The role of domain knowledge in cognitive modeling of information search

Saraschandra Karanam¹ · Guillermo Jorge-Botana² ·
Ricardo Olmos³ · Herre van Oostendorp¹

Received: 28 October 2016 / Accepted: 29 April 2017 / Published online: 24 May 2017
© The Author(s) 2017. This article is an open access publication

Abstract Computational cognitive models developed so far do not incorporate individual differences in domain knowledge in predicting user clicks on search result pages. We address this problem using a cognitive model of information search which enables us to use two semantic spaces having a low (non-expert semantic space) and a high (expert semantic space) amount of medical and health related information to represent respectively low and high knowledge of users in this domain. We also investigated two different processes along which one can gain a larger amount of knowledge in a domain: an evolutionary and a common core process. Simulations of model click behavior on difficult information search tasks and subsequent matching with actual behavioral data from users (divided into low and high domain knowledge groups based on a domain knowledge test) were conducted. Results showed that the efficacy of modeling for high domain knowledge participants (in terms of the number of matches between the model predictions and the actual user clicks on search result pages) was higher with the expert semantic space compared to the non-expert semantic space while for low domain knowledge participants it was the other way around. When the process of knowledge acquisition was taken into account, the effect of using a semantic space based on high domain knowledge was significant only for high domain knowledge participants, irrespective of the knowledge acquisition process. The implications of these outcomes for support tools that can be built based on these models are discussed.

Keywords Cognitive modeling · Information search · Domain knowledge · Knowledge acquisition · Corpus · Semantic space

✉ Saraschandra Karanam
s.karanam@uu.nl

¹ Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

² Departamento de Psicología Evolutiva, Universidad Nacional de Educación a Distancia, Madrid, Spain

³ Departamento de Psicología Social y Metodología, Universidad Autónoma de Madrid, Madrid, Spain

1 Introduction

Search systems are typically characterized as a tool to retrieve relevant information on a target page from the Internet in response to a user query. These systems are efficient only for a certain type of tasks such as look-up tasks or factoid questions (*“How much is the distance between Mars and Earth”* or *“Which is the highest mountain in Europe”*). They are not very optimal for other kind of tasks that involve knowledge discovery, comprehension and learning which require iteration, comparison and reflection by a user. Many times, important information that is needed to solve the main search problem is present on intermediate pages leading to the target page (Olston and Chi 2003). In such cases, it is important to evaluate information on each page and take decisions which hyperlink or search result to click next based on the information that is already processed. The process of information search therefore can be conceived as a process that involves learning or at least knowledge acquisition. Users acquire new knowledge not only at the end of an information search process after reaching the target page, but also during processing of intermediate search results and web-pages before they reach the target page. Learning from such contextual information as users perform search and navigation tasks on the web, involves complex cognitive processes that dynamically influence the evaluation of link texts and web contents (Fu 2013; Fu and Dong 2010). Search engines do not lay any emphasis on these intermediate steps and are largely focused only on the step involving retrieval of relevant information. They also ignore the influence of cognitive factors such as domain knowledge (Cole et al. 2011; Monchaux et al. 2015) on the cognitive processes underlying information search and navigation and follow a one-size-fits-all model.

In this paper, we focus on differences in information search behavior due to individual differences in domain knowledge of users. It is known that users with high domain knowledge have more appropriate mental representations that are characterized by higher activation degrees of concepts and stronger connections between different concepts in their conceptual space compared to users with low domain knowledge (Kintsch 1998). A number of experiments investigating the role of domain knowledge on information search and navigation performance have already been conducted (Cole et al. 2011; Duggan and Payne 2008; Held et al. 2012; Kiseleva et al. 2016; Monchaux et al. 2015; Palotti et al. 2016; White et al. 2009). For example, it is known that when searching for information on the web, users are inclined to choose a path of navigation on the basis of their prior knowledge. When the prior knowledge is incorrect, it was found that the choices made by the users were sub-optimal (Held et al. 2012). In a recent study by Monchaux et al. (2015), domain experts were found to find more correct answers in shorter time and via a path closer to the optimum path than non-experts. This difference was stronger as the difficulty of the task increased. They were also found to be less biased by the interface of the search engine result page (SERP, henceforth) and explored search results even at the bottom of a SERP (Kiseleva et al. 2016). Domain experts are also known to evaluate search results more thoroughly and click more often on relevant search results compared to non-experts. One of the possible explanations for this finding is that higher level of domain knowledge enables a user to formulate more appropriate queries and comprehend the search results and the content in the websites better, which in turn, enables them to take informed decisions regarding which hyperlink or search result to click next (Cole et al. 2011). Domain experts were also found to be more persistent (longer sessions, more page visits) (Palotti et al. 2016; White et al. 2009) and detect unfruitful search paths faster (Duggan

and Payne 2008) than non-experts. In a first simulation study we will examine the influence of differences in domain knowledge of users on information search behavior.

We also take into account different ways in which one can gain knowledge on a particular domain and its influence on information search behavior. In this paper, we focus, particularly on two types of possible exposure sequences: ‘evolutionary’ and ‘common core’ processes leading to two different types of domain experts: ‘progressive’ experts and ‘abrupt’ experts. Evolutionary exposure is similar to the gradual process of vocabulary acquisition in school which follows a standard and extended sequence until a final state of optimal knowledge is reached. Common core exposure, on the other hand, has no final state of optimal knowledge. And also, different people may have different trajectories of exposure following different sequences, though they share a common exposure at the beginning. Longitudinal studies that examined the development of knowledge over a period of time and its influence on search behavior, found significant differences in the specificity of words included in queries as users gained more expertise on a topic (Vakkari et al. 2003; Wildemuth 2004). The words became more specific and reformulations were more effective. It appears that the knowledge acquisition process impacts information search, e.g. on the specificity of query formulations. We will examine in a second simulation study the influence of the evolutionary process compared to the common core process on modeling search behavior.

We are interested in computational cognitive modeling of information search because understanding behavioral differences through laboratory experiments is not only expensive and not very scalable but also time consuming. Simulation of user interactions with information retrieval systems therefore has been an active area of research. However, among the many click models developed by researchers from the information retrieval community (Chuklin et al. 2015), only a few take into account cognitive aspects (Hu et al. 2011; Shen et al. 2012; Xing et al. 2013). Moreover, they provide only limited process description. We therefore, employ a computational cognitive model in our research because that enables us to model differences due to cognitive factors (such as domain knowledge) underlying cognitive functions such as comprehension of search results, arriving at a relevance estimate of search results and selecting one of the search results to click. Also, the focus in our computational cognitive modeling is on the process that leads to the target information. In the next section, we give a brief overview of the cognitive model of information search called CoLiDeS (Kitajima et al. 2000) that we use in our research.

2 Related work

CoLiDeS, or Comprehension-based Linked Model of Deliberate Search, developed by Kitajima et al. (2000) explains user navigation behaviour on websites. The model divides user navigation behavior into four stages of cognitive processing: parsing the webpage into high-level schematic regions, focusing on one of those schematic regions, elaboration / comprehension of the screen objects (e.g. hypertext links) within that region, and evaluating and selecting the most appropriate screen object (e.g. hypertext link) in that region. CoLiDeS is based on Information Foraging Theory (Pirulli and Card 1999) and connects to the Construction-Integration reading model of Kintsch (1998). A central design principle of CoLiDeS is the notion of *information scent*, which is defined as the subjective estimate of the value or cost of information sources represented by proximal cues (such as

hyperlinks). It is operationalized as the semantic similarity between the user goal and each of the hyperlinks. The model predicts that the user is most likely to click on that hyperlink which has the highest semantic similarity value with the user goal, i.e., the highest information scent. This process is repeated for every new page until the user reaches the target page. See Karanam et al. (2016) for more details on the model. CoLiDeS uses Latent Semantic Analysis (LSA, henceforth) introduced by Landauer et al. (2007) to compute the semantic similarities. LSA is an unsupervised machine learning technique that employs singular value decomposition to build a high dimensional semantic space using a large corpus of documents that is representative of the knowledge of the target user group. The semantic space contains a representation of terms from the corpus in a limited number of dimensions, typically between 250 and 350 and are orthogonal, abstract and latent (Landauer et al. 2007; Olmos et al. 2014). Main application area of CoLiDeS concerns web navigation. CoLiDeS has been successful in simulating and predicting user (hyper)link selections, though the websites and web-pages used were very restricted. The model has also been successfully applied in finding usability problems, by predicting links that would be unclear to users (Blackmon et al. 2007). Furthermore, the CoLiDeS model has recently been extended to predict user clicks on search result pages (Karanam et al. 2015). We use CoLiDeS in our study as it is simple and process-oriented and has features that can be exploited for our purposes.

In this paper, we focus on two main limitations of CoLiDeS:

1. *CoLiDeS modeling so far does not incorporate individual differences in domain knowledge of users* The estimation of information scent in CoLiDeS is closely connected to the prior domain knowledge of the searcher. For example, if the concepts ‘rabona’, ‘cruyff turn’, ‘rivelino’ (names of different football moves) have a strong association in a user’s long term memory, then the information scent between the concepts would be very high for the user. However, for a user who is not knowledgeable in football and football moves (and therefore the association between the concepts is not strong in his/her long term memory), the information scent between the concepts would be low. Such differences in the estimation of information scent would in turn lead to differences in information search behavior (in terms of the choice of hyperlinks clicked by a user). It is possible to use two corpora varying in the number of documents on the topic of football and using LSA to create two semantic spaces representing low and high knowledge levels of football. Although, LSA provides this option, previous research on cognitive modeling of information search, including CoLiDeS, has not systematically examined the impact of users’ prior domain knowledge and using different semantic spaces—on modeling and predicting information search behavior.
2. *CoLiDeS modeling so far does not take into account differences in the process of acquisition of domain knowledge* It is known that the associations between different concepts in long term memory are not static. They, in fact, change over time, with exposure to new material (Chi and Koeske 1983; Durso and Coggins 1990; Ferstl and Kintsch 1999). Therefore, it is natural that the strength of association between two concepts in a user’s long term memory would also be dependent on the process of knowledge acquisition that the user underwent in his/her lifetime. For example, imagine a user A, who is a Computer Science graduate, working for a software company. He is not particularly interested in the topic of health/medicine. His medical and health knowledge is based on what was taught in school and whatever he read from popular magazines and newspaper articles. We term such a user as a *non-expert*

on the topic of health and medicine. Imagine a second user B, who is a senior researcher in a pharmaceutical company. He is passionate about the topic of health/medicine from his childhood and therefore underwent a formal academic program in Biology and obtained his doctorate in Biotechnology. Please note that he also reads popular health magazines and newspapers articles just like user A. We term such a user as a *progressive expert* on the topic of health and medicine as he has undergone what we call an evolutionary exposure in developing knowledge in the domain. Imagine a third user C, who is also a Computer Science graduate, but is working as a student assistant with user B for last 6 months and reads all expert sources of health and medical information as part of his work. We term such a user as an *abrupt expert* on the topic of health and medicine as his knowledge in the domain is only based on the exposure he gets in the project. All three users have very different exposure to medical and health related information and therefore would have different conceptual networks in their semantic memory. To give a second example, consider a person who did Home Sciences during his Bachelors and Hotel Management during his Masters and is now working as a Master Chef in a business hotel. This person would be regarded as a progressive expert in cooking. On the other hand, a student who watches cooking lectures on youtube to tries out new dishes at home on weekends can be considered an abrupt expert in cooking. We can think of many such examples in other domains of expertise such as repairing a broken bike or a car (a professional mechanic vs. somebody who picks up repairing skills from youtube or other informal sources), knowledge of best agricultural practices (large-scale farming vs. gardening at home), knowledge about a popular tourist spot or a historical site (a historian or a local citizen vs. a tourist getting to know some interesting tidbits of information about a place from the Internet or from a tour guide) and knowledge about medicines (a doctor or a specialist vs. a pharmacy assistant). So, summarizing, the situation we have in mind with the abrupt expert corresponds to the pressure cooker method of quickly and intensively processing many relevant documents, without reading underlying explaining materials. It may result in being able to perform tasks without really understanding reasons why something occurs. It is possible to use two corpora varying in the number and type of documents on the topic of medicine and health, for example, and using LSA to create two semantic spaces representing the knowledge of the two types of experts in this domain. Although, LSA provides this option, previous research on cognitive modeling of information search, including CoLiDeS, has not systematically examined the influence of such differences in the process of acquiring knowledge on a domain on modeling and predicting information search behavior.

3 Research questions

Based on the limitations of CoLiDeS listed in the previous section, the following were the main research questions of this study:

1. How to incorporate differences in domain knowledge levels of users into a computational cognitive model that predicts information search behavior? More precisely, would a model that takes differences in domain knowledge into account, predict user clicks on search engine result pages better than a model that does not incorporate differentiated domain knowledge levels of users? (RQ1).

2. Does the process of acquisition of domain knowledge, when incorporated into the cognitive model, influence the predictive power of the model with regard to search behavior? (**RQ2**). More precisely, would a model that takes into account an evolutionary process of knowledge acquisition predict user clicks on search engine result pages differently than a model that takes into account a common core process of knowledge acquisition?

Please note that while information search behavior can involve activities such as generation of queries, evaluation of search results, clicking on one or more search results, evaluating the content of the website opened after clicking on a search result, evaluating the hyperlinks on the opened website, reformulation of queries, etc. we focus on modeling and predicting clicks on search engine result pages in this work. The remaining part of the paper is organized as follows. In Sect. 4, we describe an experiment conducted to collect actual behavioral data from participants varying in the amount of knowledge in the domain of health. We needed this data in order to compare actual user behavior with the simulations. Section 5 provides details of two simulation studies conducted using the cognitive model CoLiDeS and Sect. 6 concludes the paper.

4 User data collection

In this section, we describe the details of an experiment conducted to collect actual behavioral data from participants. In the experiment we examine the impact of prior domain knowledge (low vs. high) and task difficulty (simple vs. difficult) on the semantic relevance of queries generated by participants. The goal is to verify if the participants with high domain knowledge generate more appropriate queries compared to participants with low domain knowledge (Cole et al. 2011).

4.1 Method

4.1.1 Participants

48 participants (30 males and 18 females) ranging from 18 to 88 years ($M = 48.79$, $SD = 26.3$) volunteered for the study.

4.1.2 Design

We followed a 2 (Prior Domain Knowledge: Low vs. High) $\times 2$ (Task Difficulty: Simple vs. Difficult) mixed design with prior domain knowledge as between-subjects variable and task difficulty as within-subjects variable.

4.1.3 Material

The experiment was conducted with twelve simulated information search tasks (Borlund and Ingwersen 1997): six simple and six difficult, all from the domain of health. For *simple* tasks, participants in most cases could find the answer easily either in the snippets of the search engine results or in one of the websites referred to by the search engine results. For instance, for a task like “What is the main function of sweat glands?”, user could use the words like “function of sweat glands” from the task description itself as queries. One can easily find the answer

“*regulation of body temperature*” within the snippets of the corresponding search results without having to click on any of them. These tasks are similar to the lookup tasks (e.g. fact retrieval) as defined in Marchionini (2006). For *difficult* tasks, users had to frame queries using their knowledge and understanding of the task, the answer was not easily found in the snippets of search engine results and often they had to evaluate information from multiple websites. As an example, for the task “*Elbert, 76 years old, has been suffering for few years from burning sensation while passing urine. He passes urine more often than normal at night and complains of a feeling that the bladder is not empty completely. Lately, he also developed acute pain in the hip, lower back and pelvis region. He also lost 12 kilos in the last 6 months. What problem could he be suffering from?*”, users had to formulate multiple queries such as “*kidney stones pain in the back*”, “*burning sensation when urinating*”, “*urinary infection*” to find the answer. The answer to this task “*prostate cancer*” is also not found easily in the snippets of the search results of the queries, unless the query is very specific. User has to click and examine more than one search result. So these difficult tasks have an important problem solving and evaluation component compared to simple tasks. Also, they have a concrete goal and a fixed answer that needs to be found by the user and in this way, they are different from exploratory tasks (e.g. learning a new topic, serendipitous browsing) as defined in Marchionini (2006). In the simulation studies we will use these tasks. All tasks were all originally presented in Dutch and therefore participants used queries also in Dutch.

4.1.4 Procedure

Participants first answered a demographic questionnaire in which they were asked details about their age, gender, familiarity with search engines (on a Likert scale of 1(A bit) to 4 (Very Much)) and computer experience (number of years). They were next presented with a prior domain knowledge (PDK) test on the topic of health. The score on the prior domain knowledge test gives us an indication of the amount of prior knowledge on the topic of health. For the prior domain knowledge test, participants were presented with 12 questions with multiple choice options. For each question, the participants had to choose an option from four alternatives. Only one alternative was correct. Correct choices were scored 1 and wrong choices 0. Thus the maximum possible score on this test is 12 and the minimum possible score is 0.

We divided the participants into two groups of high (25 participants) and low (23 participants) prior domain knowledge (HPDK and LPDK respectively) by taking the median score on the prior domain knowledge test. The mean score of the LPDK group is 4.08 ($SD = 0.84$) and that of the HPDK group is 6.92 ($SD = 1.07$). The difference is strongly significant ($t(46) = 10.06, p < .001$). Based on the self-reported ratings, there was no significant difference in the number of years of experience with computers between the low ($M = 18.39, SD = 8.16$) and the high ($M = 21.32, SD = 12.06$) domain knowledge participants ($p > .05$). Similarly, there was no significant difference in the familiarity with search engines between the low ($M = 3.2, SD = .99$) and the high ($M = 3.16, SD = 1.06$) domain knowledge participants ($p > .05$).

After the prior domain knowledge test, participants were allowed a break of five minutes. They were then presented with twelve information search tasks (six simple and six difficult) in a counter balanced order. Participants were first shown the task and then directed to the home page of Google’s search engine. Participants were not allowed to use any other search engine. We show in Fig. 1 the main screen of our interface that participants used while solving the information search tasks. Participants could enter queries as they normally would on any browser and the corresponding search results appeared. The task description was made available to the participant at all times in the top left corner. An

Fig. 1 Interface showing the main screen in which the information search tasks are solved

empty text box was provided in the top right corner for the participant to enter his/her answer. All the queries generated by the users, the corresponding search engine result pages and the URLs opened by them were logged in the backend using Visual Basic.

Previous research on individual differences in information search behavior showed that the semantic relevance of queries with target information is a sensitive and useful metric to differentiate participants with different levels of the cognitive factor domain knowledge (Karanam and van Oostendorp 2016). Therefore, we examined the impact of prior domain knowledge on the *semantic relevance of queries (SRQ)* generated by the participants. As users with high domain knowledge have more appropriate mental representations characterized by more relevant concepts, higher activation values and stronger connections between different concepts in the conceptual space compared to users with low domain knowledge, we expect that the semantic relevance of queries generated by them with target information would be much higher than that of users with low prior domain knowledge.

4.2 Measures

Semantic Relevance of Query (SRQ) is computed as the semantic relevance between the query and the target information sought using LSA. This metric gives us an estimate of how close in semantic similarity the queries generated by the participants are to the target information. So in general, the higher the SRQ value is, the more relevant the query is.

4.3 Results

Semantic relevance of query: We divided the participants into two groups of high (25 participants) and low (23 participants) prior domain knowledge (PDK) by taking the median score on the prior domain knowledge test. A 2 (PDK: Low vs. High) $\times$ 2 (Task Difficulty: Simple vs. Difficult) mixed ANOVA was conducted with PDK as between-subjects variable and task

Fig. 2 Mean semantic relevance of queries with target information in relation to PDK and task difficulty

difficulty as within-subjects variable. The main effect of task difficulty was significant $F(1, 46) = 9.35, p < .005$. See Fig. 2. The semantic relevance of queries with target information was significantly higher for difficult tasks compared to simple tasks. This is because solving difficult tasks successfully urges participants to activate a greater number of related concepts from Long Term Memory (Kintsch 1998). The main effect of PDK was close to conventional significance $F(1, 46) = 3.68, p = .06$, indicating that participants with high prior domain knowledge generated queries with significantly higher semantic relevance to target information compared to participants with low prior domain knowledge. The interaction of PDK and task difficulty was not significant ($p > .05$). These results provide empirical evidence to the explanations given in prior work (Cole et al. 2011). Users with high domain knowledge indeed generate queries which are more appropriate (more relevant) to the target information than users with low domain knowledge. More appropriate queries generate more appropriate SERPs, leading in turn to more appropriate clicks. This outcome adds more weight to our argument that the effect of prior domain knowledge should be taken into account when modeling user clicks on SERPs.

5 Modeling analysis

In this section, we describe our simulations with the cognitive model CoLiDeS. We ran two simulation studies addressing the two research questions of this study. For each simulation, we first create semantic spaces that vary in the amount of health and medical

knowledge, evaluate the semantic spaces to check if they are indeed different in the amount of health and medical knowledge, run simulations of CoLiDeS using the semantic spaces and analyze the outcomes of simulation. The behavioral data collected in the experiment described above is used to evaluate the predictions of CoLiDeS. The evaluation is based on the number of matches between the actual user clicks and the model-predicted clicks. For analyzing the simulation outcomes, we focus only on difficult tasks the differences between high and low domain knowledge participants are known to be more prominent for difficult tasks. It also simplifies the presentation of results.

5.1 Simulation study 1

In the first simulation study, examining the first research question, we use two semantic spaces which correspond to different levels of domain knowledge on the topic of health and medicine in CoLiDeS.

5.1.1 Creation of semantic spaces

We collated two different corpora (a non-expert corpus and an expert corpus, each consisting of 70,000 articles in Dutch) varying in the amount of medical and health related information. The non-expert corpus, representing the knowledge of low domain knowledge users contained 90% news articles and 10% medical and health related articles whereas the expert corpus, representing the knowledge of high domain knowledge users contained 60% news articles and 40% medical and health related articles. After removing all the stop words, these two corpora were used to create two semantic spaces using Gallito (Jorge-Botana et al. 2013; Olmos et al. 2014): a *non-expert semantic space* using the non-expert corpus (average article size: 435 words) and an *expert semantic space* using the expert corpus (average article size: 403 words). The following settings were used to create the semantic spaces: 300 dimensions (the latent construct that best describes the variability in the use of words), entire article as the window (window is the amount of semantic context) and log-entropy weighting (a mathematical transformation to smoothen the effect of extreme word occurrences when computing the importance of a word in a document) (Pincombe 2004). Also, a word was included in the final matrix only if it occurred in at least 6 articles.

5.1.2 Evaluation of semantic spaces

We used two biomedical data sets (Hliaoutakis 2005; Pedersen et al. 2007) commonly used to evaluate measures for computing semantic relevance in the medical information retrieval community. In the first dataset (Pedersen et al. 2007), created in collaboration with Mayo Clinic experts, we have averaged similarity measures on a set of 30 medical terms assessed by a group of 3 physicians, who were experts in rheumatology and 9 medical coders who were aware about the concept of semantic similarity on a scale of 1 (low in similarity) to 4 (high in similarity). The correlation between physician judgements was 0.68, and that between the medical coders was 0.78. In the second dataset (Hliaoutakis 2005), a set of 36 word pairs extracted from MeSH repository were assessed by these authors on a scale of 0 (low in similarity) to 1 (high in similarity), by 8 medical experts. The word pairs in both datasets were translated to Dutch by 3 experts and agreement among them was very high. We dropped two word-pairs from each data set (*antibiotic-*

allergy and *cholangiocarcinoma-colonoscopy* from Pederson’s dataset and *meningitis-tricuspid atresia* and *measles-rubeola* from Hliaoutakis’s dataset) as they were not in the two corpora designed by us. So, we were left with 28 word pairs from Pedersen’s dataset and 34 word pairs from Hliaoutakis’s dataset. Next, we computed the semantic similarity between the remaining word pairs from both data sets and computed the correlation with the expert ratings. We expect the similarity values from the expert semantic space to be higher correlated with the expert ratings than the similarity values from the non-expert semantic space as the former is designed to contain a larger amount of medical and health related information. The correlation values obtained are shown in Table 1.

Analyzing the correlation values from Table 1, we found that the expert semantic space gave a significantly higher correlation with Hliaoutakis’s dataset and Pedersen’s Coders data set and a marginally higher correlation with Pedersen’s Physicians dataset, compared to the non-expert semantic space. Based on these outcomes, we were able to confirm that the expert semantic space has better representation of health and medical knowledge i.e. more detailed and more connected information—than the non-expert semantic space.

5.1.3 Model simulations

We followed the same methodology as the authors in Karanam et al. (2015) who extended the CoLiDeS model to predict user clicks on SERPs. Simulations of CoLiDeS were run using both the non-expert and the expert semantic spaces on each query and its corresponding search results using the same methodology followed by Karanam et al. (2012) on navigating in a mock-up website on the human body. We consider each SERP as a page of a website. And each of the search engine results as a hyperlink within a page of a website. The problem of predicting which search result to click is now equivalent to the problem of predicting which hyperlink to click within a page of a website. Therefore, the process of computing information scent and predicting which search result to click is similar to the process of predicting which hyperlink to click as described in Karanam et al. (2012). We used the user-generated query as representation of the local goal and the semantic similarity values were computed between the queries and the search results.

The main steps we followed are: (a) semantic similarity was computed between the query and the combination of title and snippet of a search result, (b) this was repeated for all the remaining titles and snippets on a SERP. The combination of title and snippet with the highest semantic similarity value with the query was selected by the model, and (c) finally, this process was repeated for all the queries of a task and for all the tasks of a participant and finally for all the participants (see Karanam et al. 2015 for details of the procedure). All semantic similarity values were calculated using gallitoAPI, an API that

Table 1 Correlation values obtained using expert and non-expert semantic spaces on Pedersen et al.’s and Hliaoutakis’s benchmarks

Dataset	Non-expert semantic space	Expert semantic space
Pedersen’s physicians	0.74**	0.78**
Pedersen’s coders	0.74**	0.81**
Hliaoutakis’s medical experts	0.38*	0.58**

** significant at .01 level; * significant at .05 level

exposes the functionality of the semantic spaces trained with Gallito Studio. The gallitoAPI allows one to create an API farm consisting of multiple semantic spaces.

After running the main simulation steps (a) to (c) we had available the model predictions using the two semantic spaces (the non-expert semantic space and the expert semantic space) on all the queries of all the (6 difficult) tasks and next we could compare these with the actual selections of real participants.

5.1.4 Computing the match between model predictions and actual user behavior

We evaluated the performance of the models by matching the model predictions with the actual user behavior for each task. In this next section, we elaborate on how this matching is computed.

- (a) From the behavioral data, the search results clicked by the user are noted.
- (b) If the search result clicked by user is also a search result predicted by the model (this scenario is depicted in Fig. 3), then we consider that the model prediction matches with the actual user behavior. The number of matches for this scenario is 1.
- (c) However, if the search result clicked by user is not a search result predicted by the model (this scenario is depicted also in Fig. 3), then we consider that the model prediction does not match with the actual user behavior. The number of matches for this scenario is 0.
- (d) This process is repeated for all the queries of a particular task. Every time, there is a match between the model prediction and the actual user behavior, the number of matches is incremented by 1 for that task. At the end of this step, we would have available with us the total number of matches between the user clicks and the model(s)-predicted clicks for one task.

Actual user click ✓

(a) Model prediction ✓

(b) Model prediction ✓

Psoriasis - Aandoeningen | Gezondheidsplein.nl
<https://www.gezondheidsplein.nl/aandoeningen/psoriasis/item32010> ▼
 Lees hier de begrijpelijke uitleg over psoriasis, zoals: de symptomen, oorzaak en behandeling. Lees patiëntverhalen en stel vragen.

Psoriasis - Huidarts.com
<https://www.huidarts.com/huidaandoeningen/psoriasis/> ▼
 Actuele en betrouwbare informatie door huidartsen over psoriasis. Over de oorzaken, de symptomen en moderne behandelmethoden.
 Psoriasis van de nagels · Is psoriasis te genezen · Is psoriasis besmettelijk

Psoriasis | Symptomen, oorzaken & oplossingen | Eucerin
www.eucerin.nl/informatie-over-de-huid/problemen/psoriasis ▼
 Een vervelende, chronische aandoening: psoriasis. Wat zijn precies de symptomen en oorzaken van psoriasis, en welke typen zijn er? Eucerin legt uit.

Psoriasis symptomen | Dokteronline.com
www.dokteronline.com > Home > Aandoeningen ▼
 Typerende psoriasis symptomen | Psoriasis symptomen bestrijden | Gezondheidsblog Dokteronline.

Psoriasis, behandeling van psoriasis (patientenfolder) - Huidziekten.nl
www.huidziekten.nl/folders/nederlands/psoriasis.htm ▼
 Psoriasis, behandeling van psoriasis: welke soorten psoriasis zijn er, en hoe kan psoriasis worden behandeld (patientenfolder).

Psoriasis: Ik heb psoriasis | Thuisarts.nl
<https://www.thuisarts.nl/psoriasis/ik-heb-psoriasis> ▼
 5 mrt. 2014 - Psoriasis is een huidaandoening met schilfers, roodheid en jeuk. Psoriasis is niet besmettelijk. De klachten kunnen beginnen of verergeren ...

Fig. 3 Computing match between actual user click and a model-predicted click

- (e) This process is repeated for all the tasks of a participant and finally for all the participants.

After running the simulation steps (a) to (e) using the predictions by CoLiDeS, we would have available with us the number of matches between model predictions and actual user clicks on all the queries of all the tasks for all the participants. Using this data, the mean number of matches per task between the actual user clicks and the model-predicted clicks was computed in relation to age and task difficulty for CoLiDeS. The higher the number of matches is, the better the match between the model and the actual user behavior is.

Please note that the CoLiDeS model can predict only one search result per query using this methodology because CoLiDeS does not possess a backtracking mechanism whereas users in reality click on more than one search result per query. Similar to a greedy algorithm, this version of CoLiDeS never reconsiders its choices and discards non-selected choices.

5.1.5 Simulation results

We evaluated the performance of the models by matching the model predictions with the actual user behavior for each task. The higher the number of matches is, the better the match between the model and the actual user behavior is. A 2 (Semantic Space: Non-Expert vs. Expert) $\times 2$ (Prior Domain Knowledge (PDK): High vs. Low) mixed model ANOVA was conducted with semantic space as within-subjects variable and prior domain knowledge as between-subjects variable. The main effects of semantic space and prior domain knowledge were not significant ($p > .05$). However, the interaction of semantic space and prior domain knowledge was significant $F(1, 46) = 7.5, p < .01$, (Fig. 4).

These outcomes mean that for participants with high domain knowledge, the number of matches was significantly higher with the expert semantic space whereas for participants with low domain knowledge, the number of matches was significantly higher with the non-expert semantic space. It is important to note that this interaction effect is lost when semantic space is not used as a factor in the analysis. Overall, our outcomes suggest that by incorporating differentiated domain knowledge levels into computational cognitive models, it is possible to simulate differences in the search results clicked by users due to differences in domain knowledge.

5.2 Simulation study 2

In the second simulation study, we extend the first simulation study in order to examine the second research question by also taking into account the difference in knowledge acquisition process when creating semantic spaces with differentiated knowledge on the topic of health and medicine.

5.2.1 Creation of semantic spaces

We first collated three different corpora:

1. A general corpus consisting of 100,000 news articles and 663,990 unique Dutch words. This corpus covers more or less the basic Dutch vocabulary.

Fig. 4 Mean number of matches (per task) in relation to semantic space and prior domain knowledge (PDK)

2. A non-expert medical corpus consisting of 25,000 articles (from popular health and medical magazines) and 220,749 unique Dutch words. This corpus covers popular medical terms.
3. An expert medical corpus consisting of 25,000 articles (from medical encyclopedia) and 188,126 unique Dutch words. This corpus covers specialized medical and health related terms.

Our purpose is to simulate different ways through which one can gain a larger amount of knowledge in a domain. To do this, we mimic two knowledge acquisition processes:

1. *Evolutionary* The evolutionary acquisition process assumes that word meanings undergo a natural progression from basic to expert level (the final goal state). An example of such a progression is school vocabulary acquisition. In this scenario, a students representations become adult representations over a period of time if there is a normative development. Previous studies have used this strategy to simulate a natural progression of learning, extracting maturity curves for each word of ordinary use vocabulary (Biemiller et al. 2014; Jorge-Botana et al. 2016; Kireyev and Landauer 2011). In our case, the starting context of meaning acquisition is assumed to be the same between expert and non-experts, but experts are assumed to have an additional

exposure to deeper information in the domain. In other words, non-experts have an intermediate level while experts have an optimal level. Following this criterion, a corpus representing low prior domain knowledge is constructed with the general corpus + non-expert medical corpus and next to that a corpus representing high prior domain knowledge is constructed with the general corpus + non-expert medical corpus + expert medical corpus.

2. *Common core* As in the evolutionary process, the context of acquisition is assumed to be the same between experts and non-experts only for a small set of concepts common to both. Beyond such common contexts, experts and non-experts are exposed to different contexts of specialization within a domain (such as health/medicine), resulting in different levels of expertise. Imagine for example two disciplines that share some technical terms. Both groups are exposed to the same ordinary life situations contexts, but differ in the use of their technical terms. So, the question is whether or not the representations of such technical terms are similar. Take for instance, the term “complement”. The meaning of this word for a mathematician would be different from the meaning of this word for a chemist. The common core process is involved in studies that monitor changes in the representation of some words over time, but without an optimal goal state (Balbi and Esposito 1998). Therefore, a corpus representing low prior domain knowledge is constructed based on a general corpus + a non-expert medical corpus, and next to that, a corpus representing high prior domain knowledge is constructed based on a general corpus + expert medical corpus.

To mimic both acquisition processes, we only need to compare the three corpora mentioned above: general corpus + non-expert, general corpus + expert medical corpus and general corpus + non-expert medical corpus + expert medical corpus. These three corpora were used to create corresponding semantic spaces using Gallito (Jorge-Botana et al. 2013; Olmos et al. 2014). Thus, we got three semantic spaces: one representing low prior domain knowledge (the non-expert semantic space) and two for high prior knowledge, one for each process (the so-called “abrupt” expert semantic space developed from common core process and the “progressive” expert semantic space from the evolutionary process). Following settings were used to create the semantic spaces: a special stop list that includes all Dutch stop words, 300 dimensions, entire article as the window and log-entropy weighting. Also, a word was included in the final matrix only if it occurred in at least 10 articles.

5.2.2 Evaluation of semantic spaces

We evaluated the three semantic spaces using exactly the same methodology as in the first simulation study. For each of the three semantic spaces, we first computed the semantic similarity between word pairs from two different biomedical datasets Hliaoutakis’s data set (Hliaoutakis 2005) and Pedersen’s dataset (Pedersen et al. 2007). We then computed the correlation between the semantic similarity values obtained and the expert ratings. We expect the semantic similarity values from the progressive expert semantic space to be higher correlated with the expert ratings compared to the abrupt expert semantic space and the non-expert semantic space.

Analyzing the correlation values from Table 2, we found that both the abrupt expert semantic space and the progressive expert semantic space gave a significantly higher correlation than the non-expert semantic space with all three datasets. The progressive

Table 2 Correlation values obtained using non-expert, abrupt expert and progressive expert semantic spaces on Pedersen et al.'s and Hliaoutakis's benchmarking data sets

Dataset	Non-expert semantic space	Abrupt expert semantic space	Progressive expert semantic space
Pedersen's physicians	0.66**	0.75**	0.76**
Pedersen's coders	0.65**	0.74**	0.77**
Hliaoutakis's medical experts	0.43*	0.55**	0.58**

** significant at .01 level; * significant at .05 level

expert semantic space gave a significantly higher correlation with Hliaoutakis's dataset and a marginally higher correlation with Pedersen's datasets, compared to the non-expert semantic space. Based on these outcomes, we are able to confirm that the progressive expert semantic space has an optimal level of health and medical knowledge followed by the abrupt expert semantic space and, next, the non-expert semantic space.

We also analyzed the semantic spaces by introducing a new metric called *stability index*. This index measures the change of the vector representation of a term in two semantic spaces. It can be defined as the cosine between the vector of a word in one semantic space (one knowledge model) and the vector of the same word in the other semantic space (the other knowledge model)) (formula 1)

$$Stability = \cos(v, v') \quad (1)$$

where v is the vector of a word in semantic space 1 and v' is the vector of the same word in semantic space 2. The vector of a word consists of the coordinates (numbers) that represent the semantics of a word in the high dimensional space of a Latent Semantic Analysis model. The cosine distance between the coordinates of one word to the coordinates of another word is the measure of similarity.

This measure is based on studies whose main purpose is to align semantic spaces from parallel corpora of different languages and to compare words in a single language-independent space (Balbi and Misuraca 2006b; Littman et al. 1998). For this purpose, we use Procrustes Alignment (Balbi and Misuraca 2006a), a procedure that is provided in Gallito Studio. This technique aligns both semantic spaces on the basis of a rotation angle by means of which the term matrix of one semantic space is rotated based on the references of the term matrix of the other semantic space. This rotation makes words from the two semantic spaces comparable by means of vector similarity metrics, such as cosine similarity (formula 1). The methodology starts by selecting some documents that are common in both semantic spaces and assuming that the meanings of them are the same in both semantic spaces (Kireyev and Landauer 2011). They are called *pivots*, as they act as common part of both semantic spaces. Because no medical document could be selected as pivots (the reason being that the purpose of the technique is to precisely monitor the stability of medical concepts), we selected the general corpus with its 100,000 news articles as pivots. We aligned the two semantic spaces by means of this procedure. First for the common core process, we aligned the non-expert semantic space with the abrupt expert semantic space. Second, for the evolutionary process, we aligned the non-expert semantic space with the progressive expert semantic space. As mentioned above, for both alignments, we used the general corpus documents as pivots.

We expect to find more stability in common terms than in the medical terms. In other words, the meaning of common words should not change much between the non-expert semantic space and the expert semantic spaces whereas the meaning of the medical terms should change significantly. Also, we expect the progressive expert semantic space to have higher stability of medical terms compared to the abrupt expert semantic space because it follows a more structured, cumulative and gradual inductive exposure to terms. To check this, we selected three types of terms: 35 common terms with no medical sense, 35 commonly used non-expert medical terms and 35 expert medical terms. For each process, the stability index was calculated between a term in one semantic space and the same term in the other semantic space. For instance, for the common core process, the vector of the term “heart” from the non-expert semantic space is compared with the vector of the term “heart” in the abrupt expert semantic space and for the evolutionary process, the vector of the term “heart” from the non-expert semantic space is compared with the vector of the term “heart” from the progressive expert semantic space. This procedure was repeated for all the three types of terms and for the two processes: evolutionary and common core.

A 2 (Process: Evolutionary vs. Common Core) $\times$ 3 (Type of Term: Common vs. Non-Expert Medical vs Expert Medical) mixed ANOVA with process as within subjects variable and type of term as between subjects variable and the mean stability of a term as the dependent variable was conducted. The results are presented in Fig. 5. The main effect of process was significant $F(1, 102) = 47.08, p < .001$. Stability of a term was significantly higher in the evolutionary process compared to the common core process. This is along expected lines as the evolutionary process follows a structured and cumulative exposure to terms. The main effect of type of term was significant $F(2, 102) = 33.58, p < .001$. As expected, the post-hoc multiple comparisons (with Bonferroni correction) showed that general vocabulary was the type of term with the highest stability ($p < .001$), followed by non-expert medical terms, and finally, the expert terms which showed the lowest stability.

Moreover, the interaction of process and type of term was also significant $F(1, 102) = 7.84, p < .001$. The post-hoc multiple comparisons (with Bonferroni correction and $p < .05$) showed that the drop in stability (between general vocabulary terms and non-expert or expert medical terms) is much higher for the common core process compared to the evolutionary process and this is because the evolutionary process follows a more structured and cumulative exposure to terms whereas the common core process involves a different trajectory of exposure. These outcomes contribute to the reliability of the knowledge models designed by us.

5.2.3 Model simulations

We followed exactly the same approach to run model simulations as in the first simulation study. We ran the steps (a) to (c) described in Sect. 5.1.3 for all the queries of a task and for all the tasks and for all the participants using the three semantic spaces: non-expert, abrupt expert and progressive expert. After running the main simulation steps (a) to (c) we had available the model predictions using the three semantic spaces (the non-expert, the abrupt expert and the progressive expert) on all the queries of all the (6 difficult) tasks and we could compare these with the actual selections of real participants.

5.2.4 Simulation results

We used the same metric: the number of matches between model predictions and actual user clicks as in the first simulation study. A 3 (Semantic Space: Non-Expert vs. Abrupt

Fig. 5 Mean stability index in relation to type of term and process of knowledge acquisition

Expert vs. Progressive Expert) $\times$ 2 (Prior Domain Knowledge (PDK): High vs. Low) mixed model ANOVA was conducted with semantic space as within-subjects variable and prior domain knowledge as between-subjects variable. The main effect of semantic space was significant $F(2, 46) = 4.71, p < .05$. Post-hoc comparison showed that the match between model predictions and actual user behavior was the highest with the progressive expert semantic space, followed by the abrupt expert semantic space, and finally the non-expert semantic space. The main effect of PDK was not significant ($p > .05$), however the interaction of semantic space and PDK was significant $F(2, 46) = 4.45, p < .05$, (See Fig. 6). Post-hoc tests showed that the effect of using an expert semantic space was significant ($p < .05$) only for high domain knowledge participants. That is, the match between the model predictions and the actual user behavior was significantly higher for the abrupt expert semantic space compared to the non-expert semantic space only for high domain knowledge participants and not for low domain knowledge participants. Similarly, the match between the model predictions and the actual user behavior was significantly higher for the progressive expert semantic space compared to the non-expert semantic space only for high domain knowledge participants and not for low domain knowledge participants.

6 Conclusions and discussion

In this paper, we focused on two main limitations of cognitive models of information search: 1) they do not incorporate individual differences in domain knowledge of users and 2) they also do not take into account differences in the process of acquisition of domain knowledge. However, it is known in cognitive psychology that domain knowledge does have an impact on information search performance (Cole et al. 2011; Duggan and Payne 2008; Held et al. 2012; Kiseleva et al. 2016; Monchaux et al. 2015; Palotti et al. 2016; White et al. 2009). It is also known that there are different ways of acquiring knowledge on a domain.

We used the cognitive model of information search called CoLiDeS (Kitajima et al. 2000) in our study as it is simple and process-oriented and has features that can be exploited for our purposes. For example, it uses a semantic space to represent the specific knowledge levels of a target user group and computes the information scent values from it. We exploited this feature in our cognitive model and created in our first study two semantic spaces: a non-expert semantic space and an expert semantic space that mimic low and high levels of expertise on the topic of health and medicine. We used two biomedical data sets (Hliaoutakis 2005; Pedersen et al. 2007) commonly used to evaluate measures for computing semantic relevance in the medical information retrieval community to evaluate the two semantic spaces. We found that the expert semantic space gave a significantly higher correlation with Hliaoutakis's dataset and Pedersen's Coders data set and a marginally

Fig. 6 Mean number of matches (per task) in relation to semantic space and prior domain knowledge (PDK)

higher correlation with Pedersen's Physicians dataset, compared to the non-expert semantic space (Table 1). Based on these outcomes, we were able to confirm that the expert semantic space has better representation of health and medical knowledge compared to the non-expert semantic space.

Simulations on six difficult information search tasks and subsequent matching with actual behavioral data from 48 users (divided into low and high domain knowledge groups based on a domain knowledge test) were conducted. Indeed the results showed that the modeling should take into account individual differences in domain knowledge and adapt the semantic space to these differences (Fig. 3): with high domain knowledge participants the efficacy of the modeling (in terms of the number of matches between model predictions and actual user clicks) was higher with the expert semantic space compared to the non-expert semantic space while for low domain knowledge participants it was the other way around (**RQ1**). A plausible explanation for the interaction effect is that the expert and the non-expert semantic spaces give appropriate similarity values as assessed by users with high (more precise) and low (less precise) domain knowledge respectively. It is important to note that this interaction effect is lost when semantic space is not used as a factor in the analysis. That is, if we would not have used semantic space as a factor, we would have concluded that there is no difference in the model's performance between the participants with high and low domain knowledge levels. This would have been a hasty conclusion because, when we included semantic space as a factor in the analysis, there was an effect of prior domain knowledge, but it was dependent on the type of semantic space.

In our second simulation study, we extended the first simulation study by also taking into account the process of acquisition of knowledge in a domain. We identified two types of knowledge acquisition processes: evolutionary and common core. Evolutionary exposure follows a standard and extended sequence until a final state that represents optimal knowledge, is reached (for example, the process of vocabulary acquisition in school). Common core exposure has no final state that represents an optimal knowledge state. Instead, different people have different trajectories of exposure following different sequences, but share a common core of exposure at the beginning. Based on these two types of knowledge acquisition processes, we created two types of expert semantic spaces: a progressive expert semantic space corresponding to the evolutionary process and an abrupt expert semantic space corresponding to the common core process in addition to the non-expert semantic space. We followed the same methodology as in the first simulation study to evaluate the three semantic spaces. We found that both the abrupt expert semantic space and the progressive expert semantic space gave a significantly higher correlation than the non-expert semantic space with all three datasets. The progressive expert semantic space gave a significantly higher correlation with Hliaoutakis's data set and a marginally higher correlation with Pedersen's data sets, compared to the non-expert semantic space 2). Based on these outcomes, we were able to confirm that the progressive expert semantic space has maximum health and medical knowledge followed by the abrupt expert semantic space and then the non-expert semantic space.

We introduced a new metric called stability index to further evaluate the three semantic spaces. This measure gives us an indication of the amount of change in the meaning of a word between two semantic spaces. We computed the stability index of three types of terms: common terms with no medical sense, commonly used non-expert medical terms and expert medical terms. As expected, the stability of common terms was higher than that of the medical terms. Also, the stability of medical terms in the progressive expert semantic space was higher than that of the abrupt expert semantic space (Fig. 5). These outcomes gave reliability to the semantic spaces designed by us.

Next, simulations of CoLiDeS were run using the three semantic spaces and the model clicks were matched with the behavioral data as in the first simulation study. The match between model predictions and actual user behavior was the highest with the progressive expert semantic space, followed by the abrupt expert semantic space, and finally the non-expert semantic space. The interaction effect found in the first simulation study was also significant in the second simulation study (Fig. 6). Post-hoc tests showed that the effect of using an expert semantic space was significant only for high domain knowledge participants. That is, the match between the model predictions and the actual user behavior was significantly higher between the abrupt expert semantic space and the non-expert semantic space conditions only for high domain knowledge participants and not for low domain knowledge participants. Similarly, the match between the model predictions and the actual user behavior was significantly higher between the progressive expert semantic space and the non-expert semantic space conditions only for high domain knowledge participants and not for low domain knowledge participants (**RQ2**). The difference between the abrupt expert semantic space and the progressive expert semantic space for high domain knowledge participants was not significant. One of the reasons could be that the amount of medical and health related information in the two semantic spaces was not different enough. A second possibility for the lack of significant difference between the two types of expert semantic spaces could be that we did not make a distinction between abrupt experts and progressive experts in our participants. Future research should take this into account and select real users who are abrupt experts and progressive experts for collecting actual behavioral data. A third and the least likely possibility in our view could be, of course, that the knowledge acquisition process does not matter at all.

Overall, our outcomes suggest that using appropriate semantic spaces - a semantic space with high domain knowledge represented for high domain knowledge users and a semantic space with low domain knowledge represented for low domain knowledge users - gives better prediction outcomes. This is an important finding because improved predictive capacity of these models can lead to more accurate model-generated support for search and navigation which, in turn, leads to enhanced information seeking performance, as two studies have already shown (Karanam et al. 2011; Oostendorp and Juvina 2007). In these studies, for each task, navigation support was automatically generated by recording the step-by-step decisions made by the cognitive model which in turn are based on the semantic similarity of hyperlinks to the user goal (given by a task description). The model predictions were presented online to the user in the form of visually highlighted hyperlinks during navigation. In both studies, the navigation performance of participants who received such support was found to be more structured and less disoriented compared to participants who did not receive such support. This was found to be true, especially for participants with a particular cognitive deficit, such as low spatial ability.

Model generated support for information search and navigation also contributes to the knowledge acquisition process as it helps users in efficiently filtering unnecessary information. It gives them more time to process and evaluate relevant information during the intermediate stages of clicking on search results and web-pages within websites before reaching the target page. This helps in reducing user's effort, in turn, lessening cognitive load. This can lead to better comprehension and retention of relevant material (because contextual information relevant to the user's goal is emphasized by model generated support), thereby, leading to higher incidental learning outcomes. Concerning precisising the modeling itself, we are currently running experiments with the more advanced model CoLiDeS+ (Juvina and Oostendorp 2008) which was found to be more efficient than CoLiDeS in locating the target page on real websites (Karanam et al. 2016). CoLiDeS+

incorporates contextual information in addition to information scent and implements backtracking strategies and therefore can predict more than one click on a SERP. Lastly, the domain of health has been used only as an example and we think that these results would be generalizable to any domain.

Acknowledgements This research was supported by Netherlands Organization for Scientific Research (NWO), ORA-Plus project MISSION (464-13-043).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Balbi, S., & Esposito, V. (1998). Comparing advertising campaigns by means of textual data analysis with external information. *Actes des 4es Journees Internationales d'Analyse Statistique des Donnees Textuelles*, 4, 39–47.
- Balbi, S., & Misuraca, M. (2006a). Procrustes techniques for text mining. In S. Zani, A. Cerioli, M. Riani, & M. Vichi (Eds.), *Data analysis, classification and the forward search*. Studies in classification, data analysis, and knowledge organization. Berlin, Heidelberg: Springer.
- Balbi, S., & Misuraca, M. (2006b). Rotated canonical correlation analysis for multilingual corpora. In J. M. Viprey (Ed.), *JADT 2006* (pp. 99–106). Besançon.
- Biemiller, A., Rosenstein, M., Sparks, R., Landauer, T. K., & Foltz, P. W. (2014). Models of vocabulary acquisition: Direct tests and text-derived simulations of vocabulary growth. *Scientific Studies of Reading*, 18(2), 130–154.
- Blackmon, M. H., Mandalia, D. R., Polson, P. G., & Kitajima, M. (2007). Automating usability evaluation: Cognitive walkthrough for the web puts Isa to work on real-world hci design problems. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 345–375). Mahwah, NJ: Lawrence Erlbaum Associates.
- Borlund, P., & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3), 225–250.
- Chi, M. T., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, 19(1), 29.
- Chuklin, A., Markov, I., & de Rijke, M. (2015). Click models for web search. *Synthesis Lectures on Information Concepts Retrieval and Services*, 7(3), 1–115.
- Cole, M. J., Zhang, X., Liu, C., Belkin, N. J., & Gwizdka, J. (2011). Knowledge effects on document selection in search results pages. In *Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval* (pp. 1219–1220). New York: ACM.
- Duggan, G. B., & Payne, S. J. (2008). Knowledge in the head and on the web: Using topic expertise to aid search. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 39–48). New York: ACM.
- Durso, F. T., & Coggins, K. A. (1990). Graphs in the social and psychological sciences: Empirical contributions of pathfinder. In R. W. Schvaneveldt (Eds.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 31–51). Ablex Publishing.
- Ferstl, E. C., & Kintsch, W. (1999). The construction of mental representations during reading. In H. Van Oostendorp & S. R. Goldman (Eds.), *Learning from text: Structural knowledge assessment in the study of discourse comprehension* (pp. 247–277). Mahwah, NJ: Lawrence Erlbaum.
- Fu, W. T. (2013). From plato to the world wide web: Information foraging on the internet. In T. T. Hills, W. T. Robbins, & M. P. Todd (Eds.), *Cognitive search* (pp. 283–299). Cambridge: MIT Press.
- Fu, W. T., & Dong, W. (2010). Collaborative indexing and knowledge exploration: A social learning model. *IEEE Intelligent Systems*, 1, 39–46.
- Held, C., Kimmerle, J., & Cress, U. (2012). Learning by foraging: The impact of individual knowledge and social tags on web navigation processes. *Computers in Human Behavior*, 28(1), 34–40.

- Hliaoutakis, A. (2005). Semantic similarity measures in mesh ontology and their application to information retrieval on medline. Master's Thesis, Technical Univ. of Crete, Dept. of Electronic and Computer Engineering, Crete, Greece.
- Hu, B., Zhang, Y., Chen, W., Wang, G., & Yang, Q. (2011). Characterizing search intent diversity into click models. In *Proceedings of the 20th international conference on world wide web* (pp. 17–26). New York: ACM
- Jorge-Botana, G., Olmos, R., & Barroso, A. (2013). Gallito 2.0: A natural language processing tool to support research on discourse. In *Proceedings of the 13th annual meeting of the society for text and discourse*.
- Jorge-Botana, G., Olmos, R., & Sanjosé, V. (2016). Predicting word maturity from frequency and semantic diversity a computational study. *Discourse Processes* (accepted).
- Juvina, I., & van Oostendorp, H. (2008). Modeling semantic and structural knowledge in web navigation. *Discourse Processes*, 45(4–5), 346–364.
- Karanam, S., & van Oostendorp, H. (2016). Age-related differences in the content of search queries when reformulating. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5720–5730). New York: ACM
- Karanam, S., Van Oostendorp, H., & Indurkha, B. (2011). Towards a fully computational model of web-navigation. In K. G. Mehrotra, C. K. Mohan, J. C. Oh, P. K. Varshney, & M. Ali (Eds.), *Modern approaches in applied intelligence*. IEA/AIE 2011. Lecture notes in computer science, Vol. 6703. Berlin, Heidelberg: Springer.
- Karanam, S., van Oostendorp, H., & Indurkha, B. (2012). Evaluating colides+ pic: The role of relevance of pictures in user navigation behaviour. *Behaviour and Information Technology*, 31(1), 31–40.
- Karanam, S., van Oostendorp, H., & Fu, W. T. (2016). Performance of computational cognitive models of web-navigation on real websites. *Journal of Information Science*, 42(1), 94–113.
- Karanam, S., van Oostendorp, H., Sanchiz, M., Chevalier, A., Chin, J., & Fu, W. T. (2015). Modeling and predicting information search behavior. In *Proceedings of the 5th international conference on web intelligence, mining and semantics* (p. 7). New York: ACM.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kireyev, K., & Landauer, T. K. (2011). Word maturity: Computational modeling of word knowledge. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies-volume 1, association for computational linguistics* (pp. 299–308).
- Kiseleva, J., Montes Garcia, A., Kamps, J., & Spirin, N. (2016). The impact of technical domain expertise on search behavior and task outcome. In *Proceedings of the workshop on query understanding for search on all devices*. New York: ACM
- Kitajima, M., Blackmon, M. H., & Polson, P. G. (2000). A comprehension-based model of web navigation and its application to web usability analysis. In S. McDonald, Y. Waern, & G. Cockton (Eds.), *People and computers XIV—usability or else!* (pp. 357–373). London: Springer. doi:10.1007/978-1-4471-0515-2_24.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- Littman, M. L., Dumais, S. T., & Landauer, T. K. (1998). Automatic cross-language information retrieval using latent semantic indexing. In G. Grefenstette (Ed.), *Cross-language information retrieval* (pp. 51–62). Berlin: Springer.
- Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4), 41–46.
- Monchaux, S., Amadieu, F., Chevalier, A., & Mariné, C. (2015). Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing and Management*, 51(5), 557–569.
- Olmos, R., Jorge-Botana, G., León, J. A., & Escudero, I. (2014). Transforming selected concepts into dimensions in latent semantic analysis. *Discourse Processes*, 51(5–6), 494–510.
- Olston, C., & Chi, E. H. (2003). Scentrails: Integrating browsing and searching on the web. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 10(3), 177–197.
- Palotti, J., Hanbury, A., Müller, H., & Kahn, C. E., Jr. (2016). How users search and what they search for in the medical domain. *Journal of Information Retrieval*, 19(1), 189–224.
- Pedersen, T., Pakhomov, S. V., Patwardhan, S., & Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40(3), 288–299.
- Pincombe, B. (2004). Comparison of human and latent semantic analysis (lsa) judgements of pairwise document similarities for a news corpus. Tech. rep., DTIC Document.
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643.

- Shen, S., Hu, B., Chen, W., & Yang, Q. (2012). Personalized click model through collaborative filtering. In *Proceedings of the fifth ACM international conference on web search and data mining* (pp. 323–332). New York: ACM
- Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing and Management*, 39(3), 445–463.
- Van Oostendorp, H., & Juvina, I. (2007). Using a cognitive model to generate web navigation support. *International Journal of Human Computer Studies*, 65(10), 887–897.
- White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second ACM international conference on web search and data mining* (pp. 132–141). New York: ACM
- Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology*, 55(3), 246–258.
- Xing, Q., Liu, Y., Nie, J. Y., Zhang, M., Ma, S., & Zhang, K. (2013). Incorporating user preferences into click models. In *Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge Management* (pp. 1301–1310). New York: ACM