

Age-related Differences in the Content of Search Queries when Reformulating

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

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

ABSTRACT

This study investigated the change in the content of the queries when performing reformulations in relation to age and task difficulty. Results showed that both generalization and specialization strategies were applied significantly more often for difficult tasks compared to simple tasks. Young participants were found to use specialization strategy significantly more often than old participants. Generalization strategy was also used significantly more often by young participants, especially for difficult tasks. Young participants were found to reformulate much longer than old participants. The semantic relevance of queries with the target information was found to be significantly higher for difficult tasks compared to simple tasks. It showed a decreasing trend across reformulations for old participants and remained constant for young participants, indicating that as old participants reformulated, they produced queries that were further away from the target information. Implications of these findings for design of information search systems are discussed.

Author Keywords

Information Search; Aging; Reformulation Strategies; Task Difficulty.

ACM Classification Keywords

H.5.4 Information Interfaces and Presentation: Hypertext / Hypermedia—*Navigation*; H.3.3 Information Storage and Retrieval: Information Search and Retrieval—*Search Process*; J.4 Computer Applications: Social and Behavioral Sciences—*Psychology*

INTRODUCTION

The Internet was once thought to be of not much use to older adults. Complexity and difficulty to learn its usage added to this perception [12, 26]. However, this is no longer the case, as is evident from the fact that they are now one of the fastest growing users of the Internet [1, 9]. Recent studies show that older adults are now as enthusiastic as younger adults in using the web [38]. The Internet is known to decrease isolation

by enabling alternate means of communication with near and dear, foster independence, enhance attention and memory [33] and keep the mind active, thereby increasing overall health and well-being of the elderly [25].

However, a number of barriers still exist preventing large scale adoption and usage of the Internet by the elderly. Due to the natural decline in their motor skills and fluid intelligence involving processing speed, cognitive flexibility or ability to switch processing strategies, attentional control and visuospatial span [14, 15, 36], they are known to be slow and less efficient when using the Internet. Crystallized intelligence, on the other hand, increases and/or becomes more stable with aging. Crystallized intelligence involves prior knowledge, experience and vocabulary skills. Therefore it is higher for older adults compared to younger adults [14, 15, 36].

Searching for information on the Internet can often be challenging and involves several cognitive processes such as memory (formulating relevant queries), attention (understanding the visual layout of the website or search engine result pages (SERPs, henceforth)), problem solving (information can be found in multiple locations and there could be multiple paths leading to them), comprehension (understanding the search results provided by a search engine) and decision making (choosing a relevant search result, reformulating unfruitful queries). These cognitive processes are known to be affected by the above aging-related declines in cognitive abilities. Much research has already been conducted on how aging-related decline in cognitive abilities influence either positively or negatively the performance on an information search task [5, 31]. Older adults were in general found to spend longer on search result pages, open fewer search results and reformulate queries less often than younger adults, especially for difficult tasks. Also, they were found to generate fewer queries than younger adults [21]. In another study, it was found that the performance of younger adults was better than that of older adults, especially on difficult and impossible tasks. Younger adults not only used more queries but also more keywords per query compared to older adults when solving difficult and impossible tasks. More precisely, it was difficult for older adults to reformulate unsuccessful queries [10]. Overall, older adults generate less queries, use less 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 number of times between SERPs and websites and find it difficult to reformulate unsuccessful queries [5, 29]. However, older adults are known to adapt their search

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2016, May 7–12, 2016, San Jose, California, USA.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-3362-7/16/05 ...\$15.00.

<http://dx.doi.org/10.1145/2858036.2858444>

and navigation strategies exploiting their higher crystallized knowledge to compensate for their decreased fluid capabilities when performing information search tasks using the Internet [7, 8, 10, 29]. Information search tasks on the Internet are usually carried out in two phases: first, using a search engine and formulating queries relevant to one's goal (search by query) and second, navigating on websites opened by clicking on one or more of the search results (search by navigation) [28]. We focus on the first phase, i.e., interaction with a search engine in this paper.

AGING AND SEARCH STRATEGIES

A number of researchers have looked into the search strategies employed by younger and older adults when solving information search tasks using a search engine. Youmans et al. [37] conducted a study in which older adults were found to follow a more structured and methodical approach to information search such as careful selection of query terms and spending longer time evaluating the search results whereas younger adults were found to follow a more impulsive strategy that involved more switches between search results and more clicks on irrelevant search results.

A more recent study by [4] looked at the three important phases in an information search process: planning (dividing a problem into sub-problems, formulating the first query), evaluating (assessing the relevance of search results) and controlling (modifying the query and strategy if required). In line with one of their earlier studies [3], 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 another study [6], 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 those websites visited for any given query) phase of an information search process. Older adults were found to do 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 on search strategies has so far overlooked the age related differences in reformulation strategies and the actual content of the queries during reformulations. By actual content, we mean here the semantic aspects of the query terms. These, precisely, form the main focus of this paper. Of all the different kinds of query reformulation strategies that the information retrieval domain has extensively looked at [17, 18, 32, 35], we focus on the two most common types: *generalization* and *specialization* in our study. These are concerned with modifying the content of the queries and are interesting from a cognitive perspective. Some recent studies have shown that people highly familiar with the topic of search prefer generalization whereas people less familiar with the topic of search prefer specialization as a reformulation strategy [16]. People more familiar with the topic of search were also found to use

less queries, more relevant keywords per query, select more relevant results and were more successful in solving tasks, especially the difficult ones [27, 30]. Choice of reformulation strategy employed was also shown to be influenced by the type of search task [24], however, this and the earlier studies did not incorporate any age-related differences. The two main issues that this paper addresses are:

- (a) Age-related differences in reformulation strategies in relation to difficulty of task.
- (b) Variation in the content of the queries during reformulations in relation to age, difficulty of the task and the target information sought.

Semantic relevance was used in the past as a metric to evaluate the content of hyperlink texts (to predict navigation behavior on websites) [13, 19] or the content of the search results on the SERPs (to predict interaction behavior with search engines) [21]. In both cases, semantic relevance was computed in relation to the user goal and not with the target information. As far as we know, this is the first study that investigates the content of the queries in relation to the target information sought.

Outcomes of this study would be valuable for designing and embedding support mechanisms for various reformulation strategies into search and information retrieval systems.

RESEARCH QUESTIONS

Based on the findings from previous studies, we formulated the following research questions and hypotheses for the current study:

- (a) How often do younger and older adults make use of generalization and specialization strategies when reformulating search queries? (**RQ1**). Because, reformulating a query is a cognitively effortful process, we hypothesize that younger adults would use both generalization and specialization more often than older adults (**Hyp1**).
- (b) How does the frequency of usage of generalization and specialization strategies when reformulating queries vary in relation to difficulty of a task? (**RQ2**). We hypothesize that difficult tasks would require more reformulations and therefore more generalizations and specializations than simple tasks (**Hyp2**).
- (c) How does the semantic relevance of a query with the target information sought vary in relation to age and difficulty of a task? (**RQ3**). The Construction-Integration model of text comprehension [22] states that comprehension of text happens cyclically in two phases: *construction phase* during which a representation of a new piece of text and all possible meanings of the text are generated and *integration phase* during which using reader's prior knowledge and context, a single coherent meaning is selected by filtering all those representations that do not fit to the context. We assume that this model also applies to information search and hypothesize that solving difficult search tasks successfully urges users to activate a greater number of related concepts from Long Term Memory into Working Memory

[22]. Assuming that these activated concepts are included as terms in the search queries formulated by users, we hypothesize that the semantic relevance of a query with the target information would be higher for difficult tasks compared to simple tasks (**Hyp3a**). Also, based on the outcomes of the study by [4], in which younger adults were found to be better at controlling phase of information search (modifying the query and strategy if required, which resulted in their higher performance) compared to older adults, we hypothesize that the semantic relevance of queries generated by younger adults would be higher than that of older adults (**Hyp3b**).

EXPERIMENT

We briefly report here the results of an experiment conducted to collect behavioral data from participants and check if the results are in line with prior outcomes from aging-related literature. Next to that we report the main issue of this study: the analysis of the type of reformulation strategy and the content of the queries during reformulations in relation to age and task difficulty.

Method

Participants

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

Design

We followed a 2 (Age: Young vs. Old) X 2 (Task Difficulty: Simple vs. Difficult) mixed design with age as between-subjects variable and task difficulty as within-subjects variable.

Material

The experiment was conducted with twelve simulated information search tasks [2]: six simple and six difficult, all from the domain of health. For simple tasks, participants in most cases could find the answer easily either in the snippets of the search engine results or in one of the websites referred to by the search engine results. For instance, for a task like “*Wat is de belangrijkste functie van zweetklieren? (What is the main function of sweat glands?)*”, user could use the words like “*functie van zweetklieren (function of sweat glands)*” from the task description itself as queries. One can easily find the answer “*regelen de lichaamstemperatuur (regulation of body temperature)*” within the snippets of the corresponding search results without having to click on any of them. For difficult tasks, users had to frame queries using their knowledge and understanding of the task, the answer was not easily found in the snippets of search engine results and often they had to evaluate information from multiple websites. As an example, for the task “*Elbert, 76 jaar oud, heeft al een paar jaar last van een brandend gevoel bij het plassen. Hij plast's nachts meer dan normaal en hij klaagt over het gevoel dat zijn blaas dan niet helemaal leeggemaakt wordt. Sinds kort heeft hij ook een scherpe pijn in de heup, onderrug en zijn bekken. In de laatste 6 maanden is hij 12 kilo afgevallen. Wat zou de oorzaak van zijn klachten kunnen zijn? (Elbert, 76 years old has been suffering for few years from burning sensation while*

passing urine. He passes urine more often than normal at night and complains of a feeling that the bladder is not empty completely. Lately, he also developed acute pain in the hip, lower back and pelvis region. He also lost 12 kilos in the last 6 months. What problem could he be suffering from?”, users had to formulate multiple queries such as “*niersteen pijn in rug (kidney stones pain in the back)*”, “*branderig gevoel bij plassen (burning sensation when urinating)*”, “*blaasontsteking (urinary infection)*” to find the answer. The answer to this task “*prostaatkanker (prostate cancer)*” is also not found easily in the snippets of the search results of the queries, unless the query is very specific. The tasks were all presented in Dutch.

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). Based on these self-reported ratings, old participants ($M=2.63$, $SD=1.01$) were found to be significantly less experienced with search engines than young participants ($M=3.75$, $SD=0.68$) $t(46) = 4.52$, $p < .001$. Old participants were significantly longer experienced with computers ($M=22.7$ years, $SD=9.43$) than young participants ($M=15.33$ years, $SD=4.09$) $t(46) = 3.1$, $p < .005$. They were next presented with a fluid intelligence test: a computerized version of the Trail Making Test (TMT Part B) [34]. Participants were shown 25 circles containing both numbers (1 to 13) and alphabets (A to L) on the computer screen. The participants had to click on the circles in ascending pattern alternating between numbers and alphabets (1-A-2-B-3-C and so on) starting from the number 1. If the circle clicked by a participant is right, it turned green, otherwise it turned red. We measured the time taken to finish the test correctly. In line with the traditional cognitive aging literature, significant differences were found in fluid abilities of young and old participants. Old participants took significantly longer to finish this test ($t(46)=5.13$, $p < .001$) than young participants indicating that they had significantly lower fluid abilities compared to young participants.

After the TMT test, participants were allowed a break of five minutes. They were then presented with twelve information search tasks (six simple and six difficult) in a counter balanced order. Participants were first shown the task and then directed to the home page of Google's search engine. Participants were not allowed to use any other search engine. We show in Figure 1 the main screen of our interface that participants used while solving the information search tasks. Participants could enter queries as they normally would on any browser and the corresponding search results appeared. The task description was made available to the participant at all times in the top left corner. An empty text box was provided in the top right corner for the participant to enter his/her answer. All the queries generated by the users, the corresponding search engine result pages and the URLs opened by them were logged in the back-end using Visual Basic. We first analyze the performance in terms of time, clicks, accuracy and reformulations and later analyze the type of reformulation strategy and the semantic content of the queries during reformulations more deeply.

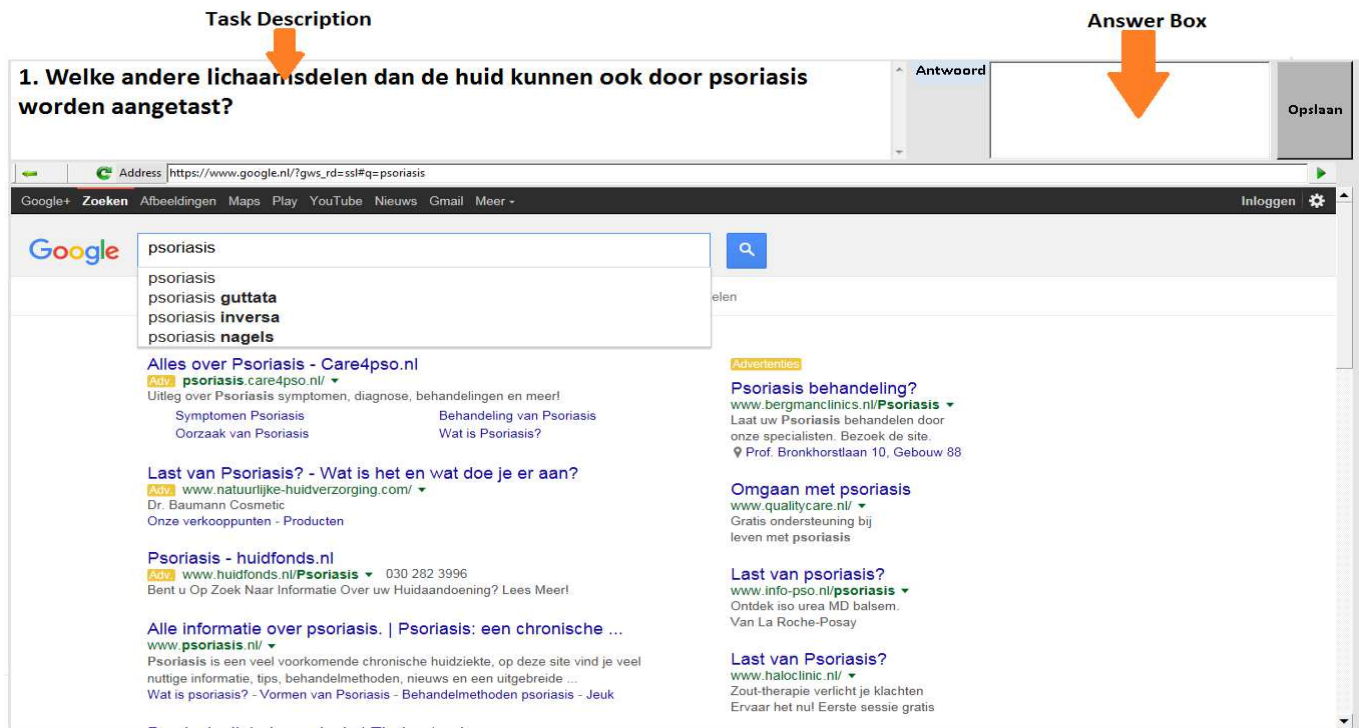


Figure 1. Interface showing the main screen in which the information search tasks are solved.

Analysis of Search Outcomes

Measures

Task-completion time is computed from the moment of opening a browser and typing in the first query to the moment of answering the question. This includes the time it takes in typing queries, evaluating search results, clicking on one of the search results, evaluating the content of websites opened from the search results and finally typing the answers.

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 is 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 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).

Data of only those tasks was included in the analysis for which the participants successfully completed the tasks. 37 data points out of a total number of (12 tasks X 48 participants) = 576 data points (6.4%) were therefore dropped. For all the four dependent variables, a 2 (Age: Young vs. Old) X 2 (Task Difficulty: Simple vs. Difficult) mixed ANOVA was

conducted with age as between-subjects variable and task difficulty as within-subjects variable.

Results

Task-completion time Figure 2a shows the mean task-completion time in relation to age and task difficulty. The main effect of age was significant $F(1,46) = 579.37, p < .001$. The main effect of task difficulty was significant $F(1,46) = 90.07, p < .001$. The interaction of age and task difficulty was also significant $F(1,46) = 11.63, p < .001$.

Number of clicks Figure 2b shows the relationship between age and task difficulty for mean number of clicks. The main effect of age was significant $F(1,46) = 15.54, p < .001$. The main effect of task difficulty was significant $F(1,46) = 21.89, p < .001$. The interaction of age and task difficulty was significant $F(1,46) = 8.46, p < .01$.

Accuracy As can be seen from Figure 2c, the main effect of age was significant $F(1,46) = 17.18, p < .001$. The main effect of task difficulty was significant $F(1,46) = 196.4, p < .001$. Interaction of age and task difficulty was not significant ($p > .05$).

Number of reformulations Figure 2d shows the mean number of reformulations in relation to age and task difficulty. The main effect of age was significant $F(1,46) = 26.92, p < .001$. The main effect of task difficulty was significant $F(1,46) = 31.35, p < .001$. The interaction of age and task difficulty was significant $F(1,46) = 12.21, p < .001$.

Summarizing the main behavioral outcomes, we found that difficult tasks took significantly longer to complete, significantly more clicks were made, were answered significantly less accu-

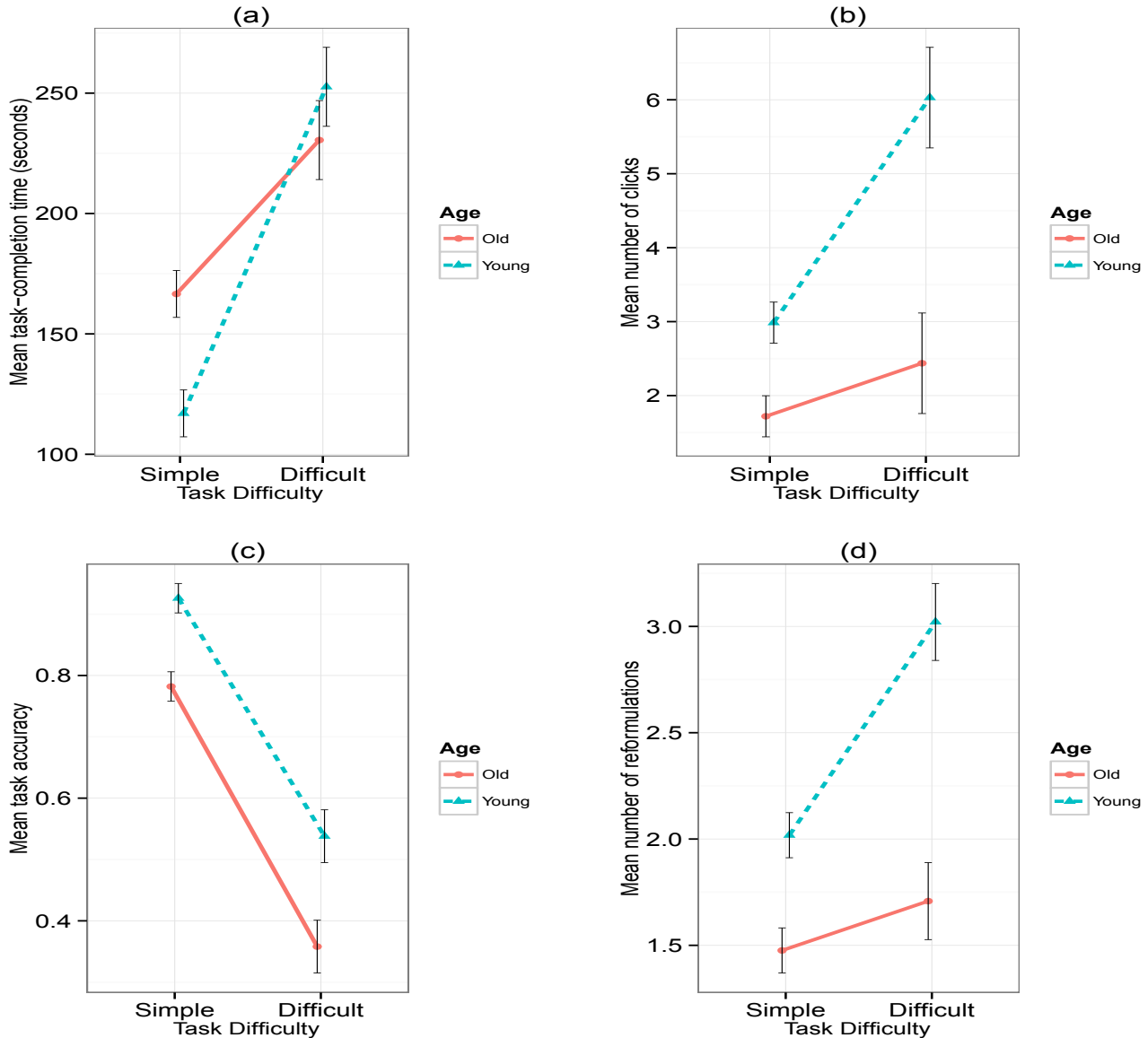


Figure 2. Analysis of search performance in terms of (a) task-completion time, (b) clicks, (c) accuracy and (d) number of reformulations in relation to age and task difficulty.

rately than simple tasks and significantly more reformulations were made than simple tasks. Young participants were found to be significantly faster in completing the tasks, especially the simple ones. Accuracy of young participants was significantly higher than that of old participants. Old participants were found to click significantly less and reformulate queries significantly less often than young participants. We next analyze the type of reformulation strategy and the semantic content of the queries during reformulations in more detail.

Analysis of Reformulation Strategies

Of all the different types of query reformulations, we focus on *generalization* and *specialization* in our study as these are concerned with modifying the content of the queries and are interesting from a cognitive perspective. We also had other

types of reformulations in our data such as spelling corrections (58), repeating same query again (7), converting a query string into its plural form (3), use of special operators such as AND or double quotes (2), reordering words in a query (2) which were dropped from the analysis as these were not many compared to the total number of reformulations we had (together 10.18%). We used a slightly modified version of generalization and specialization as defined in [24].

Measures

Generalization of Query If the user finds that the search results are too narrow, he/she normally generalizes the query by using superordinate terms or by removing some terms. Therefore, if Q_i and Q_{i+1} are two consecutive queries, Q_{i+1} is a generalization of Q_i if Q_i and Q_{i+1} have at least one query term

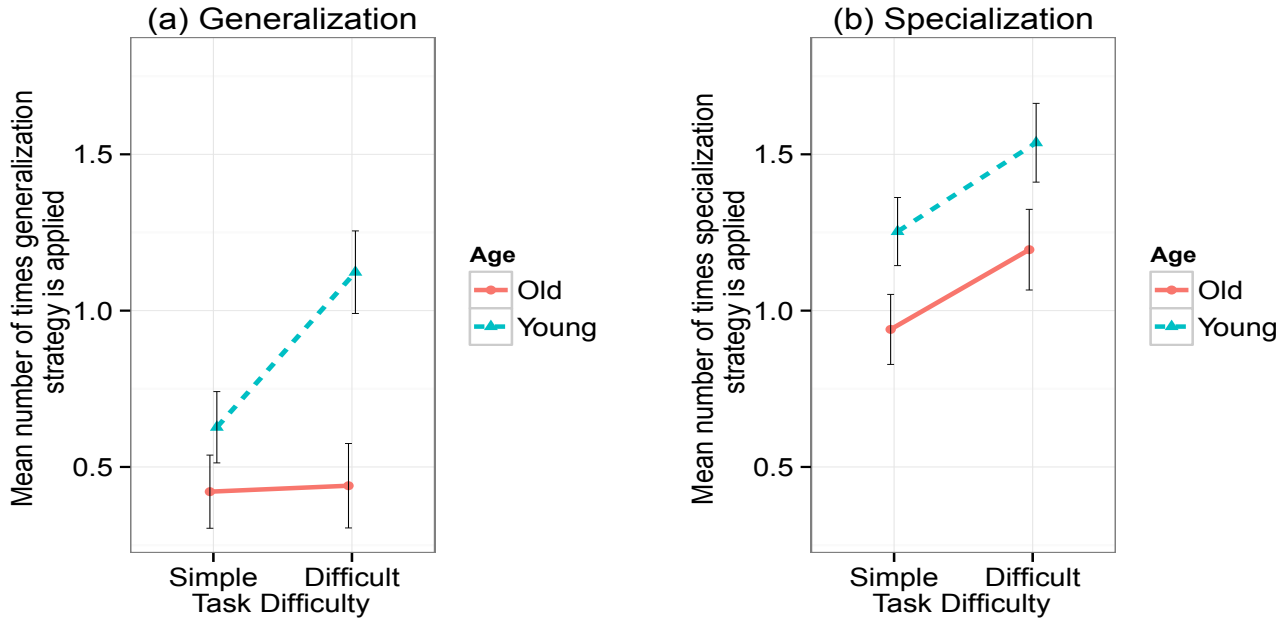


Figure 3. Mean number of times (a) generalization and (b) specialization strategies are applied by young and old participants in relation to task difficulty.

in common, Q_{i+1} contains fewer or equal number of terms than Q_i and/or Q_{i+1} contains some new terms not present in Q_i [24]. For the example task in the difficult category discussed above, if the queries are “*niersteen pijn in rug (kidney stones pain in the back)*”, “*branderig gevoel bij plassen (burning sensation when urinating)*”, “*blaasontsteking (urinary infection)*” then the number of times generalization strategy is applied is 3. However, if the queries are “*niersteen pijn in rug (kidney stones pain in the back)*”, “*pijn bij plassen (pain when urinating)*”, “*pijn bij plassen en gewichtsverlies (pain when urinating and weight loss)*”, then the number of times generalization strategy is applied is only 1.

Specialization of Query If the user is not satisfied with the search results, and finds them to be too generic, he/she normally makes the query more specific by adding new terms. Therefore, if Q_i and Q_{i+1} are two consecutive queries, Q_{i+1} is a specialization of Q_i if Q_i and Q_{i+1} have at least one query term in common, Q_{i+1} contains more terms than Q_i and/or Q_{i+1} contains some new terms not present in Q_i [24]. Taking the same example task as above, for the queries in the first case, the number of times specialization strategy was applied is 0 as all three queries were generated by using completely new terms. However, for the queries in the second case, the number of times specialization strategy is applied is 1 as the third query is generated by adding additional terms (weight loss) to the second query.

Semantic Relevance of Query (SRQ) For each task and each query corresponding to that task, semantic relevance was computed between the query and the target information sought using a technique called Latent Semantic Analysis (LSA, henceforth) [23]. LSA is a machine-learning technique that builds a semantic space representing a given user population’s under-

standing of words, short or whole texts by applying statistical computations, singular value decomposition and represents them as a vector in a multidimensional space of about 300 dimensions. The cosine value (+1 if identical and 0 if unrelated) between two vectors in this representation gives the measure of the semantic relatedness. It has been shown that higher semantic relevance between two texts relates to higher overlap in the meanings associated with those two texts. LSA has been successfully put to use in applications such as information retrieval, essay grading, synonymy tests, automatic tutors, etc [11].

We used 70,000 Dutch documents (consisting of 60% newspaper articles and 40% medical and health related articles) as a corpus to create first a semantic space in Dutch. The LSA values were then computed between a query and each of the headings and search snippets of search results from the corresponding SERP using this space. For a detailed procedure to compute semantic similarities, refer [20]. This is repeated for all queries of the task and a mean LSA value is computed. This is repeated again for all the tasks of a participant and finally for all the participants. This metric gives us an estimate of how close in semantic similarity the queries generated by the participants are to the target information.

Results

For all three dependent variables, a 2 (Age: Young vs. Old) X 2 (Task Difficulty: Simple vs. Difficult) mixed ANOVA was conducted with age as between-subjects variable and task difficulty as within-subjects variable.

Generalization of Query As shown in Figure 3a, the main effect of age was significant $F(1,41) = 9.08, p < .005$. Generalization strategy was applied significantly more often by young

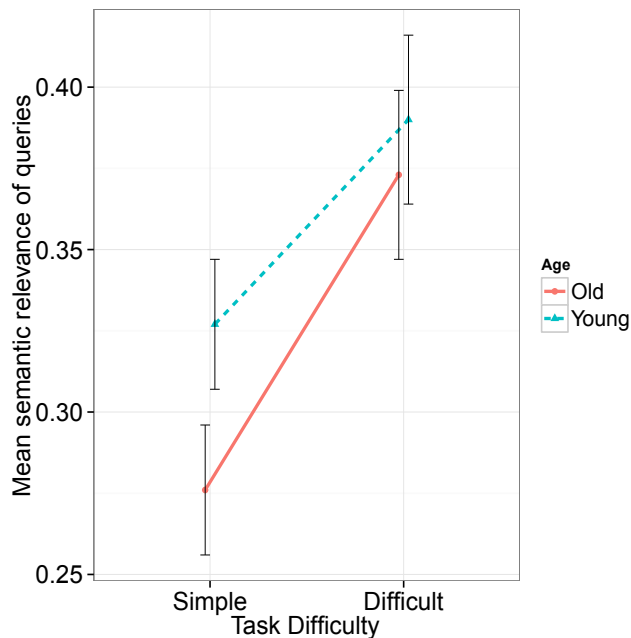


Figure 4. Mean semantic relevance of queries with target in relation to age and task difficulty.

participants compared to old participants. The main effect of task difficulty was significant $F(1,41) = 6.98, p < .05$. Difficult tasks demanded more generalizations than simple tasks. The interaction of age and task difficulty was significant $F(1,41) = 5.98, p < .005$. Post-hoc tests showed that the number of times the generalization strategy was applied by young participants was significantly more than that of old participants, especially for difficult tasks.

Specialization of Query As shown in Figure 3b, the main effect of age was significant $F(1,41) = 8.04, p < .01$. The specialization strategy was applied significantly more often by young participants compared to old participants. The main effect of task difficulty was significant $F(1,41) = 4.76, p < .05$. The number of times specialization strategy was applied for difficult tasks was significantly higher than that for simple tasks. The interaction of age and task difficulty was not significant ($p > .05$).

Semantic Relevance of Query It is evident from Figure 4 that the main effect of task difficulty was significant $F(1,46) = 9.47, p < .005$. The mean semantic relevance of queries for difficult tasks was significantly higher than that for simple tasks. The main effect of age and the interaction of age and task difficulty were not significant ($p > .05$).

We next analyzed the mean semantic relevance of the queries with the target information at a more granular level by looking at each reformulation cycle separately. First cycle corresponds to first query, second cycle corresponds to the second query, third cycle corresponds to the third query and so on. Mean semantic relevance was computed for all the queries of all the tasks of a particular type (simple and difficult separately), generated by young and old participants in each reformulation

cycle. It would have been interesting to perform this analysis separately also for generalization and specialization strategies, but we restrict ourselves to simple and difficult tasks for now because of lack of sufficient data. To achieve higher reliability, only those cycles were considered for which there were at least 5 queries (per reformulation cycle). By doing so, only 2.8% of data was excluded from the analysis. The resulting graphs are shown in Figure 5. We can observe from Figure 5 that young participants reformulated much longer than old participants.

We conducted 4 different one-way ANOVAs (for simple old, difficult old, simple young and difficult young) with cycle (reformulation cycle 1 vs 2 vs 3 and so on as repeated measures) as independent variable and mean LSA value of the SERP generated by queries in a reformulation cycle as dependent variable. The main effect of cycle on mean LSA value was weakly significant $F(2,56) = 2.55, p < .09$, for simple old and not ($p > .05$) for others. For simple old, post-hoc tests showed that the mean LSA value of the SERPs in reformulation cycle 3 was significantly less ($p < .05$) than that in the first cycle. There was also a significant linear trend $F(2,56) = 5.03, p < .05$, for simple old and a weaker trend for difficult old $F(1, 66) = 3.52, p < .07$, indicating that as old participants reformulated, they produced queries which were further away from the target information.

We next checked the difference in mean LSA value of the SERPs between the starting (first query) and ending points (last query) of a search session. We found no significant difference in the starting points of young and old participants, for both simple and difficult tasks. However, at the ending points, young participants ($M = .45, SD = .45$, for simple tasks and $M = .48, SD = .40$ for difficult tasks) end up significantly higher than old participants ($M = .14, SD = .17$ for simple tasks and $M = .12, SD = .14$ for difficult tasks), $t(20) = 2.3, p < .05$ (for simple tasks), $t(8) = 2.28, p = .06$ (for difficult tasks), indicating that the semantic relevance of queries with target information decreases with further reformulations for old participants whereas it remains constant for young participants.

Summarizing the analyses on reformulation strategies, we found that both generalization and specialization strategies were applied significantly more often for difficult tasks compared to simple tasks. Young participants were found to use the specialization strategy significantly more often than old participants. The generalization strategy was also used significantly more often by young participants, especially for difficult tasks. The semantic relevance of queries with the target information was found to be significantly higher for difficult tasks compared to simple tasks. Although the mean semantic relevance of queries with target information was overall similar for young and old participants, it remained constant for young participants and decreased for old participants with further reformulations. Also, young participants were found to reformulate much longer than old participants.

CONCLUSIONS AND DISCUSSION

This study was designed mainly to investigate deeper the query reformulation strategies and the content of the queries during



Figure 5. Mean semantic relevance of queries with target at each reformulation cycle for (a) simple and (b) difficult tasks.

reformulations in relation to age and task difficulty. A behavioral study in which 24 young and 24 old participants solved six simple and six difficult information search tasks was conducted. Analysis of behavioral outcomes showed that 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. Younger adults were found to be significantly faster in completing tasks, especially the simple ones. Accuracy of younger adults was significantly higher than that of older adults. Older adults were found to click significantly less and reformulate queries significantly less often than younger adults. These results were in line with the outcomes of previous studies [4, 10, 21] and therefore provide stronger foundation to the main contributions of this study.

We formulated three main research questions regarding reformulation strategies and the content of queries. First, how often do younger and older adults make use of generalization and specialization strategies when reformulating search queries? Because, reformulating a query is a cognitively effortful process, we hypothesized that younger adults would use both generalization and specialization more often than older adults. Second, how does the frequency of usage of generalization and specialization strategies when reformulating queries vary in

relation to difficulty of a task? We hypothesized that difficult tasks would require more reformulations and therefore more generalizations and specializations than simple tasks. Third, how does the semantic relevance of a query with the target information sought vary in relation to age and difficulty of a task? Based on the Construction-Integration model of text comprehension [22], difficult tasks would require activation of higher number of concepts (related to the task) in the Working Memory compared to simple tasks. Assuming that these activated concepts are included as terms in the search queries formulated by participants, we hypothesized that the semantic relevance of a query with the target information would be higher for difficult tasks compared to simple tasks. Also, because we think that younger adults would do much more activity in terms of number of search queries used, number of search results clicked and the number of pages of the websites opened from the search results explored, compared to older adults, we hypothesized that the semantic relevance of queries generated by younger adults would be higher than that of older adults.

Analysis of reformulation strategies showed that both generalization and specialization strategies were applied significantly more often for difficult tasks compared to simple tasks (**Hyp1**). Younger adults were found to use specialization strategy sig-

nificantly more often than older adults. Generalization strategy was also used significantly more often by younger adults, especially for difficult tasks (**Hyp2**). The semantic relevance of queries with the target information was found to be significantly higher for difficult tasks compared to simple tasks (**Hyp3a**). We could not find evidence in our data that could show that the semantic relevance of queries used by younger adults would be higher than that of the queries used by older adults (**Hyp3b**). However, we found that the semantic relevance of the queries generated by younger adults remained constant and that of the queries generated by older adults showed a decreasing trend as they reformulated further.

A deeper analysis with type of reformulation (not reported above) as a factor revealed a significant main effect of type of reformulation $F(1, 174) = 45.63, p < .001$. Specialization strategy was used significantly more often than generalization strategy. The interaction of type of reformulation with age or task difficulty was not significant ($p > .05$).

The age related differences in the number of reformulations (generalizations and specializations) may be due to the lower fluid abilities of older adults compared to younger adults based on the performance on Trail Making Test. This needs to be verified by incorporating TMT scores into the analysis. Also, these results seem to be confirming the outcomes of [4] in which they found that the younger adults were more efficient in the controlling phase (reformulation of queries, coming out of unsuccessful searches etc.) of an information search task. The age related differences in the semantic relevance of queries in terms of the increasing trend for younger adults and decreasing trend for older adults as they reformulate further, may be due to the fact that younger adults were found to use specialization and generalization strategies more often than older adults. Queries in later cycles do not bring older adults closer to the target information but take them even farther away (lower SRQ). They seem to use keywords which have, over cycles, decreasing semantic relevance to the target information, in contrast to younger adults. This applies not only to simple but also to difficult tasks. We also saw that the performance in terms of accuracy was lower for older adults. An explanation of lower accuracy of older adults could be that older adults do not succeed in generating queries relevant to the target information which in turn could be because they are not able to apply proper reformulation strategies like specialization and generalization. We did indeed find significant age related differences for both.

IMPLICATIONS FOR SYSTEM DESIGN

The results from this study have implications to the design of systems that support performing information search tasks using the Internet. It is important to take into account the difficulties faced by older adults when designing such systems. First, if the search systems could detect that search results are too narrow or too broad and accordingly give suggestions to generalize or to specialize, this would be very helpful, especially for old people. Second, one can imagine a search engine monitoring semantic relevance of SERPs with respect to the queries in a search session. Whenever it shows a decreasing trend or falls below a threshold value, the system

can automatically warn users to consider reformulating that query. A third possibility would be to provide support along the lines of [19] in which the semantic similarity between the user goal and the hyperlink text has been successfully used to generate navigation support. Highly similar links based on their LSA value between the goals and the link were emphasized to the participant. It would be interesting to complete this approach with similar support on the search result pages (by emphasizing results with high LSA value to the goal) in order to enhance search performance. These three types of interventions could enhance particularly for older adults their search performance.

ACKNOWLEDGMENTS

We would like to thank Guillermo de Jorge Botana from Department of Evolutionary and Educational Psychology, National Distance Education University, Madrid, Spain and Ricardo Olmos Albacete from Department of Social and Methodological Psychology, Autonoma University of Madrid, Madrid, Spain, for help with creating a semantic space. We thank Floortje van Oosten for assistance in testing participants and for her expertise in judging the adequacy of materials. This research was supported by Netherlands Organization for Scientific Research (NWO), ORA Plus project MISSION (464-13-043), and carried out in collaboration with University of Toulouse and University of Illinois.

REFERENCES

1. Anne Aula. 2005. User study on older adults use of the Web and search engines. *Universal Access in the Information Society* 4, 1 (2005), 67–81.
2. Pia Borlund and Peter Ingwersen. 1997. The development of a method for the evaluation of interactive information retrieval systems. *Journal of documentation* 53, 3 (1997), 225–250.
3. Aline Chevalier, Aurélie Dommès, and Jean-Claude Marquié. 2011. Information searching on the web: the cognitive difficulties experienced by older users in modifying unsuccessful information searches. In *Engineering Psychology and Cognitive Ergonomics*. Springer, 225–232.
4. Aline Chevalier, Aurélie Dommès, and Jean-Claude Marquié. 2015. Strategy and accuracy during information search on the Web: Effects of age and complexity of the search questions. *Computers in Human Behavior* 53 (2015), 305–315.
5. Aline Chevalier, Aurélie Dommès, and Daniel Martins. 2013. The effects of ageing and website ergonomic quality on internet information searching. *Ageing and Society* 33, 06 (2013), 1009–1035.
6. Jessie Chin, Evan Anderson, Chieh-Li Chin, and Wai-Tat Fu. 2015. Age differences in information search: An exploration-exploitation tradeoff model. In *Proceedings of the Human Factors and Ergonomic Society (HFES 2015)*. 85–89.
7. Jessie Chin and Wai-Tat Fu. 2010. Interactive effects of age and interface differences on search strategies and

- performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 403–412.
8. Jessie Chin, Wai-Tat Fu, and Thomas Kannampallil. 2009. Adaptive information search: age-dependent interactions between cognitive profiles and strategies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1683–1692.
 9. Jérôme Dinet, Eric Brangier, Gabriel Michel, Robin Vivian, Sophie Battisti, and Rémi Doller. 2007. Older people as information seekers: Exploratory studies about their needs and strategies. In *Universal Access in Human Computer Interaction. Coping with Diversity*. Springer, 877–886.
 10. Aurelie Dommès, Aline Chevalier, and Sarah Lia. 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, 5 (2011), 717–726.
 11. Susan T Dumais. 2004. Latent semantic analysis. *Annual review of information science and technology* 38, 1 (2004), 188–230.
 12. Jacqueline K Eastman and Rajesh Iyer. 2004. The elderly's uses and attitudes towards the Internet. *Journal of Consumer Marketing* 21, 3 (2004), 208–220.
 13. Wai-Tat Fu and Peter Pirolli. 2007. SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human-Computer Interaction* 22, 4 (2007), 355–412.
 14. John L Horn. 2012. The Theory of Fluid and Crystallized Intelligence in Relation to Concepts of Cognitive Psychology and Aging in. In *Aging and cognitive processes*. Vol. 8. Springer Science & Business Media, 237–278.
 15. John L Horn and Raymond B Cattell. 1967. Age differences in fluid and crystallized intelligence. *Acta psychologica* 26 (1967), 107–129.
 16. Rong Hu, Kun Lu, and Soohyung Joo. 2013. Effects of topic familiarity and search skills on query reformulation behavior. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–9.
 17. Jeff Huang and Efthimis N Efthimiadis. 2009. Analyzing and evaluating query reformulation strategies in web search logs. In *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 77–86.
 18. Bernard J Jansen, Amanda Spink, and Bhuvan Narayan. 2007. Query modifications patterns during web searching. In *Information Technology, 2007. ITNG'07. Fourth International Conference on*. IEEE, 439–444.
 19. Ion Juvina and Herre van Oostendorp. 2008. Modeling semantic and structural knowledge in Web navigation. *Discourse Processes* 45, 4-5 (2008), 346–364.
 20. Saraschandra Karanam, Herre van Oostendorp, and Bipin Indurkha. 2012. Evaluating CoLiDeS+ Pic: the role of relevance of pictures in user navigation behaviour. *Behaviour & Information Technology* 31, 1 (2012), 31–40.
 21. Saraschandra Karanam, Herre van Oostendorp, Mylène Sanchiz, Aline Chevalier, Jessie Chin, and Wai Tat Fu. 2015. Modeling and predicting information search behavior. In *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*. ACM, 7.
 22. Walter Kintsch. 1998. *Comprehension: A paradigm for cognition*. Cambridge university press.
 23. Thomas K Landauer, Peter W Foltz, and Darrell Laham. 1998. An introduction to latent semantic analysis. *Discourse processes* 25, 2-3 (1998), 259–284.
 24. Chang Liu, Jacek Gwizdzka, Jingjing Liu, Tao Xu, and Nicholas J Belkin. 2010. Analysis and evaluation of query reformulations in different task types. *Proceedings of the American Society for Information Science and Technology* 47, 1 (2010), 1–9.
 25. David Mellor, Lucy Firth, and Kathleen Moore. 2008. Can the Internet Improve the Well-being of the Elderly? *Ageing international* 32, 1 (2008), 25–42.
 26. Tracy L Mitzner, Julie B Boron, Cara Bailey Fausset, Anne E Adams, Neil Charness, Sara J Czaja, Katinka Dijkstra, Arthur D Fisk, Wendy A Rogers, and Joseph Sharit. 2010. Older adults talk technology: Technology usage and attitudes. *Computers in human behavior* 26, 6 (2010), 1710–1721.
 27. Sophie Monchaux, Franck Amadieu, Aline Chevalier, and Claudette Mariné. 2015. Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing & Management* 51, 5 (2015), 557–569.
 28. Christopher Olston and Ed H Chi. 2003. ScentTrails: Integrating browsing and searching on the Web. *ACM Transactions on Computer-Human Interaction (TOCHI)* 10, 3 (2003), 177–197.
 29. Richard Pak and Margaux M Price. 2008. Designing an information search interface for younger and older adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50, 4 (2008), 614–628.
 30. Ira Puspitasari, Koichi Moriyama, Ken-ichi Fukui, and Masayuki Numao. 2015. Effects of Individual Health Topic Familiarity on Activity Patterns During Health Information Searches. *JMIR medical informatics* 3, 1 (2015).
 31. Tara L Queen, Thomas M Hess, Gilda E Ennis, Keith Dowd, and Daniel Grünh. 2012. Information search and decision making: Effects of age and complexity on strategy use. *Psychology and aging* 27, 4 (2012), 817.

32. Soo Young Rieh and others. 2006. Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing & Management* 42, 3 (2006), 751–768.
33. Karin Slegers, Martin PJ Van Boxtel, and Jelle Jolles. 2012. Computer use in older adults: determinants and the relationship with cognitive change over a 6year episode. *Computers in Human Behavior* 28, 1 (2012), 1–10.
34. Esther Strauss, Elisabeth MS Sherman, and Otfried Spreen. 2006. *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press, USA.
35. Jaime Teevan, Eytan Adar, Rosie Jones, and Michael AS Potts. 2007. Information re-retrieval: repeat queries in Yahoo’s logs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 151–158.
36. Jing-Jen Wang and Alan S Kaufman. 1993. Changes in fluid and crystallized intelligence across the 20-to 90-year age range on the K-BIT. *Journal of Psychoeducational Assessment* 11, 1 (1993), 29–37.
37. Robert J Youmans, Brooke Bellows, Christian A Gonzalez, Brittany Sarbone, and Ivonne J Figueroa. 2013. Designing for the wisdom of elders: age related differences in online search strategies. In *Universal Access in Human-Computer Interaction. User and Context Diversity*. Springer, 240–249.
38. Panayiotis Zaphiris, Sri Kurniawan, and Mariya Ghiawadwala. 2007. A systematic approach to the development of research-based web design guidelines for older people. *Universal Access in the Information Society* 6, 1 (2007), 59–75.