

Integrating domain knowledge differences into modeling user clicks on search result pages

Saraschandra Karanam
Utrecht University
Utrecht, The Netherlands
s.karanam@uu.nl

Herre van Oostendorp
Utrecht University
Utrecht, The Netherlands
h.vanoostendorp@uu.nl

ABSTRACT

Computational cognitive models developed so far do not incorporate any effect of individual differences in domain knowledge of users 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 low (general semantic space) and high (special semantic space) amount of medical and health related information to represent respectively the low and high knowledge of users in this domain. Simulations on six difficult information search tasks and subsequent matching with actual behavioural data from 48 users (divided into low and high domain knowledge groups based on a domain knowledge test) were conducted. Results showed that the efficacy of modeling user selections on search results (in terms of the number of matches between users and the model and the mean semantic similarity values of the matched search results) is higher with the special semantic space compared to the general semantic space for high domain knowledge participants while for low domain knowledge participants it is the other way around. Implications for support tools that can be built based on these models are discussed.

CCS Concepts

•Information systems → Personalization; Relevance assessment;

Keywords

Cognitive Modeling, Information Search, Domain Knowledge, Corpus, Semantic Space.

1. INTRODUCTION

Search systems are typically characterized as a tool to retrieve relevant information on a target page from the Internet in response to an user query. These systems are efficient only for a certain type of tasks such as look-up tasks or factoid questions (“*What is the distance between Mars and Earth*” or “*Which is the highest mountain in Europe*”) and

are not very optimal for other kind of tasks that involve knowledge discovery, comprehension and learning. Many times, important information that is needed to solve the main search problem is present in the intermediate pages leading to the target page [19]. 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 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 [4, 5]. 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 [17, 3] on the cognitive processes underlying information search and navigation and follow a one-size-fits-all model.

In this paper, we focus on the differences in information search behavior due to the individual differences in the domain knowledge of users. It is known that users with high domain knowledge have more appropriate mental representations and higher activation degrees of concepts and stronger connections between different concepts in the conceptual space compared to users with low domain knowledge [14]. A number of experiments investigating the role of domain knowledge on information search and navigation performance have been conducted in the cognitive psychology community. For example, in a recent study by [17], 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. Higher 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 a search result to click next. Domain experts are also known to evaluate search results more thoroughly and click more often on relevant search results compared to non-experts. This is because their higher domain knowledge enables them to differentiate between a relevant and a non-relevant search result better [3].

However, understanding behavioral differences through laboratory experiments is not only expensive and not scalable but also time consuming. Simulation of user interactions with information retrieval systems therefore has been an active area of research. Among the many click models developed by researchers from the information retrieval community [2], only a few take into account cognitive aspects [24, 22, 7]. Moreover, they provide only limited process description. We therefore, employ computational cognitive models in our research which are relevant in this context as they enable us to model differences in cognitive factors (such as domain knowledge) underlying any cognitive function(s) (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 of computational cognitive models is on the process that leads to the target information and are therefore more capable of providing opportunities to incorporate behavioral differences due to variations in cognitive factors.

The main research question of the current study was: how to incorporate the differences in the domain knowledge levels of users into computational cognitive models that predict click behaviour on search results? Would such a model predict user clicks on search engine result pages (SERPs, henceforth) better than a model that does not incorporate differentiated domain knowledge levels of users? Outcomes of this study would have implications for the support tools for enhancing information search performance, that can be built based on the computational cognitive models [23, 11].

2. OUR APPROACH

We briefly introduce the computational cognitive model called CoLiDeS that we use in our research and next to that explain our approach to incorporate differentiated domain knowledge into CoLiDeS.

2.1 Cognitive model

CoLiDeS, or Comprehension-based Linked Model of Deliberate Search, developed by Kitajima et al. [15] explains user navigation behaviour on websites. It 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 [21] and connects to the Construction-Integration reading model of Kintsch [14]. The notion of *information scent*, defined as the estimate of the value or cost of information sources represented by proximal cues (such as hyperlinks), is central to CoLiDeS. 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. CoLiDeS uses Latent Semantic Analysis (LSA, henceforth) introduced by [16] 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 representation of terms from the corpus in a low number of dimensions, typically between 250 and 350 and are orthogonal, abstract and latent [16, 18]. CoLiDeS has been successful in simulating and predicting user 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 [1]. CoLiDeS model has recently been extended to predict user clicks on search result pages [13]. Please note that the CoLiDeS modeling so far does not incorporate any effect of individual differences in the domain knowledge of users and that is what we will study in the current paper.

2.2 Creation of Semantic Spaces

When using LSA, it is known that the initial corpus of documents used to create the semantic space influences the final similarity values obtained to a large extent [8]. Several factors determine the choice of the corpus and the semantic space. First and foremost, is the language of the corpus. In our case, since we are running our experiments in The Netherlands with Dutch participants, we need a Dutch semantic space. Secondly, the corpus of documents should be representative of the knowledge levels of the target user group. Since the focus of our research is modeling information search behaviour of older adults compared to younger adults, we need two corpora that could accurately characterize the difference in the knowledge levels of younger and older adults. We have seen already that older adults have higher crystallized intelligence or general knowledge and vocabulary than younger adults. Also, since older adults read more health related information and are more concerned with their health, we assume that their health and medical knowledge would be elaborated than that of younger adults. Our goal is to build two semantic spaces that are as close as possible to the above assumptions.

We collated two different corpora (general corpus and special corpus, each consisting of 70,000 articles in Dutch) varying in the amount of medical and health related information. The *general corpus*, representing the knowledge of low domain knowledge users had 90% news articles and 10% medical and health related articles whereas the *special corpus*, representing the knowledge of high domain knowledge users had 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 [18]: a *general* semantic space using the general corpus (average article size: 435 words) and a *special* semantic space using the special corpus (average article size: 403 words). Following settings were used to create the semantic spaces: 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 6 articles.

2.3 Evaluation of Semantic Spaces

We used two biomedical data sets [6, 20] commonly used to evaluate measures for computing semantic relevance in the medical information retrieval community. In the first dataset [20], 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

Table 1: Correlation values obtained using special and general semantic spaces on Pedersen et al.’s and Hliaoutakis’s benchmarks (significant at .01 level, * significant at .05 level).**

Dataset	Special Semantic Space	General Semantic Space
Pedersen’s Physicians	0.78**	0.74**
Pedersen’s Coders	0.81**	0.74**
Hliaoutakis’s Medical Experts	0.58**	0.38*

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 [6], a set of 36 word pairs extracted from MeSH repository were assessed 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 expected the similarity values from the special semantic space to be more highly correlated with the expert ratings than the similarity values from the general semantic space as the former was designed to contain greater medical and health related information. The correlation values obtained are shown in Table 1.

Analysing the correlation values from Table 1, we found that the special 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 general semantic space. Based on these outcomes, we were able to confirm that the special semantic space has health and medical knowledge better represented than the general semantic space.

2.4 Behavioral Data Collection

Actual behavioural data was collected from 48 participants (18 females, 30 males, average age: 48.79) in a laboratory experiment. Participants were first presented with a domain knowledge test on the topic of health in which they had to answer twelve multiple choice questions. A correct answer was scored 1 and a wrong answer was scored 0. They were then presented with six information search tasks in random order specifically from the domain of health in order to examine the behavioural differences in click behaviour of the participants, if any, because of the individual differences in their knowledge of the health domain. To solve these tasks, they had to formulate queries using their knowledge and understanding of the task, the answer was not present in one location or a website and often they had to evaluate information from multiple websites. For instance, 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”* was also not found easily in the snippets of the search results of the queries, unless the query was very specific.

Participants were allowed to use only Google’s search engine. All the queries generated by the users, the corresponding search engine result pages and the URLs opened by them were logged in the backend. There were in total 738 queries and 724 clicks.

3. MODEL SIMULATIONS

We followed the same methodology as authors in [13] who extended the CoLiDeS model to predict user clicks on SERPs. Simulations of CoLiDeS were run using both the general and the special semantic spaces on each query and its corresponding search results using the same methodology followed by Karanam et al., [12] 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 engine 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 remains the same as in [12]. For the time being, we used the user-generated query as a representation of local goal or the understanding of the user at any point of time and semantic similarity values were computed from it. The main steps we followed in simulating CoLiDeS on interacting with the SERPs are the following: (a) the semantic similarity between the query and the title and the snippet combination of a search result was computed, (b) this was repeated for all the remaining titles and snippets on a SERP. The title and snippet combination 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 [13] for details of the procedure).

After running the main simulation steps a) to c) we had available the model predictions on all the queries of all the tasks and we could compare these with the actual selections of real participants. 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.

4. SIMULATION RESULTS

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. We used two metrics to evaluate the efficacy of modeling: number of matches per task between the model and the actual participant behaviour and the LSA value of the matches in our analysis. For both metrics, a 2 (Semantic Space: General vs. Special) X 2 (Prior Domain Knowledge (PDK): High vs. Low) mixed model ANOVA was conducted with semantic space as within-subjects variable and prior

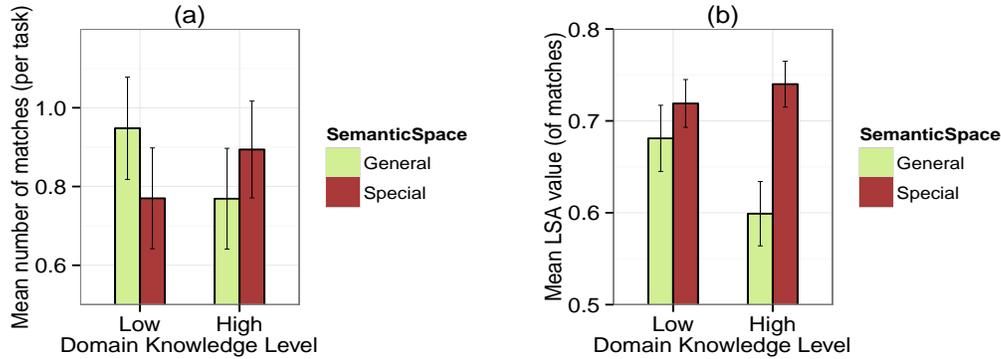


Figure 1: (a) Mean number of matches (per task) and (b) Mean LSA value (of matches) in relation to Semantic Space and Prior Domain Knowledge (PDK).

domain knowledge as between-subjects variable.

4.1 Number of matches per task

For each query and its corresponding SERP, the number of matches between the model predictions and the actual participant behavior is computed. This gives us an indication of how many of the total number of actual participant clicks per task did the model successfully predict. 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 $F(1,46) = 7.5$, $p < .01$ (Figure 1a).

4.2 LSA value of matched search result

For each match between the model and the actual participant click, the LSA value of the match is determined using the two different semantic spaces. Data of 2 participants from the low domain knowledge group and 3 participants from the high domain knowledge group had to be dropped as there were no matches with the actual behaviour for these participants. The main effect of semantic space was significant $F(1,41) = 8.88$, $p < .005$. The main effect of prior domain knowledge was not significant ($p > .05$). The interaction of semantic space and prior domain knowledge was tending towards significance $F(1,41) = 2.9$, $p < .09$ (Figure 1b).

Taking all together, Figure 1a shows that for participants with high domain knowledge, the number of matches was significantly higher with the special semantic space whereas for participants with low domain knowledge, the number of matches was significantly higher with the general semantic space. From Figure 1b, we can see that the special semantic space matched user behaviour with a significantly higher LSA value, especially for participants with high domain knowledge.

5. CONCLUSIONS

Indeed the results show that the modeling should take into account individual differences in domain knowledge and adapt the semantic space to these differences: with high domain knowledge participants the efficacy of the modeling (in terms of the number of matches and the LSA values of the matched search results) is higher with the special semantic space compared to the general semantic space while for low

domain knowledge participants it is the other way around. A possible explanation for the interaction effect is that the special and the general 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 these interaction effects are 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 PDK, *but* it was dependent on the type of semantic space.

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. Improved predictive capacity of these models would lead to more accurate model-generated support for search and navigation which, in turn, would lead to enhanced information seeking performance, as two studies have already shown [11, 23]. For each task, navigation support was generated by recording the step-by-step decisions made by the cognitive model which in turn are based on the semantic relatedness of hyperlinks to the user goal (given by a task description). The model predictions were presented to the user in the form of visually highlighted hyperlinks. 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 contributes to the knowledge acquisition process as it helps the 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 con-

textual 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+ [9] which was found to be more efficient than CoLiDeS in locating the target page on real websites [10]. 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.

6. ACKNOWLEDGMENTS

This research was supported by Netherlands Organization for Scientific Research (NWO), ORA-Plus project MISSION (464-13-043), and carried out in collaboration with University of Toulouse and University of Illinois.

7. REFERENCES

- [1] M. H. Blackmon, D. R. Mandalia, P. G. Polson, and M. Kitajima. Automating usability evaluation: Cognitive walkthrough for the web puts lsa to work on real-world hci design problems. In T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch, editors, *Handbook of Latent Semantic Analysis*, pages 345–375. Lawrence Erlbaum Associates Mahwah, NJ, 2007.
- [2] A. Chuklin, I. Markov, and M. de Rijke. Click models for web search. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 7(3):1–115, 2015.
- [3] M. J. Cole, X. Zhang, C. Liu, N. J. Belkin, and J. Gwizdka. 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*, pages 1219–1220. ACM, 2011.
- [4] W.-T. Fu. From plato to the world wide web: Information foraging on the internet. In M. T. Peter, T. H. Thomas, and W. R. Trevor, editors, *Cognitive Search*, pages 283–299. MIT Press, 2013.
- [5] W.-T. Fu and W. Dong. Collaborative indexing and knowledge exploration: A social learning model. *IEEE Intelligent Systems*, (1):39–46, 2010.
- [6] A. Hliaoutakis. 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, 2005.
- [7] B. Hu, Y. Zhang, W. Chen, G. Wang, and Q. Yang. Characterizing search intent diversity into click models. In *Proceedings of the 20th International Conference on World Wide Web*, pages 17–26. ACM, 2011.
- [8] G. Jorge-Botana, J. A. Leon, R. Olmos, and I. Escudero. Latent semantic analysis parameters for essay evaluation using small-scale corpora. *Journal of Quantitative Linguistics*, 17(1):1–29, 2010.
- [9] I. Juvina and H. van Oostendorp. Modeling semantic and structural knowledge in web navigation. *Discourse Processes*, 45(4-5):346–364, 2008.
- [10] S. Karanam, H. van Oostendorp, and W. T. Fu. Performance of computational cognitive models of web-navigation on real websites. *Journal of Information Science*, 42(1):94–113, 2016.
- [11] S. Karanam, H. van Oostendorp, and B. Indurkha. Towards a fully computational model of web-navigation. In *Modern Approaches in Applied Intelligence*, pages 327–337. Springer, 2011.
- [12] S. Karanam, H. van Oostendorp, and B. Indurkha. Evaluating colides+ pic: the role of relevance of pictures in user navigation behaviour. *Behaviour & Information Technology*, 31(1):31–40, 2012.
- [13] S. Karanam, H. van Oostendorp, M. Sanchiz, A. Chevalier, J. Chin, and W. T. Fu. Modeling and predicting information search behavior. In *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*, page 7. ACM, 2015.
- [14] W. Kintsch. *Comprehension: A paradigm for cognition*. Cambridge university press, 1998.
- [15] M. Kitajima, M. H. Blackmon, and P. G. Polson. A comprehension-based model of web navigation and its application to web usability analysis. *People and Computers*, pages 357–374, 2000.
- [16] T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch. *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum, 2007.
- [17] S. Monchaux, F. Amadiou, A. Chevalier, and C. Mariné. Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing & Management*, 51(5):557–569, 2015.
- [18] R. Olmos, G. Jorge-Botana, J. A. León, and I. Escudero. Transforming selected concepts into dimensions in latent semantic analysis. *Discourse Processes*, 51(5-6):494–510, 2014.
- [19] C. Olston and E. H. Chi. Scenttrails: Integrating browsing and searching on the web. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 10(3):177–197, 2003.
- [20] T. Pedersen, S. V. Pakhomov, S. Patwardhan, and C. G. Chute. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40(3):288–299, 2007.
- [21] P. Pirolli and S. Card. Information foraging. *Psychological review*, 106(4):643, 1999.
- [22] S. Shen, B. Hu, W. Chen, and Q. Yang. Personalized click model through collaborative filtering. In *Proceedings of the fifth ACM International Conference on Web Search and Data Mining*, pages 323–332. ACM, 2012.
- [23] H. van Oostendorp and I. Juvina. Using a cognitive model to generate web navigation support. *International Journal of Human-Computer Studies*, 65(10):887–897, 2007.
- [24] Q. Xing, Y. Liu, J.-Y. Nie, M. Zhang, S. Ma, and K. Zhang. Incorporating user preferences into click models. In *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management*, pages 1301–1310. ACM, 2013.