Cognitive Modeling of Age-Related Differences in Information Search Behavior

Saraschandra Karanam
Department of Information and Computing Sciences, Princetonplein 5, De Uithof, Utrecht, 3584 CC, The Netherlands. E-mail: s.karanam@uu.nl

Herre van Oostendorp
Department of Information and Computing Sciences, Princetonplein 5, De Uithof, Utrecht, 3584 CC, The Netherlands. E-mail: h.vanoostendorp@uu.nl

Mylene Sanchiz
Cognition, Language, Ergonomics Lab, University of Toulouse, France. E-mail: mylene.sanchiz@univ-tlse2.fr

Aline Chevalier
Cognition, Language, Ergonomics Lab, University of Toulouse, France. E-mail: aline.chevalier@univ-tlse2.fr

Jessie Chin
Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada. E-mail: chin5@illinois.edu

Wai-Tat Fu
Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. E-mail: wfu@illinois.edu

In this study, we evaluated the ability of computational cognitive models of web-navigation like CoLiDeS and CoLiDeS+ to model i) user interactions with search engines and ii) individual differences in search behavior due to variations in cognitive factors such as aging. CoLiDeS and CoLiDeS+ were extended to predict user clicks on search engine result pages. Their performance was evaluated using actual behavioral data from an experiment in which 2 types of information search tasks (simple vs. difficult), were presented to younger and older participants. The results showed that the model predictions matched significantly better with the actual user behavior on difficult tasks compared to simple tasks and with younger participants compared to older participants, especially for difficult tasks. Also, the matches were significantly better with CoLiDeS+ compared to CoLiDeS, especially for difficult tasks. We conclude that the advanced capabilities of CoLiDeS+, such as incorporating contextual information and implementing backtracking strategies enable it to predict user behavior significantly better than CoLiDeS, especially on difficult tasks. The usefulness of these modeling outcomes for the design of support systems for older adults is discussed.

Received June 15, 2016; revised March 6, 2017; accepted March 31, 2017

 • Published online 18 July 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23893

Introduction

A typical information-seeking process on the Internet usually begins in the context of a work task (Ingwerson & Järvelin, 2006) and is inherently interactive. It also involves cognitive processes such as memory, attention, problem solving, comprehension, and decision making (Sharit, Hernández, Czaja, & Pirolli, 2008), each of which is in turn affected by cognitive factors such as age, domain knowledge, internet experience, etc. This is also the reason for the wide variation observed in the efficiency of using the internet among users.

However, the dominant methodology of evaluating information retrieval (IR) systems that can be traced back to the Cranfield experiments (Cleverdon, 1991), has been to compute the efficiency of retrieving information that is most relevant to a single query, using standardized test collections and relevance assessments. It has long been recognized that developing new representations of documents, indexing, ranking, and retrieval algorithms leads to only a small incremental improvement in performance over previous models (Jones, 2005). It is also known that such incremental improvements, as evaluated by Cranfield-style TREC evaluation, rarely leads to significant improvement in a user's information search performance (Turpin & Scholer, 2006). What is needed instead is the "development of models of IR which incorporate the user as an active participant in the IR system, and which treat the person's interaction with information as a central process of IR" (Belkin, 2008, p. 53). Therefore, there has been renewed interest in incorporating user characteristics into the evaluation of information retrieval systems (Kelly & Sugimoto, 2013; Hagen, Michel, & Stein, 2016). However, the interaction between a user and a search system is non-trivial and affected by many factors (Ingwerson & Järvelin, 2006). Using laboratory studies to understand these relationships requires participation from real users, who are not always available. Also, as the number of factors in an experiment increases, the complexity of the experiment increases and the number of experiments required to investigate all possible relationships also increases dramatically. This is not only expensive but also time-consuming and difficult to scale up. Models, on the other hand, allow us to simulate user behavior without performing the experiment, which then only serves as a verification of the model (Azzopardi, Järvelin, Kamps, & Smucker, 2011).

Numerous models can be found in the field of information seeking and retrieval (ISR) that seek to understand, predict, and explain interactions of users with search systems. These models can be classified mainly into two types: the first type, which illustrates the search process, includes the ASK model (Belkin, Oddy, & Brooks, 1982), the Information Search Process (Kuhlthau, Heinström, & Todd, 2008), Ellis's model of information-seeking behaviors (1989), and the Berry-Picking Model (Bates, 1989); and the second type that describes the factors that influence the search process include the task-based information retrieval process (Vakkari, 2001) and the ISR framework proposed by Ingwerson and Järvelin (2006). Both types of models are conceptual and descriptive in nature, providing an overview of the information-seeking process and the factors that are likely to have an influence on it. They are useful in understanding the different components of an information search process and the various stages that a user may go through. Most importantly, they are useful for the generation of new hypotheses and the development of more formal models. Formal models allow us to simulate user behavior, thereby reducing the need, effort, and cost involved in running actual experiments.

In this context, a promising development in the late 1990s was the introduction of Information Foraging Theory (Pirolli & Card, 1999), which provides a formal, mathematical and predictive mechanism underlying the berry-picker's actions. It postulates that when browsing on the web, users take only those actions (such as clicking on a hyperlink) that maximize their information gain in relation to the cost of taking that action. The theory also introduces the concept of *information scent* which is the (imperfect) estimate of the value or cost of information sources represented by proximal cues (such as hyperlinks and icons). Motivated by this theory, a number of computational cognitive models of web-navigation have been developed (Kitajima, Blackmon, & Polson, 2000; Juvina & Van Oostendorp, 2008; Fu & Pirolli, 2007) that successfully predict user clicks when browsing on a website (Blackmon, Mandalia, Polson, & Kitajima, 2007). Based on theories of cognitive science, the main goal of these computational cognitive models is to use well-tested cognitive mechanisms to characterize more complex information search behavior in a precise and automated way. These computational cognitive models of web-navigation take a *process* view of information search and are therefore more capable of providing opportunities to incorporate behavioral differences due to the variations in cognitive factors. We focus, on two important gaps in the research on computational cognitive models of web-navigation in this paper.

First, computational cognitive models of web-navigation have so far been used to characterize the cognitive processes underlying navigation behavior within websites and have not been used to characterize the cognitive processes underlying user interactions with search engine result pages. Interaction with a search engine involves various steps: formulation of a query, selection of one or more search results from the search engine result page (SERP) corresponding to the query, switching between websites opened from the search results and the decision to reformulate the query. We focus primarily on the selection of search results step in this paper. In other words, the focus of this paper is on predicting which search results a user would click, for a given query and its corresponding SERP. We made first steps by extending a cognitive model of web-navigation called CoLiDeS (Kitajima et al., 2000) in this direction (Karanam et al., 2015). In the current paper, we make more progress by applying an enhanced version of CoLiDeS called CoLiDeS+ (Juvina & Van Oostendorp, 2008) and compare the accuracy of the predictions of both models.

Second, the capabilities of computational cognitive models of web-navigation in terms of modeling behavioral differences due to the variations in cognitive factors have not been fully evaluated. It is important not only to develop models of IR that treat a person's interaction with information as a central process of IR, but it is also important for such models to have the capability to explain variations in information-seeking behavior due to individual differences in cognitive factors. In this way, we can make predictions of the model more precise. Of all the cognitive factors, we focus on *age* for various reasons. The percentage of population over 65 years is increasing, especially in the OECD countries. Also, the percentage of those over 65 years who use the Internet is rapidly increasing. Although the Internet

DOI: 10.1002/asi

is growing rapidly and transforming every aspect of our lives, several barriers still exist for older adults to really make it an irreplaceable part of their daily lives, as it normally is for younger adults (Slegers, van Boxtel, & Jolles, 2009). Aging is related to a decline in fluid intelligence (such as processing speed, cognitive flexibility, attentional control, and visuospatial span) as well as in motor skills (Horn, 2012; Wang & Kaufman, 1993). Some of these abilities are directly linked to the skills required to search and process information from the Internet. However crystallized intelligence increases and/or becomes stable with aging. Crystallized intelligence involves prior knowledge, experience, and vocabulary skills. Therefore, it is generally higher for older adults compared to younger adults. In our earlier work (Karanam et al., 2015), we extended CoLiDeS to predict user clicks on search results. Actual user clicks from a behavioral experiment involving eight young and nine old participants were matched with the predictions by CoLiDeS. We could not find any significant age-related difference in the matches by CoLiDeS ($p > .05$). In this paper, we investigate if the more enhanced model CoLiDeS+ would model age-related differences better than CoLiDeS.

Cognitive Models of Web-Navigation

In this section, a detailed description of the two computational cognitive models, CoLiDeS and CoLiDeS+, is given. Next to each model's description, the methodology to implement and run the model on SERPs is also presented.

CoLiDeS

CoLiDeS, or Comprehension-based Linked Model of Deliberate Search (Kitajima et al., 2000), assumes that information seeking and navigation is driven by text-comprehension and problem-solving processes. CoLiDeS is based on the Construction-Integration (CI) model of text-comprehension (Kintsch, 1998). According to the CI model, comprehension of text happens in two phases: *construction phase*, during which a propositional representation of a new piece of text is built and all possible meanings of the text elements are generated, and *integration phase*, during which using reader's prior knowledge and context, a single coherent meaning is selected by filtering out all those representations that do not fit well to the context. A further assumption is that the processing of information occurs in a cyclic way. After processing a current sentence or part of it, information is represented in episodic memory and space is made free in working memory for processing new information (the next sentence, for instance), and so on. CoLiDeS assumes that the processes underlying navigation on the web are analogous to the processes underlying text-comprehension. CoLiDeS divides user navigation behavior into two main stages of cognitive processing*: attention cycle* and *action-selection cycle*. The attention cycle is further divided into two stages: *parsing* the webpage in high-level schematic regions, and *focusing* on one of these schematic regions. The action-selection cycle is also divided into two stages: *comprehension* of screen objects (e.g., hyperlink texts) within the focused region, and *selecting* the most appropriate screen object (such as clicking on a specific hyperlink). This process is then repeated for a new webpage that is opened by the selected hyperlink until you reach the target page. The focus of the modeling is on the navigation process based on the selection and evaluation of links (the fourth phase in the CoLiDeS model).

Notice the similarities of CoLiDeS model with the CI model: first, the process of activation of knowledge triggered by the hyperlink texts (construction phase) and the maintenance of the representation of labels that have high similarity to the goal (integration phase). Also bottom-up and top-down processes play a role just like in text-comprehension (Kintsch, 2005). Next to that common is the cyclic character of processing: construction and integration phases for the second page and third page, and so on.

In CoLiDeS, information scent is operationalized as the semantic similarity between the user goal and each of the hyperlinks. Based on the semantic similarity values, the model predicts that the user is most likely to click on that hyperlink that has the highest semantic similarity value. This process is repeated for every new page until the user reaches the target page. CoLiDeS uses Latent Semantic Analysis (LSA) (Landauer, McNamara, Dennis, & Kintsch, 2007) to compute semantic similarities. LSA is an unsupervised machine-learning technique that builds a high dimensional semantic space using a large corpus of documents that represent a given user population's knowledge and understanding of words. The meaning of a word or sentence is represented as a vector in that high dimensional space. The degree of similarity between a link and the goal of the reader is measured by the cosine value between the corresponding vectors (Martin & Berry, 2007) in the high dimensional space. Each cosine value lies between $+1$ and -1. The closer the value to $+1$ is, the higher the similarity between the two words is.

Simulation of CoLiDeS on SERPs

The methodology to run simulations of CoLiDeS on search engine result pages has been described in detail in Karanam et al. (2015). In brief, semantic similarity is computed between the query and the title and the snippet combinations of the search results on a SERP using LSA. The search result with the highest LSA value is selected. This process is repeated for all the queries of a task and for all the tasks of a participant and finally for all the participants. We will have then available the predictions made by CoLiDeS on the SERPs corresponding to all the queries of all the tasks and we can compare these with the actual selections of real participants. Note that the CoLiDeS model can predict only one search result per query using this methodology, whereas users in reality click on more than one search result per query.

CoLiDeS+

CoLiDeS+ (Juvina & Van Oostendorp, 2008) shares the main theoretical foundations of CoLiDeS and makes it more

DOI: 10.1002/asi

consistent with theoretical assumptions from work on text-comprehension that emphasizes the role of context. For example, when reading a text, contextual information helps users in comprehending sentences better, especially sentences with potentially multiple interpretations (Budiu & Anderson, 2004). Analogously, when navigating a website, users often encounter hyperlink texts varying in their degree of ambiguity. CoLiDeS+ assumes that surrounding context helps a user in building a mental representation of the information space being navigated. This representation in turn assists users in locating relevant information such as hyperlinks or target pages as well as disambiguating ambiguous hyperlink text to grasp the intended meaning. Furthermore, it is very common for users to backtrack and take an alternate route to find their target information whenever they arrive at a webpage that has no relevant hyperlinks (Cockburn & McKenzie, 2001). CoLiDeS+ incorporates this behavior by introducing backtracking strategies when the semantic similarity between the query and the goal does not increase.

According to the CI-model, context in text-comprehension is extracted from the following sources: previously read words/sentences (but still available in working memory), text read in the preceding reading session, and finally background knowledge, in that order. Similarly, CoLiDeS+ assumes that users make decisions to click or not to click, not only on the basis of information that is new, that is, an incoming new hyperlink text, but also on the basis of information that has been accessed before, that is, hyperlinks elaborated or clicked in the preceding session. This contextual information can help in reinforcing the activation of the appropriate semantic features or concepts, thereby steering the selection of the right links. CoLiDeS, on the other hand, considers only incoming hyperlink text to assess relevancy with respect to the goal and does not incorporate any contextual information. CoLiDeS+ incorporates contextual information by retaining in memory the selected links to compute the navigation path and *path adequacy* in addition to information scent. The navigation path is the sequence of hyperlinks clicked by a user at any particular moment and path adequacy is defined as the semantic similarity between the user goal and the navigation path realized until a particular moment. So, CoLiDeS+ computes the LSA value between the navigation path and the user goal.

Semantic similarity is here *cosine (NP, Q)*, where NP is the vector of the navigation path and Q is the vector of the query of the user. Both are vectors in an existing semantic space. The semantic similarity is computed between these two vectors. A Dutch semantic space was built using a corpus of 70,000 Dutch articles (60% news and 40% medical and health) and constructed using Gallito (Jorge-Botana, Olmos, & Barroso, 2013). More details of the steps involved in the construction of a semantic space can be found in Karanam and Van Oostendorp (2016a).

Only if the information from an incoming hyperlink increases in information scent is it considered for selection. If it does not increase in information scent, path adequacy is checked. If path adequacy increases, then the incoming hyperlink is selected, although it does not increase in information scent. In other words, first semantic similarity is evaluated based on information scent, and only when it is not satisfying, a more effortful evaluation of the context is performed by checking path adequacy.

If path adequacy does not increase, a latent impasse is said to have occurred and CoLiDeS+ invokes *backtracking strategies*: that is, backtracking to other regions within the same page and eventually to the previously visited pages. Therefore, on a webpage, CoLiDeS+ gives the highest activation or priority to local information or the incoming new hyperlink text followed by contextual information such as previously clicked hyperlinks (these were hyperlinks that the user thought could be relevant). CoLiDeS+ stops when the user declares the current page is the page with the target information. CoLiDeS, on the other hand, stops and declares an impasse when a page has no relevant hyperlinks. Solutions to impasses are only described and not computationally modeled. CoLiDeS+ overcomes the limitations of CoLiDeS by including contextual information and modeling backtracking strategies and solving impasses. Note that the behavior of CoLiDeS+ and CoLiDeS will be exactly the same if information scent continues to increase on every page.

When both CoLiDeS and CoLiDeS+ were tested on real navigation in websites, CoLiDeS+ was found to not only locate the target page more often than CoLiDeS, but was also found to reach closer to the target page than CoLiDeS whenever the target page was not located (Karanam, Van Oostendorp, & Fu, 2016). Also, by comparing selections made by users with the selections that CoLiDeS+ would have made for a set of tasks on a mockup website, Juvina and Van Oostendorp (2008) found that CoLiDeS+ was able to predict 54.9% of actual user clicks, slightly better than CoLiDeS, which could predict 46.9% of actual user clicks. These results support the assumptions made by CoLiDeS+ that users make use of context from previously selected hyperlinks to choose the next hyperlink. Please note that CoLiDeS+ so far has *not* been used to model any form of interaction with a search engine.

Simulation of CoLiDeS+ on SERPs

It is possible to model and predict *multiple* user clicks per query using CoLiDeS+ because of its capability to implement backtracking strategies. We assume that a user would explore more than one search result per query only if s/he is not satisfied with the information from the search result that was chosen first. Therefore, the underlying assumption is that a user clicks on a search result, explores the website corresponding to it, and backtracks to the search result page to explore other search results. Starting from the second search result, CoLiDeS+ keeps track of the contextual and historical information by computing path adequacy. To keep the simulation simple, we model only the user clicks on a SERP and ignore the clicks within the websites opened from the

SERP. For each query and the corresponding SERPs logged in the behavioral data, seven steps are followed:

a. Compute the semantic similarity between the query and the title and the snippet combination of a search result.
b. Repeat this for all the remaining titles and snippets on a SERP. The title and snippet combination with the highest semantic similarity value with the query is selected.
c. Assuming that the information was not found in the search result selected in (b), the search result with the next highest semantic similarity is chosen.
d. Starting from the selection in (c), path adequacy is computed as follows: the semantic similarity is computed between the query and the title and snippet combinations of all the search results selected by the model so far.
e. If path adequacy increases, the next best search result from (c) is selected.
f. If path adequacy does not increase, steps (c–d) are repeated, until a search result that increases path adequacy is found or all the search results are exhausted. If all search results are exhausted, the model decides to reformulate.
g. Finally, repeat this process for all the queries of a task and for all the tasks of a participant and finally for all the participants.

After running the main simulation steps (a–g), we have available the predictions made by CoLiDeS+ on all the queries of all the tasks and we can compare these with the actual selections of real participants. In the next section, we look at some of the experimental literature on the influence of age on information search performance.

Effect of Aging on Information Search

Much behavioral research has already been conducted on how age-related decline in cognitive abilities influence performance on an information search task. Older adults are known to generate fewer queries, use fewer keywords per query, reformulate less, spend longer time evaluating the search results, spend more time evaluating the content of websites opened from SERPs, switch less often between SERPs and websites, and find it difficult to reformulate unsuccessful queries (Dommes, Chevalier, & Lia, 2011; Pak & Price, 2008; Queen, Hess, Ennis, Dowd, & Grühn, 2012). However, under certain conditions, older adults can adapt their search and navigation strategies exploiting their higher crystallized knowledge to compensate for their decreased fluid capabilities (e.g., Chin, Anderson, Chin, & Fu, 2015).

For example, a recent study by Chevalier, Dommes, and Marquié (2015) looked at the three important phases in an information search process: planning (dividing a problem into subproblems, formulating the first query), evaluating (assessing the relevance of search results and information at hand in general), and controlling (modifying the query and search strategy if required). They found significant differences in the amount of time allocated to the three phases of information search in relation to age and task difficulty. Younger adults were found to control their strategy more than older adults, enabling them to perform better especially at difficult and impossible tasks. In contrast, older adults spent a lot of time to evaluate results and information at hand instead of modifying their unsuccessful strategies.

In another study (Chin et al., 2015), it was found that there are significant age differences in the amount of resources allocated to exploration (number of search results opened for any given query) phase and exploitation (number of websites and hyperlinks within the websites opened from search results for any given query) phase of an information search process. Older adults were found to do less exploration and more exploitation in terms of spending longer time and viewing more information compared to younger adults. They even found that the older adults were adaptive in adjusting the two processes depending on the difficulty of the task.

Research Questions and Hypotheses

The research questions of our study were:

1. Would the more advanced model CoLiDeS+ predict user clicks on a SERP better than CoLiDeS (**RQ1**)? We hypothesize that, because of its enhanced capabilities, such as incorporating contextual information and implementing backtracking strategies, the modeling performance of CoLiDeS+ would be better than that of CoLiDeS (**Hyp1**). If so, search behavior of users would be matched better by CoLiDeS+ compared to CoLiDeS and would lend more evidence to the assumptions made by CoLiDeS+.
2. Following up on **RQ1**, assuming that the modeling performance of CoLiDeS+ is better than that of CoLiDeS, would this enhanced performance be more prominent for difficult tasks (**RQ2**)? Difficult tasks demand users to explore multiple search results and switch often between SERPs and websites. We hypothesize that, because of the enhanced capabilities of CoLiDeS+, it would predict user behavior better than CoLiDeS, especially under difficult task conditions. Under easy task conditions, both models would perform equally (**Hyp2**).
3. Previous literature has shown an interaction effect between age and task difficulty on actual search performance of users (Karanam et al., 2015; Chevalier et al., 2015). However, in our recent study (Karanam et al., 2015), we did not find any such interaction effects of age and task difficulty on the number of matches between actual user clicks and predictions of CoLiDeS. In this paper we reexamine this issue using behavioral data from a more elaborate experiment (**RQ3**). For difficult tasks, usually the answer is not easily found in the snippets of the search engine results and often users have to evaluate information from multiple search results and websites. Younger adults are known to click more, reformulate more, and switch more often between SERPs and websites than older adults when performing difficult tasks, which enables them to perform better on difficult tasks (Chevalier et al., 2015; Karanam et al., 2015). We therefore expect a similar age X task difficulty interaction in the model predictions, that is, we expect that the

DOI: 10.1002/asi

models would match the search behavior of younger adults much better than that of the older adults, especially for difficult tasks (**Hyp3**).

The next section briefly describes the details of an experiment (which is fully reported in Karanam & Van Oostendorp, 2016b) conducted to collect actual user data and provides a brief summary of the behavioral outcomes to check if they are in line with earlier outcomes from aging-related literature. Next, the main issues of the study, that is, modeling user clicks on SERPs and modeling age-related differences in user clicks on SERPs are reported.

Experiment

Method

Participants. Twenty-four young participants (16 males and 8 females) ranging from 18 to 31 years ($M = 22.7$, $SD = 3.31$), and 24 older participants (14 males and 10 females) ranging from 65 to 88 years ($M = 73.58$, $SD = 6.74$) participated in the study.

Material

The experiment was conducted with 12 simulated information search tasks (Borlund & 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 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. The tasks were all presented in Dutch in a counterbalanced order.

Procedure

Participants first did 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 three tests: a computerized version of a Dutch vocabulary test, adapted from the Hill Mill Vocabulary (HMV) test (Raven & Court, 1998), and a fluid intelligence test: a computerized version of the Trail Making Test (TMT Part B) (Strauss, Sherman, & Spreen, 2006). The score on the vocabulary test (24 items) gives us an indication of the amount of crystallized intelligence and the score on the trail making test gives us an indication of the amount of fluid intelligence. We measured the time taken to finish the test correctly. They were next presented with a prior domain knowledge test on the topic of health with 12 multiple-choice questions. A correct answer was scored 1 and a wrong answer was scored 0. They were then presented with 12 information search tasks (six simple and six difficult) in a counterbalanced and randomized order. Participants were first shown the task and then directed to the home page of Google's search engine and were not allowed to use any other search engine.

Measures on Information Search

Task-completion time. Task-completion time was computed from the moment of opening a browser and typing in the first query to the moment of answering the question.

Number of clicks. The number of clicks is the total number of clicks made by a participant for each task. This includes the clicks made on the search results as well as the clicks made on websites opened from the search results.

Accuracy. Accuracy was measured as 1, 0.5, or 0 depending on whether the participant's answer was correct (in which case the score is 1) or partially correct (in which case the score is 0.5) or wrong (in which case the score is 0).

Number of reformulations. Number of reformulations is the total number of unique queries that a user could come up with for each task in the process of answering it (e.g., if participant added, deleted keywords, or created new ones, we counted them as reformulations of query).

Results

Briefly, the results show that the difficult tasks took significantly longer to complete, significantly more clicks were made, were answered significantly less accurately than simple tasks, and significantly more reformulations were made than simple tasks. With regard to age-related differences, we found significant main effects and interaction effects between the search performance of young and old participants in this experiment. Younger adults were found to be significantly faster in completing the tasks, especially the simple ones. Younger adults clicked on significantly more search results than older adults, especially when solving difficult tasks. Younger adults were significantly more accurate than older adults. Younger adults reformulated significantly more than older adults, especially when performing difficult tasks. In a more general way, older adults were less efficient than younger adults, one of the reasons could be their lower fluid abilities. These results were largely in line with the outcomes of previous studies (Chevalier et al., 2015; Chin et al., 2015; Dommes et al., 2011; Pak & Price, 2008; Queen et al., 2012) and therefore provided a strong foundation to the main contributions of this study.

Modeling

The queries from the behavioral experiment were used to run simulations of CoLiDeS and CoLiDeS+. We evaluated the performance of the models by computing the number of matches between the model predictions and the actual user clicks for each query and its corresponding SERP.

DOI: 10.1002/asi

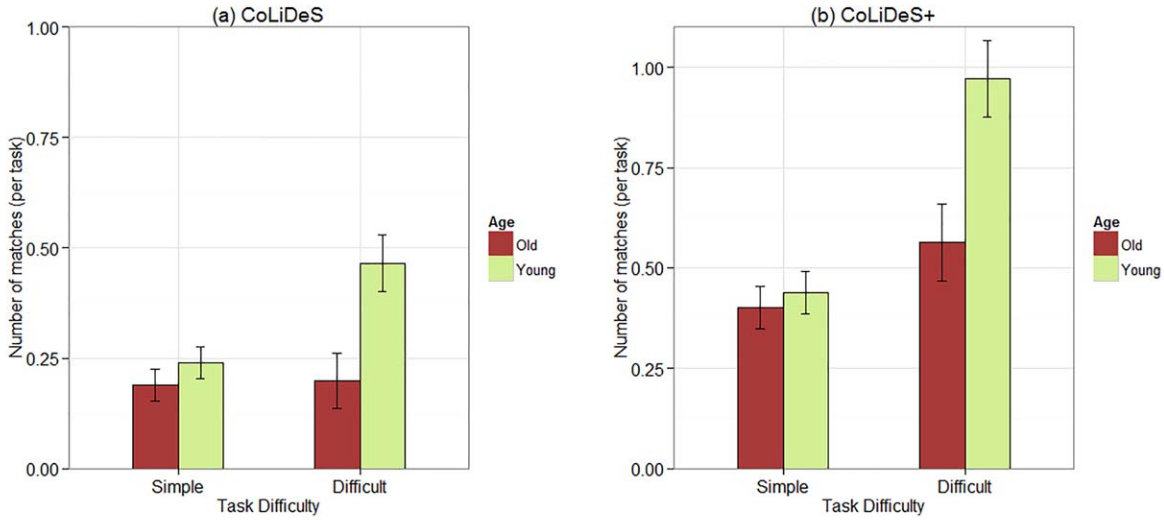

FIG. 1. Mean number of matches (and standard errors) with (a) CoLiDes and (b) CoLiDes+ in relation to age and task difficulty. [Color figure can be viewed at wileyonlinelibrary.com]

Evaluation of Model Performance

A repeated-measures analysis of variance (ANOVA) with age as the between-subjects variable, task difficulty and model as within-subjects variables, and mean number of matches per task as dependent variable was conducted. As shown in Figure 1, the main effect of task difficulty was significant, $F(1,46) = 16.66$, $p < .001$. The mean number of matches was significantly higher for difficult tasks as compared to simple tasks. The main effect of model was significant, $F(1,46) = 89.64$, $p < .001$. The mean number of matches was significantly higher for CoLiDeS+ as compared to CoLiDeS. The main effect of age was also significant, $F(1,46) = 12.42$, $p < .001$. The mean number of matches was significantly higher for young participants compared to old participants. The interaction of task difficulty and age was significant, $F(1,46) = 6.72$, $p < .05$. The mean number of matches was significantly higher for young participants compared to old participants when they performed difficult tasks. The interaction of model and task difficulty was significant, $F(1,46) = 10.43$, $p < .005$. The number of matches was significantly higher with CoLiDeS+, especially for difficult tasks. The interaction of model and age was not significant ($p > .05$). The interaction of model, task difficulty, and age was also not significant ($p > .05$).

Summarizing the outcomes of modeling, the model predictions matched significantly better with the actual user behavior on difficult tasks compared to simple tasks. Also, the model predictions matched actual behavior of young participants significantly better compared to the actual behavior of old participants, especially for difficult tasks. We also found that CoLiDeS+ matched actual user behavior significantly better than CoLiDeS, especially for difficult tasks. The age-related differences (both in the behavioral outcomes and in the modeling outcomes), which we could not find in our earlier work (Karanam et al., 2015), were found in the current experiment, probably because it is based on more participants and more tasks as compared to our earlier experiment and therefore has more statistical power.

Conclusions and Discussion

We focused on two important gaps in the research on computational cognitive models of web-navigation in this study. First, computational cognitive models of web-navigation have not been used to characterize the cognitive processes underlying user interactions with search engine result pages. Second, the capabilities of computational cognitive models of web-navigation in terms of modeling behavioral differences due to the variations in cognitive factors such as aging have not been evaluated. Our earlier work (Karanam et al., 2015) could not find significant age-related differences in user clicks on SERPs, probably due to the low number of participants (eight young and nine older participants). Also, the earlier study involved only one computational cognitive model: CoLiDeS. In this study, we extended two computational cognitive models of web-navigation, CoLiDeS and CoLiDeS+, to predict which search result a user would click given a query and its corresponding SERP. We compared the ability of the CoLiDeS and CoLiDeS+ models to predict age differences in user clicks on search engine result pages.

A behavioral experiment with older and younger adults was conducted using simple and difficult tasks (Karanam & Van Oostendorp, 2016b). Analysis of the search performance in terms of task-completion time, clicks, accuracy, and number of reformulations showed that the results obtained were largely in line with the outcomes of previous studies on aging effects (Chevalier et al., 2015; Chin et al., 2015; Dommes et al., 2011; Pak & Price, 2008; Queen et al., 2012). Real user queries from this behavioral experiment were used subsequently to run simulations of CoLiDeS and CoLiDeS+. Analysis of model performance showed that CoLiDeS+ matched actual user behavior significantly better than

DOI: 10.1002/asi

CoLiDeS (**Hyp1**). This confirmed our first hypothesis. The main reason why CoLiDeS+ seems to perform much better than CoLiDeS is because of the fact that CoLiDeS+ gives importance not only to the local cue, that is, the incoming new search results, but also to the global context, that is, the query and the search results already clicked in the preceding session. Also, CoLiDeS+ is able to go back, if necessary, to already visited pages, change route, and explore a new path to find the target page (see Karanam et al. (2016) for a visual illustration). In other words, the assumption that CoLiDeS+ made about users basing their decisions to select a particular hyperlink/search result not only on semantic relevance with the goal but also on whether an incoming hyperlink/search result is consistent with the hyperlinks/search results selected in the past or not, seems to be true. CoLiDeS, on the other hand, always focuses on current information and does not utilize any historical information. It is only capable of linear forward search.

We also found that CoLiDeS+ matched actual user behavior significantly better than CoLiDeS, especially on difficult tasks (**Hyp2**). This interaction between model and task complexity can be explained by the fact that CoLiDeS+ is more capable of simulating behavior demanded by difficult tasks (more queries, more clicks, and more switches between SERPs and websites, etc.) and therefore it is able to predict user behavior, particularly on difficult tasks, better than CoLiDeS.

Lastly, we also found evidence for our third hypothesis. The model predictions matched significantly better with the actual behavior of young participants compared to old participants, especially for difficult tasks (**Hyp3**). This interaction between age and task difficulty is probably because difficult tasks by definition require integration of information from multiple sources, which in turn requires more queries, more clicks on the SERPs generated by the queries, more switches between SERPs and websites, and overall more detours. Younger adults, owing to their higher fluid capabilities and lower switching costs, are more capable of performing all the above activities better than older adults.

Limitations

One of the main limitations of this study is that it did not explore the possibility of factors other than age such as educational background or crystallized intelligence (Pak & Price, 2008), past experience with the Internet (Crabb & Hanson, 2016); fluid ability (Crabb & Hanson, 2016; Dommes et al., 2011) could be the real reasons for the observed age-related behavioral differences in the experiment. We now discuss briefly each factor in relation to our study. We can rule out the possibility of at least one factor, that is, educational background or crystallized intelligence, as there was no significant difference between the scores of younger and older adults on both the prior domain knowledge test and the Hill Mill Vocabulary test. However, self-reported ratings indicate that older adults ($M = 2.63$, $SD = 1.01$) are significantly less familiar with search engines than younger adults ($M = 3.75$, $SD = 0.68$) $t(46) = 4.52$, $p < .001$. Also, older adults took significantly longer to finish the fluid intelligence test ($t(46) = 5.13$, $p < .001$) than younger adults, indicating that they had significantly lower fluid abilities compared to younger adults. Two separate analyses, first with the self-reported ratings on familiarity with search engine as a covariate and, second, with the scores on fluid intelligence test as a covariate were conducted. In both analyses, the main effect of age was still found to be significant ($p < .05$) for three out of four dependent variables (clicks, accuracy, and reformulations), indicating that prior experience with search engines and fluid ability were not important factors. However, we agree completely that not age per se, but the underlying and correlated cognitive and motor abilities are responsible for the effects found (Trewin et al., 2012).

A second limitation concerns assumptions in our modeling. Our aim in this work was to investigate to what extent the CoLiDeS and CoLiDeS+ models, without making any extra modeling changes, are able to simulate user interactions with search engines. We also made certain assumptions to simplify modeling. We modeled only one step involved in interacting with a search engine, that is, selecting a search result, given a query and its corresponding SERP. This step is very similar to selecting a hyperlink, given a set of hyperlinks and user goal. Therefore, the cognitive processes underlying navigation within a website and the cognitive processes underlying clicking on a search result are similar. That is, in the context of web-navigation, first the process of activation of knowledge is triggered by the hyperlink texts (construction phase) and then the representation of labels that have high similarity to the goals are maintained (integration phase). Similarly, in the context of interacting with a search engine, first the process of activation of knowledge is triggered by the title and snippet combinations of the individual search results (construction phase) and then the representation of labels that have high similarity to the goals are maintained (integration phase). Also bottom-up and top-down processes play a role just like in text-comprehension (Kintsch, 2005) both in the context of web-navigation and in the context of a search engine. Next to that, common is the cyclic character of processing: construction and integration phases for second page and third page, and so on in the context of web-navigation and 2nd query and 3rd query and so on in the context of a search engine. In the context of web-navigation, the hyperlinks clicked in the past influence the selection of an incoming new hyperlink. Similarly, in the context of a search engine, the search results clicked before influence the selection of an incoming new search result. However, interacting with a search result involves many other activities, such as generation of a query, reformulation of a query, for which there is no equivalent in the web-navigation context. These steps require further experimentation and theoretical development.

Future Directions

In the current study, our aim was to investigate to what extent the CoLiDeS and CoLiDeS+ models, without

DOI: 10.1002/asi

making any modeling changes, are able to simulate age-related differences in user interactions with search engines. We present here some preliminary ideas to modify two different parameter values in the CoLiDeS+ model that can directly reflect age differences. The first idea concerns the number of search results that are explored. In the current study we allowed the model to explore all 10 search results of a SERP before deciding to reformulate. In reality, it is known that not all search results are evaluated by users before reformulating. Moreover, the study by Chin et al. (2015) found that older adults do more exploitation (number of websites and hyperlinks within the websites opened from search results for any given query) and less exploration (number of search results opened for any given query) compared to younger adults. In order to simulate this age-related difference in the amount of exploration performed by younger and older adults, one can imagine a parameter that controls how many search results are explored. Setting a higher value to this parameter will simulate the behavior of younger adults more closely (more clicks, more switches between SERPs and websites) and setting a lower value to this parameter will simulate the behavior of older adults more closely (fewer clicks and fewer switches between SERPs and websites). The second idea concerns the needed LSA similarity value of the search results with the query, which gives an estimate of the relevancy of the search result with respect to the query. By setting a minimum threshold value of this parameter in the model, one can simulate whether or not a search result is clicked. By varying this threshold value, one can simulate different user behaviors.

Implications for System Design

Besides their theoretical value, computational cognitive models also have a practical value. They have been successfully used to generate automatic and online support for navigation within websites (Van Oostendorp & Juvina, 2007; Karanam, Van Oostendorp, & Indurkhya, 2011; Van Oostendorp & Aggarwal, 2015). The current research can lead to applications in developing support tools for interaction with a search engine. In subsequent research one can examine the influence when both types of support are combined.

In summary, we made two main contributions in this paper: first, we extended cognitive models of web-navigation CoLiDeS and CoLiDeS+ to model not only the second phase of information search (navigation within websites) but also the first phase of information search (interaction with a search engine, more precisely the step of selecting a search result from a SERP corresponding to a query). Second, we evaluated the performance of the two models on predicting age-related differences in the above task.

Acknowledgments

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

References

Azzopardi, L., Järvelin, K., Kamps, J., & Smucker, M.D. (2011). Report on the SIGIR 2010 workshop on the simulation of interaction. ACM SIGIR Forum, 44, 35–47.

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13, 407–424.

Belkin, N.J. (2008). Some (what) grand challenges for information retrieval. ACM SIGIR Forum, 42, 47–54.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. Journal of Documentation, 38, 61–71.

Blackmon, M.H., Mandalia, D.R., Polson, P.G., & Kitajima, M. (2007). Automating usability evaluation: Cognitive walkthrough for the web puts LSA to work on real-world HCI design problems. In Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (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, 225–250.

Budiu, R., & Anderson, J.R. (2004). Interpretation-based processing: A unified theory of semantic sentence comprehension. Cognitive Science, 28, 1–44.

Chevalier, A., Dommes, A., & Marquié, J.C. (2015). Strategy and accuracy during information search on the web: Effects of age and complexity of the search questions. Computers in Human Behavior, 53, 305–315.

Chin, J., Anderson, E., Chin, C.L., & Fu, W.T. (2015). Age differences in information search: An exploration-exploitation tradeoff model. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59, 85–89.

Cleverdon, C.W. (1991). The significance of the Cranfield tests on index languages. In A. Bookstein, Y. Chiaramella, G. Salton, & V. Raghavan (Eds.), Proceedings of 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Chicago, October 13–16, 1991. (pp. 3–12). New York: ACM Press.

Cockburn, A., & McKenzie, B. (2001). What do web users do? An empirical analysis of web use. International Journal of Human-Computer Studies, 54, 903–922.

Crabb, M., & Hanson, V.L. (2016). An analysis of age, technology usage, and cognitive characteristics within information retrieval tasks. ACM Transactions on Accessible Computing (TACCESS), 8, 10.

Dommes, A., Chevalier, A., & Lia, S. (2011). The role of cognitive flexibility and vocabulary abilities of younger and older users in searching for information on the web. Applied Cognitive Psychology, 25, 717–726.

Ellis, D. (1989). A behavioural approach to information retrieval system design. Journal of Documentation, 45, 171–212.

Fu, W.T., & Pirolli, P. (2007). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. Human–Computer Interaction, 22, 355–412.

Hagen, M., Michel, M., & Stein, B. (2016). Simulating ideal and average users. In Information Retrieval Technology (pp. 138–154). Cham, Switzerland: Springer.

Horn, J.L. (2012). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in. Aging and Cognitive Processes, 8, 237.

Ingwersen, P., & Järvelin, K. (2006). The turn: Integration of information seeking and retrieval in context (Vol. 18). Berlin: Springer.

Jones, K.S. (2005). Meta-reflections on TREC. In TREC. Cambridge, MA: MIT Press.

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.

Juvina, I., & Van Oostendorp, H. (2008). Modeling semantic and structural knowledge in Web navigation. Discourse Processes, 45, 346–364.

Karanam, S., & Van Oostendorp, H. (2016a). Integrating domain knowledge differences into modeling user clicks on search result pages. In Proceedings of the Workshop on Search as Learning, co-located with ACM SIGIR 2016.

Karanam, S., & Van Oostendorp, H. (2016b). 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., & Fu, W.T. (2016). Performance of computational cognitive models of web-navigation on real websites. Journal of Information Science, 42, 94–113.

Karanam, S., Van Oostendorp, H., & Indurkhya, B. (2011). Towards a fully computational model of web-navigation. In Modern approaches in applied intelligence (pp. 327–337). Berlin: Springer.

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 (WIMS), Article Number 7, ACM.

Kelly, D., & Sugimoto, C.R. (2013). A systematic review of interactive information retrieval evaluation studies, 1967–2006. Journal of the American Society for Information Science and Technology, 64, 745–770.

Kintsch, W. (1998). Comprehension: A paradigm for cognition. New York: Cambridge University Press.

Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. Discourse Processes, 39, 125–128.

Kitajima, M., Blackmon, M.H., & Polson, P.G. (2000). A comprehension-based model of Web navigation and its application to Web usability analysis. People and Computers, 357–374.

Kuhlthau, C.C., Heinström, J., & Todd, R.J. (2008). The 'information search process' revisited: Is the model still useful. Information Research, 13, 13–14.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). Handbook of latent semantic analysis. Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, D.I., & Berry, M.W. (2007). Mathematical foundation behind latent semantic analysis. In T.K. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), Handbook of latent semantic analysis. Mahwah, NJ: Lawrence Erlbaum Associates.

Pak, R., & Price, M.M. (2008). Designing an information search interface for younger and older adults. Human Factors and Ergonomics Society, 50, 614–628.

Pirolli, P., & Card, S. (1999). Information foraging. Psychological Review, 106, 643.

Queen, T.L., Hess, T.M., Ennis, G.E., Dowd, K., & Grühn, D. (2012). Information search and decision making: Effects of age and complexity on strategy use. Psychology and Aging, 27, 817.

Raven, J.C. & Court, J.H. (1998). Raven's progressive matrices and vocabulary scales. Oxford, UK: Oxford Psychologists Press.

Sharit, J., Hernández, M.A., Czaja, S.J., & Pirolli, P. (2008). Investigating the roles of knowledge and cognitive abilities in older adult information seeking on the web. ACM Transactions on Computer-Human Interaction (TOCHI), 15, 3.

Slegers, K., van Boxtel, M., & Jolles, J. (2009). Effects of computer training and internet usage on cognitive abilities in older adults: A randomized controlled study. Aging Clinical and Experimental Research, 21, 43–54.

Strauss, E., Sherman, E.M., & Spreen, O. (2006). A compendium of neuropsychological tests: Administration, norms, and commentary. Washington, DC: American Chemical Society.

Trewin, S., Richards, J.T., Hanson, V.L., Sloan, D., John, B.E., Swart, C., & Thomas, J.C. (2012, October). Understanding the role of age and fluid intelligence in information search. In Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 119–126). New York: ACM.

Turpin, A., & Scholer, F. (2006). User performance versus precision measures for simple search tasks. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 11–18). ACM.

Vakkari, P. (2001). A theory of the task-based information retrieval process. Journal of Documentation, 57, 44–60.

Van Oostendorp, H., & Aggarwal, S. (2015). Modeling and supporting web-navigation. Journal of Interaction Science, 3, 1–12.

Van Oostendorp, H., & Juvina, I. (2007). Using a cognitive model to generate web-navigation support. International Journal of Human-Computer Studies, 65, 887–897.

Wang, J.J., & Kaufman, A.S. (1993). Changes in fluid and crystallized intelligence across the 20-to 90-year age range on the K-BIT. Journal of Psychoeducational Assessment, 11, 29–37.

DOI: 10.1002/asi