	Topic Aspect	Retrieval Unit	Search Goal
ID	*Informal description*	*Where do you search*	*What aspect of the query topic is important now*	*What information do you need*	*Specificity of need*
ST1	Find a domain expert	WIW, colleagues	Expertise (knowledge)	Contact info (picture)	Specific
ST2	Find out who works on x	WIW, colleagues	Expertise (tasks)	Contact info	Specific
ST3	Explore colleagues' tasks	WIW, colleagues	Expertise (tasks)	Contact info, contact tasks	Amorphous
ST4	Find a data coach	Known people	Expertise (knowledge)	Contact info	Specific
ST5	Find most recent policy	web, Desktop, iBabs	Policy	Policy document	Specific
ST6	Find structured data	web, colleagues or giving advice	Data	Dataset, meta-data	Amorphous
ST7	Quickly (re) find a fact	web, mail	Information (Fact)	Fact	Specific
ST8	Find intranet page	Intranet	Internal info/ service	Fact or page	Specific

Two types of expertise were found as two different topic aspects: the need for a person with declarative knowledge and the need for a person with procedural knowledge. The knowledge type being sought seems to affect what expert is relevant, although further study is required to understand how this works.

The retrieval unit facet indicates the structure of the information a user is looking for in the sixth column. For policy documents, it is typically the most recent version of a given document. For expertise tasks, this typically includes the person with contact information, and possibly a picture if a face-to-face meeting is likely. Tasks involving datasets benefit from additional context, such as who created it and when, and when it was last updated. We note that some of the retrieval units in search tasks did not appear as retrieval units for work tasks, suggesting the work task in question does not determine the retrieval unit of search tasks. The same goes for the topic aspects.

A large proportion of search tasks have specific rather than amorphous goals as seen in the final column. We observe that the Who-Is-Who system on its own is not able to satisfy specific expert search tasks, which seems to lead to users to falling back to networking with colleagues. There is no adequate representation of declarative or procedural domain knowledge of colleagues, making it challenging to search for these effectively.

A categorisation of the PWs themselves based on their knowledge types [74] reveals that PWs demonstrated expertise with high levels of declarative (what is it) and procedural (how to do it) domain knowledge. They possessed significant declarative search knowledge (where to find it) for tasks within their domain and varied levels of procedural search knowledge (how to find it).

3.4.3 Overview Information Seeking

In Figure 9, we show a simplified overview of where and how policy workers search. The two main findings are that (1) half of the web search tasks involve finding human sources rather than other sources, and that (2) users face an overload of information when they perform search by themselves. We discuss these points in turn.

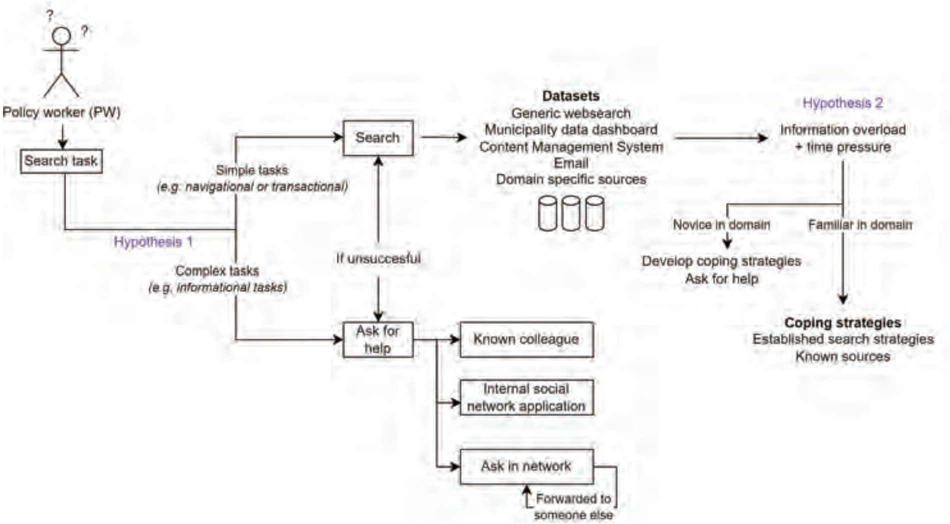


Figure 9. A simplified overview of information seeking behaviour of policy workers is presented, along with two hypotheses to improve support for their tasks. These were derived from the task analysis.

Role of Expert Search

To contextualise why half of the web search tasks involve finding human sources rather than other sources (see Table 7) we turn to the interviews. Although human sources are known to be important in organisations (e.g., [103, 135, 110, 118]), the municipality was surprised to learn that we found more tasks involving human sources than tasks involving non-human sources. Interviewees noted that this is not only an easier way to search but also yields better results for these tasks. An example of this is found in the role of data coaches. Although these know how to connect users to the correct dataset (ST4), they do not have domain expertise. Because of this, most people-finding tasks look directly for domain experts instead. The following reasons were mentioned on why PWs approach experts:

- The expert knows where relevant data are located, both inside and outside the organisation;
- The expert can give context to the data, e.g., what is trustworthy and worthwhile;
- The expert can give advice on the work task;
- The expert can help avoid performing redundant work;
- The time spent actively searching is reduced.

The reasons above help explain why finding domain experts can be more useful than searching by oneself. The decision to search for an expert has various factors, such as their perceived approachability [135] and individual preferences. We find that experts are primarily approached during complex work tasks (WT2, WT4 and WT5), whereas users do not approach them during simpler work tasks or when they themselves are the expert. We therefore hypothesise that the most important factor for policy workers is whether they perceive a task as complex. Two important aspects of task complexity are the number of subtasks [94] and the task determinability [30] (the level of uncertainty in the task processes and outcomes). If the preferable way to solve complex tasks is to search for experts, then supporting complex tasks would imply integrating expert search functionality. We note that this preference for expert advice appears more relevant in this professional context than in personal search contexts, as previous studies found that users prefer impersonal information sources (e.g., search engines) over interpersonal ones (e.g., experts) during their personal search tasks [168].

Information Overload

Another challenge is that users face an information overload once they search by themselves (PW2: "*You don't know where to search*"). This information overload is not unique to policy workers [51]. It exists because there are many information sources and because those sources often include a lot of information that is irrelevant to the PW search tasks. The best example of this is web search: although there are many relevant web resources, this is only a fragment of the whole dataset. Participants indicated the importance of authoritative sources (PW3: "*What is the official version [of this document]*?".

Novice PWs noted that the information overload is challenging, which is unsurprising because not only does it take a while to learn effective search strategies in a new domain [127] but it is also more difficult for novices to identify which search results are relevant [136]. Experienced PWs have developed search strategies and know the most pertinent sources to their domain (PW2: "*These days I know my way around*"). Over time, such coping strategies become part of the information culture of the organisation (e.g., [25] chapter 3, [79]) and are taught to novices. Even so, their information seeking still includes uncertainty, and it usually involves high time pressures [17]. For both novice users and experienced users, the combination of information overload and time pressures can be challenging. One way to deal with information overload is to remove irrelevant documents from the search index [51].

3.4.4 Conclusions: How to Better Support PW Web Search Tasks

In the first part of this paper, we characterised the tasks of PWs (RQ1), in order to find how we might better support them (RQ2). The codebook remained unchanged throughout both rounds of the analysis, indicating stability for all task-related themes. This suggests that there are no task facets missing in the task model and that the same model can be used to describe the tasks for both council members and policy workers.

Based on this analysis, we found that PWs use web search as a channel for simple (e.g., navigational/resource) tasks, and they ask colleagues for help for more complex (e.g., informational) tasks. Existing search engines are of limited use for simple search tasks because results from useful and authoritative sources are drowned out by other sources. This problem is largest when searching web sources. To better support simple tasks, we develop a web search engine containing only authoritative sources. The finding that policy workers solve complex tasks by finding experts suggest that complex tasks would be best supported by integrating expert search functionality. This leads to two hypotheses on how to improve the search experience

(RQ2):

H1 *For specific (navigational/resource) tasks during PW web search, a search engine with a focused crawl will be a more usable channel than a generic web search engine.*

H2 *For complex web search tasks (in a domain they are not familiar with), PWs will abort searching by themselves to find a colleague/ expert on the topic.*

The first part of the paper identifies challenges and opportunities in the information seeking process, resulting in these two hypotheses. Because policy workers at other municipalities have the same tasks, we expect our findings (tasks and hypotheses to support them) to be generalisable. The second part of the paper presents the development of a new search system, which is used to test these hypotheses using simulated tasks.

3.5 Search System Development

The web-based tasks of PWs were identified. Now, the development of a search engine that better supports these tasks is described. Designing a system is not an exact science, but we found that characterising the users' tasks is more useful than having the generic user stories that were typically available when designing search engines at the municipality. The approach is shown in Figure 10, where the different facets in the task model informed different aspects of the system design.

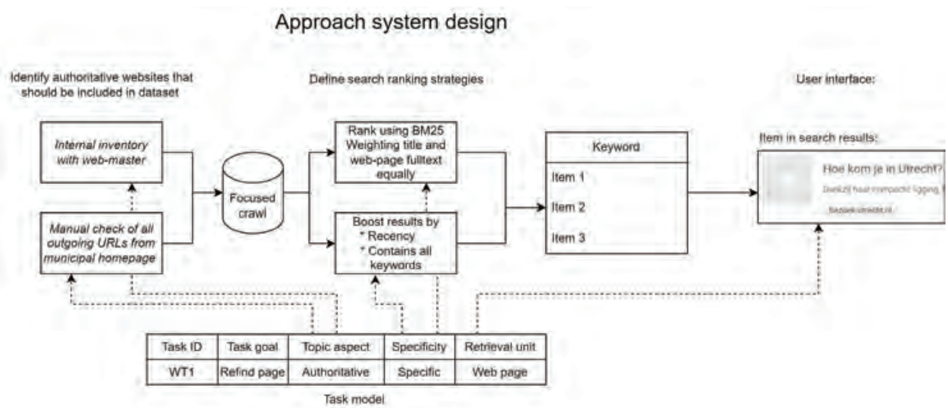


Figure 10. Development of the proposed search system was informed by the tasks previously identified. Different task facets informed different aspects of the system design.

The UI elements and ranking functions are informed by the task descriptions. The paper describes only the system elements that are relevant for testing the hypotheses.

3.5.1 Focused Crawl

A total of 40 domains were identified to be included in the focused crawl, consisting of 17 internally owned domains and 23 additional domains found by manually inspecting all 6553 outgoing links from the municipality's homepage. We consider websites linked from the city homepage authoritative. Organisations that catalogue their web resources can bypass this inventory process.

3.5.2 UI Elements

A minimal interface was developed to search the authoritative sources, where the search results display the title of the webpage, its domain, and a brief description as shown in Figure 11. The descriptions of the search results are based on the meta-data description of the webpage, which outlines its contents.

In cases where this description is unavailable, the interface provides data-driven snippets, which include the first sentence on a page containing a query term.

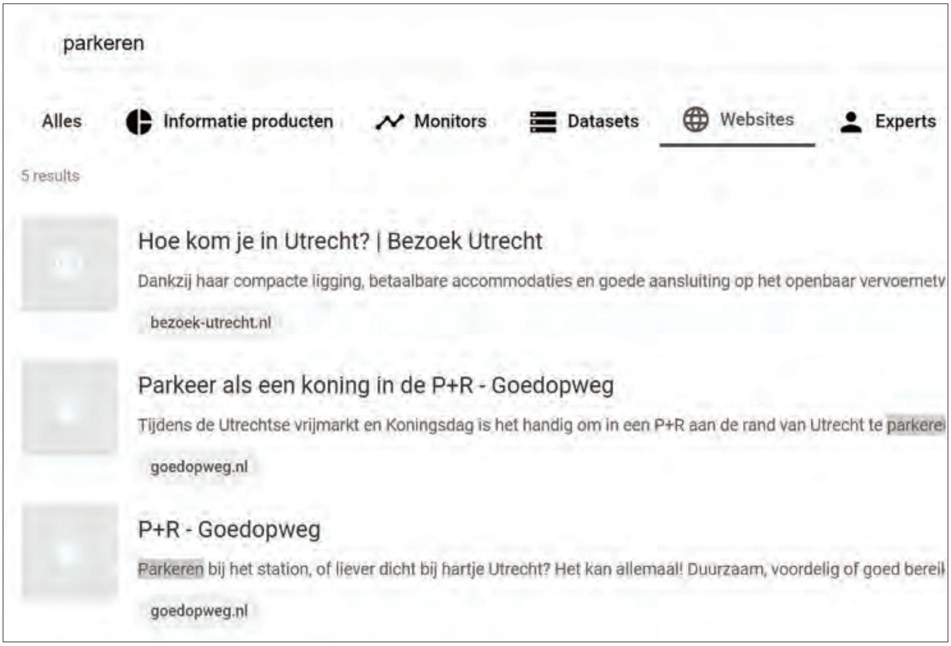


Figure 11. The document search vertical for searching in authoritative information sources on the web.

3.5.3 Ranking Functions

Authoritative sources were selected for the focused crawl. A basic BM25F ranking function was constructed based on experimentation, which weights the webpage title, keywords (if available) and fulltext content equally. Whenever the substring "municipality utrecht" was present in the query, we filtered it out (users tended to add it unnecessarily due to their prior experience with web search). To boost the score of documents based on their recency, we used a linear function. The most recent documents in the set had their score multiplied by 1.5, while the oldest documents were multiplied by 1.0.

3.6 Method

The best search engine is the one that best supports the tasks it is used for [98]. This concerns more variables than just, for example, ranking, and hence we measure the overall usability of the system for the tasks that were identified [77]. Usability consists of effectiveness, efficiency and user satisfaction [75]. These are metrics of both the task process and the task outcome as necessary for a complete task-based evaluation [144]. Additionally, basic search behaviour statistics were tracked. This methodology is similar to that of previous studies, such as the work of Arguello et al. [5]. However, unlike Arguello et al., our evaluation measured task effectiveness based on task outcomes instead of relying on self-reported measures.

Effectiveness, efficiency and search behaviour were measured by (simulated) task performance, and an anonymous post-experiment questionnaire was used to measure user satisfaction.

The effectiveness of a search engine using only authoritative sources (web domains owned by the municipality) was investigated in this study, with the potential for expanding the list of authoritative sources in the future. A challenge in this approach is the variety of systems and search methods utilised by PWs, including preference for different search engines (e.g., Google or Bing) and methods of search (e.g., web search or direct navigation to known webpages). The proposed search engine was intended to supplement the current information eco-system of the municipality and was compared to current search strategies used by PWs, such as generic web search engines, searching the municipality's homepage, or direct navigation to known pages.

3.6.1 Task Construction

Because simulated tasks should be realistic and engaging [65, 83], the tasks were adapted from real tasks that PWs have performed.

Simple Tasks (eight) were generated by asking PWs to recall and perform simple search tasks they had recently completed. Each task consisted of a sentence providing context for the work task (i.e., what the user wants to do with the information) followed by a sentence describing the search task (i.e., what information the user needs to find). For example, one task was to find a page on the municipality website with a link to the latest version of a policy document titled "Wijkaanpak Overvecht".

Complex Tasks (four) were adapted from information-seeking requests in the mailboxes of policy workers. At least one sentence of the work task context and at least one sentence of the search task context were included in each of the four tasks. These tasks were considered more complex than the simple tasks identified earlier, as they have lower a priori determinability [26, 30] and more subtasks [93]. An example of such a task is 'Suppose a new playground is planned for neighbourhood x. What is the estimated number of households and children in this area?'.

3.6.2 Task Design

Out of twelve possible tasks, six tasks were assigned to each participant. They completed half of them using the proposed system, and half using any other existing methods. These two conditions are referred to as 'Proposed system' and 'Other'. In each condition, participants completed two simple tasks and one complex task. Half the participants started with one condition, and half with the other. For every two participants, a random selection of tasks was made, where every second participant completed the ones the previous participant did not. These tasks were presented in a random order.

In the study, each task was designed to last until an answer that the participant finds satisfactory is found, a 5 min time limit is reached, or the participant gives up on searching. If the participant stops searching early, they are asked how they would proceed if they needed the information for their work.

3.6.3 Participants

Fifty people who work with policy information were invited to participate, and 16 respondents (8 female) ended up taking part in the study. Age was not recorded, but the post-experiment questionnaire showed that the participants had an average of 8.5 years of experience working on policy, with only one third of them having fewer than 6 years of experience. After the experiment, 11 of the participants completed an anonymous post-interview questionnaire.

3.6.4 Data Preparation

Results from one task were dropped due to an interruption affecting the measurements. Additionally, three further tasks were lost due to a corrupted file. The analysis presented below is based on the remaining 92 tasks (n = 92).

3.6.5 Metrics

Search task completion was investigated using effectiveness, efficiency and search satisfaction. Independent tests were conducted for each of these metrics, as they have been found to have little correlation [52, 68]. Basic search behaviour was also analysed independently.

Effectiveness was measured as the proportions of tasks that were completed correctly, incorrectly (when inaccurate results were accepted by participants), and incompletely (when participants ran out of time or stopped). In cases where a task was incomplete, participants were asked how they would proceed if they required the information for their work. Efficiency was measured as the time taken to complete a task (in seconds), with a maximum time limit of 300 s.

The system usability scale (SUS) questionnaire was used to measure user satisfaction [21]. While it does not directly measure satisfaction, it is a commonly used usability metric for test-level satisfaction, which measures usability for the entire test session rather than measuring it for each individual task [91].

Basic search behaviour was recorded, including the number of queries, clicks, and direct URL navigations. Additionally, the number of search engines used was recorded (e.g., a search result might have its own local search).

Significant effects on efficiency (in time) and satisfaction (SUS score) were tested using an analysis of variance (ANOVA). For measuring significant effects on task effectiveness, a multinomial logistic regression was used instead, as it can handle the categorical dependent variable. The independent variables in these tests were the task complexity (simple or complex) and the system used (proposed/other).

3.7 Results

Hypothesis H1, regarding whether search results are obfuscated in generic web search engines, is tested by reporting on the effectiveness and efficiency of task completion. Hypothesis H2, regarding whether users solve complex information tasks by asking colleagues for help, is investigated by asking users open questions when they give up on searching or run out of time.

3.7.1 Effectiveness

Participant effectiveness using the search engines is shown in Table 9. In some cases, users completed tasks by providing incorrect answers or approximately correct answers. An incorrect answer is defined as one where the user ended the search task, but it would not satisfy the work task upon closer inspection. For instance, users were required to identify the name of the alderman for secondary education, whereas some users found the name of the alderman for primary education instead. An approximately correct answer is when a participant used a known dataset to approximate the correct answer. Another example is when a participant found a summary of the target document rather than that document itself.

It was found that participants were less effective at completing complex tasks, which was as expected. Participants who did not complete a task were asked about their next steps to obtain the information needed for their work. All participants responded that they would ask their colleagues for help, although some considered searching for a while longer before doing so. The reasons for stopping search have been investigated in previous studies, with factors such as deadlines and the feedback of colleagues [17]. These findings are consistent with the idea that users may choose to seek assistance from colleagues to complete sub-tasks within a search.

Multinomial logistic regression was performed to compare the effectiveness of the proposed and other approaches. The results showed that both the task complexity ($p = 0.00374$, odds ratio = 0.077) and the search engine effects ($p = 0.0416$, odds ratio = 10.7) were significant factors when comparing the correctly completed tasks to the incorrectly completed tasks. When comparing the correctly completed tasks to the incomplete tasks, only the task complexity had a significant effect ($p < .001$, odds ratio = 0.0093), while the search engine did not ($p = 0.59$, odds ratio = 1.52). These findings suggest that the choice of search engine did not affect the likelihood of finding a correct result but that using the proposed system was better for avoiding incorrect results. Additionally, the odds ratios indicated that the search engine had a stronger impact on correct task completion than the task complexity.

Table 9. Results from comparing the proposed system to existing search methods. Task completion was defined as when the search ends with a correct, approximately correct or incorrect result. Additionally, tasks could also end when the user decided to ask for advice or reached the time limit. Time spent searching was reported for multiple outcomes, such as task completion and stopping search to ask an expert instead. The number of search engines and search actions were reported per task.

Condition	Simple tasks (*n* = 61)				
Effectiveness	Correct	Approx	Wrong	Expert	Time Limit
Proposed	100%	0%	0%	0%	0%
Other	87%	3%	7%	0%	0%
Efficiency	Complete	Time to correct	Avg time	Time to expert	Actions to correct
Proposed	100%	27 s	27 s	-	2.6
Other	87%	38 s	46 s	-	3.1
Behaviour	#engines	#queries	#clicks	#navs	#total
Proposed	1.00	1.19	1.04	0.38	2.61
Other	1.00	3.00	0.92	1.54	5.46
Condition	Complex tasks (*n* = 31)				
Effectiveness	Correct	Approx	Wrong	Expert	Time Limit
Proposed	33%	0%	0%	54%	13%
Other	27%	7%	20%	33%	13%
Efficiency	Complete	Time to correct	Avg time	Time to expert	Actions to correct
Proposed	16%	141 s	148 s	107 s	4.5
Other	17%	170 s	186 s	178 s	8
Behaviour	#engines	#queries	#clicks	#navs	#total
Proposed	1.00	2.00	0.75	1.75	5.00
Other	1.33	3.00	3.00	2.00	8.00

3.7.2 Efficiency

The efficiency of tasks performed in the proposed system was compared to those using other search engines to determine whether the proposed system was more efficient in supporting the tasks. Table 9 displays the results, which indicate that in the proposed system, searching was faster and required fewer actions than the other strategies. In general, complex tasks took significantly longer than simple tasks.

A factorial ANOVA was performed to determine whether tasks were performed more efficiently depending on the choice of search engine. The dependent variable was time (in seconds), and the independent variables were task complexity (simple or complex) and search engine (proposed or other). The results showed that both the task complexity ($F(1, 88) = 93$, $p << 0.001$,

η_p^2 = 0.51) and search engine ($F(1, 88) = 4.0$, $p = 0.047$, η_p^2 = 0.04) had significant effects, while the interaction effect was not significant ($F(1, 88) = 0.47$, not significant). These results indicate that using the proposed system was more efficient and that simple tasks were completed more quickly. The effect sizes suggest that changes in task complexity had a much larger impact on efficiency than the choice of search engine.

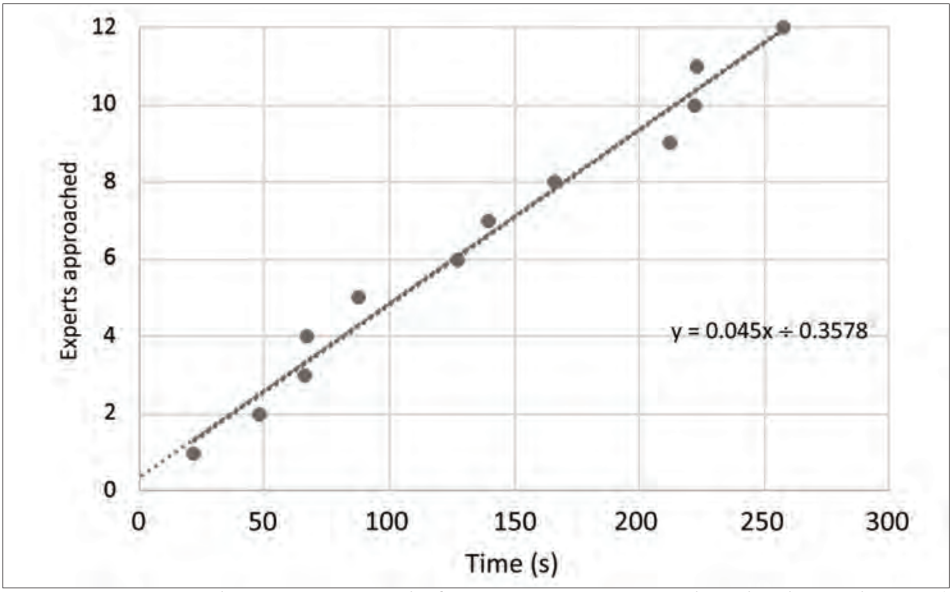


Figure 12. How much time was spent before users gave up searching by themselves and instead approached an expert.

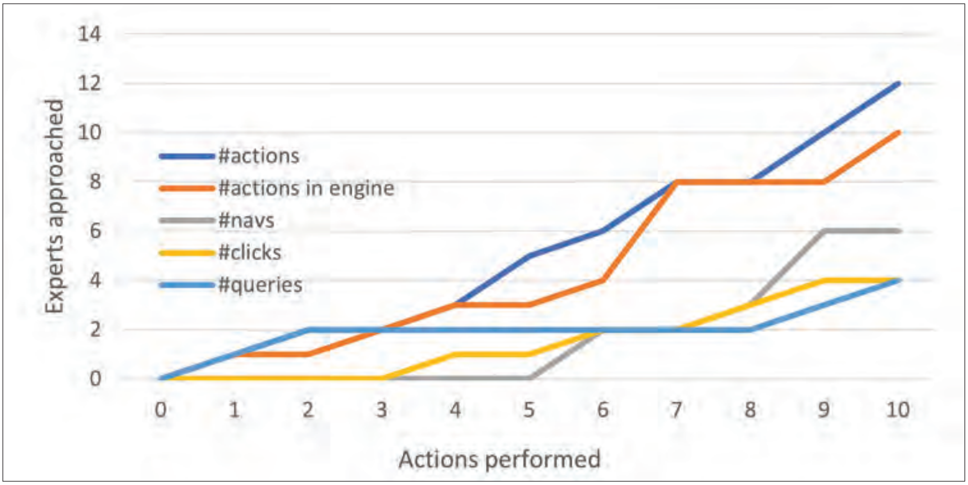


Figure 13. How many actions were performed before users gave up searching by themselves and instead approached an expert.

As time went on, users were increasingly likely to stop searching and instead ask colleagues for advice (see Figures 12 and 13), in what seems to be a linear relationship. If a task was not completed, participants were asked how they would proceed if they needed the information for their work. All said that they would find or approach an expert. It may therefore make sense to integrate expert search functionality in the web search engine, especially if a search session has been going on for some minutes. Previous studies found that participants spend longer searching during web search (a median of 10 min) [156], which may be because professional search has a lower barrier to asking for help.

3.7.3 User Satisfaction

Based on the participants' responses to the questionnaires, we found that the user satisfaction with the proposed system was average. The score on the system usability scale (SUS) was 69.25, which is an average usability score. Participants also reported an average 'overall affect' of the proposed system of 3.7 out of 5, and an 'average desire to use the system' of 3.4 out of 5.

3.7.4 Search Behaviour

The usage of basic search actions in successful search strategies was measured to consider if the system affected search behaviour. As presented in Table 9, it was found that overall fewer search actions were required when using the proposed system compared to the participants' existing methods. In general, complex tasks required twice as many actions as simple tasks, and as the tasks became more complex, there was an increase in the proportion of navigational actions on each search result.

When not using the proposed search engine, the participants used generic web search engines and/or the one on the municipality's homepage. In addition, site-scoped search engines on the webpages discovered through the SERP (Search Engine Result Page) were used. A distinction between successful search strategies using a shallow'/query-based search and a more deep'/navigational search is suggested by these results. In the former, participants rely solely on the search engine and consider only the pages on the SERP as viable results. In contrast, in the latter, participants engage more with each page by directly navigating to it or clicking links.

A factorial ANOVA was performed to test whether the search became more navigational depending on task complexity and search engine used. The dependent variable was the proportion of search actions that were navigational actions (normalised to 0–1), and the independent variables were task complexity and search engines used. The results indicated that task complexity was a significant factor ($F(1, 49) = 4.75$, $p = 0.0342$, $\eta_p^2 = 0.09$), whereas the search engine used was not significant ($F(1, 49) = 0.107$, not significant). This suggests that more complex tasks required more navigational search behaviour to complete correctly, but the search engine used did not affect this behaviour.

A significant interaction effect was found between the task complexity and the search engine, with a ($F(1, 49) = 4.83$, $p = 0.0327$, $\eta_p^2 = 0.09$. In the proposed system, a navigational strategy was a significant factor for success in completing complex tasks. It is suggested that this may be due to the limited scope of webpages available in the proposed system, which may have forced users to engage more deeply with the pages available.

3.8 Discussion

Although low sample sizes were used in both studies, the findings were consistent between two different groups of policy workers, suggesting both that findings are consistent and that we identified the most important tasks. A study conducted with council members found that tasks of a user group with the same function were consistent across municipalities [140]. If the tasks identified in this study are generalisable to other municipalities, then the quantitative findings are also expected to hold true. Additional municipalities can be supported by the search engine by including the authoritative sources relevant to those municipalities.

Overall, the findings of this study are consistent with previous literature on PW information seeking. More specific details were found about the tasks that they perform and on how to support them. It was suggested that knowledge workers, such as policy workers, would greatly benefit if their organisations improved the tools for (internal) expert search.

3.9 Conclusions

Although previous work presented both generic task models and domain-specific search engines, it is less clear how to combine these two. This paper presents how to combine the two by adapting the generic model for the purpose of developing specific applications. The search tasks of policy workers (PWs) were characterised, and a search engine was developed to improve the support provided for these tasks.

In the first part of the paper, an explorative task analysis was conducted to identify and characterise work tasks and search tasks using a faceted task model (RQ1). PWs have different challenges during simple tasks and complex tasks, and these challenges are the largest when performing web search. Existing web search engines are not effective in supporting simple tasks, as authoritative sources are drowned out by less authoritative sources. For complex informational tasks, PWs typically seek assistance from colleagues instead of attempting to search on their own. Complex tasks can be supported by integrating expert search functionality. Two hypotheses were formulated by comparing these findings to existing information systems, with the goal of improving support for PW tasks.

In the second part, a search system was developed to better support PWs, and two hypotheses were tested (RQ2). The first hypothesis proposed that simple tasks could be better supported by a search engine containing only authoritative sources. The second hypothesis states that users would abandon search during complex tasks in favour of seeking advice from colleagues. In an experiment based on real tasks, PWs performed simulated tasks to test these hypotheses.

The hypothesis for simple (navigational/resource) tasks was confirmed, as we found that the search engine with a focused crawl was overall more useful than a generic search engine. The search engine that was used did not affect the chance of obtaining a correct result, but the search engine with only authoritative sources reduced the chance of users accepting incorrect/ incomplete information. Furthermore, the proposed system resulted in lower task completion times and generally received positive user satisfaction.

The hypothesis for complex tasks was confirmed, as we found that people tend to abort searching by themselves and instead seek advice from experts. This was attributed to the added value that domain experts provide to the information found, as well as the potential time savings. Therefore, integrating expert search functionality provides valuable support for complex tasks. The municipality was surprised to find that human information sources are more important than non-human sources. All participants gave up on searching by themselves after a few minutes, and instead wanted to approach a colleague for help. The data suggests a linear relationship between the time spent searching and the likelihood of seeking help from a colleague.

As web tasks became more complex, less ‘shallow’ search behaviour was exhibited by users, and more ‘navigational’ behaviour was observed. In the former behaviour, participants mainly interacted by querying and investigating search results. In the latter, they interacted more with each page they visited, by clicking links or searching on the website of a search result.

A limitation of the work is that the qualitative studies were conducted with small sample sizes. Additionally, the search engine only addresses the web-search tasks of policy workers (because at this organisation, other search engines proved sufficient for the other tasks). There is a limit on how well the findings of this study generalise to other users and organisations because we focused on a single user group at a single organisation.

Future work could quantify how often people in other roles and at other organisations choose to abandon or skip using a search engine in favour of asking colleagues for help. If similar behaviours are found, it would emphasise the importance of properly integrating people as search results within (web) search results. An example could be to perform expert search (see, for example, [8]) and integrate the resulting experts within the (web) search results during long search tasks.

In conclusion, by conducting a task analysis of PWs, we identified tasks that lacked adequate support. A search engine was developed to better support simple (navigational/ resource) tasks of PWs, and it was found that complex (informational) tasks can be better supported by integrating expert search functionality. Search functionality for web-based search tasks can be standardised for (at least Dutch) municipalities, given that PWs at different municipalities perform similar tasks.

3.10 Chapter discussion

This chapter validated the task facets identified in the previous chapter, and explored how to determine which tasks would benefit from specialised search functionality (TRQ2). This was achieved by comparing the necessary tasks to the available functionality. Future studies could add more nuance by also identifying which tasks are important enough to warrant support based on the task's frequency, importance, and complexity, as well as the severity of the consequences if a task is not completed. The perceived complexity of task can be a useful indicator for identifying tasks that may benefit from improved support, as enhancing task support will reduce its complexity.

This chapter identified a class of tasks that could benefit from better support: complex web search tasks often lead a sub-task where people seek the assistance from colleagues with expertise. The next chapter investigates how we can better support this expert search task.

3.11 Chapter outcomes

We applied these insights during the development of a 'one-search-for-all' search engine (see Figure 1 phase 2), by first identifying all tasks of the target user group (policy workers) and then determining which tasks required further support. We found that 70 users were adequately supported in most web-based search tasks, although support could be improved for two main tasks. Consequently, the project shifted focus to support 1) web search limited to pages from trusted public sector organisations and 2) the search for internal people with expertise.

Thus, we moved from developing a one-search-for-all search engine towards creating a targeted search application (Figure 1, phase 2 to phase 3). Upon developing a prototype system, we confirmed that searching public sector web pages led to faster and more correct task completion. However, the estimated cost of maintaining the search system outweighed the estimated time savings for users[3]. Conversely, supporting expert search appeared to offer greater benefits, as the target user group is larger, more time may be saved, and there are significant benefits beyond the time that is saved. We estimated the time that would be saved with this system based on how often users reported searching, the time save within this system, and the estimated number of users.

3 We estimated the time that would be saved with this system based on how often users reported searching, the time save within this system, and the estimated number of users.

Chapter 4 Improving expert search effectiveness: Comparing ways to rank and present search results

Expert search systems help professionals find colleagues with specific expertise. These systems can present results as either a list of documents with their associated experts, or as a list of candidate experts with evidence for their expertise based on documents they authored. The type of result presentation may affect search behaviour, and therefore search task performance. Previous work has focused on ranking experts or ways to interact with the search results, but not on the effects from the result presentation

We compare the task performance of novice users using either a document-centric interface (where each search result is a document and its associated expert) or a candidate-centric interface (where each search result is a candidate expert and their associated documents). We also compare candidate-centric and document-centric ranking functions for each interface.

A post-experiment survey indicated that two variables affect participants' interface preference: the retrieval unit (candidates or documents) and the complexity (number of documents per search result). These variables influenced participants' search strategy and task performance. A quantitative analysis revealed that 1) using the candidate-centric interface results in a higher rate of correctly completed tasks, as users evaluate candidates more thoroughly, and 2) the document-centric ranking yields faster task completion. Weak evidence of a statistical interaction effect was found, preventing a straightforward combination of the most effective interface type and the most efficient ranking type.

This work resulted in a more effective, albeit less efficient, expert search engine for the municipality of Utrecht.

4.1 Chapter contribution

The present chapter considers how to design the interface for a specific task (TRQ3), focusing on an expert search task and its retrieval unit in particular (both during ranking and in the interface). This work was motivated by findings in the previous chapters, where we found that people performed expert search tasks that had 'people' as the retrieval unit, even though they said they were searching for 'the kind of document this author would write'. Changing the interface may help align the system's functionality with the user's mental model, and therefore lead to a more usable interface. The chapter is published as:

Schoegje, T., Hardman, L., De Vries, A., & Pieters, T. (2024, March). Improving expert search effectiveness: Comparing ways to rank and present search results. In *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval* (pp. 56-65). https://doi.org/10.1145/3627508.3638296

4.2 Introduction

Up to 73% of professionals in the public sector often encounter (complex) work tasks for which they seek advice from colleagues [121]. Oftentimes it is unclear for professionals where they can find a colleague with expertise on a given topic, resulting in the need to find the right expert for the right task (e.g. 'who can tell me how the sound leak in concert hall Tivoli was repaired?'). Recent work found that 59.5% of queries are conducted to find a person, based on the enterprise search logs of a large biotech company [104]. Similarly, a study on policy worker search tasks found that half of the tasks were about finding the correct person, rather than finding information directly [138]. This search strategy was employed by policy workers to solve complex search tasks, as it allowed users to acquire the information they need for less effort. Additionally, an expert could help solve one's task and contextualise the available information [138].

Previous works on expert search interfaces have considered what information is required to evaluate whether an expert is relevant (e.g. [67, 63]) and explored different ways for interacting with list of search results (e.g. [173, 59, 58, 54, 96]). However, to the authors' knowledge, no evaluation has directly considered whether expert search results should be presented as documents or as experts. We observed that, during informal think-aloud studies, participants re-framed their original search intents from a people-focused goal to an evidence-centric sub-goal: what type of documents might the person in question write? Users translated their information needs to the functionality shown in the search interface. We hypothesise that the presentation of search results affects the search strategy, and therefore task completion. In this paper, we quantify the influence of presenting search results as either documents or candidates on task performance. This can inform what is otherwise an easily overlooked and unconsidered design decision in practice.

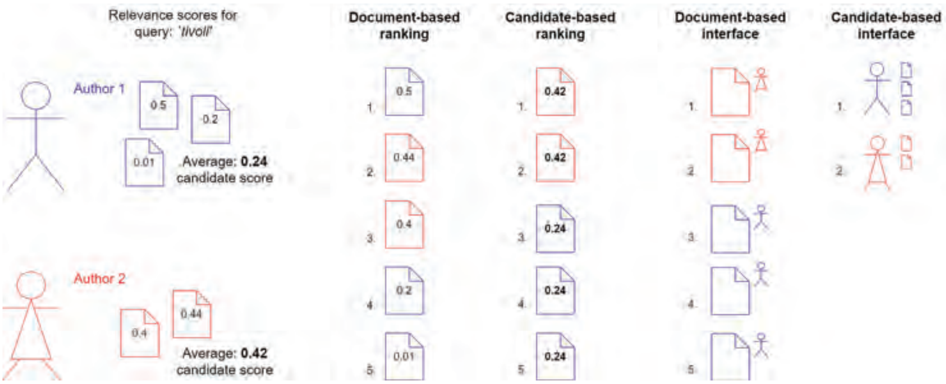


Figure 14. During expert search results both the ranking and interface can focus on the documents or on the candidates. Document-centric ranking sorts by the relevancy of individual documents, whereas the candidate-centric ranking sorts by overall relevancy of the candidate (e.g. the average of their documents).

Our scope is on expert search within the context of the municipality of Utrecht, as this lets us evaluate expert search within the enterprise search context of ‘find a colleague with expertise’. In this setting, we are interested in reproducing previous findings on how to rank experts (by document or by candidate) and then to consider how to present the search results. The two ranking types and the two interface types investigated are shown in Figure 14. Finally, we are interested in whether there are statistical interaction effects between the type of ranking function and the type of interface, and which has a larger effect size on task performance. These interests result in our research questions:

> **RQ1** *Is a document-centric result ranking or a candidate-centric result ranking preferable for finding experts who work at the municipality of Utrecht?*
> **RQ2** *Is a document-centric interface or a candidate-centric interface preferable for finding experts who work at the municipality of Utrecht?*
> **RQ3** *Are there interaction effects between the ranking type and the interface type?*
> **RQ4** *What are the relative effect sizes of the ranking type and the interface type on task completion?*

We relate this study to previous work in section 2. The dataset is characterised in Section 3, and we discuss the implementation of the system in Section 4. In section 5 our method is described which encompasses both a qualitative study and a quantitative study. Section 6 details the qualitative analysis, where two factors were found that affect how users engage with the interface: the complexity of the information and the presented retrieval unit. These variables appear to affect task performance. The quantitative results in section 7 indicate that the document-centric ranking type is faster, whereas the candidate-centric interface type leads to more tasks completed correctly. Weak evidence for an interaction effect was observed between the interface type and the ranking type, prohibiting us from combining the best interface tested with the best ranking tested. In the discussion in section 8 we argue that correct task completion is preferable over efficient task completion in this context, as approaching the incorrect expert can incur a social cost and lose time, which is not measured in this study. Based on the experiment with novice users trying to find colleagues at an organisation, the paper concludes that presenting expert search results as overviews of candidates elicits a more thorough assessment of search results, resulting in more effective task completion for novice users.

4.3 Related work

Literature on how to rank experts consists of two main approaches [8]: ranking individual documents (document-centric) or creating some model of the candidate’s expertise, and ranking these candidates directly (candidate-centric). The search behaviours of professionals can also be characterised as being either document-centric or candidate-centric. For instance, professionals perform a document-centric search strategy when they search for a relevant document and then contact the author [7]. An example where people perform a candidate-centric strategy is when they ask colleagues whether they know experts who can help them solve a task [135].

Given our domain of interest, we assume the authors of documents are experts on the topic and therefore avoid challenges in attributing expertise to the right people [8]. We note that some documents are more informative of their authors' expertise than others 75 [105], which we do not account for in this paper as it does not pertain to our research questions.

Another line of research has investigated why, and how, people search for experts [63, 173]. Such studies informed what contextual information should be included within each search result [67, 69], assisting users in their decision of whom to approach. This decision is based on both the perceived quality of the expert as well as their approachability [120]. Some of these studies note the value of presenting the search results as people as opposed to documents (e.g. [118]) and designing retrieval units suitable for the current work task (e.g. [141]).

The importance of designing interfaces has been noted in survey papers on expert search as recent as 2019 [55, 69]. Some different interface designs and functionalities have been proposed. Proposed interfaces often let users interact with the results shown (e.g. [58, 54, 96]), and sometimes deviate entirely from a traditional search engine result page (e.g. [109, 111]). There are studies that investigated exclusively ways to present a document-centric search result (e.g. [178]), candidate-centric result (e.g. [97, 120]) or entity-centric result (e.g. [61]). These are typically not directly compared, and in most works this design decision is made without explicit rationale because the research questions are focused elsewhere. However, result presentation affects how users interact with the system, as there is a relationship between the type of knowledge sought and the ideal modality of search results [119]. Studies have also found that presenting result grids or result lists affects how users examine the results [147, 126]. In addition, it was shown that the user's task affects how users engage with the interface [126], and the present authors are not aware of existing research that focused on whether results should be presented as documents or candidates for expert search tasks. Hence we reexamined this fundamental design decision of expert search interfaces, and measured the impact of this decision on the users' effectiveness, efficiency and user satisfaction while searching.

Note that precision and recall are not suitable for evaluating an interface, and hence our evaluation relies on the observation that the best search system is the one that is most useful for the work tasks of the user [13, 74, 163, 162]. This study considers which interface is most useful for expert search tasks. Usefulness is measured as useful = usability + utility [114]. All systems in our study have the same utility (i.e., they can solve the same tasks), and therefore the evaluation focuses only on which system is most usable. Usability consists of three components: the system's effectiveness, efficiency and user satisfaction [75]. This is not a novel approach to evaluating expert search interfaces (see e.g. [96]).

Effectiveness can be measured as the proportion of tasks that were completed correctly. System efficiency is typically measured in task completion time. One approach to measuring user satisfaction is the System Usability Scale (SUS) questionnaire [21]. Although it does not directly measure satisfaction, it is a widely adopted usability metric for test-level satisfaction (i.e., measuring usability for the whole test session as opposed to measuring it every task) [91].

4.4 Council Document Collection

In collaboration with one of the country's largest municipalities, we utilized a city council document collection[4]. We opted for this dataset as it is realistic to needs of expert search in an organisation, and it allows us to avoid the problem of linking a candidate expert to the evidence of their expertise. The collection comprises approximately 6000 letters and memos, which were written by around 1600 public servants who directed the documents to council members. The letters are typically two pages long and written to provide information to the city council in preparation for council meetings. Additionally, these letters may include attachments that offer more extensive and detailed information. Memos, on the other hand, are brief updates and are less informative in nature. Each document is associated with a specific sub-domain, such as public health, which users can specify in the document's metadata when uploading it. The collection of sub-domains in which a user possesses expertise is referred to as their portfolio.

The documents were written using standardized templates (created in Microsoft Word), which enables the extraction of author names from the document header with a regular expression (see Appendix B for details). Documents without extractable author names are not included in the indexing process. Documents with multiple authors are also excluded, because these documents could introduce a bias in our experimental setup (see section 5 for more detail). After grouping author aliases 1032 unique authors were found who wrote 4483 documents.

4.5 Implementing expert search

Although recent approaches to ranking (e.g. [97, 177] and presenting [16] expert search results are sophisticated, our implementation is minimalistic to maintain focus on our research questions. This section describes the design decisions that let us investigate the research questions. Further implementation details on all four combinations of ranking and interface types are in Appendix C, and the code is available at www. github.com/UtrechtUniversity/expertsearch.

4 zoek.openraadsinformatie.nl - accessed 11-9-2021

Ranking types are implemented by two elastic search[5] indexes. In the document-centric index the entries contain the full text of single documents. In the candidate-centric index each entry contains all the text of all the documents that one individual wrote.

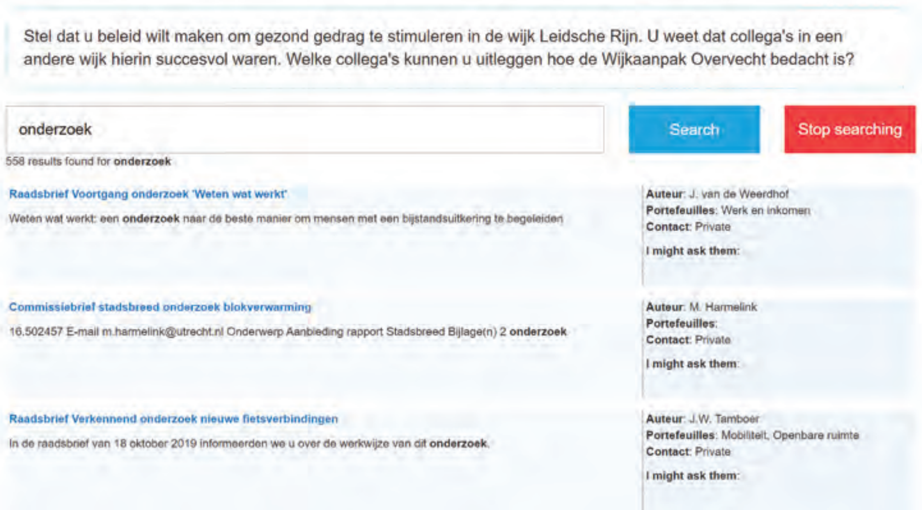


Figure 15. The expert search interface with document-centric retrieval units.

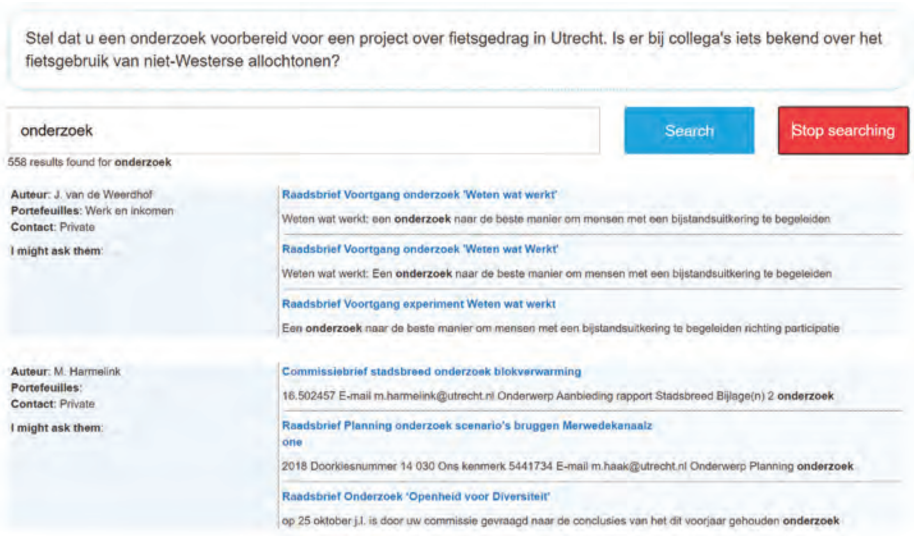


Figure 16. The expert search interface with candidate-centric retrieval units.

Interface designs are implemented by presenting each result with a document panel and an expert panel. The document panel showcases the evidence of expertise, whereas the expert panel displays the candidate's portfolio and contact information. The document-centric interface (shown in Figure 15) emphasises

5 elastic.co

the evidence of expertise, and therefore positions the document panel on the left side. If the result appears relevant, users can then locate the contact information in the expert panel. The candidate-centric interface (depicted in Figure 16) presents an overview of the candidate. Therefore the expert panel is on the left, and multiple pieces of evidence are presented on the right.

Documents are always presented using the document's title and a snippet derived from Elastic's highlight feature, limited to a maximum of 100 characters. Titles are clickable and open the corresponding documents in new tabs, to ensure users do not close the search engine tab. The author panel includes a name, contact details, and portfolio.

4.6 Method

We investigate the effects of different ranking types and interface types on task completion by having users perform simulated tasks in variations on the same search engine. This experiment was performed in person as 1) search behaviour was logged in the browser's local storage, 2) to ensure participants use the same equipment and environment, and 3) to ensure experiments were performed without distractions.

A power analysis was performed to determine the required sample size for our experiment, based on preliminary findings with the first three participants. We performed a two-factor Analysis of Variance (ANOVA) of the task completion time, while using an estimated standard deviation of 0.53 minutes, a detectable contrast of 0.5 minutes, and a desired power level of 0.946. Using these assumptions approximately 40 observations per factor combination are necessary, equivalent to 20 participants performing 8 tasks each.

Participants in the study were selected to be novices in the domain, as we observed that they faced the greatest challenges in locating both information and experts. Experienced users already know the most relevant information sources and individuals with expertise. Given that employees such as council members are elected citizens, and that no specialist training is necessary, we assume that citizens exhibit similar information behaviour as new employees.

Due to regulatory restrictions and the unavailability of public servants during the early stages of the COVID-19 pandemic, we conducted the experiment with citizens as participants. In compliance with local regulations at the time of the experiment (restriction contact outside of known social circles), we only recruited acquaintances of the first author. These were unfamiliar with the research goals beyond what was necessary for an informed consent. To ensure the safety of participants, numerous precautions were taken, including maintaining social distancing, conducting repeated self-tests, and regularly disinfecting the hardware and equipment using alcohol wipes.

Twenty participants took part in the experiment, all of whom were native speakers. Half the participants identified as women. Most participants were aged 25 to 35, with four outliers being older than 40. No participant had professional work experience in a similar domain, and none reported having any domain-specific knowledge.

Tasks were adapted from tasks policy workers reported performing at the municipality. Each of the eight simulated task starts of a work task description (i.e., the end goal of the user), which is a textual description of one or two sentences. This is followed by the search task description (i.e., information need) described in a sentence. The simulated tasks are listed in Appendix D.

Ground truth data was constructed based on the assumption that experts on a relevant sub-domain would know the answer, or would know the person to contact instead. An experienced policy worker from the municipality was available to determine which sub-domains were relevant for each task. They were not able to assess the relevance of individual experts, as they do not know the expertise of all individuals employed at the organisation.

The experimental design took into account that participants should be able to distinguish between the systems in the post-experiment questionnaire, and hence each participant tested two systems with different interfaces. The presentation order of interface types and ranking types are counter-balanced, and the task order is randomised.

If a highly relevant document was marked as relevant in the document-centric interface, all of its authors are marked as relevant. In the candidate-centric interface, the user might mark one candidate as relevant based on this highly relevant document, but not the other. To avoid this asymmetry in relevance assessments, we exclude documents authored by multiple people from the dataset.

Procedure for the experiment was to present participants with one of the interfaces and a brief introduction. After given informed consent and familiarising with the system they were instructed to imagine themselves as new employees at the municipality, tasked with assignments that required input from their colleagues. They were asked to identify and mark the candidate expert(s) whom they would consider approaching for assistance, if any. Then they performed the tasks without time limit. Each task was started and ended by pressing a button. During a task, a description is displayed and participants can check the boxes of experts they would approach. Users completed four tasks in this first system, and proceeded to complete a questionnaire to evaluate the system. Next, participants familiarised themselves with to the second system and performed an additional four tasks. After a questionnaire about this system they were presented a questionnaire that compared the two systems, and asked open-ended questions about how they search for expertise. The list of questions is published alongside the code at github.com/UtrechtUniversity/expertsearch.

Analysis of the qualitative results investigated user preferences by manually clustering and interpreting users' responses from the questionnaires. This was followed by a quantitative analysis that measures the effect of the ranking type and interface type on aspects task performance (as introduced in section 2). Systems were compared in terms of effectiveness (rate of successful task completion), efficiency (time to task completion) and user satisfaction (measured using the System Usability Scale (SUS) questionnaire [21]).

4.7 Qualitative analysis

Users' preferences for the user interface were divided, with half the users preferring one system in the questionnaire and the other half preferring the other system. We investigate the reasons for this, and whether to account for this in our quantitative analysis. One participant's data was excluded from both the qualitative and quantitative analyses as they misunderstood the instructions, and performed several tasks without issuing queries (and therefore rated the same set of results for each task).

4.7.1 User preferences

A total of 30 reasons were given by participants to support why they preferred one system over the other. After grouping similar reasons, as shown in Table 10, we found nearly all reasons pertained to the interface type. Exceptions are marked with *, but even then these were only given when participants compared two systems where only the interface (and tasks) changed.

User preferences indicated two main factors: retrieval unit complexity and perceived retrieval unit. Both interfaces represent opposites in terms of these factors, and users disagree on what is preferable. The retrieval unit complexity refers to the level of complexity involved in retrieving information, while the perceived retrieval unit relates to users' perception of the granularity and relevance of the retrieved information. Understanding these factors is crucial for designing interfaces that cater to diverse user preferences and enhance usability in expert search systems. Two main factors emerge that explain preference to one system or the other: the retrieval unit and the retrieval unit complexity.

The most reported factor to prefer the document-centric interface is that it shows less information (D1), which makes it easier to evaluate a search result (D2). This contrasts against the primary reason to prefer the candidate-centric interface: this interface gives a better overview of what a candidate expert does (C1). Participants disagree on the trade off between the amount of information needed to be confident enough of a candidate's expertise. The second factor that participants prefer the document-centric interface is that it allows them to first evaluate the document, and then use the author characteristics (i.e., their portfolio) as further evidence (D3). This contrasts with the second main reason to prefer the candidate-centric interface: these users prefer first evaluating author characteristics and using the written documents as further evidence (C2 and C3). This second factor shows how the interfaces represent two different search strategies, where one's mental model is either document- or candidate-centric.

Table 10. Reasons for preferring one search system over the other (as reported in the questionnaire). A few reasons (marked with *) are not directly about the interface, but were found when only the interface and tasks changed. These reasons could be correlated to interface type. This variable was measured between individuals, but notably none of the reasons mention different interfaces were better for different tasks.

ID	Interface	n	Reason
D1	Document	6	Simpler / not too much information
D2	Document	3	Easier to evaluate a document
D3	Document	3	I first want to evaluate the document, then the author
D4	Document	2	Less irrelevant information is combined
D5	Document	1	Focus on what one does, rather than user characteristics
D6	Document	1	More intuitive
D7	Document	1	The tasks were easier*
C1	Candidate	7	Better idea of what a user does
C2	Candidate	2	Focus on user characteristics rather than writing
C3	Candidate	1	I first want to evaluate the author, then the documents
C4	Candidate	1	Have to be less good at selecting keywords*
C5	Candidate	1	Didn't feel like I found who I wanted in the other*
C6	Candidate	1	Results were more relevant*

In a follow-up questionnaire, seven participants reported consciously modifying their search strategies. Five of these indicated they changed whether they evaluated documents or candidates first. One participant mentioned that in the candidate-centric interface, they searched for topics, whereas in the document-centric interface, they were uncertain of what to search for, and attempted searching by function titles instead. More experience with the system might have affected their search behaviour. The final participant mentioned that the candidate-centric system required them to open more documents before they were certain an author was relevant. Participant p17 succinctly remarked that ”in the [candidate-centric] interface, you find experts, and in the [document-centric] interface, you find documents”. The order of the document and candidate panels in the interface influences how users evaluate search results, as the perceived retrieval unit changes. An interesting side-note is that users who preferred simple information did not like when the candidate-centric interface presented irrelevant documents (D4). This occurred when an author had one highly relevant document and a number of tangentially related documents. Although this signals one's limited expertise, these users would prefer just not seeing it.

4.8 Quantitative study

The effectiveness, efficiency, and user satisfaction achieved with both search systems were analysed as dependent variables. The independent variables were the interface type, ranking type, and also the interface preference. This was included as the qualitative analysis indicated this is an important variable. Models were constructed in the form of *dependent_variable ~ interface_type * ranking_type * interface_preference*, with the dependent variable being the task completion rate, time spent, or SUS score. During six tasks the participants started the timer and then delayed starting the task, as they had a question to the observer. This inflated the starting time between the starting the task and the first query. To correct for this, these false starting times were substituted with the participant's average starting time.

4.8.1 Effectiveness

An overview of how many of the tasks had correct results are shown in the violin plot in Figure 17, with supplementary effectiveness metrics available in appendix E. A logistic regression tested for significant differences in the task completion rate. Interface type had a significant effect on the task completion rate (p = .044, log odds ratio = −2.25), as the comprehensive overviews in the candidate-centric interface lead to better task completion rates.

There was weak evidence of an interaction effect between the interface type and ranking type (p = .068, log odds ratio = −2.73). The candidate-centric interface performed well when there were many relevant documents per candidate (i.e., with the candidate-centric ranking) but worse when a search result included one highly relevant document as well as slightly relevant or irrelevant documents (as produced by the document-centric ranking). Participants may have ignored relevant authors when they also saw irrelevant documents.

Showing multiple documents per candidate reduces the variance in correct task completion (the distributions in Figure 17 are less tall). The candidate-centric interface appears to provide a more stable signal of a candidate's expertise, although this does not necessarily translate to more correctly completed tasks. This may be because some tasks may require a person with in-depth expertise on a topic (as evidenced by many relevant documents), whereas others require someone with experience in a very specific project (as recorded in specific documents). Further work is necessary to understand when and how conflicting information should be shown.

No significant effects were observed for the ranking type (p = .17, log odds ratio = −1.67) or other factors. The log odds ratio is greater for the interface type than for the ranking type, indicating that the interface type plays a more significant role in task completion.

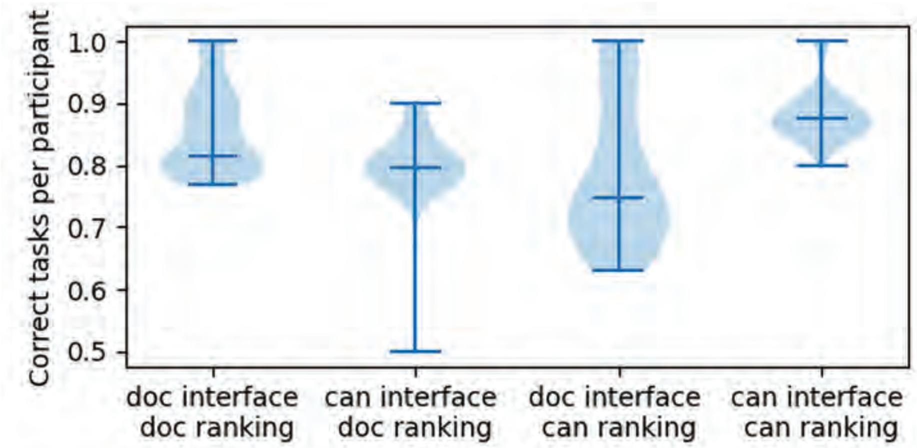


Figure 17. Violin plot showing the distribution of the correctly completed tasks. The ratio (i.e. correct tasks / total tasks) is shown per participant. For each violin plot the outliers and the median are marked.

4.8.2 Efficiency

An overview of how quickly tasks were performed is shown in the violin plot Figure 18, with additional efficiency metrics available in appendix E. An ANOVA tested for significant differences in the time to task completion. To prepare the data for analysis, we addressed a positive skew in the model residuals. The task completion times were transformed using the function *log|10(time)*. Afterwards, the task completion times no longer violated the normality and variance assumptions of the ANOVA test. The normality assumption was tested using Shapiro's test, which yielded a non-significant result ($F(3, 140) = 0.99$, $p = .32$). The variance assumption was assessed using Levene's test on the task completion times, which also resulted in a non-significant finding ($F(3, 140) = .19$, $p = .31$).

The ANOVA indicated a significant effect of the ranking type on task completion times ($F(3, 140) = 4.63$, $p = .033$, $\eta_p^2 = .035$), as the document-centric ranking lead to faster completion times. This could be attributed to finding the most relevant pieces of information, leading to more confidence during relevance assessments. This is consistent with previous research showing that document-centric rankings tend to produce more optimal rankings [86], because if the top results include more promising candidates, it can lead to quicker task completion. There is weak evidence indicating that the candidate-centric interface type slows down task completion ($F(3, 140) = 2.80$, $p = .096$, $\eta_p^2 = .022$). This could be attributed to the presence of more information that users need to parse and evaluate, potentially leading to longer completion times.

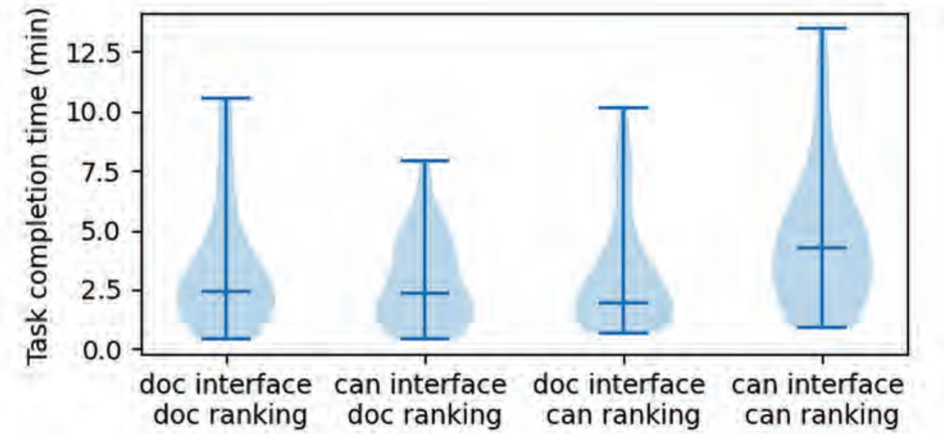


Figure 18. Violin plot showing the distribution of how many minutes each task took. For each violin plot the outliers and the median are marked.

There is also a significant interaction effect between the interface type, ranking type and user's interface preference ($F(3, 140) = 9.08$, $p = .0031$, $\eta_p^2 = .06$). This shows that users who prefer different interfaces also need different systems to search as quickly as they can. It might be that interface preference indicates which interface aligns with a user's search strategy, although future work is necessary to understand why. For example, users who want an overview of a candidate might be slower in the document-centric interface if they are looking for multiple documents by the same author. Another explanation for the same finding could be that users lose time if users lose time translating the problem 'who do I need' to the problem 'what kind of documents would this person write'. With more exposure to the system, this individual effect may reduce as users learn to employ the most effective search strategies. Hence we also consider the (non-significant) completion times of different interface and ranking types.

4.8.3 User satisfaction

The average user satisfaction as measured by the SUS was similar for all systems. An ANOVA found no significant difference on the user responses based on the interface type ($F(3, 15) = 0.047$, $p = .82$, $\eta_p^2 = .00037$), and no evidence for an effect of the ranking type ($F(3, 15) = 1.02$, $p = .31$, $\eta_p^2 = .0080$). Descriptive statistics per system are available in appendix E. Users who prefer the candidate-centric interface did provide significantly higher SUS scores ($F(3, 15) = 17.7$, $p = .000048$, $\eta_p^2 = .12$). The reason for this finding remains unclear. One possible explanation is that users who favour a comprehensive overview might feel more at ease with tasks that involve assessing the overall relevance of a candidate in general.

Additionally, there was a significant interaction effect between the user's interface preference and the ranking type on the SUS scores ($F(3, 15) = 13.16$, $p =$

.00041, η_p^2 = .094). Users who favoured the candidate-centric interface provided the highest scores for systems with the document-centric ranking, although the reason behind this remains unclear. This may be because it finds the most relevant pieces of evidence (documents). No further significant effects were found.

4.9 Discussion

Generalisability of the best ranking type and interface type can be expected at other organisations where 1) novice colleagues need the expertise of colleagues, 2) those expert colleagues document (the outcome of) their work in a shared content system, and 3) the users seek a similar type of expertise as the policy workers at the municipality of Utrecht. Our search tasks (as listed in Appendix D) seek expertise from someone with declarative knowledge ('what is it'), whereas other search tasks might require procedural knowledge ('how to do it') [73]. Future work should investigate different types of expertise, and whether searching for different types of expertise requires different types of support.

It is unclear whether current findings with novice users generalise situations where expert users need to find other experts. Although experienced employees are more likely to already 'know their way around' and find experts through e.g. their network [138], it would be interesting to see when they do struggle with expert search tasks, and whether they execute these differently. We also note that the novice users in our experiment did not make mention of whether candidate experts were still experts, or perhaps were experts in the past. We expect that experienced users would place more emphasis on finding people with recent expertise. Further work could also investigate the effect of how much experience the users have in searching (for expertise) on the preferable interface and ranking.

Limitations of analysis include that our study did not control for the type of retrieval unit and the complexity of retrieval units separately. Future work could change not only the order of the two panels, but also the number of documents shown per search result. Additionally, a post-experiment power analysis revealed that the ANOVA for efficiency was underpowered. Consequently, the analysis might have missed significant effects, and the effect sizes in statistical tests could be overestimated. This occurred because the initial findings showed a considerably lower variance in task completion time than the full dataset (0.53 minutes instead of 2.02 minutes). To achieve a power of 0.95 for interaction effects with variance we find now, the study would require a sample size of 290 participants.

User interface preferences of individuals were associated with two factors, although it is unclear what causes these preferences. This may be due to differences in cognitive styles that affect the processing of information [124] and search strategies [9, 99]. Holistic individuals, for example, tend to focus on the big picture and may be more inclined to prefer the comprehensive overview provided by the candidate-centric interface. Serialistic individuals, on the other hand, tend to approach tasks analytically in individual steps. In this study we measured the overall preferences of individuals, but we found no evidence that users preferred different interfaces for different tasks (it would be found in Table 10). The preferable user interface likely depends on the task performed, and for our set of expert search tasks user preferences appear to be stable. Future work could investigate how a task needs to change to affect user interface preferences.

Combining the optimal ranking and interface types may be impossible, as we found weak evidence of an interaction effects for effectiveness. In the tested systems we need to choose between the (more effective) candidate-centric interface or the (more efficient) document-centric ranking. We argue that, when searching for an internal colleague, approaching the correct candidate is more important, as approaching a person without expertise will lose time and potentially incur a social cost in wasting someone's time (RQ4). This argument will not hold in other expert search contexts, as others settings include hiring or selling to candidates. In these cases, the user might be more concerned with identifying true positives regardless of whether their candidate wants to be approached. This interaction effect likely followed from presenting search results that included conflicting information (one highly relevant document and additional slightly relevant documents). It may be possible to design an interface that does has no interaction effect with the ranking function, allowing for a combination of the strengths of finding highly relevant evidence (in the ranking) while also concisely displaying an overview of the author (in the interface). A step in this direction could be to show each document's relevance in the interface.

Further observations include that there are less authors than documents (as illustrated in Figure 1), and that presenting less search results in total might be preferential, especially during high-recall tasks. A final consideration is that expert search tasks can be directed at different types of expertise, such as procedural knowledge (e.g. 'how to do this') or declarative knowledge (e.g. 'what is this') (see e.g. [74]). The current study focuses on the latter, and found that a candidate-centric interface is better for finding declarative expertise. Procedural tasks tend to have different relevance criteria that can be included in the interface, such as first-hand experience [31, 33].

4.10 Conclusion

Presenting search results in interface as documents or as experts with an overall expertise affects search behaviour. Similarly, the ranking of experts by individual documents or their overall expertise affects whether the order of search results is more effective. Our study compared two types of interfaces and two types of ranking functions. The four combinations were evaluated using simulated tasks, based on task performance and questionnaires.

The document-centric interface presented documents as the retrieval unit, with the author characterised in a secondary panel. This was compared to a candidate-centric search result presentation where the retrieval unit presented was a candidate, and where the secondary panel presented up to three documents written by this author. A document-centric and a candidate-centric ranking function were implemented by indexing and searching for results at either document-level or candidate-level (the latter by appending all of an author's documents as a string).

A qualitative analysis found that users disagreed on which interface was preferable. Two variables affected this preference: the perceived retrieval unit and the complexity of the retrieval unit. Changing the retrieval unit affected how participants searched, as they first assessed the relevance of the retrieval unit in the left-hand panel and then used the right-hand panel as further evidence of a result. Whereas some users preferred the simplicity of the a single document per search result, others appreciated the overview given by the more complex candidate-centric retrieval units. Although users may prefer different retrieval units based on the users' characteristics, we suggest to instead design retrieval units that elicit desired search behaviour. The quantitative analysis investigates which interface results in successful search behaviour.

As interface preference was related to the users' search strategies it was included in the quantitative analysis. 144 tasks were analyzed, performed by eighteen participants, resulting in three main findings. The candidate-centric interface leads to higher rates of correct task completion (p = .044, log odds ratio = −2.25). The document-centric ranking leads to faster task completion ($F(3, 140) = 4.63$, p = .33, η_p^2 = .035). Finally, there are significant interaction effects between the type of interface, type of ranking and the user's interface preference.

The document-centric ranking is faster (RQ1), which may be because the top results contained more relevant candidates. The candidate-centric interface is more effective (RQ2), probably because it provides a more comprehensive overview. There was weak evidence of an interaction effect between the document type and ranking type for effectiveness, implying that it is not possible to simply match the best interface type with the best ranking type (RQ3). Instead there is a need to combine the strengths of both approaches. This would be a system that retrieves evidence with a high precision (document-centric ranking) and displays an overview of the expert (candidate-centric interface) that is not too complex. When choosing between effectiveness and efficiency, we argue that approaching appropriate candidates is arguably more important than finding a candidate expert quickly. When working with internal colleagues both saves time for the user and avoids a potential social cost (RQ4). The implications of this study for designing (expert) search systems are 1) the perceived retrieval unit and its complexity should be appropriate for the current task and user, which for expert search means that 2) users can be nodded towards a search strategy where they thoroughly evaluate candidates by presenting thorough overviews of experts, and 3) the presentation of search results appears more important than the order of search results in terms of task completion. In conclusion, our study resulted in a more effective, albeit less efficient, expert search system for the municipality of Utrecht. Similar design choices may be expected to yield similar results at other organisations where domain novice users search for colleagues with expertise.

4.11 Chapter discussion

In this chapter, we designed an expert search engine based on the task descriptions found in the previous chapter. We found that task performance was more effective when search result presentation aligned to the retrieval unit that users associated with the expert search task. This demonstrates how systems can be improved based on a better understanding of the user's task.

4.12 Chapter outcomes

We applied our research towards developing a prototype expert search engine for the municipality, allowing users to find internal colleagues based on the documents they have authored. This extends the functionality of the existing expert search system (Wie-is-Wie), which relies solely on HR-data and short biographic descriptions of expertise. While the Wie-is-Wie system excels at finding experts once you know what you are looking for specifically (e.g. 'who is the manager in this area?'), the prototype expert search system also supports non-specific questions (e.g. 'who knows about the Uithoflijn?'). Future work should focus on integrating the prototype system into the Wie-is-Wie system to combine the strengths of both systems.

Chapter 5 Reconstructing the decision-making processes of a city council based on references between documents

Council members and policy workers need to understand the long-term processes that led to council decisions. Gaining such an overview through a search engine can be challenging, as searching a complex topic can result in an overwhelming number of documents without showing how these documents are interrelated.

This study investigates how to create an overview of a decision-making process, which may be integrated into a search engine. Interviews indicate that policy workers consider documents relevant to the overview when both the document and the proposal were created in response to the same council decision document. We identify such provenance based on the co-citation of documents and textual references between documents.

In an exploratory user study, policy workers were tasked with understanding the development of policy proposals based on provided timelines. Their relevance assessments showed that our approach nearly exclusively finds relevant documents (a precision of 0.97). While the proposed approach identifies 91% of references made in documents, it only finds an exact target document in 39% of the total references. A further 52% of references finds a subset of documents including the target. A human in the loop can aid in finding the exact documents, and potentially add documents based on their domain expertise. The proposed approach creates an overview of a city council's decision making process on a given topic with high precision, and could be applied to other domains oriented around a similar decision-making process.

5.1 Chapter contribution

While identifying the tasks of knowledge workers at the municipality, we found that some tasks are not centered around retrieving individual search results, but rather on understanding how these results fit into the bigger picture (see Chapter 2). We investigate how to structure and present information when the user's task is not merely to retrieve (a subset of) information (TRQ4). This chapter demonstrates how an understanding of search tasks can result in enhanced functionalities for search engines that extend beyond typical search functionality.

The chapter is published as:

Schoegje, T., Hardman, L., De Vries, A., & Pieters, T. (2024, June). Reconstructing the decision-making processes of a city council based on references between documents. In *Proceedings of the 25th Annual International Conference on Digital Government Research* (pp. 525–533). https://doi.org/10.1145/3657054.3657116

Acknowledgements
We are grateful to the contributions of Jos Nicolai and Anne de Wildt to early discussions of this work, and the code written by Jos Nicolai to parse council information.

5.2 Introduction

City council members create policies over multiple council meetings, building upon the existing policies. Previous policies may have been created years ago, by different council members. For instance, after deciding to construct a concert hall it takes years to construct and find out that the sound leaks between its music rooms. In this case, new council members first need to understand the original construction plan. They could use search engines to find relevant fragments of information (documents and meetings), but council members also need to understand how these fragments fit together and see the bigger picture [141]. Although doing this through searching individual documents would eventually lead to a complete picture, council members need to make decisions under external constraints such as time pressure [141]. Providing an overview of council information is invaluable, as such constraints limit how much information will people gather before making a decision [17] and could therefore lead to sub-optimal decisions.

Civil servants recognise the need for an overview of information, and therefore manually create timelines of complex policy proposals. The authors of timelines have no guidelines on how to create such histories. In section 4 we show how these represent the best effort of an individual, but are typically created with subjective inclusion criteria and are incomplete. Modern technologies enable a digital transformation of how we plan, record and archive the decision-making process of city council. This results in 94 more transparent decisions and clearer accountability. In this paper we investigate how to generate timelines that complement, and eventually may replace, the manual timelines. Additionally, we investigate how to design an overview perspective (interface) based on these timelines. Our research questions are:

MRQ *How can we create timelines of policy information?*
RQ1 *Micro-level: What items should be included in the timeline?*
RQ2 *Macro-level: How should the timeline be structured?*
RQ3 *How can we algorithmically identify documents that should be included?*
RQ4 *How should generated timelines replace manual timelines?*

Drawing from informal interviews with the authors of manual timelines we develop an approach to generate timelines. We use two strategies based on extracting two types of links between documents. The first method identifies during which meetings the same (near) duplicate documents are discussed (see Figure 19). We interpret this as evidence that both documents are relevant in the same context, as determined by the staff that prepares the council meetings. The second method finds textual references in documents to other documents and meetings (see Figure 20). We interpret this as evidence that the referenced item is relevant for the current document, as determined by its author. In a user study we evaluate whether these two methods yield timelines with relevant information, and explore what qualities make for a good overview of policy information.

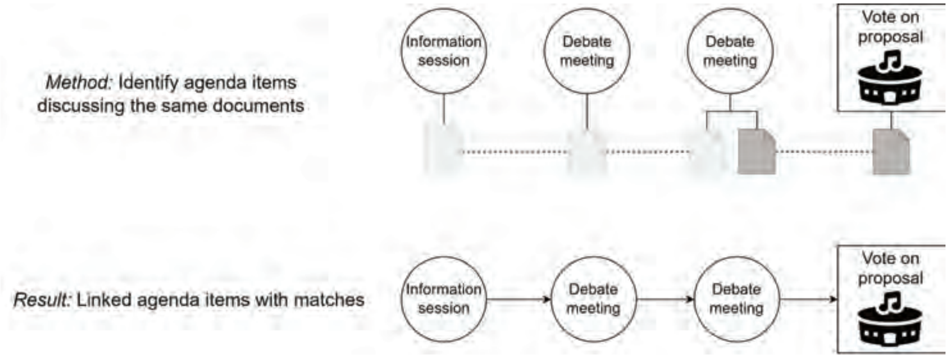


Figure 19. Plotting how document re-use can be used to generate timelines of council proposals. This example shows how the council might vote on building a concert hall, and how the light grey and dark grey documents can be used to link these meetings.

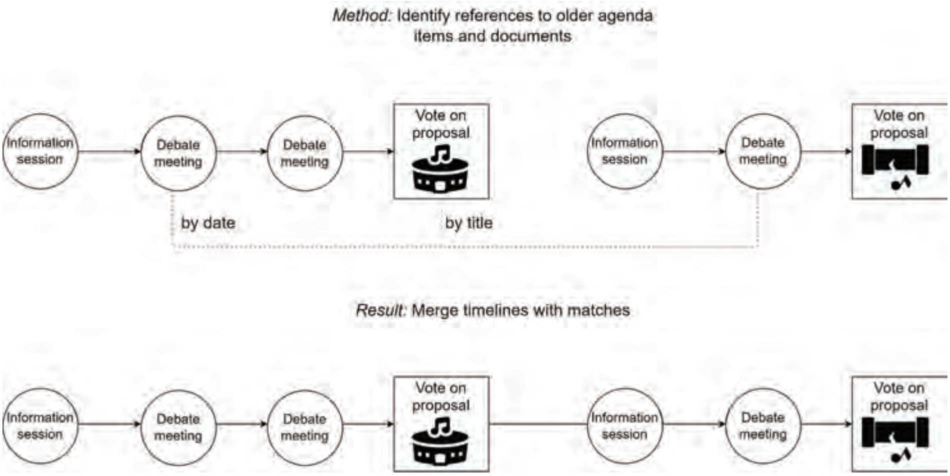


Figure 20. Plotting how temporal expressions can be used to find out which timelines should be merged. This example shows how the council might first vote on the construction of a concert hall, and later on a sound leak between the studios.

The main contribution of the paper is in characterising the need to generate timelines for council members, and in proposing a solution. This starts in section 2 by introducing related literature and introduces the council information dataset in section 3. Based on informal interviews, section 4 describes how experts manually construct timelines. Then two approaches to generate timelines are introduced (and combined) in section 5. In section 6 we present a user study where users are tasked to understand the development of a policy proposal based on a provided timeline. This study serves a dual purpose: to evaluate the relevance of documents selected by our method for a policy proposal, and to establish guidelines for timeline creation. We found that documents should be organised around their provenance (i.e. history), and that they are relevant to a proposal when both have origins in the same council decision. The generated timelines achieved a high precision (0.97) because the references between documents reflect their provenance. This precision score reflects that we do not need to account for weak links between documents, as references between documents are only created when they are directly related, according to staff with expertise.

The main limitation of our approach is that for ambiguous references we find multiple candidate documents, rather than just the intended target. This could be resolved in future work, or with a human in the loop. We conclude in section 7 that other decision making processes that use shared meeting planners may also benefit from provenance based timelines. An overview of a decision-making process makes the decisions more transparent, explainable and useful.

Transparent decision-processes tend lead to better results [72] and clarify who should be accountable for decisions. Making contextualised decision information easily accessible also lowers barriers to involving citizens [107, 43]. These factors enable an open government [18] and the quality of information available, which fosters trust [81, 92].

5.3 Related work

5.3.1 Timelines

To generate timelines we can adapt techniques from the recent survey by Norambuena et al. [115]. They show works at three different levels of resolution: sentence level, document level and cluster level. Techniques that model events at document level are of interest to our setting, as council decisions are recorded in official documents. An influential approach to map out the narrative threads in a corpus is the metro maps approach [146, 145], which constructs maps of interconnected narrative threads. Links between documents are based on the similarity of documents, determined by identifying important words in the corpus. Other approaches are based on extracting entities and/or events and then temporally ordering them (e.g. [108, 88]). The structure of council information enables two further approaches: following co-references and textual references to older documents.

5.3.2 Co-reference approaches (approach 1)

When multiple council meetings discuss the same (near) duplicate document, we can view this as one meeting referencing/building on an older meeting. Such citation patterns have been analysed in the scientific literature (see e.g. [46]). Timelines of policy proposals reason back in time (i.e. how did this political decision follow from the previous ones?), similar to the bibliographic coupling approach for citation analysis.

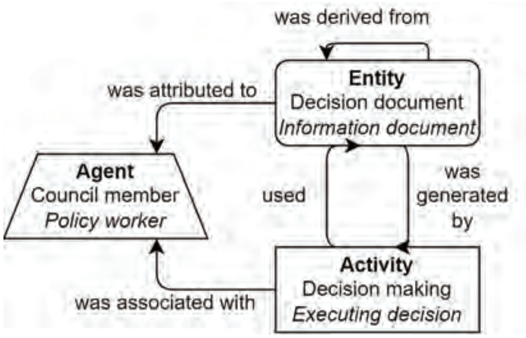


Figure 21. An overview of the key concepts in provenance, adapted from Groth et al [56]. In normal text, a council member example. In italics, a policy worker example.

A more comprehensive way to record references is to track the provenance of documents. A provenance model was proposed that records the process of how information artefacts are created, and from which information it was derived [56]. Figure 21 gives an overview of the important concepts, with two examples: the city council acts as an agent that performs the activity of making decisions, and generates their decisions which are recorded in entities such as motions and council proposals. These decisions lead to new activities performed by civil servants (agents), and outcomes are reported in documents (entities) sent to the council.

5.3.3 Identifying in-text references (approach 2)

Several approaches have used textual document references to identify how a narrative developed [19, 157, 158], often relying on URLs. Textual references can be less specific references, where multiple documents qualify (e.g. there are multiple documents on that date). This can be resolved by finding the candidate articles, and selecting the intended target based on further context (e.g. topic similarity) [19]. When establishing links between documents using both co-citation and in-text references, a graph is 98 formed where documents may be linked despite being multiple steps removed. We can model the strength of a link between documents by assigning a weight to the edges between documents is, which allows us to account for how strongly documents are linked through network analysis.

5.4 Dataset

The dataset is the public council information of the city of Utrecht in the Netherlands. This dataset reflects the two active responsibilities of council members: to shape new policies, and to oversee whether the municipality properly executes those policies. These responsibilities are recorded and carried out during weekly council meetings. The structure of council information is shown in Figure 22. During each meeting multiple items are discussed, typically concerning a variety of topics. Meeting items have documents attached which provide pertinent background information, as collected by council clerks.

Council documents are pdf files that come in various genres. The main ones being council letters, motions, memos, decision histories and council proposals. Council letters generally inform the council on upcoming matters. Motions are discussion points prepared by council members. Memos update the council on small matters. The manual decision histories were constructed by clerks for complex political topics, when it was necessary to give council members a better overview of the temporal context. The outcome of the policy making process is a policy proposal. This type of document contains the policy that was decided, and is voted on by the council. The dataset includes all data from 2013 (the origin of this system) until 2022. There are 1648 meetings held between 2013 and 2023, containing 15,314 agenda items and discussing 29,229 documents.

In addition, the municipality can send the council documents that are not tied to an agenda item, but instead uploaded as 'entries'. Each entry is a council document that may have attachments. Whereas the council has staff that select the documents attached to meetings, the entries originate from civil servants writing to the council. The dataset contains 19,156 entries that contain 22,908 documents. Entries and meeting documents can overlap, when entries are re-uploaded to discuss during a meeting item.

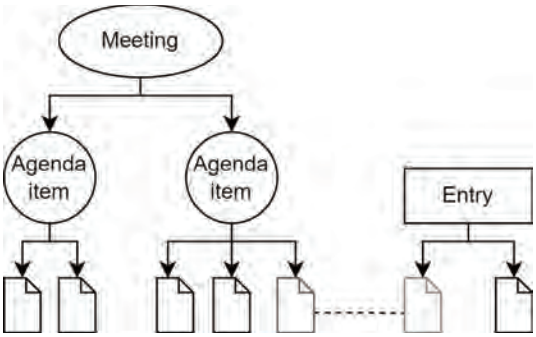


Figure 22. Documents are organised around meetings and around entries (documents from civil servants). These sets overlap when entries are discussed during a meeting.

5.5 Manual timelines

The authors of manual timelines (and the authors of this paper) are not aware of existing guidelines on constructing these timelines, and therefore create them at their own discretion. Consequently there are differences in, for example, which document genres the author includes (only council proposals, or perhaps also council letters or motions). We conducted informal interviews with six civil (three female, three male) servants who submitted a manual timeline to the council to see how and why they constructed their manual timeline. Two participants were project managers who had delegated the creation of the timeline to a colleague (30 min interviews). The remaining four were authors of the timelines (50 min interviews) themselves. Authors worked in administrative roles, supporting policy creation. All interviewees were mediors or seniors.

While every council proposal includes a brief overview of key decisions, for complex topics the authors of manual timelines can decide to provide a more comprehensive overview. This begins with the author's domain knowledge of key items, and is extended through search and asking colleagues for help. Although the authors have expertise in supporting policy makers, they are not necessarily domain experts. The reported inclusion criteria for items were subjective, as authors 1) determine what is important enough to include, 2) possess limited knowledge on what information exists and 3) may have ulterior goals when creating a timeline. For instance, one interviewee created an overly comprehensive document to emphasise how long the council had been undecided about a topic.

A recurring theme was that authors reference important documents when writing to the council. There are two types of references between documents: 1) co-citation of documents within the meeting planner, created by clerks attaching important documents to meetings, and 2) textual references in documents, used when authors describe why they are writing the council.

In summary, manual timelines are created for diverse purposes, in multiple formats and based on subjective inclusion criteria. Consequently, these do not reflect an objective or complete perspective. An algorithmic approach presents a more scalable alternative, capable of mitigating subjective factors.

5.6 Implementation

Our approach to generate timelines is based on the two types of links between documents. Both types are combined for a more comprehensive timeline.

The overall approach is to interpret each agenda-item and entry as a separate timeline, and progressively combine timelines. First we merge timelines where the same document is cited, by grouping timelines with (near) duplicate documents (approach 1). Then we identify the textual references between documents, and merge timelines that refer to each other (approach 2). Finally, we sort the timeline based on the meeting date (for agenda item) or upload date (for entries).

The source code we created for this project is available at github.com/ UtrechtUniversity/ expertsearch, and the dataset can be accessed through zoek. openraadsinformatie. nl (last accessed 11-9-2021).

5.6.1 Re-use of (near) duplicates

Co-citation of documents is found by identifying agenda items and entries containing (near) duplicate documents. Duplicate detection is performed by finding documents with the same filename and/or the same display name in the metadata fields. For this set of documents it was ensured that all documents have a similar filesize, defined as not deviating by more than 5% from the average filesize in the set. This resulted in 43507 unique documents. Before grouping, filenames and display names were normalised by removing file extensions and any trailing white spaces. The requirement for similar file sizes prevented cases where a generic filename (e.g. 'Proposal.pdf') refers to completely different documents.

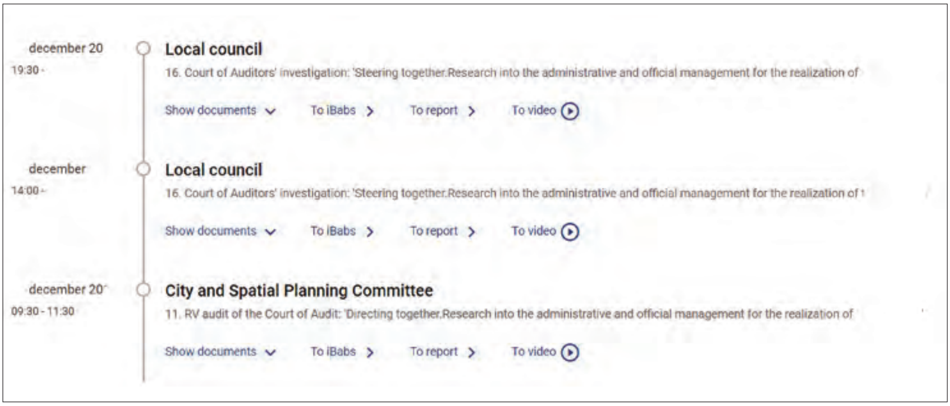


Figure 23. The resulting timeline as shown in the 'U-reka' search engine for council members (ureka.utrecht.nl/app/). All Dutch text was translated through Google translate.

The algorithm models each agenda item as a small timeline, and for each (near) duplicate document a new timeline is created that merges all timelines that co-cite it. If the original timeline contained multiple duplicates, then the same timeline will be merged into multiple new timelines. Therefore an iterative process is started to identify and merge timelines that contain new duplicates. This yielded 3,006 timelines that are especially strong at showing the history of individual policy proposals, including the weekly meetings and documents.

5.6.2 Textual references

Figure 24 shows the processing pipeline for textual references. There are three types of textual references, which we define and extract as follows:

References by ID are very specific, and refer to a unique document. ID extraction is done by detecting hyperlinks in the text using pdfminer in python [148]. The contents of the URL indicate what type of meeting or document is referenced, as well as its unique ID.

References by title are less specific, as documents may share titles or have nondescriptive and generic titles. Title extraction is performed using regular expressions to identify text strings enclosed with quotation marks (e.g. ’ ” ‘). As document titles were usually shorted, we tested if each string is a substring of an existing document title.

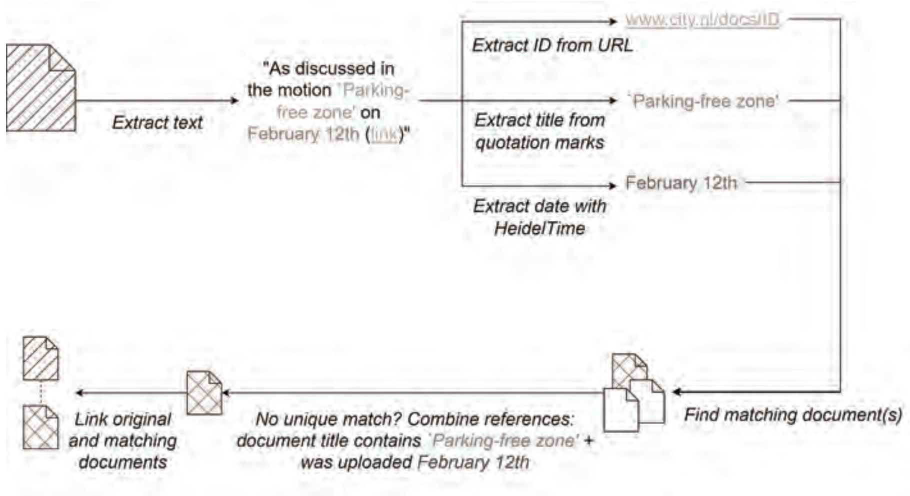


Figure 24. The processing pipeline for textual references to documents and meetings.

References by date are the least specific, as sets of documents are typically submitted as a batch on the same date. Additionally, it is ambiguous whether a date refers to a document or a meeting. Dates were extracted using HeidelTime [151]. Relative expressions such as ’yesterday’ were normalized to the document’s upload date. Because documents are uploaded in batches, the exact target is typically ambiguous. Additionally, it can be unclear whether a date refers to a document or a meeting.

Ambiguous references (e.g. multiple documents are uploaded on the same date, or share the same title) are disambiguated to the target document using other references within the same sentence. We ignore references that we cannot disambiguate to a single target to maintain a high precision (rate of true positives) in our timelines. This increases user trust, and it prevents the co-citation approach from including irrelevant documents.

The resulting timelines typically show the progression between council decisions, on a time-scale of months to years. The textual references allowed us to merge the 3,006 timelines found through co-citation into 2,751 timelines. There is an overlap in the links found by both types of references, but the co-citation approach is particularly adept at finding short-term connections whereas the textual references are better at finding long-term connections. On average the timelines consist of 4.55 agenda items and entries, spanning a period of 12.5 months. Two outliers spanned 64 months and 32 months. Two others spanned less than 1 month.

Table 11. How many unique references were annotated, how many we were able to extract, and how many we were able to disambiguate to a single target document.

	By ID	By date	By date
Total	100% (45)	100% (46)	100% (46)
Extracted	100% (45)	93% (43)	80% (37)
Disambiguated	82% (37)	22% (10)	9% (4)

5.7 Technical evaluation: identifying references

We evaluate what proportion of in-text references we successfully extracted from documents, and how many of these we disambiguated successfully. We annotated all textual references in ten randomly chosen manually constructed timelines (as other document genres contained fewer references). These annotations were compared to the extracted references.

Table 11 shows that we successfully extracted on average 91% of references from the documents, but that we were only able to disambiguate 39% of references to an exact document. References by URL were identified and trivially disambiguated, but a part of these include dead links (to pages no longer in the dataset) or link to pages outside of the dataset. References by date and title often refer to a subset of documents, including the target document. In the following user study we investigate whether the references that we found resulted in relevant documents.

5.8 Qualitative evaluation and exploration

We performed an exploratory study with a dual purpose: to investigate the qualities of a useful overview of policy information, and to assess whether the documents found by the proposed approach are relevant.

During interactive sessions users were provided with a timeline and tasked to understand what led to a policy proposal. They also assessed the relevance of each document in the timeline. Each timeline consisted of a chronologically ordered list of entries and agenda items.

Participants were presented different types of timelines to investigate the qualities of both manual and generated timelines. For four proposals, the timelines were created by combining the manual and generated timeline. One further proposal only consisted of generated items and the last one of only manually selected items. To investigate inter-rater agreement, four participants were assigned identical tasks to previous participants. These timelines were between 6-12 items long (8.5 on average), with one outlier containing only 4 items.

Participants and tasks

As council members were unavailable, we invited policy workers. These are domain experts who work with council information (for more detail, see Schoegje et al. [139]). Ten policy workers (six female, four male) were invited to participate. Four participants had over five years of experience, three had between five and one years, and three had less than a year. Nine of these sessions were conducted in person, and one over Microsoft Teams.

Participants were invited for an interactive session (30-45 minutes) where they were tasked to use a timeline to understand what led to a policy proposal. Six policy proposals were selected, each accompanied with a timeline. The proposals were randomly selected, although one proposal was replaced with another, because its timeline was too large to discuss during a single session.

Procedure

After introductions and securing informed consent, participants were directed to read the policy proposal prepared on the screen. Participants first explained to the interviewer what the proposal was about, and then chronologically started reading the documents in the timeline. Per document, the participant was first asked to clarify its contents, and what happened in between this document and its predecessor. They were asked whether the document was relevant, and whether it was useful for understanding the policy proposal (both on a three-point scale). They were asked to give reasons for these assessments.

Table 12. Summary of the main themes and findings in the qualitative study

Theme	Main findings
Inclusion criteria	Structure timelines around tracing document provenance Correct provenance leads to relevant and useful documents The proposed method finds documents from the correct provenance
Completeness & conciseness	Include a view on only decisions and a view that includes the steps in between decisions
Overview perspective	Layered view with both decisions and decision provenance perspective Show how timelines intersect and interrelate Overview perspective aids in understanding big picture Overview perspective aids in finding holes/curiosities in big picture Avoid query rephrasing challenges of drifting key words
Comparing timelines	No qualitative differences found between generated and manual timelines

Within this structure, the interviewer would allow room to discuss themes and questions that arose. In the early interviews, these themes were primarily about individual documents and what made them relevant or useful. In later interviews, discussions shifted about the timeline and policy-making process as a whole, including themes about completeness and conciseness.

Analysis

All interview responses were analyzed with a thematic analysis, by grouping responses based on recurring themes. In multiple iterations the themes were refined to be more descriptive and better reflect the participant responses.

5.8.1 Results

Three main themes emerged: inclusion criteria; completeness and conciseness; and overview perspective. These themes and the main points are briefly summarized in Table 12.

Inclusion criteria [RQ1]: In this setting it was not important to find whether a document was relevant to the topic or useful to the task. Instead, a document should be included in the timeline if both the document and the associated policy proposal can trace their provenance (as introduced in section 2.2) from the same council decision. This provenance is not explicitly tracked, but it is of such importance that authors of council documents are consistent (and trained) in referencing to previous decisions and documents. Documents typically state why the council has to read them. The importance of council decisions in particular were reflected in every interview, as shown e.g. when P3 stated "*[this document] is important because the council apparently has thoughts about this*". P5 termed the document types which contain council decisions as "*milestone documents*", and references to these milestones when persuading other civil servants to take their requests seriously (e.g. in emails).

Participant P4 illustrated the importance of provenance over relevance, stating that "*Sometimes the council asks questions about a different topic*". In these cases "*There can be a whole [internal] discussion on [which civil servant] should address that question, but we don't want to record that*". Such semi-related documents need to be included "*in the same package*", to keep track of "*who is responsible for it*". P9 was reluctant to use relevance as an organising principle as "[something] can be relevant, but it's a side issue", and "there comes a point where everything is relevant to everything".

Participants rated nearly all documents found by our approach as relevant, and no difference between any document's relevance or usefulness. The only documents that were not relevant were 1) a single entry that was incorrectly included because the duplicate detection malfunctioned (and the co-citation approach included an non-duplicate document), and 2) a few agenda items during administrative meetings. These administrative meetings should not be included in the timeline, because these same items were discussed during a separate meeting later that week. After excluding this type of meeting from the generated timelines, 97% of timeline items were found to be relevant. This precision score reflects that we do not need to account for weak links between documents, as references between documents are only created when they are directly related, according to staff with expertise. Participants always agreed in their assessments of items, with the exception of P5. P5 only considered council decisions as relevant, and other documents (those of an informative nature) as semi-relevant. All participants gave input around this theme.

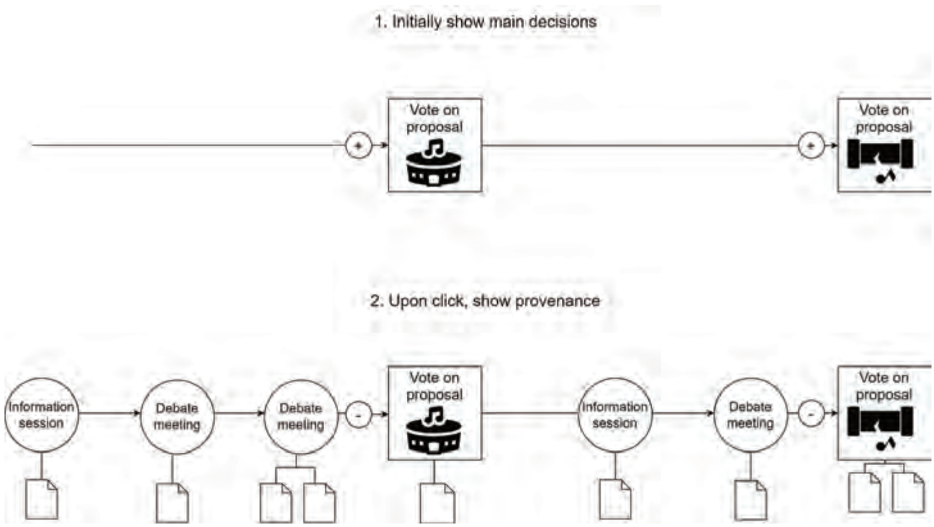


Figure 25. Example of how an interface can show information in two layers. The main decisions are initially shown, and the user can click the small circle with a '+' symbol to expand the view. The expanded view also shows the sub-steps within that decision making process. This second layer of information shows the provenance of decisions and documents.

Completeness and conciseness [RQ2]: Participants noted both the importance of conciseness (P6: "*Less is better*") and completeness (P4: "*Maybe I'm too careful, but I want the complete picture*"). P5 described how these relate: "*There is information at two ranks. Decisions are at the first rank ... informing [documents are] at the second rank*". They suggested a layered overview, which initially shows a timeline of the decision layer, which can be expanded to also show a provenance layer by toggling a button. The decision layer displays the main items, and expanding the provenance layer displays the sub-steps. This is illustrated in Figure 25 The decision layer encompasses the decisions made by the council, as recorded in motions, policy proposals, formal questions to alderpeople and formal promises by alderpeople. These result in duties and activities from municipal staff, which in turn lead to documents whose provenance is displayed in the provenance layer.

P5 suggested that the decision layer should support people who "*primarily need to know what the council decided, and the current state of affairs*". As such, the decision layer should include which decisions have been resolved, and the latest information on those unresolved. P5 noted that such information also serves as a form of accountability, where the staff shows "*we haven't been idle*". Note that 'latest information' refers to the latest official document sent to the council, as more recent working documents "*are usually not ready yet to show [to the council]*" (P10). Conversely, the provenance layer is more comprehensive, providing background information and showing how individual decisions were made. P10 noted the history of individual decisions can be important "*when something seems awkward*" about them. All participants had responses about this theme.

Overview perspective: Although some decisions follow a linear process, P8 highlighted that "*sometimes multiple [time]lines converge, and I can't see if [these documents] are in the same line*". Therefore, overviews should display the different lines, and how they interrelate (akin to the metro maps timelines [145]). P6 shared their experience tracing document references without an overview perspective, noting that "*[they] couldn't see the forest for the trees*". P10 cited an example where this approach took them 2-3 hours to understand what led to a policy proposal. P7 appreciated that a timeline draws attention to time gaps between documents, which they considered important clues that "*something might be missing*" and could prompt further investigation. P4 commended on the robustness of the timeline against changes in the keywords in documents, highlighting an example when "*the building changed names*" which would be less obvious when searching by keywords. Seven participants had responses about this theme.

Comparing types of timelines: No obvious differences emerged between participants who used different types of timeline (manual, generated, combined).

5.8.2 Discussion

We reflect upon the practical implications, and then the theoretical implications of the study. We find two themes with regards to practical implications:

How to structure council decisions: The importance of provenance explains why authors of council documents are consistent in referencing past documents in the text of their documents. This practice ensures that we can generate timelines based on references to documents (RQ3). Although previous work typically generated timelines based on document similarity, our approach can leverage the provenance information 109 that is provided explicitly in the council's work processes. The importance of provenance in this setting suggests that similarity based approaches are less appropriate for decision-making processes. Provenance ties into the main responsibilities of council members: shaping policies that generate the municipality's activities, and overseeing the execution of those activities. Hence a timeline should be presented as an overview with two layers: a decision layer that conveys these two things (decisions and current state of activities), and a provenance layer which includes more comprehensive information.

Investigating decision-making: To quantify the value of an overview perspective, future work could investigate how much information is available at the moment of making a decision (with or without the overview), and how the availability of information affected the decisions that were made. The establishment of reliable overviews of council decisions also facilitates further research on the nature of those decisions. For instance on identifying critical decision moments [34], and whether these moments are more likely to arise when the topic is discussed in information sources like the news.

With regards to the theoretical implications, we reflect:

Generalisation: The proposed method depends on the process of how decisions are made, as well as how they are recorded and archived. The decision-making process is similar for many governmental organisations at the municipal and (sub) national level in the Netherlands, although smaller organisations may record these in less detail as they have less administrative staff available. The approach likely generalises to many (Dutch) governmental organisations, especially as most municipalities in the country use the same two meeting planner systems and hence already structure their data similarly.

The method itself is domain-agnostic, and could be adapted for similar decision-making processes at other organisations. Potential future work could investigate whether this approach could be adapted broadly, specifically for use in organisations that have integrated software for their email, calendar and content management (e.g. organisations using SharePoint). Although these processes will be less structured and recorded less accurately, a timeline might still be a useful way to organise and revisit information.

Limitations: Although our approach achieved a high precision in retrieving relevant documents (97%), it is only an explorative step towards generating overviews of policy information. The main limitation is that, although 91% of references are extracted from the text, only 39% of total references find exact matches. Future work should improve the detection of true positives, both in the reference disambiguation (e.g. using the domains given in document metadata), and in the duplicate detection. One approach to disambiguate references and identify missing documents is to involve a domain expert in the loop (RQ4).

As the domain experts reference the most vital documents known to their expertise, we assume that our timelines cover the most important moments towards a decision. Future work could include documents that did not directly lead to a given council decision, but might still include contextual information. Improving the recall of these timelines could be done by first generating high-precision timelines, and using these documents as the basis for content-based recommendations. A domain expert could prepare a timeline for council members by first disambiguating references, and then expanding the timelines based on recommendations.

A limitation of the experimental setup is that each user only used one timeline. Although no obvious differences between the manual and generated timelines emerged, properly investigating these differences would require a comparative study.

Finally we wish to highlight the limited number of participants in the study (n = 10). Although this is a useful sample size for an iterative design process, later stages in this research direction will require studies with a larger participant pool to quantify how such an overview aids in task performance compared to searching individual documents.

5.9 Conclusion

Understanding the decision-making process of a city council necessitates an understanding how council documents are interrelated. In this paper we considered how a digital transformation of how we plan, record and archive the decision-making process of city council can result in more transparent decisions and clearer accountability. Specifically, we (re)constructed the timelines of the policy-making process from the existing council information. As an informal user study indicated that authors of policy documents consistently reference important documents, we proposed an approach to generate timelines based on two types of references between documents: document co-citation during meetings and textual references within council documents. We generated timelines of individual policy proposals by identifying meetings that discussed the same documents. Timelines of how policy proposals extend one another were identified by examining textual references between documents.

A user study with policy workers investigated both 1) guidelines for designing an overview interface and 2) whether the generated timelines included relevant documents. Experts considered documents relevant if the document and council decision both result from the same council decision (RQ1). Such provenance is recorded through references between documents. Policy workers need to see timelines from an overview perspective that balances conciseness with completeness by providing a decision layer and a comprehensive provenance layer (RQ2).

Creating an overview based on references between documents nearly exclusively yields relevant documents, with a precision of 97% (RQ3). The main limitation is that our approach identified the exact target document for only 39% of textual references. A further 52% of the references is ambiguous, finding a subset of documents. Although future work can enhance this aspect by extracting more context around references, we recommend involving a domain expert in the loop to select the exact matches and identify if further missing documents (RQ4).

A practical application of this work is to include overviews of council information in the municipality’s search engine, such that users can click individual search results to view that document in the context of a larger decision history. This presents a step towards better supporting council decision making, by making pertinent information more accessible. The method we proposed generates timelines of a local government’s decision-making process with high precision. As other municipalities in the Netherlands use a similar decision-making process and similar software to plan and archive their meetings, we expect that our functionality can be adapted to those organisations with minimal adjustments.

5.10 Chapter discussion

Although council members initially requested improved search functionality to satisfy their complex information seeking needs, our task analysis indicated that the underlying user need was to gain an overview of political dossiers. Hence, this chapter focused on developing an overview perspective rather than developing traditional search functionality. This illustrates how a task-based approach can identify the underlying problem, rather than merely improving the existing functionality. The resulting technical solution was straightforward, involving the use of implicit meta-information in the council documents to reconstruct the decision histories of council decisions (TRQ4).

5.11 Chapter outcomes

We applied our approach to creating an overview perspective of the decision-making process. This overview enables council members to understand how the previous city council made decisions, helping new council members to build upon those decisions more effectively. To provide this overview perspective, we initially analysed the dataset to identify which council documents and council meetings belonged to the same decision-making process. This functionality has been integrated into the search engine presented in chapter 2. Users of that system can search as usual, but can also click search results to get a secondary view. On this view the individual search result is placed in the context of the its associated decision-making process.

Chapter 6 Conclusion and Discussion

In the preceding chapters, we addressed the problem that generic search functionality does not sufficiently support the diverse information needs of council members and policy workers at the municipality of Utrecht. We investigated when and how to develop specialised search functionality by characterising the users' tasks and then designing the functionality for the tasks that were not supported well.

We focused on municipal knowledge workers, as they have complex and specific tasks. Designing task-based support requires that we address multiple Thesis Research Questions, starting with:

> **TRQ1** *How can we characterise the tasks of municipal knowledge workers for the purpose of developing (domain-)specific search functionality?*

To answer this, we conducted interviews with council members from the municipality of Utrecht across two studies, identifying the most common work tasks and search tasks (Section 2.6). We validated these findings with a third study conducted at the municipality of Hollands Kroon. By coding the task descriptions, we identified five task facets that can be used to characterise these search tasks: the task objective, topic aspect, information source, retrieval unit, and task specificity (Section 2.5). This approach can be viewed as an extension of the faceted task model [94] for the purpose of designing task-based support, where task facets inform what information should be indexed (information source); ranked (task specificity, topic aspect); and presented (retrieval unit). To demonstrate this, the task characterisation was used to informally develop a search engine for council information (Section 2.7). This is an initial step towards task-based design, where further work may be able to specify tasks in more detail and therefore yield more concrete design implications.

As it became possible to support any individual search task, it became clear that it is expensive and not scalable to design functionality around every task performed in an organisation. In addition, there are many tasks that only require generic search functionality (e.g. 'look up a known document'). This raises the following question:

> **TRQ2** *How can we determine which search tasks of municipal knowledge workers warrant the development of specialised search functionality?*

We characterised the tasks of policy workers through interviews, and then proposed functionality to improve both their simple and their complex web search tasks (Section 3.4). Tasks that warrant specialised support can be identified based on their importance, frequency, complexity and the level of frustration that they cause. To determine whether a task is sufficiently supported, we compare it to the available functionality. For instance, we discovered that the simple web-based search tasks of policy workers were unnecessarily frustrating and complex, as search results by authoritative sources are drowned out by search results from unverified sources. By filtering out such results, we reduced task complexity and sped up task performance (Section 3.7). When sixteen policy workers performed 61 simulated simple web-based search tasks, they were 1.7 times faster on average.

Reducing task complexity was an important theme. We found that policy workers tend to morph their complex web-based tasks into tasks seeking help from colleagues with expertise. When the same policy workers performed 31 complex web-based tasks, only 33% of them solved the task on their own, whereas the rest preferred to find an expert for assistance. Once we can identify which tasks require support, the next question becomes:

> **TRQ3** *How can we design specialised search functionality for municipal knowledge workers around a faceted description of their search tasks?*

We found that the type of functionality we can design to support a task depends on the level of abstraction from which we view the task. For example, in our studies with different knowledge workers, users discussed expert search tasks (searching for people with expertise) as a single broad category (see Sections 2.6 and 3.4). Consequently, we designed the search functionality around a common aspect between these expert search tasks: presenting search results around people as a retrieval unit as opposed to presenting documents.

As a use case, we developed support for the expert search tasks we identified in the previous studies. In the pre-existing expert search system, users could only look up people when they were looking for very specific criteria (e.g. a name or job description). Recognising that expert search tasks often have amorphous needs, we developed a prototype search engine that allows users to search for people based on the documents they have authored. Design decisions were based on the previously identified task facets, and a major design decision was experimentally validated: whether to rank and present documents or people as search results. We confirmed that a people-centric presentation of expert search results helped policy workers perform expert search more effectively than a document-centric presentation (Section 4.8). When twenty domain novices performed simulated expert search tasks, statistical analysis found that they completed more tasks correctly when using the people-centric interface.

The specificity of the task descriptions affects the type of functionality that can be designed. In some domains, users have different (sub)types of expert search tasks. While some expert search tasks require very specific expertise (e.g. finding a person knowledgeable about a specific aspect of a topic), other tasks may require more general expertise (e.g. finding a person who can provide broad advice on a topic). The ranking and presentation of search results can be adjusted to reflect the needs of the users in their current task. By refining our understanding of the users' tasks, we can design more tailored support.

Although this method helped us design search functionality and improve the result page, municipal knowledge workers also have search tasks which do not revolve around retrieving a specific subset of relevant information. For example, users may have longer and more complicated interactions with information when they are learning about a topic. This raises questions on how to support search tasks where the goal is not the retrieval of (a pre-definable subset of) information. To address this, we explore how to provide task-based support for these tasks we investigate one such use-case:

> **TRQ4** *How can we support council members in their task to obtain an overview perspective of how individual (council) documents are linked?*

We explored how to support council members in understanding the decision-making processes of previous city councils. This necessitates an overview perspective of the various steps during the process. To achieve this, we generated timelines of the decision making process were generated based on implicit meta-data in documents (Section 5.6). When ten experienced policy workers reviewed these timelines, we found that 97% of documents included were relevant.

Subsequently, we derived design implications to allow users to view documents in the context of the larger decision history (Section 5.8). Although the task facets tell us 116 what information is sought, they do not specify how to link and model the data, or how to present those results in the interface. For traditional retrieval tasks, we can rely on established information retrieval techniques to guide the design decisions. However, supporting less traditional search tasks is less straightforward. Therefore, an exploratory study was necessary to bridge the gap between the task characterisation and design implications.

This use-case exemplified how to characterise council member tasks (TRQ1) and identify tasks that require additional support (TRQ2). It demonstrated how to design support around a task's facets (TRQ3), even when this extends beyond typical search functionality (TRQ4). With these insights, we are prepared to address our Main Thesis Research Question:

> **MTRQ** *How can we design search functionality to support the search tasks of municipal knowledge workers?*

To borrow from the medical field: proper diagnosis precedes useful treatment. Once we adequately understand the tasks of users it becomes possible to think about tools that support those tasks. This lets the municipality move from generic functionality towards supporting the needs of different user groups (from one-search-for-all to targeted search applications in Figure 1), which leads to more search satisfaction [47]. We can contextualise user goals by characterising the users' tasks. We can identify which tasks do not have adequate support by considering whether the available search functionality satisfies all of a task's facets. One can find which tasks warrant additional support by looking at which tasks are the most important, frequent, frustrating and costly if they fail. Specialised search functionality can be designed around the task facets that are not adequately supported. Such support may extend beyond typical search functionality, and include the linking and presenting of search results in different ways.

6.1 Implications

Developing task-based search applications enables organisations to evolve from generic search functionality towards specialised search functionality. This is particularly beneficial in domain-specific search settings, where search tasks tend to be more complex and specific. We found that a task-centric approach allowed us to effectively link the user-centric perspective (objectives, intentions and experience) to the system-centric (available information and functionality) by connecting the two in concrete use cases.

By investigating the specific requirements of a search task, we could add more detail to the information needs than we had previously found through a purely user-centric approach (Section 2.4). This approach bridges the literature on information seeking (often a user-centric perspective) closer towards the practical purpose of developing search engines (often a system-centric perspective).

As we focused on supporting tasks rather than the specific organisation, we expect our results will generalise to other organisations where users perform the same tasks. In particular, we confirmed that council members at multiple municipalities in the Netherlands perform the same tasks (see Chapter 2). This implies that the proposed task-specific search functionality for Dutch municipalities could be standardised. Further studies should confirm whether the task taxonomy remains stable between different municipal user groups and municipalities of varying sizes.

The same tasks we supported may exist in domains beyond municipal search, to the extent that they can be characterised by the same task facets. In such cases there will be significant overlap in the functional requirements of the task, even if the specific implementation of functionality differs. Characterising tasks in more detail will enable future work to design specialised search functionality in a more modular way, that can be re-used in other domains.

6.2 Limitations

This work represents initial steps towards a design method for task-based search. The primary limitation of our study is that our dual purpose—understanding user tasks and developing functional search engines—necessitated an iterative and explorative approach. The number of iterations limited the number of participants we could include in each study. Future iterations could refine the proposed search functionality, and its impact on user effectiveness and efficiency could be quantified with a larger participant pool.

Each chapter in this thesis has its limitations. In Chapter 2, the study did not analyse the search engine we developed for council information, using it only as an illustrative example. Chapter 3 tested improvements for simple and complex web search tasks but did not evaluate the functionality to support complex web search tasks fully. Our study indicated that users often get frustrated and switch complex web search tasks to expert search tasks, but the details of this process should be examined in a future study where users can switch to expert search functionality.

Expert search was further investigated in Chapter 4, though we restricted our scope to novice users. While these users benefit the most from expert search, future work should test if similar findings hold for more experienced policy workers and for expert search in other domains. Finally, Chapter 5 explored how a timeline overview can be constructed and visualised but did not investigate its impact on how council members perform tasks.

Many of these limitations stem from applying our research to developing practical search engines with representative users rather than conducting laboratory studies. The advantage of these user-focused studies is that we could deliver qualitative findings and design systems that are useful for real users' tasks. After reflecting on the limitations of our studies, we now turn to promising future research directions.

6.3 Promising research directions

Research on task-based search holds significant potential to improve the development of practical search applications. In this section, we introduce several potential research directions and elaborate on the bolded ones in the following subsections.

Research on task-based search can provide practitioners with guidelines to translate abstract use cases into more concrete and practical information requirements. However, this requires addressing a paradox inherent in creating a generalised approach for specialised functionality. To create functionality that generalises across domains, we must abstract beyond domain-specific aspects. Yet, properly supporting domainspecific tasks necessitates developing specialised functionality. Consequently, our field includes many studies that either generalise across domains or are entirely specialised for specific situations.

There is opportunity for research that develops domain-specific functionality in a generic way. This is a challenging direction, as design method claims can be difficult to falsify, and individual case studies often lack generalisable evidence.

Future work could survey previous studies on specialised functionality to analyse how different types of tasks relate to different types of functionality.

A challenge in aggregating findings from different specialized settings is the inconsistent descriptions of search tasks across studies. Some studies address high-level, conceptual tasks oriented around users' goals (e.g., "How was the Uithoflijn constructed?"), while others focus on procedural and operational tasks arising during those conceptual tasks (e.g., "When was the Uithoflijn in the news, so I can filter council information by date?"). Developing a framework for task abstraction could help researchers describe search tasks more precisely and see how different studies relate.

Aggregating research on high-level search tasks could also yield insights into recurring subtasks and how to support them, potentially leading to modular search components for practical search engines. Instead of only surveying the literature, another approach could involve inventorying tasks in a practical domain, such as creating a taxonomy of municipal search tasks. By identifying which tasks occur, how often they occur, and their importance, we can pinpoint tasks needing further research. This may involve different sub-types of expertise and various expert search tasks.

Identifying tasks in this manner will highlight new research directions, revealing tasks that warrant new types of support. For example, presenting search results in an organisational context, showing documents related to the project that led to their creation, could be a valuable improvement.

We have identified future research directions in three main themes: describing tasks, inventorying tasks, and supporting tasks. We now elaborate on these research directions.

6.3.1 Describing tasks

There are aspects of tasks and task-based functionality which could be described in more detail.

Surveying specialised functionality In our work we approached task-based support from a user-centered perspective and moved towards designing functionality. It would also be valuable to approach tasks from the other direction. Current descriptions of task-based support are highly situation-specific, making individual solutions difficult to re-use. However, a survey of specialised search support may be able to identify recurring themes in functionality, and investigate whether sub-types of functionality map on different task facets. Highlighting recurring characteristics would close the gap between the domain-agnostic literature on search tasks and specialised search functionality, as it would elucidate how specific task-based support could be re-used in new settings.

Framework for task abstraction The academic literature has approached tasks from multiple perspectives, and for different purposes (as illustrated in [22]), and are developing increasingly sophisticated ways to describe them (e.g. [28]). As a result, there is a lot of variety in how papers describe tasks. Some tasks are described at a very high and conceptual level (e.g. ‘exploratory search’), whereas others are as specific as concrete search intents (e.g. ‘a query’). Future work could develop a framework that describes different levels of abstraction of a task. This would help disambiguate whether a paper is discussing a class of search task (e.g. ‘exploratory search’), an instantiated user objective (e.g. ‘explorative literature search on topic x’), or an operational task that will spawn in fulfilling that objective (e.g. ‘looking up a known item to find out when topic x was relevant, letting you filter your search results on date’). An accurate way to conceptualise tasks at different levels of abstraction would help academics be more specific in how they describe and design task-based research.

Recurring sub-tasks There are recurring sub-tasks between tasks, especially as tasks become more complex. This might be the case because tasks require similar cognitive processes (e.g. synthesise information), or because they require similar operational tasks (e.g. query reformulation). Although we focused on tasks starting from high-level user goals, another avenue would consider the lower level operational sub-tasks that recur between tasks (e.g. by supporting note taking during complex tasks [32, 42]). Understanding how these tasks recur may lead to different ways to support information seeking. For example, supporting functionality could be integrated into the internet browser, such as an interactive note-taking/data-analysis tool, or AI-based personal assistant. It would also inform, for example, how users want to navigate between different views in the interface (see e.g. [131], chapter 4).

6.3.2 inventory tasks

There are a few domains where an inventory of tasks would be particularly valuable.

Taxonomy of (municipal) tasks We found that working with one target user group at a time allowed us to design specific solutions, but that we were not able to gain a broader perspective on what tasks are the priority within a (municipal) organisation. Both practitioners and academics would benefit from an inventory of tasks that accounts for how often they are performed and by whom, as well as what kind of functionality is available to support these tasks. Such an overview would enable the organisation to formulate a more targeted information strategy. Practitioners could then quantify how often certain tasks are performed, how much time is spent on them, and the cost of failing to complete them. This enables them to roughly estimate an economic value of supporting those search tasks, which is useful during conversations with project stakeholders (decision-makers such as project managers or the problem owners). An overview of (municipal) tasks would also provide academics a dataset to continue developing existing task taxonomies (e.g. [130, 104]). For example, by quantifying how much certain task types occur and whether there are frequently co-occurring sub-tasks.

Sub-types of expert search In chapter 4 we found that policy workers tend to have one class of expert search tasks, but future work could explore whether expert search in other domains requires different sub-types of expert search as well as different types of expertise.

6.3.3 Supporting tasks

Finally, we consider a particular task that would be worth investigating.

Search results in organisational context Although many search tasks target specific information or documents, other tasks require an understanding of how information fits into a larger context. Future work could expand the method used in chapter 5, where we created an overview of council decisions based on council documents. A similar approach could be applied to projects in an organisation's content management system (in Utrecht that would be Sharepoint). Documents in this system are typically produced in the context of a larger project, and it could be useful to view individual search results within the context of those projects. To construct such overviews we could use a number of signals, including the metadata of who accessed certain documents (and when), the document file structure, and (co-citation and textual) references between documents. A data model of project-context may also 'clean up' search results, as it should be possible to, e.g., filter outdated versions of documents.

To navigate the wealth of research directions at the municipality we can consider which opportunities provide the most practical value to the municipality. To consider this practical aspect we now turn to some of our more practical reflections and insights around developing search at the municipality of Utrecht.

6.4 Applying research to practice

We reflect on how to perform applied research within an operational environment, and then share insights on the development on the search engines at the municipality.

6.4.1 Reflections

During our studies, we aimed to bridge the gap between research and practice by developing search engines for specific user groups. However, our approach evolved throughout the course of the studies. In Chapters 2 and 3 of this thesis, we prioritized aligning research objectives with the current stakeholder interests. This introduced several challenges.

One challenge was that stakeholder interests shifted over time, necessitating reactive adjustments in research focus, which made it difficult to plan research activities effectively. Another challenge was that the practical requirements of projects often threatened to expand the research scope: designing a complete search engine involves changing numerous variables simultaneously, making it difficult to measure the effect of any individual variable when comparing the new system to the old one.

This issue could be avoided by comparing multiple versions of the same proposed system, but doing so would undermine our goal of measuring the value of task-based search in a practical setting rather than in an unrealistic 'laboratory' setting.

In Chapters 4 and 5, we established research projects with well-defined scopes, focusing on poorly supported but important tasks identified in Chapters 2 and 3. By narrowing our research scope, we were unable to stay aligned with the shifting interests of stakeholders, who focused on solving technical problems and delivering products. Consequently, the research project lost visibility and traction within the organisation.

Future projects may have more success at combining research and practice if the research efforts are 1) separated from direct (often short-to-mid-term) organisational interests, but still 2) aligned with highly stable organisational interests. This approach allows the research to develop while ensuring continued stakeholder investment.

Firstly, this type of research should be separated from short-to-mid-term interests because its strengths lie in identifying product requirements and extending a product based on unmet needs. It is less suited to developing a minimum viable product, during which interests and priorities may shift. Secondly, research should align with stable organisational interests to ensure continued support. To identify such interests, it is useful to start from the core business of the organisation.

The municipality's main responsibilities are providing local services and enacting local legislation. To identify research opportunities around information retrieval, we can consider instances where the municipality struggles to fulfil services that involve finding and providing information. For the municipality of Utrecht, potential research opportunities might involve improving how they handle information for WOB (Wet Openbaarheid van Bestuur) requests.[6]

Although expert search is arguably very useful for the municipality, it indirectly contributes to the municipality's main responsibilities. Building support for search functionality that does not directly contribute to the organisation's core business would require an exceedingly robust information strategy from the organisation. As we move towards discussing the practical lessons learned about developing search functionality at Utrecht, we may consider elements of such an information strategy.

6 These are analogous to FOIA (Freedom of Information Act) requests, in that they are requests that oblige a governmental organisation to provide their records on a given topic.

6.4.2 Practical insights

Throughout our studies we had a number of insights on how we can effectively bring search engines to practice. In this section we introduce those insights, and we elaborate on these (bolded) concepts in the following subsections.

Our insights reflect the main theme of thesis, where a better understanding of user requirements translates to be more effective, efficient and satisfactory search engines. This starts with inventorying the target user groups, their needs and the available functionality in the organisation. This allows us to create a **unified information strategy**. Once it is more clear which needs are unmet, it is possible to plan a path towards satisfying those needs. To ensure that the needs of users are being met, it may be beneficial to **approach search vertically** by focusing on one user group at a time. As different user groups may have similar tasks, this approach could focus on developing **modular search functionality** to solve specific sub-tasks. This facilitates the **Re-use of functionality**. For example, we identified a recurring **need for people search** between different groups of municipal knowledge workers.

Translating an information strategy to practical systems is challenging, and requires expertise as well as a good process. As an organisation that often relies on the expertise of vendors, the municipality could consider how to **build the internal expertise** necessary to operationalise an information strategy. Empowering that expertise may involve **approaching search as a process** rather than a short term project, where **end users are involved** and the team is not afraid to **fail often and fail quickly**. It may be especially beneficial for the development process to include safeguards where **a party with expertise can represent the project requirements** independently from other (short-term) organisational pressures.

This brief overview mentioned insights in three main themes: strategic, technological, and developmental:

Strategic insights

Define a unified information strategy To develop a structured approach to supporting the information needs at the municipality, we can formulate an overall strategy to information and findability. The municipality has taken steps into this direction, but could take further steps by accounting for factors across five dimensions: the business, the users, the information, the organisation and the technology [49]. This would involve a more detailed overview of concepts such as:

- What user groups have different information needs (e.g. council members, policy workers)?
- What information systems are available?
- Which search tasks does each user group need to perform?

- What are the most prominent information-related complaints each user group?
- Which information-management tools are available (e.g. software for retrieval, classification, anonimisation)?
- Which information-related projects are ongoing (both current and previous)?
- What internal expertise is available around information management?

Quantifying how often certain tasks are performed and estimating the cost of doing these inefficiently allows us to roughly estimate how much value would be gained by supporting specialised support for that task (or the larger group of users). The risk of not refining and communicating this strategy is that a lot of effort will be shifted towards putting out fires in the short term, rather than improving the information architecture in the long term.

Approach search vertically Once it is more clear what information requirements need to be addressed, it is possible to create a more detailed plan to address those needs. Findability can be improved vertically (focusing on one user group at a time) or horizontally (focusing on one technology at a time). Focusing on one user group at a time has two main benefits. Firstly, problems with the new system will not affect all users in the organisation. Secondly, focusing on a more specific target user group results in better addressing their requirements.

As support is developed for additional user groups, it will become evident that some tasks are already supported by functionality developed for the first user group. Hence the investment to support each user group is likely to diminish over time.

Technological insights

Develop modular functionality Even though task-based search can be used to design specialised functionality, we should not neglect the overlap between similar tasks in different domains (e.g. exploring council information or exploring the content management system). We should keep this in mind while in how we implement functionality by designing components that can be re-used in new contexts. For example, by disclosing information or search results via API's, or creating interfaces that can adapt to different types of input.

Re-use functionality The tasks performed by council members and policy workers at the municipality of Utrecht are also performed at other public sector organisations. As shown for council members in chapter 2, council members at other municipalities in the Netherlands perform the same tasks and could therefore benefit from the same functionality. The main difference would be in how information has been recorded following slightly different processes, which might require pre-processing when the data is ingested. A viable strategy would therefore be to pool resources in a collaboration between governmental organisations, to develop re-usable search functionality for those common processes. Such pooling of resources and expertise would benefit the individual organisations by saving them time and money on redundant investments.

Develop people search Whenever we investigated the tasks of users, we found that a around half of the tasks are completed by asking colleagues for assistance. Similarly, recent studies found that around half of the queries in the SharePoint of a large biotechnology company involved the names of people [104, 152]. People search can act as a backup for solving complex search tasks (see chapter 3), and can add more context and value to the information being found (see chapters 3 and 4). Although the municipality's existing system (Wie-is-Wie) is good in cases where the user already knows which specific search terms will be useful, future work could enable users to find people with expertise based on the documents that those experts have authored, and the projects those experts have contributed to.

Development insights

Build internal expertise strategically It can be difficult to build internal expertise in a new area, as it can require a culture shift to leverage that expertise. For example, in a company without internal technical expertise it makes sense to shift the ownership of technical problems towards an external party (vendor or consultant). As an organisation builds internal expertise, those experts come into projects and processes which shift technical problems away from them. The risk is that technical problems leave the organisation before the expert can contribute, and then return once the expert missed their chance to contribute. The organisation could view building expertise as a culture shift, rather than as the process of hiring individuals. This could be approached by recruiting experienced experts, who know how to claim ownership and space within a project's workflow. In addition to technical expertise, they may need consultancy skills to get the project to a point where they can have an internal discussion about the technical aspects. Specifically, they will need to ask the right questions, and the ability to guide the project towards a new process.

Develop search as a process Achieving high search satisfaction is done by approaching search as a process, not a project [169], and requires ongoing investments in a dedicated search team [170]. As information management has effects across organisational departments, it can be unclear where the ownership and accountability end up. As a result, it can be challenging to ensure sufficient resources are available to fundamentally address these challenges. It can be useful to make someone accountable for doing this at the strategic level [170].

Involve end users Many decision-makers will not have the time to understand all technical aspects of their project, as they work towards multiple projects in a bigger picture. Although decision-makers will know where the largest bottlenecks are, they may not have time to investigate the cause. Therefore the focus for decision-makers can go towards high-level user stories or a focus on technological solutions. Projects can benefit from understanding the user's problems at in more detail, to reduce the risk of not solving the underlying problem. Projects can benefit from involve experts who can interview and represent the end users' needs.

Fail faster, fail more often During our studies we prioritised only developing functionality once the user task was clear, and avoided projects where these tasks were unclear. However, even projects that never provide value to end users are useful as they provide perceived value, experience and visibility in the organisation. This can be an approach towards asking the right questions.

Third party representing project requirements There is a risk during the development of search systems that can be illustrated using a design thinking perspective. Figure 26 shows the Stanford Design Thinking Process. Organisations often have the knowledge and expertise necessary for the initial steps of this process. However, they then contact vendors, who offer solutions and tend to have expertise in the latter steps. While vendors possess significant technical expertise, they often lack insight into the specific challenges faced by the organisation. A critical risk arises at the Define step, where these two meet.

Figure 26. Overview of the steps in the design thinking process. Image adapted under creative commons from the Interaction Design Foundation [38].

The organisation might not have the technical or domain expertise to identify the underlying problems leading to a specific use case, whereas the vendor is incentivised to show how their existing product can be adapted for any use case presented. Without the right expertise involved, neither party is fully empowered or incentivised to specify the project requirements in sufficient detail. This increases the risk of developing a product that does not address the underlying problem and will therefore need replacement in a few years.

To resolve this, a third party could take ownership of representing the project requirements and help connect the organisation and vendor. This could be an internal or external consultant with relevant expertise—in the case of search engines, this would involve skills in consultancy, information retrieval, and user-centered design. This consultant may face challenges, as the organisation's stakeholders might already be thinking in terms of solutions and may be hesitant to return to defining the problem. Ideally, the consultant would be involved while the project is still centered around understanding the problem.

6.5 Final Thoughts

As new technologies continue to emerge and user tasks evolve, the goal remains to support users in achieving their tasks. The research we presented complements emerging (AI-based) trends in literature that focus on providing increasingly relevant information through learning to rank or generating information. New technologies will only be useful if they solve user tasks, underscoring the necessity for a method to characterise user tasks.

The strengths of task-based functionality lie in enabling users to 1) interact with information in useful ways and 2) view information in a meaningful context, both defined in relation to the users' tasks. Supporting the tasks of knowledge workers will help both researchers and practitioners design more effective search functionality, both at municipalities and in other domains.

Appendices

Appendix A: Task-based term expansion (Supplementary chapter)

A.1 Chapter foreword

A crucial aspect of search engines is in properly ranking the search results, but the chapters presented in this thesis did not provide much guidance on how to rank results. The main finding was to search different subsets of information when searching for a different topic aspect. Previous work has similarly noted how different tasks are associated with different document genres [50]. In this chapter we explore a task-based ranking approach that does not rely on filtering results. We investigate how a search task can improve the ranking of search results by using the task to add context to a query. In this appendix we present a brief summary of the work that is available as:

Schoegje, T., Kamphuis, C., Dercksen, K., Hiemstra, D., Pieters, T., & de Vries, A. P. (2020). Exploring term expansion for task-based retrieval at the TREC-COVID track. arXiv preprint https://doi.org/10.48550/arXiv.2010.12674

A.2 Introduction

We explore how to generate effective queries based on search tasks. Our approach has three main steps: 1) identify search tasks based on research goals, 2) manually classify search queries according to those tasks, and 3) compare three methods to improve search rankings based on the task context. The most promising approach is based on expanding the user's query terms using task terms, which slightly improved the NDCG@20 scores over a BM25 baseline. Further improvements might be gained if we can identify more specific search tasks.

A.3 Approach

Our approach broadly follows the same three steps as used in the seminal taxonomy of web search [20], which was later extended [125]: identifying the tasks, mapping queries into tasks, and (re)ranking based on the current task.

A.3.1 Identifying tasks

COVID-related search tasks were collected by combining the research goals from two main sources: Kaggle's "COVID-19 Open Research Dataset Challenge" [1] and the WHO's "Roadmap to COVID-19 research" [122]. We found that the Kaggle goals are a superset of the WHO goals. The WHO roadmap contains less goals, and these are more specific (e.g. the Kaggle 'the vaccines and therapeutics' goal corresponds to the WHO's 'vaccines' and 'therapeutics' goals).

A.3.2 Query-task mapping

A manual and automatic task classification are compared. In both approaches each query is classified into one task. The manual classification was performed by the first author by matching words in the topic fields to those in the task descriptions (with some liberty taken with regard to hypernyms and synonyms). We noted that manually annotating search tasks based on a query alone can be difficult due to the ambiguous nature of search intentions. In the current TREC setting however, the annotator was able to use the information available in the topic questions and narrative in addition to the query terms.

Völske et al. compared methods for query-task mapping based on query logs [166]. They found that the most effective method was to index tasks in a small search engine, and then rank these tasks by a query using BM25. A query was classified as belonging to the top ranking task. We found a 66% agreement between the automatic and manual classifications. Because this agreement is fairly low and we wish to focus on the potential of our task-based approach, we used the manual classifications for the remainder of the paper.

A.3.3 Improving search rankings

We tested three approaches for task-based re-ranking:

doc2vec Re-rank search results based on the similarity in vector space between a paper's abstract and the task description

journal We used relevance assessments from previous rounds to identify which journals were relevant. Two variants were explored. In the **journal.prior** version, a prior likelihood was computed based on the proportion of papers from a given journal were relevant in previous rounds. The likelihood is then normalized such that journals with only irrelevant papers get a score of -1, and journals with only relevant papers get a score of 1. Journals without prior information get a score of 0.
The task-dependent **journal.task** variant is similar. The same procedure is repeated for each task, but this time only using relevance assessments of topics that were manually classified into the current task. The task-based prior scores have some intuitive validity - some high scoring journals for the 'risk factors' task include journals about diabetes and cardiovascular research. These are indeed some of the risk factors in the task description.

term expansion The task-based approach is to perform query expansion using task terms extracted from the Kaggle task descriptions. There is a **query+task** variant and a **query+udel+task** variant. The udel terms were created by the University of Delaware by using SciSpacy to lemmatize and remove non-stop words from the combined query, question, and narrative fields for each topic.

Of these, the task-based term expansion would turn out to be most successful. This was done by selecting task terms from the Kaggle task descriptions. First, biomedical entities are extracted from the task descriptions using ScispaCy's biomedical entity recognition. In order to keep the query short and specific, a selection of these task terms is made based on their TF-IDF score. This is based on the TF in the task description and the IDF in the collection of paper abstracts. The top n terms are then used, and appended to the query string. In order to weight the query terms more than task terms, we added duplicates of the original query terms to the new query string. Choosing how many terms should be added, and how these should be weighted was done by tuning on the relevance assessments available from the previous rounds.

A.4 Findings

During the first round of the competition, the baseline significantly (NDCG@20 = 0.2490) outperformed the doc2vec re-ranked run (NDCG@20 = 0.0964). It seems that a proximity between task descriptions and paper abstract in doc2vec space does not imply a semantic similarity.

During the third TREC-COVID round we switched our baseline ranker to the Anserini r3.rf run [175]. The baseline (NDCG@20 = 0.5800) outperformed the journal.priors approach (NDCG@20 = 0.3228) and did slightly better than the journal.task approach (NDCG@20 = 0.5406). The task-based variant of the journal performs much better than the variant based on journal priors. This suggests that there is no objectively better journal when it comes to ranking results, but that it depends on the context of the information need. The tasks may have been able to capture some but not enough of this context.

In the fifth round we tested our task-based term expansion, and Table 13 shows that finding that adding task terms to query terms marginally improves results over just using the query. However, if the udel terms are available the task terms slightly reduce precision. The udel terms were were taken from a written description of the underlying information need for each query, and may not be available in a realistic search scenario. However, they highlight the limitations of our current approach.

Table 13. Comparing fusion runs with various query term selection on the cumulative round 4 assessments.

	30 topics		45 topics	
	NDCG@20	MAP	NDCG@20	MAP
query	0.4290	0.1626	0.4316	0.1910
query+udel	0.5073	0.2082	0.4956	0.2300
query+task	0.4433	0.1620	0.4446	0.1893
query+udel+task	0.4929	0.2080	0.4907	0.2293

We suspect the Kaggle task descriptions describe the search intent at a higher/more abstract level than the topic question. When considering the findings in TREC's precision medicine track we find a potential explanation, as it was shown that using hypernyms during term expansion has a negative effect on search rankings [45]. The search tasks we identified may be too generic, and having more specific tasks may improve the efficacy of our approach. The scores of the **query+udel** run show a clear potential for improvements available by formulating better queries. The scores remained consistent, and even slightly improved as new topics were introduced. This suggests that the tasks identified are stable and complete enough to deal with new topics.

A.5 Conclusion

A successful strategy during early rounds of the TREC-COVID track was to extend a user's query with other relevant terms. We explored task-based search for the scientific COVID-19 literature, which allowed us to generate task terms that the user might not have entered. Our approach slightly improved NDCG@20 scores compared to using only query terms. Our approach to query generation did not yet reach the potential that others have shown when using terms from a topic's question field, although question terms may not be available in a real life search situation. Our approach is a step towards achieving similar scores without requiring users to input additional terms.

A.6 Chapter afterword

The task-based term expansion achieved a slight improvement over a BM25 baseline, but other approaches within the TREC track achieved significantly better improvements. One reason that limited results may be that the search tasks were defined fairly broadly. Tasks could be formulated that are more specific than the abstract Kaggle task descriptions, yet more generic than the specific TREC search intents. Another limiting factor could be the way we selected and weighted task terms. Yet if either of these was the limiting factor, we would have still hoped to achieve a larger improvement over the baseline system. As it stands, we have found that designing task-based search is more useful for revealing how users want to interact with information than it is for ranking the information. A possible exception, based on the previous thesis chapters and other work [50], may be to automatically filter irrelevant document genres when a task is detected.

Appendix B: Extracting author names (Chapter 4)

Author names are extracted using a regular expression from standardised document templates. Text is extracted from between text fragments between the markers ”Behandeld door” and ”datum”. The validity of author names is checked in the following ways: the name is between 4 and 30 characters in length, does not contain a number or special character and contains a space (both a first and last name). Documents authored by multiple individuals identified by the symbols ‘/’, ‘&,’ or ‘,’ in the name. Only documents with valid names and single authors are used in the experiment (as described in section 5).

Finding author aliases

Authors may spell their names in different ways, and hence a disambiguation is necessary. Common titles are removed from all names that contain them (such as ”Mr.” and ”Mw.”) and normalize accented characters to their basic ASCII counterparts (e.g., ¨e -¿ e)). Names where the surname and the first letter of the first name match are assumed to be the same person. This reduces approximately 1600 author names to 1032 unique authors.

Appendix C: Implementation (Chapter 4)

Building the index

The document-centric index is constructed by indexing documents using their title, fulltext content, author name, upload date, and sub-domain information. The candidatecentric index is created by indexing individual candidates with their name, all the text they have written, and the sub-domains in which they have contributed. The written text of each candidate is obtained by concatenating the titles and full-text of their associated documents into a single text string, an approach similar to the P@noptic search system [36]. While more recent studies have found that shorter candidate representations lead to improved retrieval effectiveness, our research questions do not necessitate state-of-the-art candidate representations. A candidate's portfolio is constructed by concatenating all sub-domains in which they have written a document. Although this portfolio is not used during the ranking process, it is displayed within the search results. A hand full of documents lacked a sub-domain, and documents that were marked as relevant were manually inspected and assigned the most relevant domain. This was typically based on an explicit reference to the domain in its content.

Ranking functions

Integrating the document-level index and candidate-level index with the two interfaces results in four distinct scenarios that the ranking functions, as described below. In the following discussion, we refer to ”document search” as the process of searching the document index. This is accomplished using the default Elastic search functionality, employing a *query string* query that searches the title and text fields within the document index. Conversely, ”candidate search” refers to searching the candidate index by utilising a *query string* query targeting the ”written texts” field. Although better ranking functions exist (e.g., language models over bag-of-words approaches [179]), our study focuses on these basic versions to simplify the discussion.

Document interface - document ranking

Documents are ranked using a document search, and results are presented in the document-centric interface.

Document interface - candidate ranking

Experts are ranked using a candidate search. This informs the expert panels of the document interface. Then the single most relevant document to this query is retrieved for each expert, by performing a document search while filtering on author. The result is presented in the corresponding document-panel.

Candidate interface - candidate ranking

A candidate search determines the ranking and contents of the author panels in the interface. Per expert a document search is performed with a filter on the author, to identify up to three documents written by this author that are the most relevant to the query.

Candidate interface - document ranking

A document search is performed which populates the author panel as well as the most relevant document in the document panel. Then another document search is performed with a filter on the author, to find up to two more relevant documents by this author that match this query. The document ranking can produce the same author multiple times in the ranking. Because these would contain an identical set of documents we hide duplicate candidates.

Appendix D: List of simulated tasks (Chapter 4)

Table 14 displays the tasks performed during the experiment.

Table 14. The tasks as presented to the users.

Task	Description
T1	Stel dat u een onderzoek voorbereid voor een project over fietsgedrag in Utrecht. Is er bij collega's iets bekend over het fietsgebruik van niet-Westerse allochtonen?
T2	Stel dat u het aantrekkelijk wil maken voor bedrijven om te vestigen in een bepaalde wijk. Hebben collega's data over het aantal bedrijven en het aantal arbeidsplaatsen in de verschillende wijken van Utrecht? Weten ze waarom bedrijven voor deze plekken kiezen?
T3	Stel dat u een nieuwe speelplek kunt laten bouwen, en wil controleren of er genoeg belangstelling voor is. Is er bij collega's iets bekend over hoeveel kinderen er zijn in de wijk Overvecht, en of we meer jonge huishoudens kunnen verwachten in de toekomst?
T4	Stel dat u Utrecht aantrekkelijker wilt maken voor toeristen. Weten collega's hoeveel overnachtigen toeristen jaarlijks maken in Utrecht? Waarom kiezen toeristen voor Utrecht?
T5	Als u een woning koopt zit er een anti-speculatiebeding op om te voorkomen dat mensen huizen kopen om ze vervolgens door te verkopen. Welke collega's weten hoe effectief deze maatregel blijkt te zijn om huizen meer betaalbaar te maken?
T6	Stel dat u beleid wilt maken om gezond gedrag te stimuleren in de wijk Leidsche Rijn. U weet dat collega's in een andere wijk hierin succesvol waren. Welke collega's kunnen u uitleggen hoe de Wijkaanpak Overvecht bedacht is?
T7	Stel dat u de tijdlijn wil schetsen van de bouw van de Uithoflijn, vanaf de planning tot de huidige status. Wie kan u hierbij helpen?
T8	Stel dat u wilt weten of corona invloed gaat hebben een bouwproject. Wie kan u vertellen of corona invloed heeft op de bouwplannen van Zorgcentrum Rosendael?

Appendix E: Descriptive statistics (Chapter 4)

Table 15 shows effectiveness metrics that are not pertinent to the main analysis. Similarly, Table 16 shows supplementary metrics for efficiency and user satisfaction.

Table 15. Effectiveness of the four search engines compared. 'Completion' is the average rate of task completion. 'No can' is short for no candidates selected during a task. 'Cans / task' is short for number of candidates per task. 'Rank' is the (zero-indexed) rank of the first candidate that was selected.

Search engine		Effectiveness		Candidates selected				
Interface	Ranking	Completion	Precision	No can	Cans/ task	Docs/ can	Rank	Avg rank
doc	doc	0.81	0.75	0.00	2.22	1.00	1.75	3.82
doc	can	0.75	0.74	0.06	1.58	1.00	1.56	2.60
can	doc	0.79	0.73	0.06	2.11	2.63	0.61	3.38
can	can	0.86	0.76	0.03	2.03	2.48	1.26	3.29

Table 16. Efficiency, user satisfaction and pairwise preferences of the four search engines compared. User preferences measures the preferability of the search engine in the row to the one in the column.

Search engine		Efficiency				User satisfaction	
Interface	Ranking	Time (min)	Queries	Clicks	Sum actions	SUS	Confidence
doc	doc	3.39	4.08	1.72	5.81	71.94	4.00
doc	can	3.02	3.08	1.14	4.22	70.45	3.73
can	doc	2.88	4.42	0.36	4.78	72.05	3.55
can	can	4.59	4.28	1.75	6.03	72.5	3.89

References

[1]. A. I. F. AI. *COVID-19 Open Research Dataset Challenge (CORD-19)*. kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks, Accessed: 18-10-2020.

[2]. D. Alexander, W. Kusa, and A. P. de Vries. “ORCAS-I: Queries Annotated with Intent using Weak Supervision”. In: *CoRR* abs/2205.00926 (2022). doi: 10.48550/arXiv.2205.00926. arXiv: 2205.00926. url: https://doi.org/10.48550/arXiv.2205.00926.

[3]. M. Aliannejadi, M. Harvey, L. Costa, M. Pointon, and F. Crestani. “Understanding Mobile Search Task Relevance and User Behaviour in Context”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. CHIIR 2019. Glasgow, Scotland: ACM, 2019, pp. 143–151.

[4]. M. J. Anderson. “A comparative analysis of information search and evaluation behavior of professional and non-professional financial analysts”. In: *Accounting, Organizations and Society* 13.5 (1988), pp. 431–446.

[5]. J. Arguello, W. Wu, D. Kelly, and A. Edwards. “Task complexity, vertical display and user interaction in aggregated search”. In: *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR ’12, August 12-16, 2012*. Ed. by W. R. Hersh, J. Callan, Y. Maarek, and M. Sanderson. Portland, OR, USA: ACM, 2012, pp. 435–444. doi: 10.1145/2348283.2348343. url: https://doi.org/10.1145/2348283.2348343.

[6]. A. H. Awadallah, R. W. White, P. Pantel, S. T. Dumais, and Y. Wang. “Supporting Complex Search Tasks”. In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014*. Ed. by J. Li, X. S. Wang, M. N. Garofalakis, I. Soboroff, T. Suel, and M. Wang. ACM, 2014, pp. 829–838. doi: 10.1145/2661829.2661912. url: https://doi.org/10.1145/2661829.2661912.

[7]. K. Balog, L. Azzopardi, and M. de Rijke. “Formal models for expert finding in enterprise corpora”. In: *SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, USA, August 6-11, 2006*. Ed. by E. N. Efthimiadis, S. T. Dumais, D. Hawking, and K. Järvelin. ACM, 2006, pp. 43–50. doi: 10.1145/1148170.1148181. url: https://doi.org/10.1145/1148170.1148181.

[8]. K. Balog, Y. Fang, M. de Rijke, P. Serdyukov, and L. Si. "Expertise Retrieval". In: *Found. Trends Inf. Retr.* 6.2-3 (2012), pp. 127–256. doi: 10.1561/1500000024. url: https://doi.org/10.1561/1500000024.

[9]. D. Bawden and L. Robinson. "Individual differences in information-related behaviour: what do we know about information styles?" In: *New directions in information behaviour* 1 (2011), pp. 127–158.

[10]. E. M. A. L. Beauxis-Aussalet et al. "Statistics and Visualizations for Assessing Class Size Uncertainty". PhD thesis. Utrecht University, 2019.

[11]. K. Beelen, T. A. Thijm, C. Cochrane, K. Halvemaan, G. Hirst, M. Kimmins, S. Lijbrink, M. Marx, N. Naderi, L. Rheault, et al. "Digitization of the Canadian parliamentary debates". In: *Canadian Journal of Political Science/Revue canadienne de science politique* 50.3 (2017), pp. 849–864.

[12]. N. J. Belkin. "On the evaluation of interactive information retrieval systems". In: *The Janus Faced Scholar* (2010).

[13]. N. J. Belkin. "People, Interacting with Information". In: *SIGIR* 49.2 (2015), pp. 13–27.

[14]. N. J. Belkin. "Salton Award Lecture: People, Interacting with Information". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015*. Ed. by R. Baeza-Yates, M. Lalmas, A. Moffat, and B. A. Ribeiro-Neto. ACM, 2015, pp. 1–2. doi: 10.1145/2766462.2767854. url: https://doi.org/10. 1145/2766462.2767854.

[15]. J. R. Benetka, J. Krumm, and P. N. Bennett. "Understanding Context for Tasks and Activities". In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, CHIIR 2019, Glasgow, Scotland, UK, March 10-14, 2019*. Ed. by L. Azzopardi, M. Halvey, I. Ruthven, H. Joho, V. Murdock, and P. Qvarfordt. ACM, 2019, pp. 133–142. doi: 10.1145/3295750.3298929. url: https://doi.org/10.1145/3295750.3298929.

[16]. M. Berger, J. Zavrel, and P. Groth. "Effective distributed representations for academic expert search". In: *Proceedings of the First Workshop on Scholarly Document Processing, SDP@EMNLP 2020, November 19, 2020*. Ed. by M. K. Chandrasekaran, A. de Waard, G. Feigenblat, D. Freitag, T. Ghosal, E. H. Hovy, P. Knoth, D. Konopnicki, P. Mayr, R. M. Patton, and M. Shmueli-Scheuer. Online: Association for Computational Linguistics, 2020, pp. 56–71. doi: 10. 18653/v1/2020.sdp-1.7. url: https://doi.org/10.18653/v1/2020.sdp- 1.7.

[17]. J. M. Berryman. "What defines 'enough' information? How policy workers make judgements and decisions during information seeking: preliminary results from an exploratory study". In: *Inf. Res.* 11.4 (2006). url: http ://www . informationr.net/ir/11-4/paper266.html.

[18]. F. Birghan, R. Hettenhausen, C. Meschede, and T. Siebenlist. "Informing Citizens via Council Information Systems". In: *20th Annual International Conference on Digital Government Research, DG.O 2019, June 18-20, 2019*. Ed. by Y. Chen, F. Salem, and A. Zuiderwijk. Dubai, United Arab Emirates: ACM, 2019, pp. 280–286. doi: 10.1145/3325112.3325220. url: https://doi.org/10.1145/3325112.3325220.

[19]. T. Bögel and M. Gertz. "Time will Tell: Temporal Linking of News Stories". In: *Proceedings of the 15th ACM/IEEE-CE Joint Conference on Digital Libraries, June 21-25, 2015*. Ed. by P. L. B. II, S. Allard, H. Mercer, M. Beck, S. J. Cunningham, D. H. Goh, and G. Henry. Knoxville, TN, USA: ACM, 2015, pp. 195–204. doi: 10.1145/2756406.2756919. url: https://doi.org/10.1145/2756406.2756919.

[20]. A. Z. Broder. "A taxonomy of web search". In: *SIGIR Forum* 36.2 (2002), pp. 3–10. doi: 10.1145/792550.792552. url: https://doi.org/10.1145/792550.792552.

[21]. J. Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7.

[22]. K. Byström. "Approaches to task in contemporary information studies". In: *Inf. Res.* 12.4 (2007). url: http://www.informationr.net/ir/12-4/colis/colis26.html.

[23]. K. Byström. "Information and information sources in tasks of varying complexity". In: *J. Assoc. Inf. Sci. Technol.* 53.7 (2002), pp. 581–591. doi: 10.1002/ASI.10064. url: https://doi.org/10.1002/asi.10064.

[24]. K. Byström and P. Hansen. "Conceptual framework for tasks in information studies". In: *Journal of the American Society for Information science and Technology* 56.10 (2005), pp. 1050–1061.

[25]. K. Byström, J. Heinström, and I. Ruthven. *Information at work: information management in the workplace*. Facet Publishing, 2019.

[26]. K. Byström and K. Järvelin. "Task Complexity Affects Information Seeking and Use". In: *Inf. Process. Manag.* 31.2 (1995), pp. 191–213. doi: 10.1016/0306- 4573(95)80035-R. url: https://doi.org/10.1016/0306-4573(95)80035-R.

[27]. K. Byström and S. Kumpulainen. "Vertical and horizontal relationships amongst task-based information needs". In: *Inf. Process. Manag.* 57.2 (2020), p. 102065. doi: 10.1016/j.ipm.2019.102065.

[28]. K. Byström and S. Kumpulainen. "Vertical and horizontal relationships amongst task-based information needs". In: *Inf. Process. Manag.* 57.2 (2020), p. 102065. doi: 10.1016/J.IPM.2019.102065. url: https://doi.org/10.1016/j.ipm. 2019.102065.

[29]. B. B. Cambazoglu, L. Tavakoli, F. Scholer, M. Sanderson, and W. B. Croft. "An Intent Taxonomy for Questions Asked in Web Search". In: *CHIIR '21: ACM SIGIR Conference on Human Information Interaction and Retrieval, Canberra, ACT, Australia, March 14-19, 2021*. Ed. by F. Scholer, P. Thomas, D. Elsweiler, H. Joho, N. Kando, and C. Smith. ACM, 2021, pp. 85–94. doi: 10 .1145/3406522.3446027. url: https://doi.org/10.1145/3406522.3446027.

[30]. R. Capra, J. Arguello, H. O'Brien, Y. Li, and B. Choi. "The Effects of Manipulating Task Determinability on Search Behaviors and Outcomes". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*. Ed. by K. Collins-Thompson, Q. Mei, B. D. Davison, Y. Liu, and E. Yilmaz. ACM, 2018, pp. 445–454. doi: 10.1145/3209978.3210047. url: https://doi.org/10. 1145/3209978.3210047.

[31]. B. Choi, J. Arguello, and R. Capra. "Understanding Procedural Search Tasks "in the Wild"". In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, CHIIR 2023, Austin, TX, USA, March 19-23, 2023*. Ed. by J. Gwizdka and S. Y. Rieh. ACM, 2023, pp. 24–33. doi: 10 .1145/3576840.3578302. url: https://doi.org/10.1145/3576840.3578302.

[32]. B. Choi, J. Arguello, R. G. Capra, and A. R. Ward. "OrgBox: A Knowledge Representation Tool to Support Complex Search Tasks". In: *CHIIR '21: ACM SIGIR Conference on Human Information Interaction and Retrieval, Canberra, ACT, Australia, March 14-19, 2021*. Ed. by F. Scholer, P. Thomas, D. Elsweiler, H. Joho, N. Kando, and C. Smith. ACM, 2021, pp. 219–228. doi: 10.1145/3406522.3446029. url: https://doi.org/10.1145/3406522.3446029.

[33]. B. Choi, S. Casteel, R. Capra, and J. Arguello. "Procedural Knowledge Search by Intelligence Analysts". In: *CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval, March 14 - 18, 2022*. Ed. by D. Elsweiler. Regensburg, Germany: ACM, 2022, pp. 169–179. doi: 10 . 1145/3498366 . 3505810. url: https://doi.org/10.1145/3498366.3505810.

[34]. M. D. Cohen, J. G. March, and J. P. Olsen. "A garbage can model of organizational choice". In: *Administrative science quarterly* 17 (1972), pp. 1–25.

[35]. M. J. Cole, C. Hendahewa, N. J. Belkin, and C. Shah. "User Activity Patterns During Information Search". In: *ACM Trans. Inf. Syst.* 33.1 (2015), 1:1–1:39.

[36]. N. Craswell, D. Hawking, A.-M. Vercoustre, and P. Wilkins. "P@noptic expert: Searching for experts not just for documents". In: *Ausweb Poster Proceedings*. Vol. 15. Queensland, Australia: CSIRO Australia, 2001, p. 17.

[37]. J. Crawford and C. Irving. "Information literacy in the workplace: A qualitative exploratory study". In: *JOLIS* 41.1 (2009), pp. 29–38. doi: 10.1177/0961000608099897. url: https://doi.org/10.1177/0961000608099897.

[38]. R. F. Dam and Y. S. Teo. What is Design Thinking and Why Is It So Popular? https://www.interaction-design.org/literature/article/what-is- design-thinking-and-why-is-it-so-popular. Accessed: 11-04-2024. 2024.

[39]. P. Deacon, J. B. Smith, and S. Tow. “Using metadata to create navigation paths in the HealthInsite Internet gateway”. In: *Health Information & Libraries Journal* 18.1 (2001), pp. 20–29.

[40]. J. T. DeCuir-Gunby, P. L. Marshall, and A. W. McCulloch. “Developing and using a codebook for the analysis of interview data: An example from a professional development research project”. In: *Field methods* 23.2 (2011), pp. 136– 155.

[41]. A. van Deursen and J. van Dijk. “Civil Servants’ Internet Skills: Are They Ready for E-Government?” In: *Electronic Government, 9th IFIP WG 8.5 International Conference, EGOV 2010, Lausanne, Switzerland, August 29 - September 2, 2010. Proceedings*. Ed. by M. A. Wimmer, J. Chappelet, M. Janssen, and H. J. Scholl. Vol. 6228. Lecture Notes in Computer Science. Springer, 2010, pp. 132–143. doi: 10.1007/978-3-642-14799-9_12. url: https://doi.org/10.1007/978-3-642-14799-9%5C_12.

[42]. D. Donato, F. Bonchi, T. Chi, and Y. S. Maarek. “Do you want to take notes?: identifying research missions in Yahoo! search pad”. In: *Proceedings of the 19th International Conference on World Wide Web, WWW 2010*. Releigh, USA: ACM, 2010, pp. 321–330. doi: 10.1145/1772690.1772724. url: https://doi.org/10.1145/1772690.1772724.

[43]. N. Edelmann and V. Albrecht. “Designing public participation in the digital age: Lessons learned from using the policy cycle in an Austrian case study”. In: *Proceedings of the 24th Annual International Conference on Digital Government Research, DGO 2023, July 11-14, 2023*. Ed. by D. Duenas-Cid, N. Sabatini, L. Hagen, and H. Liao. Gdan´sk, Poland: ACM, 2023, pp. 300–308. doi: 10.1145/3598469.3598502. url: https://doi.org/10.1145/3598469.3598502.

[44]. T. Erjavec, M. Ogrodniczuk, P. Osenova, N. Ljubeˇsi´c, K. Simov, A. Panˇcur, M. Rudolf, M. Kopp, S. Barkarson, S. Steingrímsson, et al. “The ParlaMint corpora of parliamentary proceedings”. In: *Language resources and evaluation* (2022), pp. 1–34.

[45]. E. Faessler, M. Oleynik, and U. Hahn. “What Makes a Top-Performing Precision Medicine Search Engine?: Tracing Main System Features in a Systematic Way”. In: *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*. Ed. by J. Huang, Y. Chang, X. Cheng, J. Kamps, V. Murdock, J. Wen, and Y. Liu. ACM, 2020, pp. 459–468. doi: 10.1145/3397271.3401048. url: https://doi.org/10.1145/3397271.3401048.

[46]. S. Ferilli, D. Redavid, and D. D. Pierro. “Holistic graph-based document representation and management for open science”. In: *Int. J. Digit. Libr.* 24.4 (2023), pp. 205–227. doi: 10.1007/S00799-022-00328-Z. url: https://doi.org/10.1007/s00799-022-00328-z.

[47]. Findwise. *Enterprise Search and Findability Survey 2015*. 2015. url: https://findwise.com/en/enterprise-search-and-findability-survey-2015.

[48]. Findwise. *Enterprise Search and Findability Survey 2015*. Tech. rep. Findwise, 2015. url: https://findwise.com/en/enterprise-search-and-findability-survey-2015.

[49]. Findwise. *Findability strategy - how to get started*. 2016. url: http://www.findwise.com/en/strategy-guide-get-started.

[50]. L. Freund, E. G. Toms, and C. L. Clarke. “Modeling task-genre relationships for IR in the workplace”. In: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. 2005, pp. 441–448.

[51]. L. Freund, E. G. Toms, and J. Waterhouse. “Modeling the information behaviour of software engineers using a work - task framework”. In: *Sparking Synergies: Bringing Research and Practice Together - Proceedings of the 68th ASIS&T Annual Meeting, ASIST 2005, Charlotte, North Carolina, USA, October 28 - November 2, 2005*. Vol. 42. Proc. Assoc. Inf. Sci. Technol. Wiley, 2005. doi: 10.1002/meet.14504201181. url: https://doi.org/10.1002/meet. 14504201181.

[52]. E. Frøkjær, M. Hertzum, and K. Hornbæk. “Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?” In: *Proceedings of the CHI 2000 Conference on Human factors in computing systems, The Hague, The Netherlands, April 1-6, 2000*. Ed. by T. Turner and G. Szwillus. The Hague, The Netherlands: ACM, 2000, pp. 345–352. doi: 10.1145/332040.332455. url: https://doi.org/10.1145/332040.332455.

[53]. D. Garigliotti, K. Balog, K. Hose, and J. Bjerva. “Recommending tasks based on search queries and missions”. In: *Nat. Lang. Eng.* 30.3 (2024), pp. 577– 601. doi: 10.1017/S1351324923000219. url: https://doi.org/10.1017/s1351324923000219.

[54]. S. D. Gollapalli, P. Mitra, and C. L. Giles. “Similar researcher search in academic environments”. In: *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL ’12, June 10-14, 2012*. Ed. by K. B. Boughida, B. Howard, M. L. Nelson, H. V. de Sompel, and I. Sølvberg. Washington, DC, USA: ACM, 2012, pp. 167–170. doi: 10.1145/2232817.2232849. url: https://doi.org/10.1145/2232817.2232849.

[55]. R. Gonçalves and C. F. Dorneles. “Automated Expertise Retrieval: A Taxonomy-Based Survey and Open Issues”. In: *ACM Comput.*

Surv. 52.5 (2019), 96:1–96:30. doi: 10.1145/3331000. url: https://doi.org/10.1145/3331000.

[56]. P. Groth and L. Moreau. *PROV-overview. An overview of the PROV family of documents*. 2013. url: https://www.w3.org/TR/prov-overview/.

[57]. J. T. Hackos and J. Redish. *User and task analysis for interface design*. Vol. 1. New York, NY: Wiley New York, 1998.

[58]. S. Han, D. He, J. Jiang, and Z. Yue. "Supporting exploratory people search: a study of factor transparency and user control". In: 22nd ACM International Conference on Information and Knowledge Management, CIKM'13, October 27 - November 1, 2013. Ed. by Q. He, A. Iyengar, W. Nejdl, J. Pei, and R. Rastogi. San Francisco, CA, USA: ACM, 2013, pp. 449–458. doi: 10.1145/2505515.2505684. url: https://doi.org/10.1145/2505515.2505684.

[59]. S. Han, D. Zhang, D. He, and Q. Cheng. "User exploration of slider facets in interactive people search system". In: *IConference 2016 Proceedings* 1.1 (2016).

[60]. P. Hansen, A. Järvelin, G. Eriksson, and J. Karlgren. "A Use Case Framework for Information Access Evaluation". In: *Professional Search in the Modern World - COST Action IC1002 on Multilingual and Multifaceted Interactive Information Access*. Vol. 8830. Lecture Notes in Computer Science. Springer, 2014, pp. 6–22.

[61]. F. Hasibi, K. Balog, and S. E. Bratsberg. "Dynamic Factual Summaries for Entity Cards". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*. Ed. by N. Kando, T. Sakai, H. Joho, H. Li, A. P. de Vries, and R. W. White. ACM, 2017, pp. 773–782. doi: 10.1145/3077136. 3080810. url: https://doi.org/10.1145/3077136.3080810.

[62]. M. Heidari, S. Zad, B. Berlin, and S. Rafatirad. "Ontology creation model based on attention mechanism for a specific business domain". In: *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE. 2021, pp. 1–5.

[63]. M. Hertzum. "Expertise seeking: A review". In: *Inf. Process. Manag.* 50.5 (2014), pp. 775–795. doi: 10 .1016 /j.ipm .2014 .04 .003. url: https://doi.org/10.1016/j.ipm.2014.04.003.

[64]. D. Hienert, M. Mitsui, P. Mayr, C. Shah, and N. J. Belkin. "The Role of the Task Topic in Web Search of Different Task Types". In: *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval, CHIIR 2018*. New Brunswick, NJ: ACM, 2018, pp. 72–81.

[65]. D. Hienert, M. Mitsui, P. Mayr, C. Shah, and N. J. Belkin. "The Role of the Task Topic in Web Search of Different Task Types". In: *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval, CHIIR 2018, New Brunswick, NJ, USA, March 11-15, 2018*. Ed. by C. Shah, N. J. Belkin, K. Byström, J. Huang, and F. Scholer. NJ, USA: ACM, 2018, pp. 72–81. doi: 10. 1145/3176349.3176382. url: https://doi.org/10.1145/3176349.3176382.

[66]. M. Hildebrand et al. "End-user support for access to heterogeneous linked data". PhD thesis. SIKS, 2010.

[67]. K. Hofmann, K. Balog, T. Bogers, and M. de Rijke. "Contextual factors for finding similar experts". In: *J. Assoc. Inf. Sci. Technol.* 61.5 (2010), pp. 994–1014. doi: 10.1002/asi.21292. url: https://doi.org/10.1002/asi.21292.

[68]. K. Hornbæk and E. L. Law. "Meta-analysis of correlations among usability measures". In: *Proceedings of the 2007 Conference on Human Factors in Computing Systems, CHI 2007, San Jose, California, USA, April 28 - May 3, 2007*. Ed. by M. B. Rosson and D. J. Gilmore. CA, USA: ACM, 2007, pp. 617–626. doi: 10. 1145/1240624.1240722. url: https://doi.org/10.1145/1240624.1240722.

[69]. O. Husain, N. Salim, R. A. Alias, S. Abdelsalam, and A. Hassan. "Expert finding systems: A systematic review". In: *Applied Sciences* 9.20 (2019), p. 4250.

[70]. H. Huurdeman. *Supporting the complex dynamics of the information seeking process*. University of Amsterdam, 2017.

[71]. E. Hyvönen, P. Leskinen, L. Sinikallio, M. L. Mela, J. Tuominen, K. Elo, S. Drobac, M. Koho, E. Ikkala, M. Tamper, R. Leal, and J. Kesäniemi. "Finnish Parliament on the Semantic Web: Using ParliamentSampo Data Service and Semantic Portal for Studying Political Culture and Language". In: *Proceedings of the Digital Parliamentary Data in Action (DiPaDA 2022) Workshop co-located with 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022), Uppsala, Sweden, March 15, 2022*. Ed. by M. L. Mela, F. Norén, and E. Hyvönen. Vol. 3133. CEUR Workshop Proceedings. CEUR- WS.org, 2022, pp. 69–85. url: https://ceur-ws.org/Vol-3133/paper05. pdf.

[72]. O. Ibrahim and A. Larsson. "Intelligibility and Transparency in Model-based Collaborative Governance". In: *20th Annual International Conference on Digital Government Research, DG.O 2019, June 18-20, 2019*. Ed. by Y. Chen, F. Salem, and A. Zuiderwijk. Dubai, United Arab Emirates: ACM, 2019, pp. 214–226. doi: 10.1145/3325112.3325247. url: https://doi.org/10.1145/3325112. 3325247.

[73]. P. Ingwersen and K. Järvelin. *The Turn - Integration of Information Seeking and Retrieval in Context*. Vol. 18. The Kluwer International Series on

Information Retrieval. Dordrecht, the Netherlands: Springer, 2005. isbn: 978-1-4020-3850-1.

[74]. P. Ingwersen and K. Järvelin. *The turn: Integration of information seeking and retrieval in context*. Vol. 18. Springer Science & Business Media, 2006.

[75]. W. Iso. "9241-11. Ergonomic requirements for office work with visual display terminals (VDTs)". In: *The international organization for standardization* 45.9 (1998).

[76]. K. Järvelin and T. D. Wilson. "On conceptual models for information seeking and retrieval research". In: *Inf. Res.* 9.1 (2003).

[77]. H. Joho. "Diane Kelly: Methods for evaluating interactive information retrieval systems with users - Foundation and Trends in Information Retrieval, vol 3, nos 1-2, pp 1-224, 2009, ISBN: 978-1-60198-224-7". In: *Inf. Retr.* 14.2 (2011), pp. 204–207. doi: 10.1007/s10791-010-9149-1. url: https://doi.org/10. 1007/s10791-010-9149-1.

[78]. T. Jonasen. "Automatic indexing in e-government: Improved access to administrative documents for professional users?" English. PhD thesis. Aalborg Universitet, Oct. 2012.

[79]. W. Jones, R. Capra, A. Diekema, J. Teevan, M. A. Pérez-Quin˜ones, J. D. Dinneen, and B. M. Hemminger. ""For Telling" the Present: Using the Delphi Method to Understand Personal Information Management Practices". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, Seoul, Republic of Korea, April 18-23, 2015*. Ed. by B. Begole, J. Kim, K. Inkpen, and W. Woo. ACM, 2015, pp. 3513–3522. doi: 10. 1145/2702123.2702523. url: https://doi. org/10.1145/2702123.2702523.

[80]. M. R. Kamdar, D. Zeginis, A. Hasnain, S. Decker, and H. F. Deus. "ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research". In: *J. Biomed. Informatics* 47 (2014), pp. 112–130. doi: 10.1016/J. JBI.2013.10.001. url: https://doi.org/10.1016/j.jbi.2013.10.001.

[81]. A. Kanaan, A. AL-Hawamleh, A. Abulfaraj, H. Al-Kaseasbeh, and A. Alorfi. "The effect of quality, security and privacy factors on trust and intention to use e-government services". In: *International Journal of Data and Network Science* 7.1 (2023), pp. 185–198.

[82]. R. Kaptein and M. Marx. "Focused retrieval and result aggregation with political data". In: *Inf. Retr.* 13.5 (2010), pp. 412–433. doi: 10.1007/s10791-010- 9130-z. url: https://doi.org/10.1007/s10791-010-9130-z.

[83]. D. Kelly, J. Arguello, A. Edwards, and W. Wu. "Development and Evaluation of Search Tasks for IIR Experiments using a Cognitive Complexity Framework". In: *Proceedings of the 2015 International Conference on The*

Theory of Information Retrieval, ICTIR 2015, Northampton, Massachusetts, USA, September 27-30, 2015. Ed. by J. Allan, W. B. Croft, A. P. de Vries, and C. Zhai. ACM, 2015, pp. 101–110. doi: 10.1145/2808194.2809465. url: https://doi.org/10.1145/2808194.2809465.

[84]. D. Kelly, J. Arguello, A. Edwards, and W. Wu. “Development and Evaluation of Search Tasks for IIR Experiments using a Cognitive Complexity Framework”. In: *Proceedings of the 2015 International Conference on The Theory of Information Retrieval, ICTIR 2015*. Northampton, MA: ACM, 2015, pp. 101–110.

[85]. K. Kovar, J. Fürnkranz, J. Petrak, B. Pfahringer, R. Trappl, and G. Widmer. “Searching for Patterns in Political Event Sequences: Experiments with the Keds Database”. In: *Cybern. Syst.* 31.6 (2000), pp. 649–668. doi: 10.1080/01969720050143184. url: https://doi.org/10.1080/01969720050143184.

[86]. U. Kruschwitz and C. Hull. “Searching the Enterprise”. In: *Found. Trends Inf. Retr.* 11.1 (2017), pp. 1–142. doi: 10.1561/1500000053. url: https://doi.org/10.1561/1500000053.

[87]. C. C. Kuhlthau. “Inside the search process: Information seeking from the user’s perspective”. In: *J. Am. Soc. Inf. Sci.* 42.5 (1991), pp. 361–371. doi: 10.1002/(SICI)1097-4571(199106)42:5\<361::AID-ASI6\>3.0.CO;2-\%23. url: https://doi.org/10.1002/(SICI)1097- 4571(199106)42:5%5C%3C361:: AID-ASI6%5C%3E3.0.CO;2-%5C%23.

[88]. E. Laparra, I. Aldabe, and G. Rigau. “From TimeLines to StoryLines: A preliminary proposal for evaluating narratives”. In: *Proceedings of the First Workshop on Computing News Storylines*. Beijing, China: Association for Computational Linguistics, 2015, pp. 50–55.

[89]. G. J. Leckie, K. E. Pettigrew, and C. Sylvain. “Modeling the information seeking of professionals: a general model derived from research on engineers, health care professionals, and lawyers”. In: *The Library Quarterly*. Vol. 66, No 2. Chicago, IL: University of Chicago Press, 1996, pp. 161–193.

[90]. E. Lee, M. Dobbins, K. DeCorby, L. McRae, D. Tirilis, and H. Husson. “An optimal search filter for retrieving systematic reviews and meta-analyses”. In: *BMC medical research methodology* 12 (2012), pp. 1–11.

[91]. J. R. Lewis. “The System Usability Scale: Past, Present, and Future”. In: *Int.J. Hum. Comput. Interact.* 34.7 (2018), pp. 577–590. doi: 10.1080/10447318.2018.1455307. url: https://doi.org/10.1080/10447318.2018.1455307.

[92]. Y. Li and H. Shang. “How does e-government use affect citizens’ trust in government? Empirical evidence from China”. In: *Information & Management* 60.7 (2023), p. 103844.

[93]. Y. Li. “Exploring the relationships between work task and search task in information search”. In: *JASIST* 60.2 (2009), pp. 275–291. doi: 10.1002/asi.20977. url: https://doi.org/10.1002/asi.20977.

[94]. Y. Li and N. J. Belkin. “A faceted approach to conceptualizing tasks in information seeking”. In: *Inf. Process. Manage.* 44.6 (2008), pp. 1822–1837. doi: 10.1016/j.ipm.2008.07.005. url: https://doi.org/10.1016/j.ipm.2008.07.005.

[95]. Y. Li and N. J. Belkin. “An exploration of the relationships between work task and interactive information search behavior”. In: *J. Assoc. Inf. Sci. Technol.* 61.9 (2010), pp. 1771–1789. doi: 10.1002/ASI.21359. url: https://doi. org/10.1002/asi.21359.

[96]. R. Liebregts and T. Bogers. “Design and Evaluation of a University-Wide Expert Search Engine”. In: *Advances in Information Retrieval, 31th European Conference on IR Research, ECIR 2009, April 6-9, 2009. Proceedings*. Ed. by M. Boughanem, C. Berrut, J. Mothe, and C. Soulé-Dupuy. Vol. 5478. Lecture Notes in Computer Science. Toulouse, France: Springer, 2009, pp. 587–594. doi: 10.1007/978-3-642-00958-7_54. url: https://doi.org/10.1007/978-3-642-00958-7%5C_54.

[97]. R. C. Lima and R. L. T. Santos. “On Extractive Summarization for Profile-centric Neural Expert Search in Academia”. In: *SIGIR ’22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, July 11 - 15, 2022*. Ed. by E. Amig´o, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, and G. Kazai. Madrid, Spain: ACM, 2022, pp. 2331– 2335. doi: 10.1145/3477495.3531713. url: https://doi.org/10.1145/3477495.3531713.

[98]. C. Liu, Y. Liu, J. Liu, and R. Bierig. “Search Interface Design and Evaluation”. In: *Found. Trends Inf. Retr.* 15.3-4 (2021), pp. 243–416. doi: 10.1561/1500000073. url: https://doi.org/10.1561/1500000073.

[99]. H.-C. Liu. “Investigating the impact of cognitive style on multimedia learners’ understanding and visual search patterns: an eye-tracking approach”. In: *Journal of Educational Computing Research* 55.8 (2018), pp. 1053–1068.

[100]. T. Liu. *Learning to Rank for Information Retrieval*. Springer, 2011. isbn: 978- 3-642-14266-6. doi: 10.1007/978-3-642-14267-3. url: https://doi.org/10.1007/978-3-642-14267-3.

[101]. A. Livne, V. Gokuladas, J. Teevan, S. T. Dumais, and E. Adar. “CiteSight: supporting contextual citation recommendation using differential search”. In: *The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’14, Gold Coast , QLD, Australia - July 06 - 11, 2014*. Ed. by S. Geva, A. Trotman, P. Bruza, C. L. A.

Clarke, and K. Järvelin. ACM, 2014, pp. 807–816. doi: 10 . 1145/2600428 . 2609585. url: https://doi.org/10.1145/2600428.2609585.

[102]. G. Luo and C. Tang. "On iterative intelligent medical search". In: *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008, Singapore, July 20-24, 2008*. Ed. by S. Myaeng, D. W. Oard, F. Sebastiani, T. Chua, and M. Leong. ACM, 2008, pp. 3–10. doi: 10.1145/1390334.1390338. url: https://doi.org/10.1145/1390334.1390338.

[103]. M. Lykke, A. Bygholm, L. B. Søndergaard, and K. Byström. "The role of historical and contextual knowledge in enterprise search". In: *Journal of Documentation* (2021).

[104]. M. Lykke, A. Bygholm, L. B. Søndergaard, and K. Byström. "The role of historical and contextual knowledge in enterprise search". In: *J. Documentation* 78.5 (2022), pp. 1053–1074. doi: 10.1108/JD-08-2021-0170. url: https://doi.org/10.1108/JD-08-2021-0170.

[105]. V. Mangaravite and R. L. T. Santos. "On Information-Theoretic Document-Person Associations for Expert Search in Academia". In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, July 17-21, 2016*. Ed. by R. Perego, F. Sebastiani, J. A. Aslam, I. Ruthven, and J. Zobel. Pisa,Italy: ACM, 2016, pp. 925–928. doi: 10.1145/2911451.2914751. url: https://doi.org/10.1145/2911451. 2914751.

[106]. R. Marcella, G. Baxter, S. Davies, and D. Toornstra. "The information needs and information-seeking behaviour of the users of the European Parliamentary Documentation Centre: A customer knowledge study". In: *Journal of Documentation* 63.6 (2007), pp. 920–934.

[107]. A. Marzouki, S. Mellouli, and S. Daniel. "Spatial, temporal and semantic contextualization of citizen participation". In: *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, DG.O 2018, May 30 - June 01, 2018*. Ed. by M. Janssen, S. A. Chun, and V. Weerakkody. Delft, the Netherlands: ACM, 2018, 63:1–63:8. doi: 10.1145/3209281.3209385. url: https://doi.org/10.1145/3209281.3209385.

[108]. A. Minard, M. Speranza, E. Agirre, I. Aldabe, M. van Erp, B. Magnini, G. Rigau, and R. Urizar. "SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering". In: *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, June 4-5, 2015*. Ed. by D. M. Cer, D. Jurgens, P. Nakov, and T. Zesch. Denver, Colorado, USA: The Association for Computer Linguistics, 2015, pp. 778–786. doi: 10.18653/v1/s15-2132. url: https://doi.org/10.18653/v1/s15-2132.

[109]. J. Mori, N. Basselin, A. Kröner, and A. Jameson. “Find me if you can: designing interfaces for people search”. In: *Proceedings of the 13th International Conference on Intelligent User Interfaces, IUI 2008, January 13-16, 2008*. Ed. by J. M. Bradshaw, H. Lieberman, and S. Staab. Gran Canaria, Canary Islands, Spain: ACM, 2008, pp. 377–380. doi: 10.1145/1378773.1378834. url: https://doi.org/10.1145/1378773.1378834.

[110]. M. R. Morris, J. Teevan, and K. Panovich. “What do people ask their social networks, and why?: a survey study of status message q&a behavior”. In: *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010*. Ed. by E. D. Mynatt, D. Schoner, G. Fitzpatrick, S. E. Hudson, W. K. Edwards, and T. Rodden. Atlanta, Georgia, USA: ACM, 2010, pp. 1739–1748. doi: 10.1145/1753326.1753587. url: https://doi.org/10.1145/1753326.1753587.

[111]. H. Murakami, H. Ueda, S. Kataoka, Y. Takamori, and S. Tatsumi. “Summarizing and Visualizing Web People Search Results”. In: *ICAART 2010 - Proceedings of the International Conference on Agents and Artificial Intelligence, Volume 1 - Artificial Intelligence, January 22-24, 2010*. Ed. by J. Filipe, A. L. N. Fred, and B. Sharp. Valencia, Spain: INSTICC Press, 2010, pp. 640–643.

[112]. H. Murakami, H. Ueda, S. Kataoka, Y. Takamori, and S. Tatsumi. “Summarizing and visualizing web people search results”. In: *International Conference on Agents and Artificial Intelligence*. Vol. 2. SCITEPRESS. 2010, pp. 640–643.

[113]. T. S. Network. *Search Insights 2018*. Ed. by intranetfocus.com. [Online; posted 25-March-2018]. Mar. 2018. url: %5Curl % 7Bhttp ://intranetfocus.com/search-insights-2018-from-the-search-network/%20%7D.

[114]. J. Nielsen. *Usability 101: Introduction to usability*. Jan. 2012. url: https://www.nngroup.com/articles/usability-101- introduction-to-usability/.

[115]. B. K. Norambuena, T. Mitra, and C. North. “A Survey on Event-based News Narrative Extraction”. In: *CoRR* abs/2302.08351 (2023). doi: 10 . 48550/arXiv.2302.08351. arXiv: 2302.08351. url: https://doi.org/10.48550/arXiv.2302.08351.

[116]. E. NV. *Elasticsearch 7.16.2 [computer software]*. 2022. url: https://www . elastic.co/.

[117]. D. Odijk, C. Gârbacea, T. Schoegje, L. Hollink, V. de Boer, K. Ribbens, and J. van Ossenbruggen. “Supporting Exploration of Historical Perspectives Across Collections”. In: *Research and Advanced Technology for Digital Libraries 19th International Conference on Theory and Practice of Digital Libraries, TPDL 2015, Poznan´, Poland, September 14-18, 2015. Proceedings*. Ed. by S. Kapidakis, C. Mazurek, and M. Werla. Vol. 9316.

Lecture Notes in Computer Science. Springer, 2015, pp. 238–251. doi: 10.1007/978-3-319-24592-8_18. url: https://doi.org/10.1007/978-3-319-24592-8%5C_18.

[118]. A. Oeldorf-Hirsch, B. J. Hecht, M. R. Morris, J. Teevan, and D. Gergle. "To search or to ask: the routing of information needs between traditional search engines and social networks". In: *Computer Supported Cooperative Work, CSCW '14, Baltimore, MD, USA, February 15-19, 2014*. Ed. by S. R. Fussell, W. G. Lutters, M. R. Morris, and M. Reddy. MD, USA: ACM, 2014, pp. 16–27. doi: 10.1145/2531602.2531706. url: https://doi.org/10.1145/2531602.2531706.

[119]. G. Pardi, S. Gottschling, P. Gerjets, and Y. Kammerer. "The moderating effect of knowledge type on search result modality preferences in web search scenarios". In: *Computers and Education Open* 4 (2023), p. 100126.

[120]. S. A. Paul. "Find an Expert: Designing Expert Selection Interfaces for Formal Help-Giving". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, May 7-12, 2016*. Ed. by J. Kaye, A. Druin, C. Lampe, D. Morris, and J. P. Hourcade. San Jose, CA, USA: ACM, 2016, pp. 3038–3048. doi: 10.1145/2858036.2858131. url: https://doi.org/10.1145/2858036. 2858131.

[121]. M. Peacock. *The search for expert knowledge continues*. May 2009. url: https://www . cmswire . com/cms/enterprise-cms/the-search-for-expert-knowledge-continues-004594.php.

[122]. W. H. O. R&D Blue Print. *A coordinated Global Research Roadmap*. https://www.who.int/blueprint/priority- diseases/key- action/Roadmap- version-FINAL-for-WEB.pdf?ua=1, Accessed: 18-10-2020.

[123]. G. Ramirez and A. P. de Vries. "Relevant contextual features in XML retrieval". In: *Proceedings of the 1st International Conference on Information Interaction in Context, IIiX 2006*. Copenhagen, Denmark: ACM, 2006, pp. 56–65.

[124]. R. Riding and I. Cheema. "Cognitive Styles—an overview and integration". In: *Educational Psychology* 11.3-4 (1991), pp. 193–215. doi: 10 . 1080/0144341910110301. eprint: https://doi.org/10.1080/0144341910110301. url: https://doi.org/10.1080/0144341910110301.

[125]. D. E. Rose and D. Levinson. "Understanding user goals in web search". In: *Proceedings of the 13th international conference on World Wide Web, WWW 2004, New York, NY, USA, May 17-20, 2004*. Ed. by S. I. Feldman, M. Uretsky, . Najork, and C. E. Wills. ACM, 2004, pp. 13–19. doi: 10.1145/988672.988675. url: https://doi.org/10.1145/988672.988675.

[126]. N. Roy, D. Maxwell, and C. Hauff. "Users and Contemporary SERPs: A (Re) Investigation". In: *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, July 11 - 15, 2022*. Ed. by E. Amig´o, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, and G. Kazai. Madrid, Spain: ACM, 2022, pp. 2765–2775. doi: 10.1145/3477495.3531719. url: https://doi.org/10.1145/3477495.3531719.

[127]. T. Russell-Rose and J. Chamberlain. "Expert search strategies: the information retrieval practices of healthcare information professionals". In: *JMIR medical informatics* 5.4 (2017), e33.

[128]. T. Russell-Rose, J. Chamberlain, and L. Azzopardi. "Information retrieval in the workplace: A comparison of professional search practices". In: *Inf. Process. Manag.* 54.6 (2018), pp. 1042–1057. doi: 10.1016/J.IPM.2018.07.003. url: https://doi.org/10.1016/j.ipm.2018.07.003.

[129]. T. Russell-Rose, J. Chamberlain, and L. Azzopardi. "Information retrieval in the workplace: A comparison of professional search practices". In: *Inf. Process. Manag.* 54.6 (2018), pp. 1042–1057. doi: 10.1016/j.ipm.2018.07.003.

[130]. T. Russell-Rose, J. Lamantia, and M. Burrell. "A Taxonomy of Enterprise Search". In: *Proceedings of the 1st European Workshop on Human-Computer Interaction and Information Retrieval*. Ed. by M. L. Wilson, T. Russell-Rose, B. Larsen, and J. Kalbach. Vol. 763. CEUR Workshop Proceedings. Newcastle, UK: CEUR-WS.org, 2011, pp. 15–18. url: http://ceur-ws.org/Vol- 763/paper4.pdf.

[131]. T. Russell-Rose and S. Makri. "Designing for Consumer Search Behaviour". In: Proceedings of HCIR 2012 (2012).

[132]. T. Russell-Rose and T. Tate. *Designing the search experience - the information architecture of discovery*. Morgan Kaufmann, 2012. isbn: 978-0-12-396981-1. url: https://www.elsevier.com/books/designing-the-search-experience/russell-rose/978-0-12-396981-1.

[133]. M. Saastamoinen and K. Järvelin. "Queries in authentic work tasks: the effects of task type and complexity". In: *J. Documentation* 72.6 (2016), pp. 1114–1133. doi: 10.1108/JD-09-2015-0119. url: https://doi.org/10.1108/JD-09-2015-0119.

[134]. M. Saastamoinen and K. Järvelin. "Queries in authentic work tasks: the effects of task type and complexity". In: *Journal of Documentation* 72.6 (2016), pp. 1114– 1133.

[135]. M. Saastamoinen and S. Kumpulainen. "Expected and materialised information source use by municipal officials: intertwining with task complexity". In: *Inf. Res.* 19.4 (2014).

[136]. T. Saracevic. “Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance”. In: *J. Assoc. Inf. Sci. Technol.* 58.13 (2007), pp. 1915–1933. doi: 10.1002/asi.20682. url: https://doi.org/10.1002/asi.20682.

[137]. S. Sarkar, M. Amirizaniani, and C. Shah. “Representing Tasks with a Graph- Based Method for Supporting Users in Complex Search Tasks”. In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, CHIIR 2023, Austin, TX, USA, March 19-23, 2023*. Ed. by J. Gwizdka and S. Y. Rieh. ACM, 2023, pp. 378–382. doi: 10.1145/3576840.3578279. url: https://doi.org/10.1145/3576840.3578279.

[138]. T. Schoegje, A. de Vries, L. Hardman, and T. Pieters. “Improving the Effectiveness and Efficiency of Web-Based Search Tasks for Policy Workers”. In: *Information* 14.7 (2023), p. 371.

[139]. T. Schoegje, A. P. de Vries, L. Hardman, and T. Pieters. “Improving the Effectiveness and Efficiency of Web-Based Search Tasks for Policy Workers”. In: *Inf.* 14.7 (2023), p. 371. doi: 10.3390/info14070371. url: https://doi.org/10.3390/info14070371.

[140]. T. Schoegje, A. P. de Vries, and T. Pieters. “Adapting a Faceted Search Task Model for the Development of a Domain-Specific Council Information Search Engine”. In: *Electronic Government - 21st IFIP WG 8.5 International Conference, EGOV 2022, Linköping, Sweden, September 6-8, 2022, Proceedings*. Ed. by M. Janssen, C. Csáki, I. Lindgren, E. N. Loukis, U. Melin, G. V. Pereira, M. P. R. Bolívar, and E. Tambouris. Vol. 13391. Lecture Notes in Computer Science. Springer, 2022, pp. 402–418. doi: 10.1007/978-3-031-15086-9_26. url: https://doi.org/10.1007/978-3-031-15086-9%5C_26.

[141]. T. Schoegje, A. P. de Vries, and T. Pieters. “Adapting a Faceted Search Task Model for the Development of a Domain-Specific Council Information Search Engine”. In: *Electronic Government - 21st IFIP WG 8.5 International Conference, EGOV 2022, September 6-8, 2022, Proceedings*. Ed. by M. Janssen, C. Csáki, I. Lindgren, E. N. Loukis, U. Melin, G. V. Pereira, M. P. R. Bolívar, and E. Tambouris. Vol. 13391. Lecture Notes in Computer Science. Linköping, Sweden: Springer, 2022, pp. 402–418. doi: 10.1007/978-3-031-15086-9_26. url: https://doi.org/10.1007/978-3-031-15086-9%5C_26.

[142]. S. C. Segura-Rodas. “What tasks emerge from Knowledge Work?” In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, CHIIR 2023, Austin, TX, USA, March 19-23, 2023*. Ed. by J. Gwizdka and S. Y. Rieh. ACM, 2023, pp. 495–498. doi: 10.1145/3576840.3578328. url: https://doi.org/10.1145/3576840.3578328.

[143]. C. Shah, R. White, P. Thomas, B. Mitra, S. Sarkar, and N. J. Belkin. “Taking Search to Task”. In: *Proceedings of the 2023 Conference on Human*

Information Interaction and Retrieval, CHIIR 2023, Austin, TX, USA, March 19-23, 2023. Ed. by J. Gwizdka and S. Y. Rieh. ACM, 2023, pp. 1–13. doi: 10.1145/3576840.3578288. url: https://doi.org/10.1145/3576840.3578288.

[144]. C. Shah and R. W. White. *Task Intelligence for Search and Recommendation*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2021. doi: 10.2200/S01103ED1V01Y202105ICR074. url: https://doi.org/10.2200/S01103ED1V01Y202105ICR074.

[145]. D. Shahaf, C. Guestrin, and E. Horvitz. “”Metro maps of information” by Dafna Shahaf, Carlos Guestrin and Eric Horvitz, with Ching-man Au Yeung as coordinator”. In: *SIGWEB Newsl.* 2013.Spring (2013), 4:1–4:9. doi: 10.1145/2451836.2451840. url: https://doi.org/10.1145/2451836.2451840.

[146]. D. Shahaf, C. Guestrin, and E. Horvitz. “Metro maps of science”. In: *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’12, August 12-16, 2012*. Ed. by Q. Yang, D. Agarwal, and J. Pei. Beijing, China: ACM, 2012, pp. 1122–1130. doi: 10.1145/2339530.2339706. url: https://doi.org/10.1145/2339530.2339706.

[147]. Y. Shao, J. Mao, Y. Liu, M. Zhang, and S. Ma. “From linear to non-linear: investigating the effects of right-rail results on complex SERPs”. In: *Advances in Computational Intelligence* 2.1 (2022), p. 14.

[148]. Y. Shinyama. *Elasticsearch 7.16.2 [computer software]*. 2022. url: https://pypi.org/project/pdfminer/.

[149]. J. Singh and A. Anand. “Designing Search Tasks for Archive Search”. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, CHIIR 2017*. Oslo, Norway: ACM, 2017, pp. 361–364.

[150]. T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan. “A diary study of mobile information needs”. In: *Proceedings of the 2008 Conference on Human Factors in Computing Systems, CHI 2008*. Florence, Italy: ACM, 2008, pp. 433–442.

[151]. J. Strötgen and M. Gertz. “Multilingual and Cross-domain Temporal Tagging”. In: *Language Resources and Evaluation* 47.2 (2013), pp. 269–298. doi: 10.1007/s10579-012-9179-y.

[152]. T. Svarre, M. Lykke, and A. Bygholm. “Searching for people in the workplace: aims, behaviour, and challenges”. In: *Information Research an international electronic journal* 29.2 (2024), pp. 573–588.

[153]. G. Tanhaei, I. Troost, L. Hardman, and W. Hürst. “Designing a Topic-Based Literature Exploration Tool in AR - An exploratory study for neuroscience”. In: *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, Singapore, October 17-21, 2022*.

IEEE, 2022, pp. 471–476. doi: 10.1109/ISMAR-ADJUNCT57072.2022.00099. url: https: //doi.org/10.1109/ISMAR-Adjunct57072.2022.00099.

[154]. R. S. Taylor. "Information use environments". In: *Progress in communication sciences* 10.217 (1991), p. 55.

[155]. R. S. Taylor and R. S. Taylor. *Value-added processes in information systems*. Norwood, NJ, USA: Greenwood Publishing Group, 1986.

[156]. J. Teevan, K. Collins-Thompson, R. W. White, S. T. Dumais, and Y. Kim. "Slow Search: Information Retrieval without Time Constraints". In: *Symposium on Human-Computer Interaction and Information Retrieval, HCIR '13*. Vancouver, BC, Canada: ACM, 2013, 1:1–1:10.

[157]. M. Tikhomirov and B. V. Dobrov. "News Timeline Generation: Accounting for Structural Aspects and Temporal Nature of News Stream". In: *Data Analytics and Management in Data Intensive Domains - XIX International Conference, DAMDID/RCDL 2017, October 10-13, 2017, Revised Selected Papers*. Ed. by L. A. Kalinichenko, Y. Manolopoulos, O. Malkov, N. A. Skvortsov, S. A. Stupnikov, and V. Sukhomlin. Vol. 822. Communications in Computer and Information Science. Moscow, Russia: Springer, 2017, pp. 267–280. doi: 10.1007/978- 3-319-96553-6_19. url: https://doi.org/10.1007/978-3-319-96553-6%5C_19.

[158]. G. B. Tran, T. A. Tran, N.-K. Tran, M. Alrifai, and N. Kanhabua. "Leveraging learning to rank in an optimization framework for timeline summarization". In: *SIGIR 2013 Workshop on Time-aware Information Access (TAIA*. Coast, Queensland, Australia: ACM, 2013.

[159]. J. R. Trippas, D. Spina, F. Scholer, A. H. Awadallah, P. Bailey, P. N. Bennett, R. W. White, J. Liono, Y. Ren, F. D. Salim, and M. Sanderson. "Learning About Work Tasks to Inform Intelligent Assistant Design". In: *CHIIR 2019*. Glasgow, Scotland: ACM, 2019, pp. 5–14.

[160]. J. R. Trippas, D. Spina, F. Scholer, A. H. Awadallah, P. Bailey, P. N. Bennett, R. W. White, J. Liono, Y. Ren, F. D. Salim, and M. Sanderson. "Learning About Work Tasks to Inform Intelligent Assistant Design". In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, CHIIR 2019, Glasgow, Scotland, UK, March 10-14, 2019*. Ed. by L. Azzopardi, M. Halvey, I. Ruthven, H. Joho, V. Murdock, and P. Qvarfordt. ACM, 2019, pp. 5–14. doi: 10 .1145/3295750 .3298934. url: https://doi.org/10 .1145/3295750.3298934.

[161]. P. Vakkari. "The Usefulness of Search Results: A Systematization of Types and Predictors". In: *CHIIR '20: Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, March 14-18, 2020*. Ed. by H. L. O'Brien, L. Freund, I. Arapakis, O. Hoeber, and I. Lopatovska. ACM, 2020, pp. 243–252. doi: 10.1145/3343413.3377955.

[162]. P. Vakkari. "The Usefulness of Search Results: A Systematization of Types and Predictors". In: *CHIIR '20: Conference on Human Information Interaction and Retrieval, March 14-18, 2020*. Ed. by H. L. O'Brien, L. Freund, I. Arapakis, O. Hoeber, and I. Lopatovska. Vancouver, BC, Canada: ACM, 2020, pp. 243–252. doi: 10.1145/3343413.3377955. url: https://doi.org/10.1145/3343413. 3377955.

[163]. P. Vakkari, M. Völske, M. Potthast, M. Hagen, and B. Stein. "Modeling the usefulness of search results as measured by information use". In: *Inf. Process. Manag.* 56.3 (2019), pp. 879–894. doi: 10.1016/j.ipm.2019.02.001. url: https://doi.org/10.1016/j.ipm.2019.02.001.

[164]. E. Vassilakaki, E. Garoufallou, F. Johnson, and R. J. Hartley. "Users' information search behavior in a professional search environment". In: *Professional search in the modern world*. Springer, 2014, pp. 23–44.

[165]. S. Verberne, J. He, G. Wiggers, T. Russell-Rose, U. Kruschwitz, and A. P. de Vries. "Information search in a professional context - exploring a collection of professional search tasks". In: *CoRR* abs/1905.04577 (2019). arXiv: 1905. 04577. url: http://arxiv.org/abs/1905.04577.

[166]. M. Völske, E. Fatehifar, B. Stein, and M. Hagen. "Query-Task Mapping". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*. Ed. by B. Piwowarski, M. Chevalier, E´. Gaussier, Y. Maarek, J. Nie, and F. Scholer. ACM, 2019, pp. 969–972. doi: 10.1145/3331184.3331286. url: https://doi.org/10.1145/3331184.3331286.

[167]. L. Walter, N. M. Denter, and J. Kebel. "A review on digitalization trends in patent information databases and interrogation tools". In: *World Patent Information* 69 (2022), p. 102107.

[168]. Y. Wang, S. Sarkar, and C. Shah. "Juggling with Information Sources, Task Type, and Information Quality". In: *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval, CHIIR 2018, New Brunswick, NJ, USA, March 11-15, 2018*. Ed. by C. Shah, N. J. Belkin, K. Byström, J. Huang, and F. Scholer. ACM, 2018, pp. 82–91. doi: 10.1145/3176349.3176390. url: https://doi.org/10.1145/3176349.3176390.

[169]. M. White. *Achieving Enterprise Search Satisfaction*. 2018. url: http :// intranetfocus.com/wp-content/uploads/2018/10/Achieving-enterprise-search-satisfaction-Oct-2018.pdf.

[170]. M. White. *Enterprise search: enhancing business performance (2nd edition)*. " O'Reilly Media, Inc.", 2016.

[171]. J. L. Wildman, A. L. Thayer, M. A. Rosen, E. Salas, J. E. Mathieu, and S. R. Rayne. "Task types and team-level attributes: Synthesis of team

classification literature”. In: *Human resource development review* 11.1 (2012), pp. 97–129.

[172]. J. O. Wobbrock and J. A. Kientz. “Research contributions in human-computer interaction”. In: *Interactions* 23.3 (2016), pp. 38–44. doi: 10.1145/2907069. url: https://doi.org/10.1145/2907069.

[173]. D. Wu, S. Fan, and F. Yuan. “Research on pathways of expert finding on academic social networking sites”. In: *Inf. Process. Manag.* 58.2 (2021), p. 102475. doi: 10.1016/j.ipm.2020.102475. url: https://doi.org/10.1016/j.ipm.2020.102475.

[174]. B. Xu. “Supporting Neuroscience Literature Exploration by Utilising Indirect Relations between Topics in Augmented Reality”. In: *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval, CHIIR 2024, Sheffield, United Kingdom, March 10-14, 2024*. Ed. by P. D. Clough, M. Harvey, and F. Hopfgartner. ACM, 2024, pp. 457–460. doi: 10. 1145/3627508.3638312. url: https://doi.org/10.1145/3627508.3638312.

[175]. P. Yang, H. Fang, and J. Lin. “Anserini: Reproducible Ranking Baselines Using Lucene”. In: *ACM J. Data Inf. Qual.* 10.4 (2018), 16:1–16:20. doi: 10.1145/3239571. url: https://doi.org/10.1145/3239571.

[176]. Y. Yang and R. Capra. “Nested Contexts of Music Information Retrieval: A Framework of Contextual Factors”. In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, CHIIR 2023, Austin, TX, USA, March 19-23, 2023*. Ed. by J. Gwizdka and S. Y. Rieh. ACM, 2023, pp. 368–372. doi: 10.1145/3576840.3578322. url: https://doi.org/10.1145/3576840. 3578322.

[177]. Z. Yang, S. Yan, A. Lad, X. Liu, and W. Guo. “Cascaded Deep Neural Ranking Models in LinkedIn People Search”. In: *CIKM ’21: The 30th ACM International Conference on Information and Knowledge Management, November 1 - 5, 2021*. Ed. by G. Demartini, G. Zuccon, J. S. Culpepper, Z. Huang, and H. Tong. Virtual Event, Queensland, Australia: ACM, 2021, pp. 4312–4320. doi: 10.1145/3459637.3481899. url: https://doi.org/10.1145/3459637.3481899.

[178]. J. Yao, X. Wan, and J. Xiao. “Recent advances in document summarization”. In: *Knowl. Inf. Syst.* 53.2 (2017), pp. 297–336. doi: 10.1007/s10115-017-1042-4. url: https://doi.org/10.1007/s10115-017-1042-4.

[179]. J. Zhu, X. Huang, D. Song, and S. M. Rüger. “Integrating multiple document features in language models for expert finding”. In: *Knowl. Inf. Syst.* 23.1 (2010), pp. 29–54. doi: 10.1007/s10115-009-0202-6. url: https://doi.org/10.1007/s10115-009-0202-6.

Summary

As the municipality of Utrecht manages an ever-increasing amount of information, its employees require effective search tools to find the necessary data. With numerous information systems and tools in use, there emerged a need for a central interface to access all data storages. However, using generic functionality for diverse user groups is insufficient, as the same user input may require different results based on the user's needs and tasks. This becomes most noticeable when user needs are complex or domain-specific, necessitating specialised functionality. Therefore, organisations like the municipality must understand the varying needs of different user groups and how to support them effectively.

We investigated how to design specialised search tools that cater to domain-specific information needs. We characterised search tasks and their functional requirements to develop specialised search functionality, employing several task facets. By identifying when tasks require functionalities absent in existing information systems, we determined which search tasks need additional support. This approach was applied to characterise the tasks of municipal knowledge workers, leading to improved task performance through task-specific search functionalities.

This manuscript presents four main scientific studies. The initial study focused on characterising tasks to develop search functionality. We conducted semi-structured interviews with council members from two municipalities to identify and characterise their work and search tasks. By coding their responses, we identified five task facets representing domain-specific information requirements.

In the second study, we explored when tasks warrant specialised search functionality. Through semi-structured interviews with policy workers, we confirmed that the task facets identified are stable across different groups of municipal knowledge workers. We examined which tasks necessitate specializsed search functionality by identifying mismatches between users' tasks and the search functionality in existing systems. Two classes of tasks were identified for which support could be improved: simple web-based search tasks and complex web-based search tasks. We developed two types of search functionality to support these tasks: limiting the search index to authoritative information sources for simple web-based tasks and providing expert search functionality for complex web-based tasks. When policy workers performed representative simulated tasks, limiting the search index enabled them to perform simple tasks 1.7 times faster. Additionally, we found that users consistently resort to consulting a colleague with expertise when unable to complete a complex task by searching for a few minutes.

The third study exemplifies the application of our task-based approach to search system design, particularly for expert search tasks. Based on the characterisation of expert search tasks in previous studies, we adjusted and tested how we rank and present search results. We compared people-centric and document-centric approaches to ranking and presenting search results. Statistical analysis revealed that domain novices performed more tasks correctly when the interface presented people-centric search results, and achieved faster task completion with a document-centric ranking.

The final study demonstrates that designing functionality around search tasks allows us to transcend typical search functionalities. We supported council members with an overview perspective on council decisions. Previously, these overviews were manually constructed by policy workers, proving to be unscalable, incomplete, and subjective. Instead, we provided a method to reconstruct decision histories of policy dossiers using two methods: identifying when multiple meetings discuss the same document and identifying when a document refers to a previous document or meeting. This approach allowed us to construct approximately 2,500 decision histories of policy dossiers. When experts used these histories to understand how previous decisions were made, they indicated that 97% of the documents included were relevant. This high level of precision was achieved because our inclusion criterion was based on references made by experts with domain knowledge. Reflecting on what made these timelines useful, we derived design implications for presenting council decision histories in the interface. The primary finding was that a layered interface would enable users to initially see a concise overview (only the main decisions), while still allowing a complete overview when expanded, to display the provenance of these decisions. These findings were integrated into a search engine for council information, where council members can search for documents as usual, but clicking a search result presents the document within its broader context.

Although the search engines and functionalities presented in this manuscript are domain-specific and tailored to the municipality of Utrecht, we identified examples where the same user group at a different municipality in the Netherlands performed the same tasks. This suggests that search functionality for Dutch municipalities could be standardised, although further studies should confirm this. It is also plausible that variations of the same task exist across multiple domains. When two tasks in different domains can be characterised with the same task facets, we anticipate that a functionality designed for one domain can be adapted for another.

Samenvatting

Omdat de gemeente Utrecht werkt met een groeiende hoeveelheid informatie, hebben haar medewerkers effectieve zoekinstrumenten nodig. Naarmate de gemeente meerdere informatiesystemen begon te gebruiken, ontstond de behoefte aan een centrale locatie om alle gegevensopslagen te doorzoeken. Generieke functionaliteit is echter niet voldoende voor diverse gebruikersgroepen, omdat dezelfde gebruikersinvoer kan leiden tot verschillende resultaten, afhankelijk van de behoeften en taken van de gebruiker. Dit is vooral merkbaar wanneer gebruikers complexe of (domein-)specifieke behoeften hebben en gespecialiseerde functionaliteit nodig hebben. Daarom moeten organisaties begrijpen wat verschillende gebruikersgroepen nodig hebben en hoe deze behoeften kunnen worden ondersteund.

We hebben onderzocht hoe gespecialiseerde zoekinstrumenten ontwikkeld kunnen worden voor de (domein-)specifieke behoeften van gebruikers. We hebben zoektaken en hun functionele benodigdheden gekarakteriseerd door middel van taak-eigenschappen, met het doel om gespecialiseerde zoekfunctionaliteiten te ontwikkelen. Door te identificeren wanneer taken functionaliteiten nodig hebben die niet aanwezig zijn in bestaande informatiesystemen bepalen we welke zoektaken meer ondersteuning nodig hebben. We hebben de taken van gemeentelijke kenniswerkers gekarakteriseerd, en hun prestaties in verschillende taken verbeterd door zoekfunctionaliteiten aan te laten sluiten op de eigenschappen van taken. Dit manuscript presenteert hoofdzakelijk vier wetenschappelijke studies. De eerste studie focust op het karakteriseren van taken voor het doel om zoekfunctionaliteit te ontwikkelen. We hebben semi-gestructureerde interviews afgenomen met raadsleden van twee gemeentes om hun werktaken en zoektaken te identificeren en karakteriseren. Door hun antwoorden te coderen hebben we vijf taakeigenschappen gevonden waarmee we domein-specifieke informatiebehoeften kunnen representeren.

In de tweede studie hebben we onderzocht wanneer taken gespecialiseerde zoekfunctionaliteiten vereisen. We begonnen met het onderzoeken van de taken van beleidsmedewerkers door middel van semi-gestructureerde interviews en stelden vast dat de geïdentificeerde taak-eigenschappen stabiel zijn voor meerdere groepen van gemeentelijke kenniswerkers. We bekeken welke taken gespecialiseerde zoekfunctionaliteit nodig hebben door te kijken naar de discrepanties tussen de taken van de gebruiker en de functionaliteiten in de beschikbare systemen. Twee klassen zoektaken werden geïdentificeerd die beter ondersteund konden worden: simpele web-gebaseerde zoektaken en complexe web-gebaseerde zoektaken. Twee typen zoekfunctionaliteit werden ontwikkeld als ondersteuning: het beperken van de zoekindex tot documenten van betrouwbare bronnen voor simpele zoektaken en het bieden van zoekfunctionaliteit voor het vinden van experts bij complexe zoektaken. Bij representatieve (gesimuleerde) taken vonden we dat beleidsmedewerkers

1,7 keer sneller waren met een beperkte zoekindex. Daarnaast vonden we dat gebruikers vaak terugvallen op het raadplegen van een collega met expertise wanneer ze een complexe zoektaak niet binnen een paar minuten kunnen oplossen.

De derde studie laat zien hoe onze taak-gebaseerde aanpak kan worden toegepast op het ontwerp van systemen, specifiek voor expert zoektaken. Gebaseerd op de karakterisering van expert zoektaken in eerdere studies, hebben we de ordening en presentatie van zoekresultaten aangepast en getest. We vergeleken persoons-centrische en document-centrische methoden om zoekresultaten te ordenen en presenteren. Statistische analyse toonde aan dat nieuwelingen in een domein meer taken correct uitvoerden met een persoons-centrische presentatie en sneller waren met een document-centrische ordening.

De laatste studie toont aan dat het ontwerpen van functionaliteit rondom zoektaken ons in staat stelt verder te gaan dan typische zoekfunctionaliteiten door raadsleden een overzichtsperspectief te bieden op raadsbesluiten. Voorheen werden deze overzichten handmatig samengesteld door beleidsmedewerkers, wat onschaalbaar, incompleet en subjectief bleek te zijn. We ontwikkelden een methode om besluithistories van politieke dossiers te reconstrueren door te herkennen wanneer hetzelfde document in meerdere vergaderingen werd besproken en wanneer een document verwijst naar een eerder document of vergadering. Deze aanpak stelde ons in staat ongeveer 2500 besluithistories te reconstrueren. Experts die deze histories gebruikten, gaven aan dat 97% van de opgenomen documenten relevant waren. Deze hoge precisie is mogelijk omdat de documenten werden opgenomen op basis van verwijzingen door domein-experts. We hebben ontwerp-implicaties afgeleid voor het presenteren van besluithistories in de interface, waarbij een gelaagd interface gebruikers een bondig overzicht biedt dat kan worden uitgebreid tot een compleet overzicht van de beslissingen.

Hoewel de zoekmachines en functionaliteiten in dit manuscript specifiek zijn voor de gemeente Utrecht, hebben we ook gezien dat gebruikersgroepen in verschillende gemeenten dezelfde taken uitvoeren. Dit suggereert dat zoekfunctionaliteit voor Nederlandse gemeenten gestandaardiseerd kan worden, hoewel dit verder onderzocht moet worden. Het is ook mogelijk dat variaties van dezelfde taak bestaan in meerdere domeinen. Wanneer twee taken in verschillende domeinen op dezelfde manier kunnen worden gekarakteriseerd met taak-eigenschappen, verwachten we dat een functionaliteit voor de ene taak kan worden aangepast om de andere taak te ondersteunen.

Acknowledgements

This thesis comprises the scientific papers we have produced, which were only possible through my formation as an independent researcher. It was the following people who have influenced my scientific direction and perspective. I am grateful to professor Egon van den Broek, who solidified my interest in understanding the contextual factors surrounding information use. Through his influence I have become more meticulous about identifying and controlling potential confounding variables. Working with Lazo Bozarov, MSc, helped me prioritise the value of solving the right problem. This collaboration also taught me to work with others by building common interests. It is through professor Arjen de Vries that I have seen freedom in science, and paths crossing between academia and industry. Throughout all this, professor Toine Pieters demonstrated how to be precise as a scientist, teaching both how to highlight the limitations of one's work while emphasising its contributions. Additionally, I must mention professor Lynda Hardman, who inspires me in many ways. She has shown me how and when to claim space, and when to give it to others. She has taught me how to communicate to an audience, and how to navigate thinking at multiple levels of abstraction. I am also grateful to ChatGPT, for its patient assistance in editing my texts. Many others have helped shape my perspective, and I would also like to acknowledge the influence of Marike, Tanja, Manel, Joanne, Ramiro, Joyce, Shelina, Anna, Eleonora, Miguel, Rianne, Nadine, Jeroen, Theo, Petra and Robin.

Thomas Martin Schoegje

Amsterdam, August 2024

Curriculum vitae

Thomas Schoegje was born on December 30th, 1990, in Zaandam, the Netherlands. Their academic background started with a Bachelor's degree in Computer Science from the University of Amsterdam in 2013, with a minor in practical philosophy and participation in a honours program. This interdisciplinary foundation initiated an trend of diverse research interests, with a focus on extracting insights from data.

In 2017, Thomas completed a Master's degree in Computer Science at the University of Utrecht, specialising in Algorithmic Data Analysis. This program was complented by an extensive extra curricular program at both the Vrije Universiteit Amsterdam and University of Amsterdam, focusing on advanced techniques for machine learning and data mining. During this period, their interest extended beyond developing the models themselves and to understanding how to provide meaningful signals to the models. Working in such academic and explorative settings fostered a growing enthusiasm for research.

Thomas' exposure to practical applications began with their master thesis and a subsequent research project, leading them to realise that data is only meaningful with respect to some purpose or context. Motivated by this, Thomas embarked on research as a PhD student at the Municipality of Utrecht, where they aimed to quantify the context of users' search queries to improve their search results. Over time, they found there are more meaningful ways to interact with information than to improve the relevancy of search results. They found that the user's tasks determine what interactive functionality is useful, and how information can be presented in a meaningful way.

The same applies at a meta level, as doing meaningful research (or work) also depends on understanding which problem is relevant. Thomas developed an appreciation for integrating diverse stakeholders perspectives, emphasizing the importance of solving the right problems. This approach has informed their holistic methodology for designing innovative systems, which balances an understanding of people, their challenges and their technology.

Upon completing their PhD in 2024, Thomas will continue to find and address the right problems at the Municipality of Utrecht.

FI Scientific Library

(formerly published as CD-β Scientific Library)

122 Waard, E. de (2024). *Into the cycle of sustainability - Fostering students' life cycle reasoning in secondary chemistry education.*

121 Linden, A. van der (2024). *Chasing Newton – Designing and implementing an intrinsically integrated game on Newtonian mechanics.*

120 Hattinga van 't Sant, E.A. (2023). *De mythe van de alfaman - De dominantie van dominantie in de behavioural sciences, 1920-2020. Een historische reconstructie.*

119. Bachtiar, R. W. (2023). *Animated reasoning - Supporting students' mechanistic reasoning in physics by constructing stop-motion animations*

118. Harskamp, M. van (2023). *Ask, find out, and act: Fostering environmental citizenship through science education*.

117. Boels, L. (2023). *Histograms - An educational eye*.

116. Huang, L. (2022). *Inquiry-based learning in lower-secondary mathematics education in China (Beijing) and the Netherlands*.

115. Jansen, S. (2022). *Fostering students' meta-modelling knowledge regarding biological concept-process models*.

114. Pieters, M.L.M. (2022). *Between written and enacted: Curriculum development as propagation of memes. An ecological-evolutionary perspective on fifty years of curriculum development for upper secondary physics education in the Netherlands*.

113. Veldkamp, A. (2022). *No Escape! The rise of escape rooms in secondary science education*.

112. Kamphorst, F. (2021). *Introducing special relativity in secondary education*.

111. Leendert, A.-M. J. M. van (2021). *Improving reading and comprehending mathematical expressions in Braille*.

110. Gilissen, M. G. R. (2021). *Fostering students' system thinking in secondary biology education*.

109. Dijke-Droogers, M.J.S. van (2021). *Introducing statistical inference: Design and evaluation of a learning trajectory*.

108. Wijnker, W. (2021). *The unseen potential of film for learning. Film's interest raising mechanisms explained in secondary science and mathematics education*.

107. Groothuijsen, S. (2021). *Quality and impact of practice-oriented educational research.*

106. Wal, N.J. van der (2020). *Developing techno-mathematical literacies in higher technical professional education.*

105. Tacoma, S. (2020). *Automated intelligent feedback in university statistics education.*

104. Zanten, M. van (2020). *Opportunities to learn offered by primary school mathematics textbooks in the Netherlands*
103. Walma, L. (2020). *Between Morpheus and Mary: The public debate on morphine in Dutch newspapers, 1880-1939*
102. Van der Gronde, A.G.M.P. (2019). *Systematic review methodology in biomedical evidence generation.*
101. Klein, W. (2018). *New drugs for the Dutch republic. The commodification of fever remedies in the Netherlands (c. 1650-1800)*.
100. Flis, I. (2018). *Discipline through method - Recent history and philosophy of scientific psychology (1950-2018).*
99. Hoeneveld, F. (2018). *Een vinger in de Amerikaanse pap. Fundamenteel fysisch en defensie onderzoek in Nederland tijdens de vroege Koude Oorlog*.
98. Stubbé-Albers, H. (2018). *Designing learning opportunities for the hardest to reach: Game-based mathematics learning for out-of-school children in Sudan*.
97. Dijk, G. van (2018). *Het opleiden van taalbewuste docenten natuurkunde, scheikunde en techniek: Een ontwerpgericht onderzoek*.
96. Zhao, Xiaoyan (2018). *Classroom assessment in Chinese primary school mathematics education.*
95. Laan, S. van der (2017). *Een varken voor iedereen. De modernisering van de Nederlandse varkensfokkerij in de twintigste eeuw*.
94. Vis, C. (2017). *Strengthening local curricular capacity in international development cooperation*.
93. Benedictus, F. (2017). *Reichenbach: Probability & the a priori. Has the baby been thrown out with the bathwater*?
92. Ruiter, Peter de (2016). *Het mijnwezen in Nederlands-Oost-Indië 1850- 1950.*
91. Roersch van der Hoogte, Arjo (2015). *Colonial agro-industrialism. Science, industry and the state in the Dutch Golden Alkaloid Age, 1850- 1950.*
90. Veldhuis, M. (2015). *Improving classroom assessment in primary mathematics education.*
89. Jupri, Al (2015). *The use of applets to improve Indonesian student performance in algebra.*
88. Wijaya, A. (2015). *Context-based mathematics tasks in Indonesia: Toward better practice and achievement*.
87. Klerk, S. (2015). *Galen reconsidered. Studying drug properties and the foundations of medicine in the Dutch Republic ca. 1550-1700*.
86. Krüger, J. (2014). *Actoren en factoren achter het wiskundecurriculum sinds 1600*.
85. Lijnse, P.L. (2014). *Omzien in verwondering. Een persoonlijke terugblik op 40 jaar werken in de natuurkundedidactiek*.

84. Weelie, D. van (2014). *Recontextualiseren van het concept biodiversiteit*.

83. Bakker, M. (2014). *Using mini-games for learning multiplication and division: a longitudinal effect study.*

82. Ngô Vũ Thu Hăng (2014). *Design of a social constructivism-based curriculum for primary science education in Confucian heritage culture*.

81. Sun, L. (2014). *From rhetoric to practice: enhancing environmental literacy of pupils in China.*

80. Mazereeuw, M. (2013). Th*e functionality of biological knowledge in the workplace. Integrating school and workplace learning about reproduction.*

79. Dierdorp, A. (2013). *Learning correlation and regression within authentic contexts.*

78. Dolfing, R. (2013). *Teachers' professional development in context-based chemistry education. Strategies to support teachers in developing domain-specific expertise.*

77. Mil, M.H.W. van (2013). *Learning and teaching the molecular basis of life*.

76. Antwi, V. (2013). *Interactive teaching of mechanics in a Ghanaian university context.*

75. Smit, J. (2013). *Scaffolding language in multilingual mathematics classrooms*.

74. Stolk, M. J. (2013). *Empowering chemistry teachers for context-based education. Towards a framework for design and evaluation of a teacher professional development programme in curriculum innovations.*

73. Agung, S. (2013). *Facilitating professional development of Madrasah chemistry teachers. Analysis of its establishment in the decentralized educational system of Indonesia.*

72. Wierdsma, M. (2012). *Recontextualising cellular respiration*.

71. Peltenburg, M. (2012). *Mathematical potential of special education students*.

70. Moolenbroek, A. van (2012). *Be aware of behaviour. Learning and teaching behavioural biology in secondary education*.

69. Prins, G. T., Vos, M. A. J., & Pilot, A. (2011). *Leerlingpercepties van onderzoek & ontwerpen in het technasium.*

68. Bokhove, Chr. (2011). *Use of ICT for acquiring, practicing and assessing algebraic expertise.*

67. Boerwinkel, D. J., & Waarlo, A. J. (2011). *Genomics education for decision- making. Proceedings of the second invitational workshop on genomics education, 2-3 December 2010.*

66. Kolovou, A. (2011). *Mathematical problem solving in primary school*.

65. Meijer, M. R. (2011). *Macro-meso-micro thinking with structure-property relations for chemistry. An explorative design-based study*.

64. Kortland, J., & Klaassen, C. J. W. M. (2010). *Designing theory-based teaching-learning sequences for science. Proceedings of the symposium in honour of Piet Lijnse at the time of his retirement as professor of Physics Didactics at Utrecht University.*

63. Prins, G. T. (2010). *Teaching and learning of modelling in chemistry education. Authentic practices as contexts for learning*.

62. Boerwinkel, D. J., & Waarlo, A. J. (2010). *Rethinking science curricula in the genomics era. Proceedings of an invitational workshop*.

61. Ormel, B. J. B. (2010). *Het natuurwetenschappelijk modelleren van dynamische systemen. Naar een didactiek voor het voortgezet onderwijs*.

60. Hammann, M., Waarlo, A. J., & Boersma, K. Th. (Eds.) (2010). *The nature of research in biological education: Old and new perspectives on theoretical and methodological issues – A selection of papers presented at the VIIth Conference of European Researchers in Didactics of Biology.*

59. Van Nes, F. (2009). *Young children's spatial structuring ability and emerging number sense.*

58. Engelbarts, M. (2009). *Op weg naar een didactiek voor natuurkunde-experimenten op afstand. Ontwerp en evaluatie van een via internet uitvoerbaar experiment voor leerlingen uit het voortgezet onderwijs*.

57. Buijs, K. (2008). *Leren vermenigvuldigen met meercijferige getallen*.

56. Westra, R. H. V. (2008). *Learning and teaching ecosystem behaviour in secondary education: Systems thinking and modelling in authentic practices*.

55. Hovinga, D. (2007). *Ont-dekken en toe-dekken: Leren over de veelvormige relatie van mensen met natuur in NME-leertrajecten duurzame ontwikkeling*.

54. Westra, A. S. (2006). *A new approach to teaching and learning mechanics*.

53. Van Berkel, B. (2005). *The structure of school chemistry: A quest for conditions for escape.*

52. Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education: The case of water quality.*

51. Doorman, L. M. (2005). *Modelling motion: from trace graphs to instantaneous change.*

50. Bakker, A. (2004). *Design research in statistics education: on symbolizing and computer tools.*

49. Verhoeff, R. P. (2003). *Towards systems thinking in cell biology education*.

48. Drijvers, P. (2003). *Learning algebra in a computer algebra environment. Design research on the understanding of the concept of parameter*.

47. Van den Boer, C. (2003). *Een zoektocht naar verklaringen voor achterblijvende prestaties van allochtone leerlingen in het wiskundeonderwijs*.

46. Boerwinkel, D. J. (2003). *Het vormfunctieperspectief als leerdoel van natuuronderwijs. Leren kijken door de ontwerpersbril*.

45. Keijzer, R. (2003). *Teaching formal mathematics in primary education. Fraction learning as mathematising process*.

44. Smits, Th. J. M. (2003). *Werken aan kwaliteitsverbetering van leerlingonderzoek: Een studie naar de ontwikkeling en het resultaat van een scholing voor docenten*.

43. Knippels, M. C. P. J. (2002). *Coping with the abstract and complex nature of genetics in biology education – The yo-yo learning and teaching strategy*.

42. Dressler, M. (2002). *Education in Israel on collaborative management of shared water resources.*

41. Van Amerom, B.A. (2002). *Reinvention of early algebra: Developmental research on the transition from arithmetic to algebra*.

40. Van Groenestijn, M. (2002). *A gateway to numeracy. A study of numeracy in adult basic education.*

39. Menne, J. J. M. (2001). *Met sprongen vooruit: een productief oefenprogramma voor zwakke rekenaars in het getallengebied tot 100 – een onderwijsexperiment*.

38. De Jong, O., Savelsbergh, E.R., & Alblas, A. (2001). *Teaching for scientific literacy: context, competency, and curriculum.*

37. Kortland, J. (2001). *A problem-posing approach to teaching decision making about the waste issue*.

36. Lijmbach, S., Broens, M., & Hovinga, D. (2000). *Duurzaamheid als leergebied; conceptuele analyse en educatieve uitwerking.*

35. Margadant-van Arcken, M., & Van den Berg, C. (2000). *Natuur in pluralistisch perspectief – Theoretisch kader en voorbeeldlesmateriaal voor het omgaan met een veelheid aan natuurbeelden.*

34. Janssen, F. J. J. M. (1999). *Ontwerpend leren in het biologieonderwijs. Uitgewerkt en beproefd voor immunologie in het voortgezet onderwijs*.

33. De Moor, E. W. A. (1999). *Van vormleer naar realistische meetkunde Een historisch-didactisch onderzoek van het meetkundeonderwijs aan kinderen van vier tot veertien jaar in Nederland gedurende de negentiende en twintigste eeuw.*

32. Van den Heuvel-Panhuizen, M., & Vermeer, H. J. (1999). *Verschillen tussen meisjes en jongens bij het vak rekenen-wiskunde op de basisschool – Eindrapport MOOJ-onderzoek.*

31. Beeftink, C. (2000). *Met het oog op integratie – Een studie over integratie van leerstof uit de natuurwetenschappelijke vakken in de tweede fase van het voortgezet onderwijs*.

30. Vollebregt, M. J. (1998). *A problem posing approach to teaching an initial particle model.*

29. Klein, A. S. (1998). *Flexibilization of mental arithmetics strategies on a different knowledge base – The empty number line in a realistic versus gradual program design.*

28. Genseberger, R. (1997). *Interessegeoriënteerd natuur- en scheikundeonderwijs – Een studie naar onderwijsontwikkeling op de Open Schoolgemeenschap Bijlmer*.

27. Kaper, W. H. (1997). *Thermodynamica leren onderwijzen*.

26. Gravemeijer, K. (1997). *The role of context and models in the development of mathematical strategies and procedures*.

25. Acampo, J. J. C. (1997). *Teaching electrochemical cells – A study on teachers' conceptions and teaching problems in secondary education*.

24. Reygel, P. C. F. (1997). *Het thema 'reproductie' in het schoolvak biologie*.

23. Roebertsen, H. (1996). *Integratie en toepassing van biologische kennis– Ontwikkeling en onderzoek van een curriculum rond het thema 'Lichaamsprocessen en Vergift'*.

22. Lijnse, P. L., & Wubbels, T. (1996). *Over natuurkundedidactiek, curriculumontwikkeling en lerarenopleiding*.

21. Buddingh', J. (1997). *Regulatie en homeostase als onderwijsthema: een biologie-didactisch onderzoek.*

20. Van Hoeve-Brouwer G. M. (1996). *Teaching structures in chemistry – An educational structure for chemical bonding*.

19. Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education.*

18. Klaassen, C. W. J. M. (1995). *A problem-posing approach to teaching the topic of radioactivity.*

17. De Jong, O., Van Roon, P. H., & De Vos, W. (1995). *Perspectives on research in chemical education.*

16. Van Keulen, H. (1995). *Making sense – Simulation-of-research in organic chemistry education.*

15. Doorman, L. M., Drijvers, P. & Kindt, M. (1994). *De grafische rekenmachine in het wiskundeonderwijs.*

14. Gravemeijer, K. (1994). *Realistic mathematics education*.

13. Lijnse, P. L. (Ed.) (1993). *European research in science education*.

12. Zuidema, J., & Van der Gaag, L. (1993). *De volgende opgave van de computer*.

11. Gravemeijer, K., Van den Heuvel-Panhuizen, M., Van Donselaar, G., Ruesink, N., Streefland, L., Vermeulen, W., Te Woerd, E., & Van der Ploeg, D. (1993). *Methoden in het reken-wiskundeonderwijs, een rijke context voor vergelijkend onderzoek*.

10. Van der Valk, A. E. (1992). *Ontwikkeling in energieonderwijs*.

9. Streefland, L. (Ed.) (1991). *Realistic mathematics education in primary schools*.

8. Van Galen, F., Dolk, M., Feijs, E., & Jonker, V. (1991). *Interactieve video in de nascholing reken-wiskunde.*
7. Elzenga, H. E. (1991). *Kwaliteit van kwantiteit*.
6. Lijnse, P. L., Licht, P., De Vos, W., & Waarlo, A. J. (Eds.) (1990). *Relating macroscopic phenomena to microscopic particles: a central problem in secondary science education.*
5. Van Driel, J. H. (1990). *Betrokken bij evenwicht*.
4. Vogelezang, M. J. (1990). *Een onverdeelbare eenheid*.
3. Wierstra, R. F. A. (1990). *Natuurkunde-onderwijs tussen leefwereld en vakstructuur*.
2. Eijkelhof, H. M. C. (1990). *Radiation and risk in physics education*.
1. Lijnse, P. L., & De Vos, W. (Eds.) (1990). *Didactiek in perspectief*.

Utrecht University

Faculty of Science
Freudenthal Institute

The municipality of Utrecht works with an overwhelming amount of information. It is important that citizens and public servants have effective tools to search for relevant information. A challenge when creating search engines is that users such as citizens have different information needs than users such as council members. This is most noticeable when users have very complex and (domain-)specific needs. These require specialised functionality.

A standard search engine cannot satisfy the diverse and specific information needs of all user groups at the same time. Therefore, we characterise the search tasks of different user groups and develop specialised search functionality to support those tasks. We investigate which tasks are already supported by existing systems, and which require a new kind of support.

Over the course of four studies, we implemented multiple search engines and tested whether these help employees perform their tasks more effectively and efficiently. First, we developed a search engine for council information to support the tasks of council members. Second, we introduced a search engine that supports the web-based search tasks of policy workers, where our results indicated that many of their tasks are solved by getting help from the right person. Third, we supported this search behaviour by developing a search engine that helps employees get into contact with colleagues who have relevant expertise. Finally, we enhanced the original search engine for council information by automatically generating timelines of political dossiers. This let us present individual search results in their political context.

The proposed solutions are domain-specific and have been evaluated at one municipality only. However, we found that user groups at other municipalities perform the same tasks, and these tasks are also found in other large organisations. Hence, we expect the specific functionality could be standardised for municipalities, and adapted to support similar tasks in other domains.

https://doi.org/10.33540/2620