

# Mental Models: Have Users' Mental Models of Web Search Engines Improved in the Last Ten Years?

Sifiso Mlilo and Andrew Thatcher

University of the Witwatersrand Psychology, Wits, 2050, South Africa  
sifiso.mlilo@gmail.com, Andrew.Thatcher@wits.ac.za

**Abstract.** This study investigated the accuracy and completeness of mental models users have of Web search engines in the context of a comparison of matched data obtained from samples from 2000 and 2010. The performance measures time, steps and accuracy were assessed along with 17 salient features of Web search engines identified in the study conducted in 2000. The results indicated that the 2010 sample had improved significantly across all performance measures. The two samples did, however, identify an equal number of salient features ( $N=7$ ). It was clear from the detailed analyses of the salient features though, that the accuracy and completeness of users' mental models of search engines had demonstrably improved from 2000. So, while users' mental models of Web search engines still remain largely inaccurate and incomplete, their alignment with designer's conceptualisations has improved.

**Keywords:** Mental models; search engine; time; steps; accuracy; salient features.

## 1 Introduction

Web search engines (WSEs) are now recognised as the dominant information seeking tool utilised by people using the Web [1][2]. A number of studies interested in the information retrieving methods employed by users have long recognised this fact [2]. As such, there has been a shift from the focus on users' mental models of the Web in general [3] towards an active interest in the kinds of mental models users have of online library databases [2][4] and WSEs more specifically [1][5][6][7][8][9]. A number of studies that have looked at mental models of WSEs have either done so in a rudimentary fashion [5][7][10] or the mental models were inferred indirectly from log-file data on user queries submitted to WSEs [1][11][12][13]. The log-file analysis (e.g. [13]) found that users' query formulations (and re-formulations) indicated a poor understanding of the manner in which WSEs work. Similar results have been found by researchers like Muramatsu and Pratt [7] with query-based analyses in user studies. Muramatsu and Pratt [7] concluded that users' mental models were "naïve and erroneous" (p. 223) and suggested that WSE interface designers make the query transformations more transparent for their users. Efthimiadis and Hendry [11] found a wide variation in users' mental models of WSEs but in general found them to be

simple with few users understanding complex WSE concepts such as link analysis and query parsing. Crudge and Johnson [5] provided a fairly detailed summary of 10 users' mental models of WSEs using a repertory grid and a laddering technique. However, they did not provide any value judgement as to the completeness or accuracy of these mental models or their ability to aid the search process. Zhang [9] found that undergraduates' mental models of WSEs also varied in complexity and accuracy with most users showing fairly naïve mental models that were frequently incorrect. Thatcher and Greyling [8], examining a sample of 80 users, concluded that users' mental models of WSEs were incomplete and inaccurate. Users in their sample frequently misunderstood how search terms worked, used search terms inappropriately (e.g. used Boolean logic operators incorrectly or searched using whole phrases), misunderstood relevance feedback, and rarely used different WSEs to verify information. The general sense from these studies is that users' interactions with WSEs indicate a poor understanding of how they work and hence poor mental models.

When an individual interacts with a particular system they form ideas about how that system works. The more they learn about the system the more likely it is that they will figure out how that system works and respond to certain commands, actions or behaviours. In figuring out how the system works individuals are able to anticipate system responses prior to even engaging the system or giving it the relevant commands. The more a user is exposed to a particular system the better the opportunity for mental models to become more accurate and complete, although a lot depends on the type of exposure and the transparency of the system interface. Experience or knowledge of systems in existence prior to direct exposure also helps users in understanding the new systems they are faced with. All other studies of mental models of WSEs have used a cross-sectional design. This study sought to explore whether mental models of WSE systems have, over time, become more aligned with the designers' conceptualisations of how they work. Applying the same methodology used by Thatcher and Greyling [8], with data collected in 2000, this study aimed to determine how time (and by inference, greater experience) has influenced WSE mental model formation over the last decade. Much has changed in the domain of Web searching since data were collected in 2000. The Web now has far more webpages and websites; this means that webspace is more complex but it also means there are more points where information can be found. Connectivity, for most parts of the World, is considerably faster, making it quicker for users to find information without getting frustrated by slow download speeds. Importantly for mental model formation, users are more likely to have been exposed to WSEs for longer periods of time. Finally, the WSE landscape has changed considerably where no single WSE dominated the Web search environment in 2000. Since the exposure people have to the Web is likely to have increased quite significantly since the Thatcher and Greyling [8] investigation, a meaningful platform was set to assess whether the changes since 2000 have seen a closer alignment of users' mental models with designers' conceptual models of WSEs.

## 2 Method

### 2.1 Sample

For the 2010 sample a total of 80 students, scholars, professionals and semi-professionals, were selectively targeted for voluntarily participation in the study. This purposive sampling technique was used to attempt to match the 2010 sample as closely as possible to the 80 participants in the 2000 sample. Respondents were matched as closely as possible on age, language, gender, and occupation (e.g. student, scholar and type of occupation). In the 2000 sample there were 50 male participants (48 male respondents in 2010), a mean age of 23.28 years (25.78 years in 2010), 51 respondents who spoke English (50 in 2010), 25 who spoke an African language from South Africa (24 in 2010), and 2 who spoke a language from elsewhere in Africa (5 in 2010). Since the sample in 2000 was anonymous it was not possible to trace the exact same participants which would have been ideal. The 2010 sample must therefore be considered as a contrast group.

### 2.2 Procedure

Upon their arrival at the location where the data was being collected and following their completion of the biographical and composite WWW experience questionnaire, participants were asked to complete a directed search task while having their search actions recorded by HyperCam3 onscreen capture software. Because there was a possibility that some participants (expert users or subject experts) would not use a WSE for the directed search task, a secondary task (i.e. a general purpose browsing task) requiring them to use a WSE was also put in place. Only two participants each in 2000 and 2010 did not use a WSE for the directed search task and therefore also completed the secondary task. In these two cases the directed search task was used for the performance measure and the general purpose task was used to help establish participants' mental models. The primary, directed search task was: "Find Bill Clinton's mother's maiden name" (the secondary, general purpose browsing task was: "Find all the information on the relationship between carbon monoxide and desertification"). Following the completion of the task/s, participants had their onscreen search actions played back to them and they were asked why they had engaged in these actions (focusing on the WSE actions). Notes of their responses were taken by the researchers during this process. All participants started the task from the same webpage. After each participant had completed the task the researcher cleared all browsing data (browsing history, download history, emptying the cache, etc) to make sure the following participant could not follow the search path. Participants were then asked to provide an illustration of how they thought a WSE worked accompanied by a brief written description to complement their illustrations. Active participation in this process (from the researcher briefing participants through to the drawing of how a WSE works) ranged from 15 to 55 minutes.

### 2.3 Measures and Analysis

As this research study was comparing performance for the two samples from 2000 (T1) to 2010 (T2) the number of steps and time taken to complete the directed search task were analysed using a t-test [14]. A chi-squared test was used to assess the accuracy of the answer given by participants comparing T1 to T2 [14]. The salient features were identified from a combination of content analysis of the retrospective verbal protocols and participant drawings/descriptions, Marchionini's [15] conceptual framework of the electronic information-seeking process (i.e. choose search system → formulate query → execute search → examine results), the emergent features from the 2000 data [8], and empirical studies on WSE functioning (e.g. [5][16][17][18]). Two raters working independently assigned the 17 salient features to participants (the 17 salient features are provided in the results). The weighted Kappa coefficient of 92% (91% in 2000) indicated a high level of agreement between the two raters. All remaining discrepancies in the allocation of salient features was done by consensus. In order to determine the mental model clusters for the 2010 sample Wards cluster analysis was conducted using the same salient features identified in the original study as no new features had emerged. As the clusters that emerged were not identical at T1 and T2 a descriptive analysis using chi-squared comparisons of the individual salient features of the respective samples (i.e. potential differences in salient feature 1 at T1 vs. T2) was used.

## 3 Results

### 3.1 Performance Measures

Significant differences were found for all the performance measures, with the T2 sample requiring significantly fewer steps and less time to complete the directed search task. Since the T2 sample was significantly more experienced using the Web, experience was added as a covariate. The performance effects were still significant after accounting for experience, suggesting that the performance improvements were due to some other factor such as improved mental models or improvements in WSE functioning rather than increased experience directly. The T2 sample was also significantly more accurate in the answers they provided.

**Table 1.** Performances differences on the search task from T1 to T2

	Mean T1 (SD)	Mean T2 (SD)	t-statistic
Time since using Web (months)	36.86	119.42	14.82**
Weekly usage of Web (hours)	7.35	16.44	5.49**
Self-rated Web exp. (1-5 scale)	3.32	3.70	2.86*
Steps	14.25 (7.80)	8.48 (4.97)	5.59**
Time	497.10 (272.6)	222.20 (155.9)	7.83**
Correct answers	N=46	N=59	4.68*

\*\*  $p < 0.01$ ; \*  $p < 0.05$ .

The WSE landscape was vastly different from T1 to T2. Participants at T2 either used Google (N=69), Bing (N=8), or Yahoo (N=1) as their WSE to complete the task. At T1 participants used a much wider range of WSEs including Yahoo (N=29), Altavista (N=9), Looksmart (N=9), Infoseek (N=9), Askjeeves (N=8), Google (N=3), Lycos (N=2), Hotbot (N=1), Excite (N=1), Dogpile (N=1), and Metacrawler (N=1). Of course, some of these WSEs were no longer in operation at T2.

### 3.2 Cluster Analyses of Salient Features (T1 vs. T2)

The 3 clusters for the T2 sample shared a few similarities with the 4 clusters found in the T1 sample. Like in the T1 clusters, a clear indication of an increment in the quality of mental models across the T2 clusters was present. The clusters (1 and 2 at T1, N=17 and N=16 respectively and 1 at T2, N=27) with the simplest mental models in each respective sample showed little understanding of most of the salient features, with a poor understanding of a WSE as a database and the existence of multiple WSEs. However, unlike clusters 1 and 2 in the T1 sample, where there were no participants who showed an understanding that WSE results were ranked, close to 50% of cluster 1 participants in the T2 sample demonstrated this understanding. That is, even in the cluster with the simplest mental models at T2, the T2 sample had slightly more complex mental models. Indeed, there were fewer participants in this cluster at T2 (N=27) compared to the two simplest mental model clusters at T1 (N=33). Cluster 2 of the T2 sample (N=34), which demonstrated more advanced mental models than cluster 1, indicated a substantial improvement regarding the grasp of salient features. Cluster 2 of the T2 sample represents a greater appreciation for Boolean logic and a greater understanding of search terms matching a webpage or database over and above the qualities shared with cluster 1 of the T2 sample. Cluster 3 of the T1 sample (N=21) emphasised the fact that different WSEs produced different results. Cluster 3 at T1 was thus analogous to cluster 2 at T2.

Regarding the most advanced clusters for both samples, a clear appreciation of WSEs as databases that collect information from the Web was apparent. Also, the importance of keywords and knowledge that WSEs rank results was present in both samples. Cluster 4 of the T1 sample (N=26) emphasised the importance of keywords matching with databases (as did a large proportion of the cluster 2 T2 sample). These qualities were also emphasised in cluster 3 of the T2 sample (N=18), along with more advanced features such as the importance of changing/modifying search terms to influence the quality of the search outcome and the recognition of multiple WSEs (although this latter feature was not as strongly present as for clusters 3 and 4 in the T1 sample).

### 3.3 Presence of Each Salient Feature: T1 to T2

**Salient features more prevalent at T2.** Significantly more participants at T2 chose a WSE instead of allowing the Web browser to decide. At T1 a large proportion of the respondents did not have a WSE preference and instead clicked on the browser's "search" button. At T1 this had the effect of the browser randomly assigning one of four different WSEs to the user. Significantly more participants at T2 showed an understanding of the Web as the collection for a database. The participants recognised

that the WSE did not search the entire Web (perhaps because the Web is known to be practically too large to search in its entirety and because it changes so rapidly) but gave a representative sample of the Web. More participants at T2 chose key words or phrases to perform the search task. This is because at T1 more participants chose WSE categories and did not enter search terms at all (effectively allowing the WSE to refine the search using predefined categorisation algorithms). At T2 the emphasis was more on user-defined search terms. Similarly, a greater number of participants at T2 showed an appreciation of the impact reformulating search terms had on modifying (widening or narrowing) the search. A larger number of participants at T2 indicated an understanding that WSEs matched search terms and phrases to indicators/tags that the WSE had defined. The use of bold and highlighted terms in WSE summary results probably explains why this occurs. More participants at T2 showed an understanding that WSE algorithms rank results. This was evident in the fact that very few participants went beyond the first page of results and often did not scan below the first four or five search results. More participants from T2 showed an understanding that different WSEs have different ranking algorithms even if this number was, technically speaking, quite small ( $N=7$  at T2 and  $N=0$  at T1). A small number of respondents recognised the differences between WSEs and were more likely to use different WSEs to verify results.

**Salient features more prevalent at T1.** More participants at T1 understood that different WSEs have different database algorithms (i.e. use different WSE algorithms to collect records from the Web to form a database). Participants at T1 were also more likely to browse the WSE categories to widen/narrow the search domain before submitting search queries (and in some instances did not even submit search query terms). WSEs at T2 largely hide the WSE categories, placing them near the top of the screen and away from the main searching area. Participants at T1 were more likely to use Boolean logic and other operators to change their search parameters. This is primarily because the most common Boolean logic term (“AND” was automatically built into most WSEs at T2). T1 participants were more likely to understand that WSEs match terms and phrases to actual words or phrases in webpages themselves (i.e. not to words and phrases that WSEs have catalogued and tagged). More participants at T1 understood that WSEs match terms and phrases with the webpage title. Participants at T2 appeared to focus more on the search terms being prevalent in the actual document content rather than just in the headings (or meta-tags). Significantly more participants from T1 indicated an understanding that WSEs allow users to narrow search results by finding “similar” results (other relevant/connected results). This might be because WSEs have improved their database collection methods to the point where results are found without having to resort to “similar” search terms. Some WSEs also have a scroll-down bar with search term hints. It could be that these functionalities replace the need for searching for “similar” results. More participants from T1 understood that WSEs display a hyperlink to the original location of the information. The number of participants who knew this at T1 was still relatively small ( $N=22$ ), but significantly larger than at T2 ( $N=1$ ). This could be because highlighting/bolding the search terms captures greater visual attention and less attention is directed towards the hyperlink.

**Salient features the same at T1 and T2.** Three salient features remained statistically unchanged. Participants were equally unlikely to recognise that a WSE actually searches a collection from the Web (rather than the whole Web). Likewise, respondents were equally unlikely to recognise that a WSE would use a different collection algorithm to construct the database from the Web (although in both these instances, respondents at T2 reflected slightly more instances of this salient feature than at T1). Finally, no respondents at T1 or T2 understood that a WSE also searches for alternative combinations and extensions of search terms. This is a surprising finding given that the dominant WSE often suggests alternative spellings to search terms with the phrase “Do you mean ... ?”.

**Table 2.** Salient features T1 vs. T2

Salient feature	Frequency		$\chi^2$
	T1	T2	
Chooses a WSE rather than letting the web browser decide (SF1)	36	74	42.00**
Recognises that a WSE is a database (SF2)	10	17	N.S.
Understanding of the web as the collection place for a database (SF3)	26	61	30.86**
Understands that WSEs use a specific algorithm to collect the database (SF4)	8	12	N.S.
Recognises that different WSEs have different databases (SF5)	73	9	102.46**
Participants choose keywords or phrases to search (SF6)	22	73	67.39**
Participants browse WSE categories to widen/narrow the search domain (SF7)	11	3	5.00*
Participants use Boolean logic or other operators to change search parameters (SF8)	52	16	33.14**
Understanding that search terms can be changed/modified to widen/narrow the search of the Web or database (SF9)	6	39	33.66**
WSE matches terms/phrases to the identifiers on the webpage or database (SF10)	4	41	42.32**
WSE matches terms/phrases to words in webpage or Web document (SF11)	54	2	74.28**
WSE matches terms/phrases with the webpage title (SF12)	12	1	10.13**
WSE looks for terms/phrases and related terms/phrases or extensions (SF13)	0	0	N.S.
WSE algorithm ranks results (SF14)	4	30	25.24**
Different WSEs have different ranking algorithms (SF15)	0	7	7.32**
WSEs allow users to narrow search results by finding “similar” results (SF16)	18	1	17.26**
WSE displays a hyperlink to the original location of the information (SF17)	22	1	22.39**

\*\*  $p < 0.01$ .

## 4 Discussion

While seven of the seventeen salient features were more frequent at T1 and the same number were more frequent at T2 there have been clear improvements regarding the completeness and accuracy of WSEs. First, the marked improvements in performance by the 2010 sample provide indirect evidence that mental models of WSEs have improved over the last decade. These performance improvements give an indication that users' interactions with the system have truly become more efficient partly as a function of greater exposure and frequent use; a finding consistent with that of Muramatsu and Pratt [7]. Regarding the equal distribution of salient features, there has been a clear, significant improvement in the comprehension of the various salient features that make up a WSE a decade later. Thatcher and Greyling [8] identified four clusters, two of which displayed highly incomplete and inaccurate mental models of WSEs, only the first cluster of the 2010 sample shared that quality, although even in that cluster, mental models were slightly more complete. The remaining two clusters (52 participants in total at T2 compared to 47 participants in clusters 3 and 4 at T1) showed more accurate and complete mental models of WSEs. The chi-squared comparisons are an indication of the state of WSEs in their respective eras and the demands they placed on their users. A large proportion of users in more recent times have gravitated towards Google with considerable success [19]. Thatcher and Greyling [8] commented on how WSEs did not make many of their salient features transparent and thus made it difficult for users to develop accurate and complete mental models. It would seem not much has changed in that regard as the number of salient features the respective samples were able to identify remained the same. This concern is minor and distracts from a far more important point. It would seem that even though WSEs still hide a great number of their features, they have configured their design specifications to such a high level that this need has been largely neutralised. It would seem that users have become faster, more efficient, and more accurate in their search processes despite still holding relatively naive mental models. It would appear that the salient features that occurred less frequently at T2 are either primarily technical or they are simply unimportant for search performance.

The poor user query formulations noted by a number of authors (e.g. [2][9][12][13]) from which they inferred that inaccurate information was consistently being retrieved was not demonstrated as a concern because of the simpler, if still fairly ambiguous, nature of WSE interfaces. Previous research has been critical of WSE interfaces, suggesting that they fail to deal with cultural and situational contexts as well as the naturally iterative (i.e. users changing, refining, or expanding their search criteria) nature of the search process [19]. Rose [20] suggests that WSE interfaces should be redesigned to invite the refinement of search terms, the use of exploratory searching, and should allow users to select search contexts depending on their search goals. Vaughan and Resnick [21] propose that WSE interfaces should also allow users to easily view their search history as this helps them understand what new search queries to try. As was reported in the introduction, Muramatsu and Pratt [7] found that making the WSEs' transformation of users' queries more transparent assisted in more accurate mental models and improved search performance. WSE designers have already made several modifications (e.g. crawlers driven by 'relevance' principles, drop-down menus, larger databases, automatic search



term/phrase correction feedback, etc) that have meant users get the most out of searching without fully comprehending much about the underlying mechanics of WSEs. The mental models held by users in the T2 sample, whilst reasonably better than those of the T1 sample, were still not accurate and complete. It is likely that the decreased time taken to find answers and the reduced number of steps is related to these WSE interface modifications. However, increased download speeds and the increased amount of information on the Internet would also have also contributed in this regard. We would argue, that despite these improvements there are still opportunities for WSE interface designs to become more transparent, to allow more accurate and complete mental models and thereby improve search performance further. There are a number of implications both practically and theoretically from these results.

First, it would seem that having a better mental model of a particular system (even if only moderately better) does in fact help in improving performance in using that system. Second, the results also suggest that systems can be designed to negate the need for highly accurate and complete mental models. However, considering that systems can seemingly overcome the need for effective user mental models for optimised system use there is a theoretical question concerning the real importance of mental models. A large body of literature spanning multiple disciplines has consistently proposed that accurate and complete mental models are essential for effective system engagement [2][22][23]. Indeed, the mediocre search performance of the T1 sample would also give credence to that particular argument. However, the marked differences in performance, even in the presence of the relatively modest improvements in mental models of WSEs, indicate that this is not necessarily true. So, it is possible that high level mental model formation is not always necessary for effective system use, at least regarding directed search tasks. Third, the results show that users' mental models have increasingly aligned better with designers' conceptual models, albeit marginally.

While every attempt has been made to match the T1 and T2 samples the best that can be achieved with this research design is a contrast group. It would have been ideal, from a longitudinal research design perspective, to have sampled the same 80 people ten years later. However, given that participants were anonymous at T1, such a longitudinal design was not possible. It is possible that any variations in performance and mental models were due to using different participants at T1 and T2. The sample size, while relatively good for this type of laboratory-based investigation, is still relatively small in comparison to the population of WSE users. The generalisation of the study findings to other users therefore requires further verification. This study used only one type of search task to determine search performance (i.e. a directed search task determined by the researchers). It is likely that search behaviour varies according to the type of search task and whether it is internally or externally assigned. Further research would have to be conducted on the performance of Web searching using a greater variety of search tasks. Finally, it is worth noting that it is difficult to causally link Web search performance to mental model complexity and accuracy in this study. Search performance may just as easily have improved due to drastically increased download speeds, the enormous increase of available Web content, and

improvements in the algorithms that WSEs use to collect, categorise, and rank their databases. Despite these limitations, the study clearly demonstrates that users' mental models of WSEs are now more complex, complete, and accurate.

## References

1. Jansen, B.J., Spink, A.: How Are We Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs. *Inf. Proc. Manag.* 42, 248–263 (2006)
2. Slone, D.J.: The Influence of Mental Models and Goals on Search Patterns During Web Interaction. *Journal of the J. Am. Soc. Inf. Sc. Techn.* 53, 1152–1169 (2002)
3. Thatcher, A., Greyling, M.: Mental Models of the Internet. *Int. J. Ind. Erg.* 22, 299–305 (1998)
4. Makri, S., Blanford, A., Gow, J., Rimmer, J., Warwick, C., Buchanan, G.: A Library or Just Another Information Resource? A Case Study of Users' Mental Models of Traditional and Digital Libraries. *J. Am. Soc. Inf. Sc. Techn.* 58, 433–435 (2007)
5. Crudge, S.E., Johnson, F.C.: Using the Repertory Grid and Laddering Technique to Determine the User's Evaluative Model of Search Engines. *J. Document* 63, 259–280 (2007)
6. Efthimiadis, E.N., Hendry, D.G.: Search Engines and How Students Think They Work. In: *Proceedings of the 28th Annual international ACM SIGIR Conf. Res. Dev. in Inf. Ret.* (2005)
7. Muramatsu, J., Pratt, W.: Transparent Queries: Investigating Users Mental Models of Search Engines. In: *Proceedings of the 24<sup>th</sup> annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 217–224 (2001)
8. Thatcher, A., Greyling, M.: Mental Models of Search Engines: How Do Search Engines Work? In: Harris, D., Duffy, V., Smith, M., Stephanidis, C. (eds.) *Human-centred Computing: Cognitive, social and ergonomics aspects*. Lawrence Erlbaum Associates Inc., Mahaw (2003)
9. Zhang, Y.: Undergraduate Students Mental Models of the Web as an Information Retrieval System. *J. Am. Soc. Inf. Sc. Techn.* 59, 2087–2098 (2008)
10. Liaw, S., Huang, H.: Information Retrieval From the World Wide Web: A User-Focused Approach Based on Individual Experience With Search Engines. *Comp. Hum. Beh.* 22, 501–517 (2006)
11. Spink, A., Jansen, B.J., Blakely, C., Koshman, S.: A Study of Results Overlap and Uniqueness Among Major Web Search Engines. *Inf. Proc. Manag.* 42, 1379–1391 (2006)
12. Spink, A., Jansen, B.J., Ozmutlu, H.C.: Use of Query Reformulation and Relevance Feedback by Excite Users. *Internet Res.: Elect. Net. App. Pol.* 10, 317–328 (2000)
13. Spink, A., Wolfram, D., Jansen, B.J., Saracevic, T.: Searching the Web: The Public and Their Queries. *J. Am. Soc. Inf. Sc. Techn.* 52, 226–234 (2001)
14. Huck, S.W.: *Reading Statistics and Research*, 4th edn. Pearson, Boston (2004)
15. Marchionini, G.: *Information Seeking in Electronic Environments*. Cambridge University Press, New York (1995)
16. Bar-Ilan, J., Keenoy, K., Yaari, E., Levene, M.: User Rankings of Search Engine Rankings. *J. Am. Soc. Inf. Sc. Techn.* 58, 1254–1266 (2007)
17. Crystal, A., Greenberg, J.: Relevance Criteria Identified by Health Information Users During Web Searches. *J. Am. Soc. Inf. Sc. Techn.* 57, 1368–1382 (2006)

18. Sullivan, D.: *How Search Engines Work* (2007),  
<http://searchenginewatch.com/2168031>
19. Search Engine Watch. *Top Search Providers for September 2010* (2010),  
<http://searchenginewatch.com/3634991>
20. Rose, D.E.: Reconciling Information-Seeking Behavior With Search User Interfaces for the Web. *J. Am. Soc. Inf. Sc. Techn.* 57, 797–799 (2006)
21. Vaughan, M.W., Resnick, M.L.: *Search User Interfaces: Best Practices and Future Visions*. *J. Am. Soc. Inf. Techn.* 57, 777–780 (2006)
22. Doyle, J.K., Ford, D.N.: *Mental Models Concepts Revisited: some Clarifications and a Reply to Lane*. *Syst. Dyn. Rev.* 15, 411–415 (1999)
23. Senge, P.M.: *The Fifth Discipline*. Currency Doubleday, New York (1990)