

Performance of computational cognitive models of web-navigation on real websites

Journal of Information Science
2016, Vol. 42(1) 94–113
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0165551515615842
jis.sagepub.com


Saraschandra Karanam

Utrecht University, the Netherlands

Herre van Oostendorp

Utrecht University, the Netherlands

Wai Tat Fu

University of Illinois, USA

Abstract

Computational cognitive models of web-navigation developed so far have largely been tested only on mock-up websites. In this paper, for the first time, we compare and contrast the performance of two models, CoLiDeS and CoLiDeS+, on two real websites from the domains of technology and health, under two conditions of task difficulty, simple and difficult. We found that CoLiDeS+ predicted more hyperlinks on the correct path and had a higher path completion ratio than CoLiDeS. CoLiDeS+ found the target page more often than CoLiDeS, took more steps to reach the target page and was more ‘disoriented’ than CoLiDeS for difficult tasks. Difficult tasks in general for both models had less task success and lower path completion ratio, predicted less hyperlinks on the correct path, visited pages with lower mean LSA and took more steps to complete compared with simple tasks. Overall, inclusion of context from previously visited pages and implementation of backtracking strategies (which are both part of CoLiDeS+) led to better modelling performance. Suggestions to further improve the performance of these computational cognitive models on real websites are discussed.

Keywords

Computational cognitive modelling; information scent; real websites; task difficulty; web-navigation

1. Introduction

Information on the internet is increasing at a rapid pace, leading to what is called information overload. Humans have limited capacity to process information and they are often known to get lost when surfing the internet [1]. Many cognitive processes such as perception, reading, comprehension, decision making, reasoning and problem solving are involved when navigating on the web, making it a complex process that requires considerable effort. Information retrieval systems or search engines that filter and present relevant information to the user have been developed to help users surf through this abundance of information. However, the focus of such information retrieval systems to date has been primarily on retrieving relevant results. These systems are more useful in answering factoid questions such as ‘What is the distance between Mars and Earth’ or ‘Which is the highest mountain in Europe’, and are not very useful when it comes to questions that involve knowledge acquisition or learning, such as ‘Your doctor advised you to sleep on your side and not on your back to avoid snoring. Find out how sleeping on your side prevents snoring’. Tasks that are solved using a search engine are called ‘search by query’ tasks and the tasks that require one to move sequentially through a set of pages are called ‘search by navigation’ tasks [2]. Frequently, important information that is needed to solve the main search problem

Corresponding author:

Saraschandra Karanam, Princetonplein 5, Department of Information and Computing Sciences, Utrecht University, 3584CC Utrecht, the Netherlands.
Email: S.Karanam@uu.nl

is present in the intermediate pages leading to the target page [3]. In such cases, it is important to evaluate information on each page and take a decision on which hyperlink to click next based on the information that is already processed. Users acquire new knowledge not only at the end of an information search process after reaching the target page, but also during the intermediate pages through which they reach the target page. Search engines do not lay any emphasis on these intermediate steps and are largely focused only on the final step, that is, retrieval of target information that is sought by the user. The search by query approach neglects the acquisition of contextual information, and in contrast the search by navigation approach keeps track of this and can contribute in this way to searching as learning. The strict query approach also ignores many cognitive factors such as aging [4], domain knowledge [5] and past experience with the internet [6] that influence information search performance and follows a one-size-fits-all model. In this paper we focus on search by navigation tasks.

Learning from contextual information as users navigate on the web through hyperlinks involves complex cognitive processes that dynamically influence the evaluation of link texts and web contents [7, 8]. This is particularly the case when the user is engaged in a less well-defined search goal, in which the user is not certain about what it is that they will find, and they have to incrementally update their knowledge to evaluate the relevance of information as they encounter it. This form of search can thus be characterized as a learning process, as knowledge is incrementally gained as new information is assimilated into existing knowledge [8]. In fact, cognitive search has been shown to be an important process underlying human learning in various cognitive tasks, in which new knowledge is acquired or generated through the search process as new information is encountered [9].

Unlike traditional information retrieval systems, computational cognitive models of web-navigation take into account cognitive factors that influence the information search process to predict information search performance in a precise and automated way. For both search by query and search by navigation tasks, the focus of these models is on the process that leads to target information and not only on retrieving the correct answer. In results from earlier studies we did show that information-seeking performance in terms of the probability of finding what you are looking for and how efficiently you find it, is enhanced considerably when support based on cognitive models of web-navigation is provided [10–13]. It is important to note here that this support, delivered by emphasising links highly relevant to the task, was derived in an automatic way. The support that was presented to participants helped them to perform their tasks much quicker; they were also more accurate and less disoriented. The accuracy of support generated using such cognitive models depends on how robust the models themselves are. However, except for one model [14], we find that largely computational cognitive models of web-navigation have so far never been tested on real websites. In this paper, we, for the first time, test the performance of two cognitive models of web-navigation CoLiDeS and CoLiDeS + on real websites under varying conditions of task difficulty. The focus of this paper is on the performance of the models: we want to examine to what extent the models perform successfully their search task, how far they come with navigating through the website and how they behave in terms of disorientation. We call this the efficacy of the models. It is important to stress here that we examine the efficacy of the models, irrespective of the match with the behaviour of real participants. The fit between the models and user behaviour will be the focus of future work. The remainder of this paper is organized as follows: we first review the two models in Section 2. Section 3 provides the implementation details of the two models on real websites. Section 4 describes the experimental setup used to test the models on real websites. In Section 5, we describe our findings. Section 6 concludes the paper by summarizing the main outcomes and providing some preliminary suggestions to improve the performance of cognitive models of web-navigation on real websites.

2. Computational cognitive models of web-navigation

Computational cognitive models [10, 14–17] are useful to characterize search as a learning process. For example, existing models of semantic memories can be used to characterize how information cues (e.g. link texts, social tags or text snippets of web pages) are interpreted to reflect their semantic relevance with respect to the user's existing semantic knowledge [10, 14–18], which incrementally guides their web navigation behaviour. Based on theories of cognitive psychology and cognitive science, the main goal of these computational cognitive models is to use well-tested cognitive mechanisms to characterize more complex information search behaviour [19]. For example, CoLiDeS [15] is based on the well-tested Construction-Integration (CI) theory of text comprehension [20], which predicts how people interpret text snippets during web search and navigation. One important challenge for these computational cognitive models, however, is whether they can handle the complexities involved in a realistic environment, such as websites that involve complex link structures and contents. Simulation of user behaviour with these models in realistic websites is therefore important for establishing the validities of the predictions of the computational cognitive models.

In addition to a strong theoretical basis, another important advantage of using computational cognitive models to analyse and predict web behaviour is that the models can be used to capture the effects of individual differences in cognitive

abilities, domain-specific knowledge and internet experience on search and learning behaviour and outcomes. Given that computational cognitive models are often implemented as computer programs to simulate behaviour, they can be used to conduct automatic assessment of websites and simulations of web behaviour in different web designs by individual users with varying cognitive abilities and background knowledge. By comparing the simulated behaviour of different computational cognitive models, one can gain much insight into how variations in cognitive abilities and knowledge will impact their search and learning performance.

2.1. Comprehension-based Linked Model of Deliberate Search

The Comprehension-based Linked Model of Deliberate Search (CoLiDeS) proposed by Kitajima et al. [15] assumes that information seeking and navigation is driven by text comprehension and problem solving processes. CoLiDeS is based on the CI model of text-comprehension [20]. According to the CI model, comprehension of text happens in two phases: the *construction* phase, during which a propositional representation of the new piece of text is built and all possible meanings of the text elements are generated; and the *integration* phase, during which, using the reader's prior knowledge and context, a single coherent meaning is selected by filtering all those representations that do not fit well to the context. Construction is local (context-free) whereas integration is global (context-dependent). A further assumption in the CI model is that 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 reading comprehension.

The CoLiDeS model divides user navigation behaviour into two main stages of cognitive processing: the *attention cycle* and the *action-selection cycle*. The attention cycle is further divided into two stages: *parsing* the web page in high-level schematic regions and *focusing* on one of these schematic regions. Every web page consists of approximately 200 screen objects (buttons, icons, etc.). Each object competes for the user's attention. These objects are first categorized into high-level schematic regions such as a navigation pane, a content pane, window/browser controls and pictures. After high-level schematic regions are formed, one of the regions is chosen to *focus*. A combination of both bottom-up and top-down processes plays an important role in the attention cycle. Low-level perceptual features such as colour, orientation and top-down factors such as past experience with internet, knowledge of layout conventions, frequency of usage of a particular widget and the current task determine how many schematic regions are formed and which one is selected to focus. The action-selection cycle is also divided into two stages: *comprehension* of screen objects (e.g. hyperlink texts) within that region; and *selecting* the most appropriate screen object. The focused region from the attention cycle contains actual screen objects that need to be comprehended and understood by the user. Every screen object is thus understood by utilizing the information from that object and elaborating it with already existing information in the user's long-term memory. Once all objects are comprehended, a final step of selecting (e.g. clicking on a hyperlink) is performed. This process is then repeated for the new web page that is opened by the selected hyperlink until you reach the target page.

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 goals (integration phase). Also bottom-up and top-down processes play a role here just like in text comprehension [21]. Also in common is the cyclic character of processing: construction and integration phases for the second and third pages, and so on. Figure 1 shows a schematic diagram of the processes involved in CoLiDeS.

CoLiDeS is also connected to Information Foraging Theory, proposed by [19]. According to Information Foraging Theory, just as animals use their sense of smell when foraging for food, humans also look for a 'information scent' on the websites and click on those objects that have the maximum information scent. Information scent is an approximate estimation of how much useful information users are likely to get on a given page, based on information from proximal cues (such as hyperlinks). In general, a hyperlink with a higher information scent cue (e.g. text snippet) tends to have more users follow that cue and click on the corresponding hyperlink. At any point of time, users will assess the net information gain he/she would obtain from each screen object (such as a hyperlink) and click on the object that maximizes the net information gain. Based on their theory, Pirolli and Card [19] demonstrated that hyperlinks that have a high information scent with the goal of the reader are more likely to be selected first. CoLiDeS operationalizes information scent as the semantic similarity between the user goal and the screen object (hyperlink). Semantic similarity is computed with Latent Semantic Analysis (LSA) [22, 23].

The focus of the modelling is on the navigation process based on the selection and evaluation of links (the fourth phase in the CoLiDeS model). Of course, a complete modelling of the information-seeking process should also incorporate the reading and comprehension process of the web pages itself, that is, of the content of a website that is inspected by a reader. At this moment that is clearly a bridge too far. However, for the kind of tasks we are involved with, it is

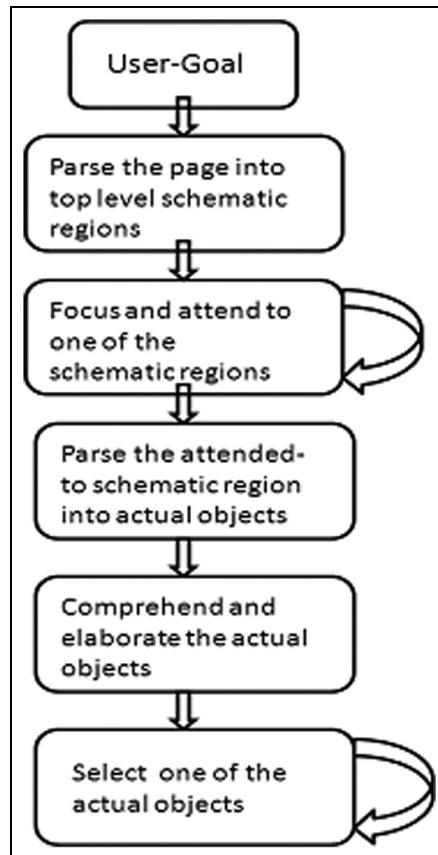


Figure 1. Schematic diagram of processes involved in CoLiDeS.

justified to isolate and focus on the fourth phase and model only the navigation process. Actually, we showed before that under these conditions availability of main content does not impact navigation behaviour much [24]. It appeared in that study that user navigation behaviour in terms of number of clicks, task accuracy and time spent on a page was the same for a condition where the main content was removed compared with a normal condition with content.

The CoLiDeS model described above has been successful in simulating and predicting hyperlink selections made by users during navigation [15]. The model has also been successfully applied within the domain of Human–Computer Interaction for finding usability problems, by predicting links that will be unclear to users [25–27]. However, one main limitation of this model is that it has so far been tested only on restricted, simplistic websites and not on real websites.

2.2. CoLiDeS +

CoLiDeS + [10, 11] shares the main theoretical foundations on which it is based with CoLiDeS. It further augments CoLiDeS and makes it more consistent with its theoretical assumptions by incorporating the concept of path adequacy and navigation strategies as a complement to the concept of information scent. It draws inspiration from several works on text comprehension that lay emphasis on the role of context. For example, when reading a text, it has been shown that contextual information helps users in comprehending sentences better, especially those with potentially multiple interpretations. It has also been shown that lack of supportive context leads to fast reading and low comprehension [28]. Analogously, when navigating on a website, users often encounter hyperlink texts that are varying in their degree of ambiguity. CoLiDeS + assumes that surrounding context helps the user in building a mental representation of the information space being navigated. This mental representation in turn will help in locating and finding relevant information such as hyperlinks or target pages and disambiguating ambiguous hyperlink text to grasp the intended meaning.

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, on a web page, CoLiDeS + gives the highest activation or priority to the incoming new hyperlink text followed by previously elaborated or clicked hyperlinks (these were hyperlinks that the user

thought could be relevant). 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+ assumes that users make decisions to click or not to click, not only on basis of information that is new, that is, an incoming new hyperlink text, but also on information that has been accessed before, that is, hyperlinks elaborated or clicked on in the preceding session. CoLiDeS, on the other hand considers only incoming hyperlink text to assess relevancy with respect to the goal (which is determined on the basis of semantic similarity between concepts, which is approximately equivalent to the overlap in patterns of their activation values) and does not incorporate any contextual information.

CoLiDeS+ incorporates this contextual information into modelling by defining path adequacy as the semantic similarity between the user goal and the navigation path. The navigation path is the sequence of hyperlinks clicked by a user at any given moment. Only if the information from an incoming hyperlink increases in information scent (or semantic similarity with the user-goal) is it considered for selection. If it does not increase in information scent, path adequacy is computed. 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 locally evaluated based on information scent, and only when it is not satisfying is a more effortful evaluation of the context performed by checking the path adequacy.

When path adequacy also does not increase (this scenario is termed as a latent impasse), then CoLiDeS+ tries to incorporate navigation strategies – backtracking to other regions within the same page and even to previously visited pages. Whenever CoLiDeS encounters a page with no relevant hyperlinks, it stops and declares an impasse. Solutions to impasse are only described and not computationally modelled. In this way, CoLiDeS+ overcomes the limitations of CoLiDeS by including contextual information and modelling backtracking and solving impasses. Note that the behaviour of CoLiDeS+ and CoLiDeS will be exactly the same if the information scent continues to increase on every page. Figure 2 shows a schematic diagram of the processes involved in CoLiDeS+. Notice the similarities of CoLiDeS+

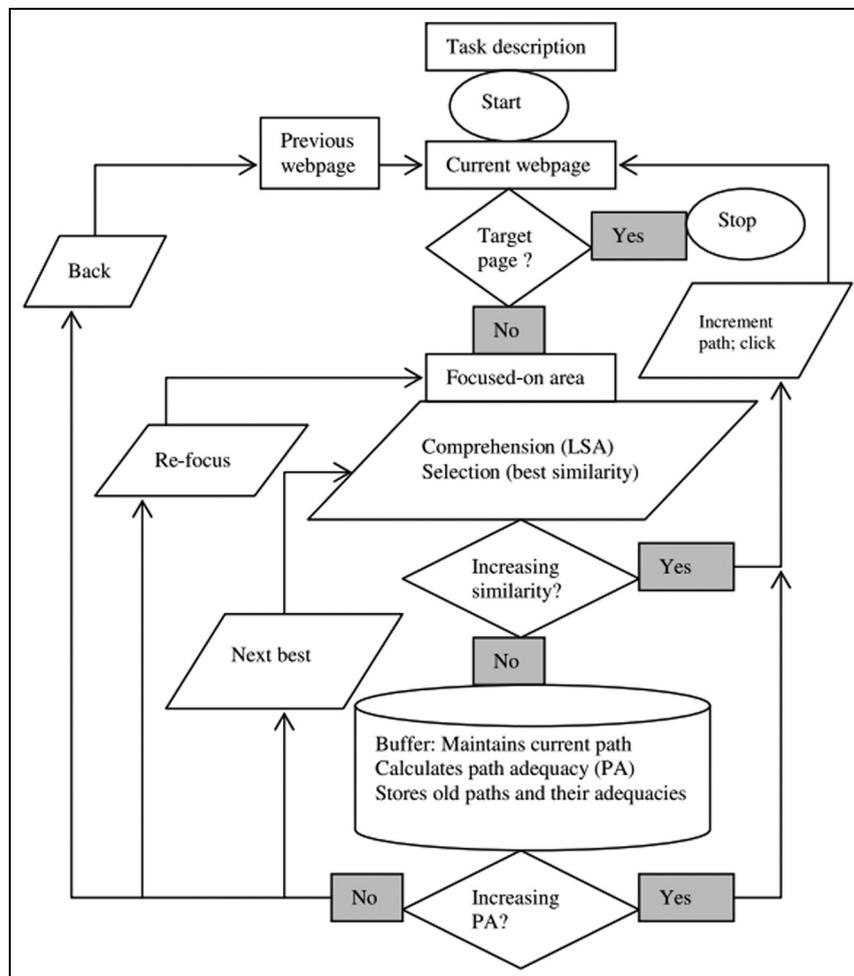


Figure 2. Schematic diagram of processes involved in CoLiDeS+ (reproduced from Juvina and Van Oostendorp [10]).

model with the CI model: the construction and integration phases are exactly the same as in the CI model. Next to that, the inclusion of context in the cyclic character of processing is common: retaining what is selected from the first page and connecting that to the information from the subsequent page(s).

By comparing selections made by users with the selections that CoLiDeS+ would have made for a set of tasks on a mock up website, Juvina and Van Oostendorp [10] 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. CoLiDeS+, just like CoLiDeS, has been tested only on mock-up websites under laboratory conditions and not with real websites.

Before introducing our research questions, we will first look at the characteristics of real websites and what makes modelling on real websites interesting, and then provide details about how we implemented CoLiDeS and CoLiDeS+.

3. Implementation of CoLiDeS and CoLiDeS+ on real websites

3.1. Characteristics of real websites

Why are real websites interesting? What characteristics of real websites make modelling navigation behaviour on them challenging? Below, we list some of the typical characteristics of a real website.

- (1) *Number of links per page* – the number of links per page is very high. Although there have been studies [29] that indicate that users do not process all the hyperlinks before deciding which one to click, for the sake of simplicity, we assume for now that all hyperlinks are processed before taking a decision.
- (2) *Information architecture* – the information architecture of a real website need not be nicely hierarchical and well organized.
- (3) *Location of target information* – it is possible to find information about a topic in multiple locations of a real website.
- (4) *Path to the target information* – there can be multiple ways of reaching the multiple locations of target information.
- (5) *Closely related competing and confusing links* – the number of competing and confusing links [27] can be very high.

The above characteristics make modelling on real websites both interesting and challenging.

3.2. Assumptions

Before getting into the actual implementation details, we first describe some of the assumptions under which both the models were implemented. For both CoLiDeS and CoLiDeS+, only the selection phase in which a hyperlink is chosen is implemented. It is assumed that the remaining phases before the selection phase are all completed and the models take as input a set of hyperlinks on a given web page or a portion of a web page. Also, for simplicity we assume that the simulation always starts from the home page of a website and not from an intermediate location within it. We observed that there are many hyperlinks that are repeated across multiple pages on real websites. This could cause the models to run in a loop indefinitely. To avoid this, we restricted our implementation in such a way that, at any point of time, only those hyperlinks that are new and were not seen in the previous iterations are considered for evaluation by the models.

We made certain manual interventions to ensure that the model does not fail in an irrelevant way and runs smoothly on a real website. Whenever a hyperlink points to a page with all pictures, a video, a puzzle, a quiz (without any other hyperlinks of interest) or an external website, the model ignores that hyperlink and selects the next best hyperlink on that page. If there are no more hyperlinks on the current page available for evaluation, then the next best from the previously clicked web page is chosen, and so on.

We also defined when and how each model would stop. Both CoLiDeS and CoLiDeS+ are stopped when the target page is reached. However, in real websites, to ensure that the model does not run indefinitely, we defined certain additional conditions under which the model stops running. For example, we set a maximum limit of 100 web pages that the model can evaluate in the process of finding the target page. If the model evaluates 100 web pages and has still not reached the target page, we stop it. Also, CoLiDeS stops whenever it encounters a page with no new hyperlinks or when all the hyperlinks on a page have a semantic similarity of < 0.1 to the goal. CoLiDeS+, under these two conditions (page with no new hyperlinks and a page with all hyperlinks with LSA value < 0.1) considers the next best hyperlink

on the current page if available, or the next best hyperlink from the previously visited web page if there are no new hyperlinks on the current page. CoLiDeS + stops when the probability of reaching the target page is zero. There can be two possible scenarios in which this can happen. First, if the target page has a very low LSA value (< 0.1), CoLiDeS + goes back one level and selects the next best hyperlink from the previous page. Second, if the LSA value of the hyperlink leading to the target page is less than the LSA value of current page and does not contribute to an increase in path adequacy, CoLiDeS + goes back one level and selects the next best hyperlink from the previous page. In both scenarios, CoLiDeS + misses the target page and therefore can be stopped.

3.3. Implementation of CoLiDeS

As discussed earlier, CoLiDeS draws analogies between cognitive processes behind text comprehension and cognitive processes in web-navigation. Therefore, to simulate what goes on inside human memory when comprehending a hyperlink text, CoLiDeS uses the process of elaborations. The process of elaboration simulates the cognitive processes of activation of semantically related terms to a piece of text that happens in our working memory (through spreading activation mechanism). These elaborations are known to assist in better comprehension of what is being read [20]. When reading a piece of text, similar and familiar words highly associated with the words in the text are activated in our working memory. Which words are activated depends on the strength of association between the words in the text and words in our memory. The greater the strength of association is, the greater the probability of the word getting activated is. One way of characterising the strength of association is to check how often the two words co-occur and how closely related they are semantically. Therefore, CoLiDeS uses LSA's near-neighbour analysis to extract words that are close in semantic similarity to the target text. For this purpose as well as for computing the semantic similarity between goals and hyperlinks, we used the semantic space 'tasaALL' provided by the LSA website from the University of Colorado, Boulder (<http://lsa.colorado.edu>). This space is meant to represent the knowledge and vocabulary levels of first year American college going students. The implementation of CoLiDeS consists of taking the following steps.

- (1) *Select a user task* – we will take the following example task to illustrate the different steps involved in the implementation of CoLiDeS: 'Birthmarks are those coloured skin spots that are either present at birth or develop shortly after. Some birthmarks are only colorations of the surface of the skin; others are raised above the surface of the skin or extend into the tissues under the skin. How many types of birthmarks are there and what are they?' The website used to find an answer to this question is <http://www.webmd.com> (WebMD). The answer to the question is two types: vascular and pigmented. The optimum navigation path to reach the target page where this information can be found is Home Page > Skin Problems > Skin Discolorations > Red Birthmarks, Haemangiomas, and Your Skin. Thus, the user has to make three clicks to find the information.
- (2) *Check if the current page is the target page* – if the current page is the target page, the model stops. If not, the model generates a list of new hyperlinks present on the current page. All those hyperlinks that were already seen on previous pages are ignored. Assuming we are on the home page of WebMD and since it is not the target page, the model generates a list of all hyperlinks from it. A total of 198 hyperlinks were extracted from the home page of WebMD. For the sake of brevity, we list here only the top five hyperlinks: 'Skin Problems', 'Psoriasis', 'Allergy', 'Allergy Relief' and 'Tenderness to touch'.
- (3) *Elaborate the user task* – the user task is elaborated with words that are its nearest semantic neighbours. We used LSA's nearest neighbour analysis to pick those words that occurred at least 15 times in the corpus and that had a semantic similarity value of at least 0.3 with the goal. Using this method, we got the following elaborated representation of the task: 'birthmarks are those coloured skin spots that either are present at birth or develop shortly after some birthmarks are only colorations of the surface of the skin others are raised above the surface of the skin or extend into the tissues under the skin how many types of birthmarks are there and what are they (skin dermis epidermis perspiration melanin follicles sweat keratin pores blackheads pimples acne follicle layer surface oily blisters smooth shampoos creases waterproof fingernails deeper epithelial hairs drying cosmetic pigment penetrate soap pore massage creams folds irritation covering protective powders microbial beneath grit soaps defences layers oils peels rough prevents antiseptic swelling cuticle cloths dries waxy dirt cosmetologist surfaces nails bumps thicker allergy reddened smoothing perspire protects underneath redness emery soaking outer openings wrinkles clog cleansing fingertips)'. Words in brackets are new words added after elaboration.
- (4) *Elaborate the hyperlinks* – using the methodology described in (3), all the hyperlinks are elaborated. For the sake of brevity, we provide the elaboration for the first hyperlink, that is, Skin Problems: 'skin problems (skin dermis perspiration epidermis melanin sweat keratin acne blackheads pimples problems pores follicles blisters oily shampoos creams follicle solve solving problem cosmetic solved massage creases fingernails hairs irritation

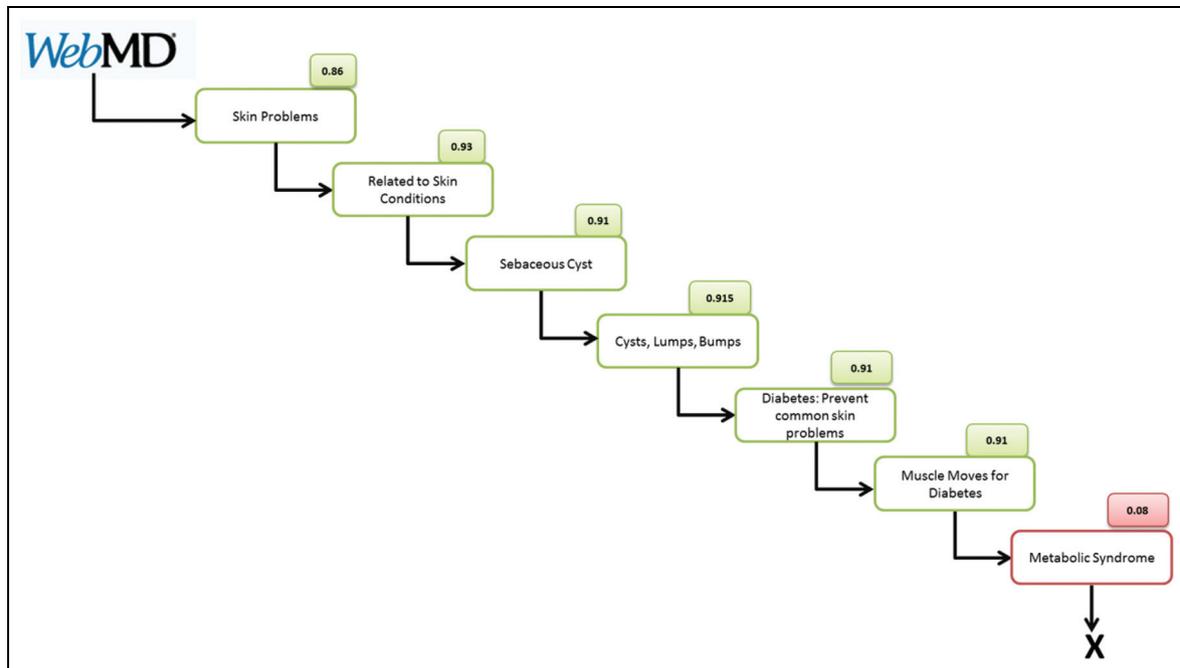


Figure 3. Navigation path predicted by CoLiDeS for the example goal (clicked hyperlinks in green, not clicked in red, X indicates that the model stopped (reached a page with LSA < 0.1)). Figure is reproduced in colour in online version.

peels drying soap waterproof perspire oils pore cosmetologist dirt redness pigment scalp lotion cuticle wrinkles reddened deeper fingertips layer antiseptic soaps shampoo microbial cloths defences soaking excessive cosmetics bumps epithelial powders lice protective lather sweating nails penetrate clogged clog swelling complexion). Again, words in brackets are new words added after elaboration.

- (5) *Compute semantic similarity between elaborated task and the elaborated hyperlinks* – using LSA, the model computes the semantic similarity between the elaborated task and the elaborated hyperlinks. For our example task and the set of five hyperlinks, the model obtained the following semantic similarity values: 0.865, 0.4, 0.3, 0.215 and 0.178. These values are computed for all the 198 hyperlinks.
- (6) *Prediction* – CoLiDeS selects the hyperlink that has the highest semantic similarity to the task. From the 198 LSA values obtained in step (5), the hyperlink with the highest LSA value is considered as the prediction. In our example task, this is obtained for the hyperlink: Skin Problems with an LSA value of 0.865.
- (7) *Repeat* – the model clicks on the hyperlink predicted in step (6) and steps (1)–(7) are repeated on every new page until one of the following three happens:
 - (7.1) the target page is reached;
 - (7.2) a leaf node, that is, a node with no new hyperlinks to be evaluated, is reached.
 - (7.3) a page with all hyperlinks with < 0.1 LSA value with the task is reached.

Following the methodology described above, for our example task, the model predicted path is shown in Figure 3. As can be seen from the figure, CoLiDeS clicks on the following six hyperlinks: ‘Skin Problems’ (0.865), ‘Related to Skin Conditions’ (0.93), ‘Sebaceous Cyst’ (0.91), ‘Cysts, Lumps & Bumps’ (0.915), ‘Diabetes: Prevent Common Skin Problems’ (0.91) and ‘Muscle Moves for Diabetes’ (0.91). The next hyperlink is ‘Metabolic Syndrome’ (0.08), which has an LSA value < 0.1 and therefore CoLiDeS stops. CoLiDeS fails to reach the target page. Notice that it could only predict the first hyperlink (Skin Problems) on the correct path.

3.4. Implementation of CoLiDeS +

We take the same example task as we took for CoLiDeS to illustrate the steps involved in the implementation of CoLiDeS +. For the first two web pages, LSA continues to increase and therefore the predictions of both CoLiDeS and

CoLiDeS+ are the same. We illustrate the steps involved in the implementation of CoLiDeS+ from the third web page, that is, from the page on ‘Related to Skin Conditions’ with an LSA value of 0.93.

- (1) *Check if the current page is the target page* – check if the current page is the target page. If yes, the model is stopped. If no, a list of new hyperlinks present on this page is generated. Since we are on the Related to Skin Conditions page of WebMD and since it is not the target page, the model generates a list of new hyperlinks from it. Twenty new hyperlinks were found. For the sake of brevity, we list here only the top three hyperlinks: ‘Sebaceous Cyst’, ‘Rocky Mountain Spotted Fever’ and ‘Skin Tags’.
- (2) *Compute semantic similarity between elaborated task and elaborated hyperlinks from step (1)* – CoLiDeS+ uses exactly the same methodology as CoLiDeS to generate elaborated representations of the goal and the hyperlinks. Semantic similarity between the elaborated goal and the elaborated hyperlinks is then computed. The LSA values of the three respective hyperlinks from step (1) after elaborations are 0.91, 0.907 and 0.897. The LSA values are computed for all of the 20 new hyperlinks generated in step (1).
- (3) *Prediction* – prediction in CoLiDeS+ is as follows:
 - (3.1) *Select the hyperlink with the maximum information scent* – of all the new hyperlinks from step (1), the hyperlink with the maximum LSA value forms the first level prediction. In our example, this is ‘Sebaceous Cyst’ with a LSA value of 0.91.
 - (3.2) *Check if information scent is increasing or not* – the information scent in our example is not increasing. The information scent of the previously clicked hyperlink was 0.93, corresponding to the page on ‘Related to Skin Conditions’, and the hyperlink with maximum information scent now is 0.91.
 - (3.3) *Compute path adequacy* – if information scent is not increasing, path adequacy is computed. That is, the semantic similarity between the elaborated task and the elaborated path (the set of hyperlinks clicked so far: ‘Skin Problems’ (0.865) and ‘Related to Skin Conditions’ (0.93)) is computed. We obtain path adequacy as 0.904.
 - (3.4) *Compute path adequacy with the new hyperlink* – path adequacy is computed again with the new hyperlink from step (3.1) added to the navigation path. Semantic similarity between the elaborated task and the new elaborated path (the set of hyperlinks predicted so far, ‘Skin Problems’ (0.865) and ‘Related to Skin Conditions’ (0.93), and the new hyperlink, ‘Sebaceous Cyst’ (0.91)) is computed. Path adequacy with the new hyperlink added is 0.911.
 - (3.5) *Check if path adequacy increases with the addition of new hyperlink* – in our example, path adequacy increases from 0.904 to 0.91. So, even though information scent decreases, path adequacy is still increasing. Therefore, CoLiDeS+ clicks on ‘Sebaceous Cyst’.
- (4) *Repeat* – this process is repeated until one of the following happens:
 - (4.1) the target page is reached;
 - (4.2) one of the two possible scenarios (discussed in Section 3.2) when the probability of CoLiDeS+ reaching the target page is 0 occurs.

The example above describes a scenario in which information scent decreases, but path adequacy continues to increase. We now continue the simulation to describe a scenario in which both information scent and path adequacy decrease. From the page on ‘Sebaceous Cyst’, the hyperlink with the highest information scent is ‘Cysts, Lumps & Bumps’ (0.915). As information scent is increasing from 0.91 to 0.915, CoLiDeS+ continues further. From the page on ‘Cysts, Lumps & Bumps’, the hyperlink with the highest information scent is ‘Diabetes: Prevent Common Skin Conditions’ (0.912). Since information scent is decreasing from 0.915 to 0.912, CoLiDeS+ computes path adequacy. Path adequacy between the goal and the hyperlinks clicked so far (‘Skin Problems’, ‘Related to Skin Conditions’, ‘Sebaceous Cyst’, ‘Cysts, Lumps & Bumps’) is 0.914. Path adequacy between the goal and the hyperlinks clicked so far along with the new incoming hyperlink (‘Skin Problems’, ‘Related to Skin Conditions’, ‘Sebaceous Cyst’, ‘Cysts, Lumps & Bumps’, ‘Diabetes: Prevent Common Skin Conditions’) is also 0.914. Path adequacy is not increasing. This is a latent impasse: a scenario in which both information scent and path adequacy do not increase. CoLiDeS+ backtracks one step to the previous page and selects the next best hyperlink from it. In our example, CoLiDeS+ backtracks to the page ‘Sebaceous Cyst’ and selects ‘Epidermoid & Sebaceous Cysts’ with an LSA value of 0.63 as the next best hyperlink. This process is repeated. All the steps followed by CoLiDeS+ for our example task are illustrated in Figure 4. For this example task, CoLiDeS+ reaches the target page in 10 clicks and backtracks seven times.

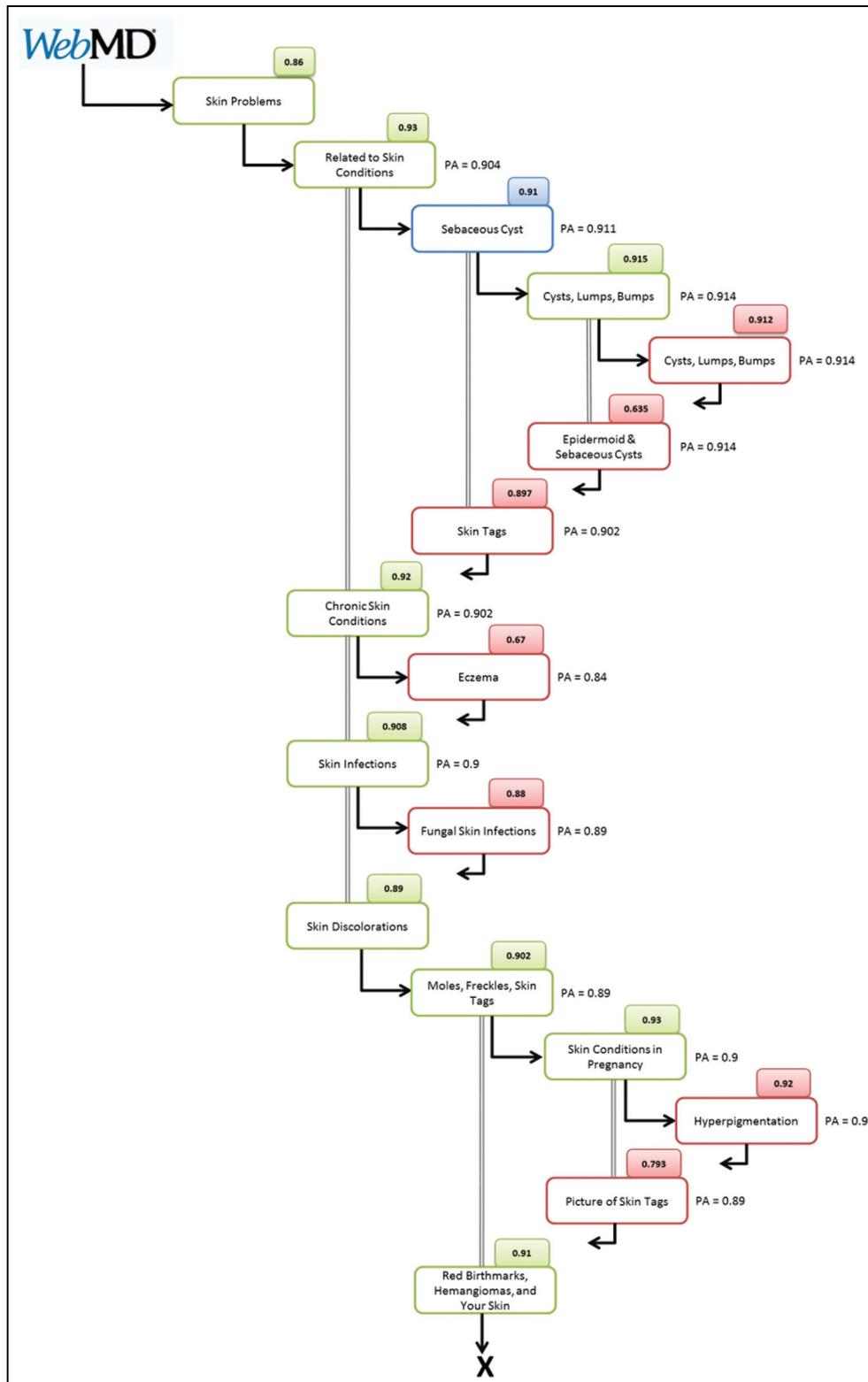


Figure 4. Navigation path predicted by CoLiDeS+ for the example goal (clicked hyperlinks in green, clicked after checking for path adequacy in blue, not clicked in red, X indicates that the model stopped (found the target page in this case), double line indicates hyperlinks from the same page). Figure is reproduced in colour in online version.

We can already see some performance differences between the two models CoLiDeS and CoLiDeS + .

- CoLiDeS does not reach the target page whereas CoLiDeS + does.
- CoLiDeS manages to find only one out of the three hyperlinks on the optimum path whereas CoLiDeS + finds all three.
- CoLiDeS clicked on six hyperlinks before stopping whereas CoLiDeS + clicked on 10.

CoLiDeS continues to run even if the hyperlinks under evaluation are getting more and more irrelevant to the task. We can see from Figure 3 that pages like ‘Diabetes: Prevent Common Skin Problems’ and ‘Muscle Moves for Diabetes’ are not relevant to the task but CoLiDeS continues to click on them. CoLiDeS + on the other hand, detects that these pages are not relevant and backtracks to evaluate alternative routes. CoLiDeS + finds such irrelevant pages correctly on seven different occasions, as can be seen from Figure 4.

In summary, CoLiDeS models the ideal situation of forward linear navigation. At any point of time, CoLiDeS evaluates the relevancy of incoming hyperlinks with the user goal (information scent) and selects that hyperlink that has maximum relevancy. If a page has no relevant hyperlinks, it declares an impasse and stops. Solutions to impasses are only described but not computationally modelled.

CoLiDeS + , on the other hand, implements both forward linear navigation and backtracking. It incorporates contextual information in the form of path adequacy as a complement to information scent. CoLiDeS + selects an incoming hyperlink if it increases in information scent. Else, it checks for consistency with previously selected hyperlinks (path adequacy). If the incoming hyperlink is consistent with previously selected hyperlinks, it is chosen, even if it does not increase information scent. If both information scent and path adequacy are found to decrease, then CoLiDeS + declares a latent impasse and fires backtracking strategies, that is, it goes back to previous pages and selects the next best hyperlink.

3.5. Research questions

Our first research question is to see how the computational cognitive models fare with real websites and which model performs better in terms of efficacy of navigation. As mentioned earlier, we do not compare here the model performance with the behaviour of real participants. We focus on the behaviour of the models and compare their behaviour to one another in taking decisions on which link to click while navigating. Given the complexity of real websites, we think that backtracking and detecting dead ends or impasses would play a major role in finding one’s way to the target information and, therefore, we hypothesize that CoLiDeS + will outperform CoLiDeS under real conditions. That is, CoLiDeS + would find the target page more often, would reach closer to the target page and would predict more hyperlinks on the optimum path than CoLiDeS. However, because of its additional capability of backtracking and employing navigation strategies, it would take more steps and would be more ‘disoriented’, that is, show more exploration, than CoLiDeS.

Our second research question deals with the difficulty of tasks for two reasons. First, modelling itself is more involved and challenging for difficult navigation tasks as opposed to simple tasks. Second, navigation support is more meaningful when task difficulty is high; therefore we are interested in testing how the models perform under two levels of task difficulty – low and high. We hypothesize that CoLiDeS + will perform better than CoLiDeS under difficult task conditions. Under easy task conditions, both models would perform equally.

Task complexity or task difficulty can be conceptualized in different ways: Gwizdka and Spence [30] distinguished several components of navigation path complexity: navigation path length (the length of the navigation path leading to the target information); page complexity (complexity of the navigation choices on each web page); and page information assessment (difficulty in judging the relevance of information contained in the page). We added to these components of navigation complexity the concept of path relevance: the degree to which the description of the task is pointing out the optimal path to the target information [31]. Descriptions of a task differ in the degree of guidance they provide towards the optimal path to the required information. More specifically we defined path relevance as the semantic similarity between the information goal (the task) and the optimal navigation path. We showed in these studies that navigation tasks with low path relevance decreased task accuracy and increased task completion time and lostness in navigation compared with tasks with high path relevance. The addition of navigation path relevance to the concept of navigation path complexity is important as it includes semantic aspects during search navigation contrary to, for example, path length, which is more related to structural aspects of a website.

Now that we have described the implementation of CoLiDeS and CoLiDeS + model and presented our research questions, we present in the next section the details of an experiment set up to run simulations of the two models to get a more precise picture of the power of the models and answers to our research questions.

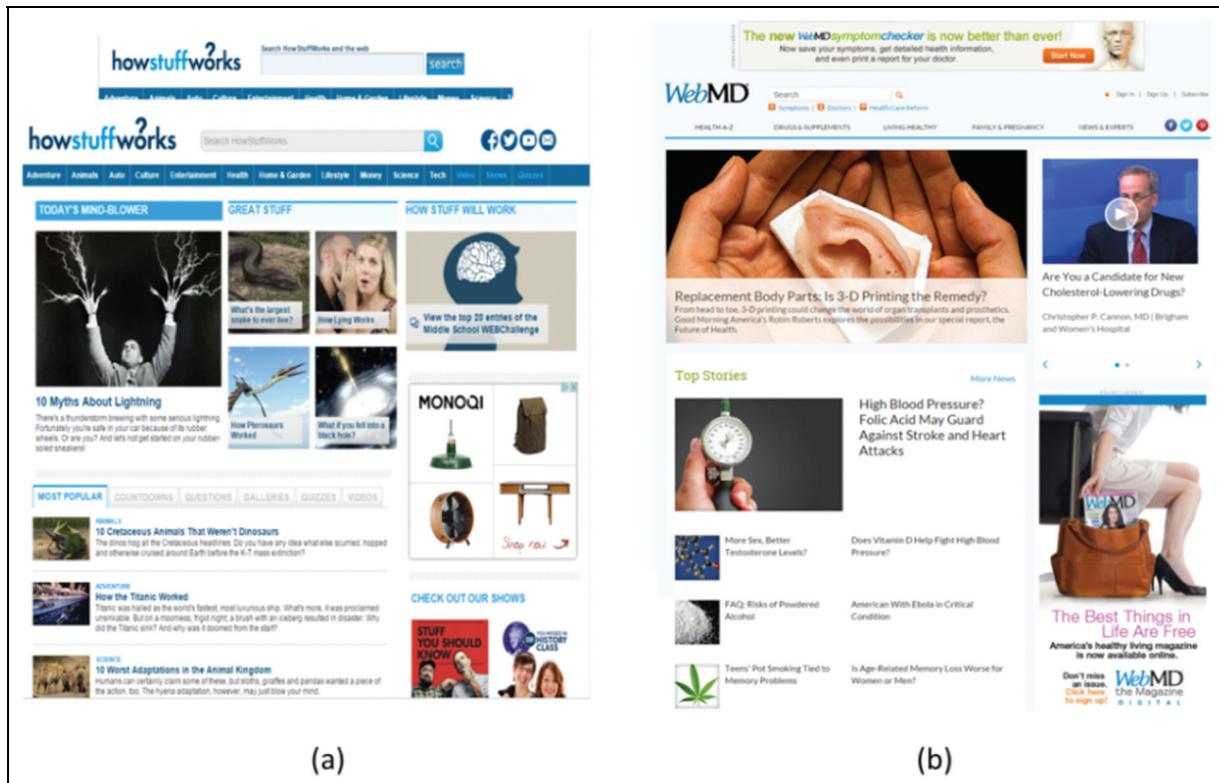


Figure 5. Screenshots of the top portion of both websites (a) HowStuffWorks and (b) WebMD as taken on 18 March 2015.

4. Experiment

4.1. Method

4.1.1. Design. A 2 (Model: CoLiDeS + vs CoLiDeS) \times 2 (Task Difficulty: Simple vs Difficult) design was used to run our simulations. Task difficulty was defined in terms of path relevance [31]. As indicated, path relevance is computed as the semantic similarity between a task and the *optimum* path to the target page. It is an indicator of the degree of information that the description of the task itself has about the optimum path. If path relevance is low, the description of the goal does not give you much information about the path to the target page and therefore it is difficult for users to find it. If path relevance is high, the description of the goal gives enough information about the path to the target page and therefore it is easy for users to find it. In this way, path relevance is more a characteristic of the website structure and is different from path adequacy, which is more a characteristic of the user behaviour.

4.1.2. Material. We used two real websites: <http://www.howstuffworks.com/> (HSW henceforth, for General Knowledge, 47,280 static web pages as of 1 July 2015) and <http://www.webmd.com/> (WebMD henceforth, for Health, 46,153 static web pages as of 3 July 2015). Both are general purpose websites frequently used to find out more about technology or health. Figure 5 shows screenshots of the top portions of both websites. Both websites are hierarchically organized. General topics are listed on the home page and more specific topics within each general topic are listed deeper in the website. Both websites require the user to click two to three links to reach a content page. A total of 20 tasks were used, 10 on each website. The tasks required the user to navigate within the websites (HSW or WebMD), find the page that contains the target information, read the content of the page and come up with an answer. Appendix I provides a list of the tasks used on both websites. We computed path relevance values for all 20 tasks and used the median to categorize them into five simple and five difficult on each website. The path relevance values for HSW were 0.73, 0.45, 0.79, 0.84 and 0.86 for simple, and 0.18, 0.12, 0.01, 0.25 and 0.431 for difficult (median was 0.44). Similarly the path relevance values for WebMD were 0.83, 0.62, 0.79, 0.69 and 0.9 for easy and 0.54, 0.59, 0.57, 0.19 and 0.53 for difficult (median was 0.6).

4.1.3. Measures. We measured the performance of both models using the following metrics:

Task success is defined as the number of times the model reaches the page with the target information. If the model reaches the target page, task success is defined as 1 and if it does not, then it is defined as 0. Note that the target page could be any page with the target information and not necessarily the target page identified by us.

Path completion ratio is defined as the ratio of the number of hyperlinks predicted by the model on the optimum path and the total number of hyperlinks on it. This measure is especially useful when a model fails in reaching the target page. If the optimum path for a goal has four hyperlinks and the model could reach up to one hyperlink, then path completion ratio is $1/4 = 0.25$; up to two hyperlinks, then the path completion ratio is $2/4 = 0.5$; up to three hyperlinks, then path completion ratio is $3/4 = 0.75$. If the model reaches the target page successfully, then path completion ratio becomes $4/4 = 1$. If the model reaches the target page or one of the intermediate pages on the optimum path via routes other than that identified by us, we use that as the final state. For example, if the model finds the first and third hyperlinks without going through the second hyperlink (because there exists an alternative route to the third hyperlink without going through the second hyperlink), we consider the path completion ratio as $3/4 = 0.75$.

Number of links predicted on the optimum path – for each task, and for each hyperlink on the optimum path, if the model predicts that hyperlink, it is counted as 1, else 0. This is repeated for all hyperlinks on a task and for all the tasks. The minimum possible score is 0, when the model fails to predict any of the hyperlinks on the optimum path and the maximum possible score is equal to the number of hyperlinks on the optimum path. That is, if the optimum path for a task has three hyperlinks, then the maximum possible score for that task is 3.

Number of steps to task completion is defined as the number of steps the model takes to reach the target page. All hyperlinks that were ignored because they were quizzes, puzzles, videos or picture galleries were not counted in this metric. Also, hyperlinks for which CoLiDeS+ decides not to click because both information scent and path adequacy decrease are also not counted. Only those hyperlinks that the model decides to click are used.

Mean LSA value per task – for each task, the sum of LSA values of all the links on the pages predicted by a model divided by the number of pages is taken. For example, if for a task, the model predicts four pages with LSA values of 0.46, 0.58, 0.64 and 0.7, then the mean LSA value is computed as $(0.46 + 0.58 + 0.64 + 0.7)/4 = 0.59$.

Disorientation – an objective measure of disorientation was used. It was computed using Smith (1996)'s L measure [32].

$$L = \sqrt{\left(\frac{N}{S} - 1\right)^2 + \left(\frac{R}{N} - 1\right)^2}$$

where R is the number of nodes required to finish the task successfully (i.e. the number of nodes on the optimal path); S is the total number of nodes visited while searching; and N is the number of different or unique nodes visited while searching.

This disorientation measure was originally proposed for human behaviour. We are using the same expression now to assess a model's behaviour. If the model takes exactly the same number of steps as the number of hyperlinks on the optimum path (no deviations, no repeat visits to already visited pages), then $N/S = 1$ and $R/N = 1$ and therefore $L = 0$. If the model makes more repeat visits to already visited pages (backtracking), N/S tends to 0 and L increases. Also, if the model stops with $S < R$ (that is number of steps is less than the number of nodes on the optimum path), then there are two possibilities: either the model is genuinely lost or the model took a short cut (which is possible on a real website) and reached the target page. In the latter case, R is adjusted to reflect the shorter path as the model cannot be described as lost. In the former case, $R/N > 1$ and L increases. Thus, L can range from 0 to ∞ . The higher the L measure, the greater the disorientation. One can think of this measure as an indicator of how perseverant a model is in finding the target page or how much of the website the model explores to find the target page. Therefore, higher values of disorientation in a model's context would actually mean more perseverance and more exploration.

4.2. Results

We used tasks instead of subjects as cases for our analysis. The performance of all 10 tasks on both websites (20 in total) was analysed together. Therefore, we report 2 (Model: CoLiDeS vs CoLiDeS+) \times 2 (Task Difficulty: Simple vs Difficult) ANOVAs in our results.

4.2.1. Task success. A 2 (Model: CoLiDeS+ vs CoLiDeS) \times 2 (TaskDifficulty: Simple vs Difficult) mixed ANOVA was conducted with model as within-cases variable, task difficulty as the between-cases variable and mean task success as the dependent variable. The main effect of model was significant $F(1,18) = 14.4$, $p < 0.001$. CoLiDeS+ was

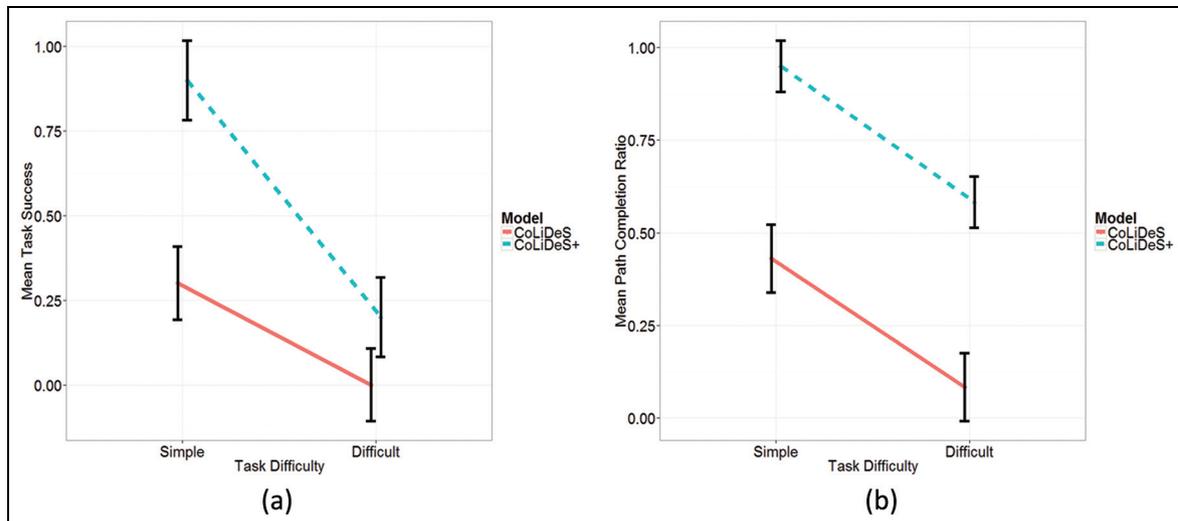


Figure 6. (a) Mean task success and (b) mean path completion ratio in relation to model and task difficulty.

significantly more successful in reaching the target page than CoLiDeS. The main effect of task difficulty was significant $F(1,18) = 17.3$, $p < 0.001$. Task success was significantly higher on simple tasks compared with difficult tasks. Interaction of model and task difficulty was not significant $p > 0.05$. Figure 6a shows the means.

4.2.2. Path completion ratio. A 2 (Model: CoLiDeS+ vs CoLiDeS) \times 2 (TaskDifficulty: Simple vs Difficult) mixed ANOVA was conducted with model as the within-cases variable, task difficulty as the between-cases variable and path completion ratio as the dependent variable. Twice on HSW tasks, CoLiDeS+ reached the target page via a path that was different from the one identified by the authors. The path completion ratio was updated as 1.0 for these two tasks. The main effect of model was significant $F(1,18) = 12.8$, $p < 0.005$. CoLiDeS+ was significantly more successful in reaching closer to the target page than CoLiDeS. The main effect of task difficulty was significant $F(1,18) = 25.34$, $p < 0.001$. Both CoLiDeS and CoLiDeS+ reached significantly closer to the target page for simple tasks compared with difficult tasks. Interaction of model and task difficulty was not significant ($p > 0.05$). See Figure 6b.

4.2.3. Number of hyperlinks predicted on the optimum path. A 2 (Model: CoLiDeS+ vs CoLiDeS) \times 2 (TaskDifficulty: Simple vs Difficult) mixed ANOVA was conducted with model as within-cases variable, task difficulty as the between-cases variable and mean number of hyperlinks per task predicted on the optimum path as the dependent variable. The main effect of the model was significant $F(1,18) = 31.56$, $p < 0.001$. CoLiDeS+ predicted significantly more hyperlinks on the optimum path compared with CoLiDeS. The main effect of task difficulty was significant $F(1,18) = 12.41$, $p < 0.005$. Significantly more hyperlinks on the optimum path were predicted for simple tasks compared with difficult tasks. Interaction of model and task difficulty was not significant ($p > 0.05$). Figure 7a shows the means.

4.2.4. Number of steps to task completion. A 2 (Model: CoLiDeS+ vs CoLiDeS) \times 2 (TaskDifficulty: Simple vs Difficult) mixed ANOVA was conducted with model as the within-cases variable, task difficulty the between-cases variable and mean number of steps to task-completion as the dependent variable. The main effect of model was significant $F(1,18) = 12.95$, $p < 0.005$. CoLiDeS+ took significantly more steps to reach the target page than CoLiDeS. The main effect of task difficulty was not significant at a conventional level, $p > 0.05$ ($F(1,18) = 4.06$, $p = 0.059$), but there was a tendency for difficult tasks to take more steps to complete than simple tasks. Interaction of model and task difficulty was significant $F(1,18) = 11.57$, $p < 0.005$. *Post-hoc* tests showed that CoLiDeS+ took significantly more steps than CoLiDeS to complete difficult tasks. For simple tasks, there was no difference between the two models ($p > 0.05$). Figure 7b shows the means.

4.2.5. Mean LSA value per page. A 2 (Model: CoLiDeS+ vs CoLiDeS) \times 2 (TaskDifficulty: Simple vs Difficult) mixed ANOVA was conducted with model as the within-cases variable, task difficulty as the between-cases variable and mean

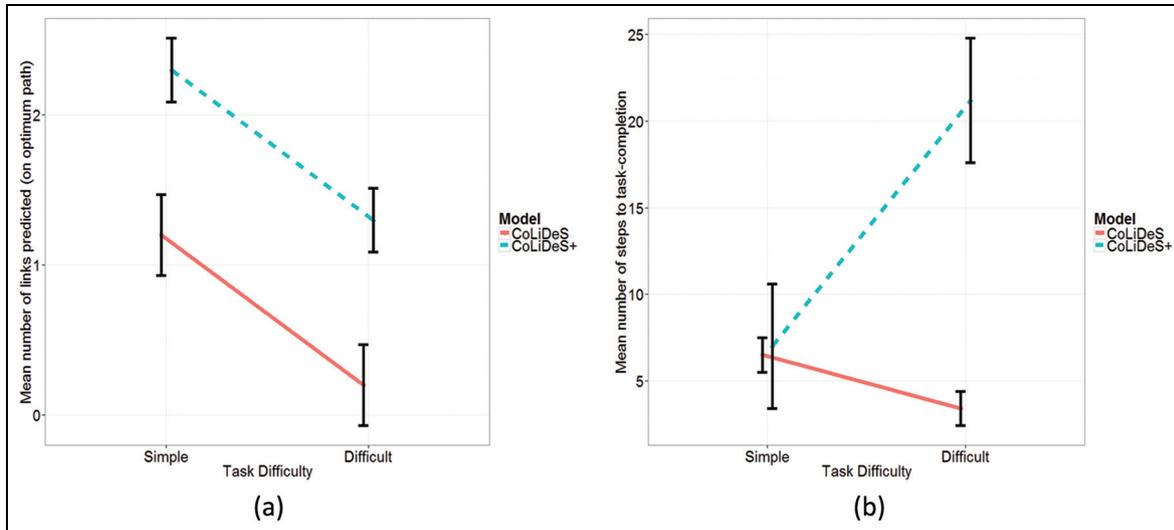


Figure 7. (a) Mean number of hyperlinks predicted on the optimum path and (b) mean number of steps to task completion in relation to model and task difficulty.

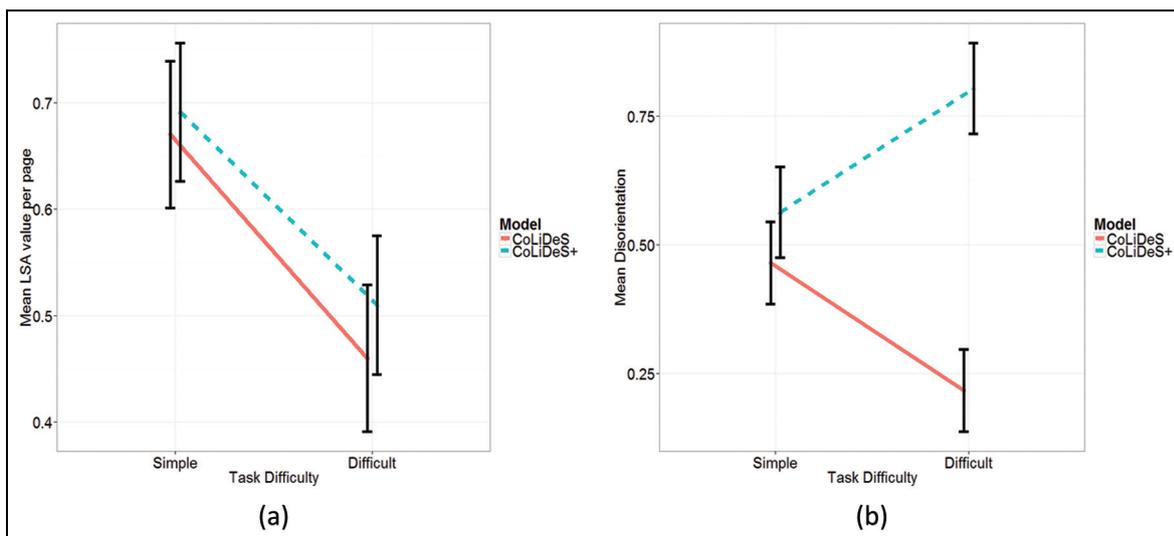


Figure 8. (a) Mean LSA value per page and (b) mean disorientation in relation to model and task difficulty.

LSA value of the links on the pages visited by the models as the dependent variable. The main effect of model was not significant ($p > 0.05$). The main effect of task difficulty was significant $F(1,18) = 4.6$, $p < 0.05$. Both CoLiDeS and CoLiDeS+ visited pages with significantly higher LSA values for simple tasks, compared with difficult tasks. Interaction of model and task difficulty was not significant ($p > 0.05$). See Figure 8a.

4.2.6. Disorientation. A 2 (Model: CoLiDeS+ vs CoLiDeS) \times 2 (TaskDifficulty: Simple vs Difficult) mixed ANOVA was conducted with model as the within-cases variable, task difficulty as the between-cases variable and mean disorientation as the dependent variable. The main effect of the model was significant $F(1,18) = 25.33$, $p < 0.001$. CoLiDeS+ was significantly more disoriented than CoLiDeS. The main effect of task difficulty was not significant ($p > 0.05$). Interaction of model and task difficulty was significant $F(1,18) = 12.94$, $p < 0.005$. *Post-hoc* tests showed that CoLiDeS+ was significantly more disoriented than CoLiDeS only for difficult tasks and not for simple tasks. Figure 8b shows the means. A summary of the findings is presented in Table 1.

Table 1. Summary of main outcomes of simulations of CoLiDeS and CoLiDeS +

	Model	Task difficulty	Model × task difficulty
Task success	CoLiDeS + > CoLiDeS	Simple > difficult	No interaction
Path completion ratio	CoLiDeS + > CoLiDeS	Simple > difficult	No interaction
Number of links on the optimum path	CoLiDeS + > CoLiDeS	Simple > difficult	No interaction
Number of steps	CoLiDeS + > CoLiDeS	Tendency difficult > simple	CoLiDeS + > CoLiDeS for difficult tasks
LSA per task	No difference	Simple > difficult	No interaction
Disorientation	CoLiDeS + > CoLiDeS	No difference	CoLiDeS + > CoLiDeS for difficult tasks

Table 2. Improvements in performance of CoLiDeS + in relation to each improvement in modelling

	Task success	Path completion ratio	Mean LSA	Number of links predicted on optimum path	Number of steps	Disorientation	Number of revisits
CoLiDeS +	0	0.5	0.62	1	48	1.03	31
CoLiDeS + plus improvement 1	1	1	0.62	2	40	1.0	19
CoLiDeS + plus improvements 1 and 2	0	0.5	0.675	1	28	0.99	15
CoLiDeS + plus improvements 1–3	1	1	0.9	2	11	0.92	8

Summarizing the main outcomes, difficult tasks had in general for both models significantly less task success, significantly low path completion ratio, significantly fewer hyperlinks predicted on the correct path, significantly more steps to complete and significantly lower mean LSA value of visited pages. Models were equally disoriented irrespective of the difficulty of task. The difference in performances between the two models in general was in line with our hypothesis. We found that CoLiDeS + predicted significantly more hyperlinks on the correct path, had significantly higher task success and had significantly higher path completion ratio than CoLiDeS. CoLiDeS + took significantly more steps to reach the target page and was significantly more disoriented than CoLiDeS for difficult tasks. There was no difference in the mean LSA value of pages visited by both models. CoLiDeS + was more disoriented than CoLiDeS because it is capable of detecting latent impasses and backtracking to explore alternative paths, whereas CoLiDeS is not. That is, CoLiDeS navigates linearly (Figure 3), picking the hyperlink with the highest LSA value on every page until it reaches the target page or reaches a page in which LSA value of all hyperlinks is less than 0.1. Because of this approach, even if the path being taken is irrelevant, CoLiDeS continues along the irrelevant path. CoLiDeS +, on the other hand navigates more non-linearly (Figure 4) because it is capable of detecting an irrelevant path by checking for path adequacy whenever information scent decreases. If both path adequacy and information scent decrease, CoLiDeS + backtracks to a page previously visited and selects the next best hyperlink from it. In this way, it avoids continuing along the irrelevant path. However, this additional capability comes at a cost: it takes longer and is seemingly more disoriented, in the sense that it explores more pages by taking alternative paths before stopping.

5. Conclusions and discussion

Computational cognitive models of web-navigation so far have practically all been tested on mock-up materials and not on real websites. Real websites pose several challenges, for example, target information can be found in multiple locations, there can be multiple paths to reach the same target page and there can be many hyperlinks that are close in semantic relevance to the target page. Therefore, we were interested in testing the computational cognitive models of web navigation under such adverse conditions and evaluating their performance. In this paper, we compared the performance of two computational cognitive models of web-navigation: CoLiDeS and CoLiDeS + on two real websites from the domains of technology (HSW) and health (WebMD), under two conditions of task difficulty: easy and difficult. We used two websites to get a precise picture of the performance of both models on the tasks in more than one environment. Our two main research questions in this study were:

- (1) How do computational cognitive models of web-navigation CoLiDeS and CoLiDeS+ perform on real websites? We hypothesized that, because CoLiDeS+ incorporates context by including information from previously clicked hyperlinks and implements backtracking strategies, it would perform much better than CoLiDeS, which relies only on local cues, that is, it includes information only from the incoming hyperlinks. CoLiDeS+ predicted more hyperlinks on the correct path and had higher path completion ratio than CoLiDeS. CoLiDeS+ took more steps to reach the target page and was more ‘disoriented’ than CoLiDeS for difficult tasks. CoLiDeS+, because of its additional capability of detecting latent impasses, backtracking and finding alternative paths, explores more web pages than CoLiDeS. There was no difference in the mean LSA of pages visited by both models.
- (2) Which of the two models CoLiDeS or CoLiDeS+ performs better on difficult tasks? We hypothesized that CoLiDeS+ will perform better than CoLiDeS under difficult task conditions. Under easy task conditions, both models would perform equally. We found that difficult tasks in general for both models had less task success and lower path completion ratio, predicted less hyperlinks on correct path and visited pages with lower mean LSA value. Models were equally disoriented irrespective of the difficulty of task. 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 hyperlinks, but also to the global context, that is, the goal and the hyperlinks already clicked on 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. Notice the similarity between local and global coherence processes occurring during reading as, for instance, in the CI model of Kintsch [21]. In other words, the assumption that CoLiDeS+ made about users basing their decisions to select a particular hyperlink not only on semantic relevance with the goal but also on whether an incoming hyperlink is consistent with the hyperlinks 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.

While it is encouraging to see that the computational cognitive models are able to run on real websites, we think that their performance is affected because of some of the challenges posed by real websites listed in Section 3.1. By making minor modifications to how the models run, based on the observations we made from the simulations on real websites, it is possible to further enhance the performance of CoLiDeS+. This is an important outcome for us to be able to make precise and directed improvements to the power of the modelling. We take the example goal: ‘Your doctor advised you to sleep on your side and not on your back to avoid snoring. Find out how sleeping on your side prevents snoring’ to explain these improvements. The optimum path for this goal is Home > Sleep Disorders > Snoring. This goal under normal conditions took 48 steps to complete, the model could not find the target page, the path completion ratio was 0.5, it predicted one hyperlink on the optimum path and the number of revisits, that is, the number of times CoLiDeS+ backtracked to a page visited in the previous step because both information scent and path adequacy were decreasing, was 31. We look at three possible enhancements:

- (1) *Minimum increment in path adequacy* – frequently, we observed that the increase in path adequacy is < 0.01 . This causes CoLiDeS+ to explore a hyperlink even though it is not relevant. We can set a threshold to a minimum of 0.01. Experiments are needed to arrive at a good threshold value that improves performance of the model. It can be varied depending on how the material of the website is organized. If there are many related concepts, the threshold should be low, else it can be high. By setting it to 0.01, for our example goal, CoLiDeS+ took 40 steps, it could find the target page and therefore path completion ratio was 1.0 and the number of revisits was 19.
- (2) *Path adequacy should be higher than information scent* – whenever path adequacy is checked, impose an additional stricter constraint that the hyperlink is clicked only if path adequacy not only increases but is also higher than the information scent of the page under consideration. This will reduce many detours. For the example goal, by incorporating both improvements (1) and (2), it took 28 steps, did not find the target page, path completion ratio was 0.5 and the number of revisits was 15. CoLiDeS+ could reach the target page, but did not decide to click on it because the path adequacy was lower than information scent.
- (3) *Using only relevant words from the task for elaboration* – users generally use keywords and do not type the entire question when using a search engine. If a user uses the entire task description as a query, it is normally an indication that the user is lost or did not understand the task. Motivated by this behaviour, we suggest a similar methodology to enhance the modelling performance. Instead of using the entire text from a task, we suggest using only the important keywords from it. How to identify the main keywords from a task can be a different research problem altogether, but for now, we used a topic modelling tool¹ provided by Google. For our example goal, the keywords identified by this tool were: ‘advised sleep side back avoid snoring find sleeping side prevents’. We used this as our modified task. Subsequently, we ran the CoLiDeS+ model along with improvements (1)–(3), respectively.

Now, with the example task the model ran for 11 steps, found the target page and therefore the path completion ratio was 1.0 and the number of revisits was 8. Table 2 summarizes the performance improvements in CoLiDeS+ and includes also the variable number of revisits.

One of our motivations to conduct this research was to provide accurate support during navigation, which, studies have already shown, is beneficial in enhancing information-seeking performance [8–10]. In these studies we developed a support tool consisting of emphasising the most relevant link (by an extra visual arrow) on each page for a certain goal. The link selection was automatically indicated by decisions of the computational cognitive model. Participants were faster and more accurately solving the tasks and less disoriented to a condition without these hints. However for the support to be really useful, the models have to be tested under real conditions, especially with difficult tasks. Model-generated navigation support contributes to the knowledge acquisition process as it helps the users in efficiently filtering unnecessary information and gives them more time to process and evaluate relevant information during the intermediate stages of clicking and reaching the target page. This helps in reducing the 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 learning outcomes. So in this way, 'search as learning', conceived as solving a 'search by navigation' task, can have positive outcomes. Practically, we expect that these modelling results support features such as 'breadcrumbs', especially for older adults with impaired cognitive functions such as working memory. Providing breadcrumbs will make it easy for older adults to understand their location in the website hierarchy and will avoid efforts to remember the hyperlinks they clicked in the preceding session.

6. Future work

We made several assumptions to make modelling on real websites practical. For example, we defined when a model would stop. We also stopped the model manually when it reached the target page. To make this decision automatically requires further experimentation and research, that is, can the model decide on its own if it has reached the target page or not? We used path relevance to create two levels of task difficulty but there could be many other ways, such as *page information assessment* or *difficulty* to judge the relevance of the information contained in the page in relation to the information goal and *page complexity* or complexity of the navigation choices characterized by aspects such as the number of links in a page or the visual design [33]. It would be interesting to look at how the performance of the computational cognitive models varies in relation to the other ways of characterizing task difficulty. Another point that requires attention is the influence of websites on model's performance. Would differences in information architecture and layout of website influence the performance of computational cognitive models? We plan to look at backtracking in more detail in the context of aging effects on navigation behaviour. It is known that older adults, or those with lower executive function (e.g. working memory), tend to pay attention to only local cues (e.g. text snippets of the current page, similar to what CoLiDeS does), while young people or those with better executive function tend to pay attention to both local and global cues (similar to what ColiDeS+ does) [34]. So the amount of previous histories taken into account by CoLiDeS+ could be, for example, a parameter that has the potential to fit individual differences. The improvements to the computational cognitive models suggested in this paper require further research and experimentation, for example, on how to set the threshold for increment in path adequacy.

Funding

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

Note

1. <https://code.google.com/p/topic-modelling-tool/>

References

- [1] Cockburn A and McKenzie B. What do Web users do? An empirical analysis of Web use. *International Journal of Human-Computer Studies* 2001; 54(6): 903–922.
- [2] Jul S and Furnas GW. Navigation in electronic worlds. A CHI 97 workshop. *SIGCHI Bulletin* 1997; 29: 44–49.
- [3] Olston C and Chi EH. ScentTrails: Integrating browsing and searching on the Web. *ACM Transactions on Computer-Human Interaction* 2003; 10(3): 177–197.

- [4] Pak R and Price MM. Designing an information search interface for younger and older adults. *Human Factors* 2008; 50(4): 614–628.
- [5] Liu C, Liu J, Cole M, Belkin NJ and Zhang X. Task difficulty and domain knowledge effects on information search behaviors. *Proceedings of the American Society for Information Science and Technology* 2012; 49(1): 1–10.
- [6] Van Deursen AJAM, Van Dijk JAGM and Peters O. Re-thinking Internet skills. The contribution of gender, age, education, Internet experience, and hours online to medium- and content-related Internet skills. *Poetics* 2011; 39(2), 129–144.
- [7] Fu WT. From Plato to the WWW: Exploratory information foraging. In: Todd PM and Robbins T (eds), *Cognitive search*. Cambridge, MA: MIT Press, 2012.
- [8] Fu WT and Dong W. Collaborative indexing and knowledge exploration: A social learning model. *IEEE Intelligent Systems* 2012; 39–46.
- [9] Fu WT, Hills T and Todd PM. Interfacing mind and environment: The central role of search in cognition. *Topics in Cognitive Science* 2015; 7(3), 384–390.
- [10] Juvina I and Van Oostendorp H. Modeling semantic and structural knowledge in Web navigation. *Discourse Processes* 2008; 45(4–5): 346–364.
- [11] Van Oostendorp H and Juvina I. Using a cognitive model to generate web navigation support. *International Journal of Human–Computer Studies* 2007; 65(10); 887–897.
- [12] Karanam S, Van Oostendorp H and Indurkha B. Towards a fully computational model of web-navigation. In: *Modern approaches in applied intelligence*. Berlin: Springer, 2011, pp. 327–337.
- [13] Aggarwal S, Van Oostendorp H and Indurkha B. Automating web-navigation support using a cognitive model. In: *Proceedings of the 4th international conference on web intelligence, mining and semantics*. New York: ACM, 2014.
- [14] Fu WT and Pirolli P. SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human–Computer Interaction* 2007; 22(4): 355–412.
- [15] Kitajima M, Blackmon MH and Polson PG. A comprehension-based model of web navigation and its application to web usability analysis. In: Mc Donald S, Waern S and Cockton G (eds), *People and computers XIV—Usability or else!* London: Springer, 2000, pp. 357–373.
- [16] Van Oostendorp H, Karanam S and Indurkha B. CoLiDeS + Pic: A cognitive model of web-navigation based on semantic information from pictures. *Behaviour and Information Technology* 2012; 31(1): 17–30.
- [17] Karanam S, Van Oostendorp H and Indurkha B. Evaluating CoLiDeS + Pic: The role of relevance of pictures in user navigation behaviour. *Behaviour and Information Technology* 2012; 31(1): 31–40.
- [18] Fu WT. The microstructures of social tagging: A rational model. In: *Proceedings of the ACM conference of computer supported cooperative work (CSCW)*, 2008, pp. 229–238.
- [19] Pirolli P and Card S. Information foraging. *Psychological Review* 1999; 106(4): 643–675.
- [20] Kintsch W. *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press, 1998.
- [21] Kintsch W. An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes* 2005; 39(2–3): 125–129.
- [22] Landauer TK, Foltz PW and Laham D. An introduction to latent semantic analysis. *Discourse Processes* 1998; 25(2–3): 259–284.
- [23] Dumais ST. Latent semantic analysis. *Annual Review of Information Science and Technology* 2004; 38(1): 188–230.
- [24] Karanam S, Van Oostendorp H and Indurkha B. The role of content in addition to hyperlinks in user-clicking behavior. In: *Proceedings of the 28th annual european conference on cognitive ergonomics*, 2010, pp. 125–131.
- [25] Kitajima M, Blackmon MH and Polson PG. Cognitive architecture for website design and usability evaluation: Comprehension and information scent in performing by exploration. In: *Proceedings of HCI-International*, 2005.
- [26] Blackmon MH, Kitajima M and Polson PG. Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. New York: ACM, 2005, pp. 31–40.
- [27] Blackmon MH, Mandalia DR, Polson PG and Kitajima M. Automating usability evaluation: Cognitive walkthrough for the web puts LSA to work on real-world HCI design problems. In: Landauer T, McNamara D, Dennis S and Kintsch W (eds), *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum, 2007, pp. 345–375.
- [28] Budiu R and Anderson JR. Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science* 2004; 28: 1–44.
- [29] Fu WT and Gray WD. Suboptimal tradeoffs in information seeking. *Cognitive Psychology* 2006; 52(3): 195–242.
- [30] Gwizdzka J and Spence I. Implicit measures of lostness and success in web navigation. *Interacting with Computers* 2007; 19: 357–369.
- [31] Van Oostendorp H, Madrid RI and Puerta Melguizo MC. The effect of menu type and task complexity on information retrieval performance. *Ergonomics Open Journal* 2009; 2: 64–71.
- [32] Smith P. Towards a practical measure of hypertext usability. *Interacting with Computers* 1996; 8(4): 365–381.
- [33] Gwizdzka J and Spence I. What can searching behavior tell us about the difficulty of information tasks? A study of web navigation. In: *Proceedings of the 69th annual meeting of the American Society for Information Science and Technology 2006 (ASIS&T)*, Vol. 43, Austin, TX.

- [34] Dommes A, Chevalier A and Lia S. The role of cognitive flexibility and vocabulary abilities of younger and older users in searching for information on the web. *Applied Cognitive Psychology* 2011; 25(5): 717–726.

Appendix I

Table AI. Information retrieval tasks

Simple	Difficult
<p><i>How stuff works</i> You are working on a project on environmental science and your topic is green technologies. Find out more about green highways.</p> <p>You are fond of reading fiction and adventure novels. You would like to go for an adventure yourself and choose to go hunting for wolves and deer. But you are unable to decide the kind of rifle you should take with you. List at least two factors to consider when choosing a rifle for hunting.</p> <p>You are fond of reading fiction and adventure novels. You would like to go for an adventure yourself and choose to go on a salmon fishing journey. You have with you a rowing boat, net, hook and all necessary gear but you do not know where to go. Find out the top four salmon fishing locations.</p> <p>You have dog as your pet animal and it ate some chocolate accidentally. Find out if chocolate is poisonous to dogs?</p> <p>What is the reason for aquatic mammals like whales and dolphins not going to sleep completely?</p> <p><i>WebMD</i> You have decided go for an early morning walk on a daily basis as part of your exercise routine. It is important to work out at a certain intensity. Find out what the ideal heart rate should be when walking.</p> <p>What name is given to the type of arthritis in which crystals collect in your joints?</p> <p>You have just adopted a puppy for the first time. Your puppy refuses to eat any food and you are distressed. What medical name is given to the complete loss of appetite in dogs?</p> <p>You have an ulcer on the inside of your cheek. Your oral health practitioner diagnosed it as a mouth sore. What are the two types of mouth sores and what is the common medical term given to mouth sores?</p> <p>Birthmarks are those coloured skin spots that either are present at birth or develop shortly afterwards. Some birthmarks are only colorations of the surface of the skin; others are raised above the surface of the skin or extend into the tissues under the skin. How many types of birthmarks are there and what are they?</p>	<p>United Nations is an inter-governmental organization set up in 1945 to resolve disputes between two different countries and foster greater cooperation among people of different cultures. According to the norms laid down by UN, which country's citizenship would a child born on board a flight get?</p> <p>You have been working on editing a movie for quite some time, yet you are not satisfied with your output. You heard about a technique using which you can gather outside opinion on the good and bad parts. Name the technique and also three famous movies that were altered using this approach.</p> <p>A performance on stage involves much more than rehearsals and acting: costumes, music, lights, special effects etc. One such special effect is to create artificial fog. How is artificial fog created in the theatre?</p> <p>You applied and received financial aid from the US government to pursue your graduation studies. Find out four different possibilities of paying back your debt.</p> <p>Which hardware connector synchronizes the alternating images on a 3D television screen with LCD lenses in 3D glasses?</p> <p>Your baby is 10 months old and going through teething. It is causing much discomfort with lots of drooling and irritation. Find at least one tip to make your baby comfortable</p> <p>You have two children; the younger one is 3 years and has health problems in developing speech and language. He does not interact or play with your older child. Clinical examination has diagnosed him with autism. Check for other symptoms of autism that you should look out for in your child.</p> <p>Your doctor advised you to sleep on your side and not on your back to avoid snoring. Find out how sleeping on your side prevents snoring.</p> <p>Your baby has watery eyes, is sensitive to bright light and keeps her eyes closed most of the time. Check if these are symptoms of glaucoma.</p> <p>You feel palpitation, a feeling of pounding in your chest, as if your heart is beating too fast or too slow. Your doctor has told you that your heart is pumping out of rhythm. What medical name is given to this heart disease?</p>