"Irrational" searchers and IR-rational researchers

Nils Pharo, Faculty of Journalism, Library and Information Science, Oslo University College, PO Box 4, N-0130 Oslo, Norway. Tel.: +47 22 45 26 84, Fax: +47 22 45 26 05, email: nils.pharo@jbi.hio.no (corresponding author)

&

Kalervo Järvelin, University of Tampere, Department of Information Studies, FIN-33014 University of Tampere, Finland; tel. +358-3-2156953, fax +358-3-2156560; email: kalervo.jarvelin@uta.fi

Abstract

In this article we look at the prescriptions advocated by Web search textbooks in the light of a selection of empirical data of real Web information search processes. We use the strategy of disjointed incrementalism, which is a theoretical foundation from decision making, to focus on how people face complex problems, and claim that such problem solving can be compared to the tasks searchers perform when interacting with the Web. The findings suggest that textbooks on web searching should take into account that searchers only tend to take a certain number of sources into consideration, that the searchers adjust their goals and objectives during searching, and that searchers reconsider the usefulness of sources at different stages of their work tasks as well as their search tasks.

1. Introduction

Web information searching (WIS) is quite complex due to a number of factors, including characteristics of the searcher, the environment in which the searcher works, the work task, the search tasks generated by the work task, and the characteristics of the search process. In order to help the searchers a large number of guides to web searching have been developed. Such guides are in online (e.g., www.searchengineshowdown.com and www.pandia.com) as well as in textbook formats (e.g., Ackermann & Hartman, 2003; Chowdhury & Chowdhury, 2001; Hock, 2001; Sherman & Price, 2001). The present article seeks to analyse how the search situations described in textbooks correspond with the actual search behaviours of users performing real work tasks. In particular we will focus on *how textbooks' prescriptions of search strategies relate to descriptions of the searchers' actual performance.* The article's title addresses the inconsistence between the textbook authors' (or researchers') advocacy of rational planning of search strategies and the actual behaviour of ordinary searchers, which may often seem irrational from the point of view of the IR specialists.

This paper is structured in the following way: In Part 1 (Section 2) we discuss WIS from the textbook perspective. It includes a presentation of methods, an analysis of how a selected set of textbooks address the domain, and a summary of our findings. In Part 2 (Section 3) we present some results of an earlier study (Pharo & Järvelin, 2004) on real Web interaction, which is analysed from the perspective of a theoretical foundation from Management Science called *disjointed incrementalism*. The section includes a summary of the earlier study (Pharo & Järvelin, 2004). In Section 4 we discuss the findings of the two studies and point out implications of the empirical analysis for WIS textbooks. Section 5 is the conclusion.

2. Part 1: the textbook approach to Web searching

2.1. The different Web IR systems

The rise of the Web as a tool in people's professional and everyday information searching has generated a large number of books and online guides on the subject. These books have different perspectives, in terms of their target groups and the type of sources they focus on. Some books target a general audience while others are aimed at specific professions, such as the Library and Information Science (LIS) professionals. The books also differ with respect to the pre-understanding they expect from the searchers; some books are meant for novices while others are for more experienced users.

In Table 1 we present a taxonomy of different IR systems with examples. In the text we refer to the different categories described in the table, e.g., Category number 1 represents a search engine indexing a local site whereas Category 11 is known as a subject index. We have filled in URLs which represent examples of the different kinds of IR services. An asterisk in a taxonomy cell

means that we have not found any examples of such services belonging to the cell although some might exist.

Table 1. Web IR systems taxonomy

Online guides to search engines primarily focus on query features of the best-known general search engines. Two examples are Searchengineshowdown and Pandia, both of which emphasise a user focus in their marketing, although Pandia, like e.g. Searchenginewatch, also focuses on "search engine optimisation", i.e., giving advice to webmasters on how to increase their sites' ranking in search engines' result lists.

The nature of online guides necessarily gives them the potential of always being up-to-date, which is obviously an advantage when dealing with the dynamic nature of the Web. In addition they may implement "hands-on" help features, e.g., include search engine query forms next to the guiding text. On the other hand the textbooks are in many ways handier, especially for users who are offline and would like to plan their search strategies beforehand. The advantage of the paper format with respect to e.g. browsing pages may to a large degree compensate for their static format.

Research on WIS behaviour has primarily focused on searchers' use of general search engines, i.e., Category 9 in Table 1 (see, e.g., Jansen & Pooch, 2001). A broader perspective, however, reveals that searchers switch between using different kinds of IR services during their WIS processes (Pharo, 2002). In addition they may use several other interaction techniques, such as following links in pages, direct access entering the site's URL, and using their own bookmarks.

The WIS processes thus can be said to consist of interaction characterised by switching between meta resources, such as the ones presented in Table 1, and "real" information sources, i.e., pages containing potential solutions to the searchers' problems.

In the present Part 1 we analyze four respected textbooks on Web information searching. The books take somewhat different perspectives, but all have in common that they are quite general in their choice of the type of Web IR systems they address. This study aimed at learning more about the perspectives chosen in the WIS textbooks. In what degree do they focus on different types of sources, and to what extent are the textbooks analytical in their advice on how searchers should plan their search strategies?

Our basic hypothesis was that guides to information searching and retrieval on the Web to a large degree model users and their information needs as static and objective entities. Such a perspective is in our view problematic because the textbooks prescribe rational search strategies to deal with information needs, which can be very unclear and unarticulated (Belkin, Oddy & Brooks, 1982; Bates, 1989; Ingwersen, 1992).

Section 2.2 presents the methods used in textbook analysis, and Section 2.3 the findings.

2.2. Analysis of textbooks on Web information searching

In order to identify how online guides and textbooks on web searching advice searchers to formulate their strategies we picked out textbooks that were recommended in various sources, such as the reviews in Amazon, IR specialists' own web pages, and presentations made by the publishers. We did not systematically try to identify all existing books on the topic, rather we aimed at books that were reasonably new and that represented different perspectives on web searching. For example, we wanted the books to cover not only searching in standard search

engines, but also more specific web-based databases, including the "invisible" web, as well as how to use human-indexed subject indices and portals.

We initially also compared the textbooks to a selection of online guides (www.searchengineshowdown.com, www.searchenginewatch.com, and www.pandia.com) and found that textbooks in paper-format were more thorough in their coverage of our topic. Thus we choose to use the latter for our analysis. Some of the books also introduce subject-specific databases, such as e.g., Medline, available through the Web. These may have much in common with Category 12 resources, but since they mainly index documents that exist outside of the Web we have not included books specifically focusing on such search tools in our discussion.

Four textbooks were selected which represented guides to searching a broad spectrum of web information systems:

- Ackermann and Hartman (2003): Searching and researching on the Internet and the World Wide Web

- Chowdhury and Chowdhury (2001): Information sources and searching on the World Wide Web

- Hock (2001): The extreme searcher's guide to Web search engines

- Sherman & Price (2001): The invisible Web.

We analyzed the textbooks with respect to their approach, i.e., to what degree they were descriptive versus analytical in their advice on how searchers might deal with their information needs. An analytical approach would give advice based on the kind of problem or search task faced by the searcher. This could be contrasted with a more descriptive perspective, which would be the case of books emphasising how to use particular search services, regardless of the context

of the search session. In the next section we present each of the four books individually before we compare them with one another.

2.3. Findings of the textbook study

Ackerman and Hartman have quite a general target group, among which are teachers and students at various levels as well as librarians, researchers, and "others interested in tapping the Web and the Internet for information". Compared to the other books investigated this book clearly has the broadest scope; it can be used to introduce the Web to novice users as well as providing tips for experts. It has a general section on searching, including advice on how to design search strategies. In addition there are sections treating different types of search tools (search engines, special databases etc.) as well as sections on specific search tools and how to make optimal use of these (e.g., Google, Yahoo).

Chowdhury and Chowdhury's book has students of LIS as well as practising librarians as its main target group. It includes short introductions to basic web technology and search techniques as well as chapters dedicated to overviews on individual search engines (e.g., Google), subject indices (e.g., Yahoo), subject gateways (e.g., SOSIG), reference and information services (e.g., "Ask a librarian"), e-journal providers, and digital libraries. The text is very descriptive and richly illustrated, but it has no focus on analytical search strategies.

Hock's guide focuses, as implied by the title, on search engines (although there is a chapter on Yahoo, which primarily is a subject index). The target group is experienced searchers who use the web for professional purposes. It contains two general chapters on search engines and query formulation; the remaining chapters treat each major search engine individually. The chapters are quite descriptive, but there are also some analytical advices on how to develop search strategies.

Sherman and Price's book on searching the "invisible Web" has a scope which makes it an important supplement to the other books discussed here. It has a general target group and is aimed at novices as well as experienced searchers. The invisible Web represents the part of the Web that is not indexed by general search engines. The book thoroughly describes why and how a lot of web resources cannot be accessed via search engines (e.g., because they are dynamically generated by databases), and it contains sound advice on when to use invisible web resources, primarily databases dedicated to specialised content such as "public company fillings", "telephone numbers", and "customized maps and driving directions"[1], to satisfy specific information needs.

One characteristic of the textbooks is the focus on the differences between general search engines and subject indices (Categories 9 and 11 in Table 1) and their advantages in satisfying information needs of different specificity. They may, e.g., prescribe that "search engines are the best tools to use when you are looking for very specific information or when your research topic has many facets" (Ackermann & Hartman, 2003, p. 167) whereas subject indices are effective for recovering "general information on a subject" (Ackermann & Hartman, 2003, p. 100). Also indices organised by experts on specific subjects (Category 12) are treated as separate services by most authors (e.g., Chowdhury & Chowdhury, 2001, p. 51-79).

In addition to the general search tools some textbooks (e.g., Ackerman & Hartman, 2003) also present different kinds of search tools to use for specific kinds of material (e.g., Usenet News articles, FTP files, email addresses, and telephone numbers).

The advices in Web searching textbooks generally are very prescriptive. Much space is dedicated to presentations of individual search engines and how their search facilities should be used. Hock's (2001) book, which is one of the most comprehensive guides to the advanced use of

---

[1] These are the top three categories of the book's "Top 25 Invisible Web Categories"

search engines, gives a very thorough introduction to the major search engines (plus Yahoo). This book is structured so that each chapter apart from the introduction is devoted to the presentation of specific search engines.

Sherman and Price (2001) include a chapter with a set of case studies that exemplify when invisible web sites are good information sources. These cases are also very good at describing real-life information behaviour, and can inspire searchers to think creatively when they develop search strategies.

Ackerman and Hartman (2003) have chosen another approach. They choose to focus on various materials available on the web and how this material is retrievable with specific kinds of search tools. This book gives very specific advice on how to use search tools systematically. The following is their 10-step-prescription for developing a basic Web search strategy:

1. "Identify the important concepts of your search.

2. Choose the keywords that describe these concepts.

3. Determine whether there are synonyms, related terms, or other variations of the keywords that should be included.

4. Determine which search features may apply, including truncation, proximity operators, Boolean operators, and so forth.

5. Choose a search engine.

6. Read the instructions on the search engine's home page. Look for sections entitled "Help," "Advanced search," "Frequently Asked Questions," and so forth.

7. Create a search expression using syntax that is appropriate for the search engine.

8. Evaluate the results. How many hits were returned? Were the results relevant to your query?

9. Modify your search if needed. Go back to Steps 2 through 4 and revise your query accordingly.

10. Try the same search in a different search engine, following Steps 5 through 9 above."

(Ackermann & Hartman, 2003, pp. 173-174)

Chowdhury and Chowdhury (2001) refer to secondary literature for search strategy development, but they also summarise the literature. Their summary-based advices are, however, very general and far from as systematic as those by Ackerman and Hartman. Some examples of their advice are "it is better to do many narrow searches than to make a search that is too broad", "use Boolean AND and phrase options whenever possible to make a search more specific", and "put the most important word first in your query" (Chowdhury & Chowdhury, 2001, p. 33).

The two latter books are in a sense complementary. Chowdhury and Chowdhury (2001) provide some good advice based on heuristics whereas Ackerman and Hartman (2003) focus on the individual steps of the Web IR process. In this sense Ackerman and Hartman's book resembles guides to traditional (i.e., pre-Web/non-Web) online searching (e.g., Hartley et al, 1990) as well as IR textbooks (e.g., Lancaster, 1979; Henry et al., 1980; Belkin & Vickery, 1985; Ingwersen, 1992; Marchionini, 1995).

Lancaster's (1979) steps of the IR operation (or search process) can easily be identified in the strategy described above. He identified six separate steps, i.e., information need, stated request, selection of database, search strategy (sic), search in database, and screening of output. In Table 2 we compare the two approaches.

Table 2 Comparing Ackerman and Hartman's web search strategy with Lancaster's steps

The comparison in Table 2 suggests that the view of information searching and information search processes has been fairly stable the last 25 years. Although new information systems have developed we seem to prescribe the same formula for interacting with them. During this period, however, there has been a huge growth in research focusing on searchers and their interaction with information systems. This research has identified many factors influencing the searcher-system-interaction (e.g., Fidel & Soergel, 1983). We find it strange that this is not reflected upon or discussed in the "modern" guides. Sherman and Price's cases are probably the best attempt at emphasising how normal information needs often need to be solved using trial-and-error, serendipitous findings, and creative solutions. This is particularly interesting since we shall see that there are clear discrepancies between the textbook prescriptions and descriptions of real life information searching.

In Table 3 we summarise our findings from the textbook study. We see that the books have different target groups; that three out of four books claim to address searchers on all levels of expertise; that most books focus on general search tools, and in particular on search engines; and that the books mainly focus on describing how to enter queries in search engines, rather than advocating a more holistic perspective giving advice based on knowledge of real search performance.

Table 3. Summary of the textbooks features

This is not to say that we disagree with the guidelines given in the above mentioned books, rather we find them usable for rationally planning one's searching. Nevertheless we want to point out some problems related to their correspondence to how real searchers behave when they execute their search strategies. We believe it is necessary to complement these prescriptions with knowledge about how people actually behave when they develop and execute their search strategies. The IR community needs to pay attention to real-life behaviour in addition to carefully analysed "optimal behaviour" which may be very rare in real life.

3. Part 2: Real-life web information searching

In order to model real-life web searching we found it worthwhile to go to Management Science, where the perspective on how complex tasks are performed changed quite drastically during the 1960-70s. We believe that it is fruitful to look at how searchers perform (complex) search tasks in light of a "non-rationalistic" viewpoint on Management Science, and we chose the Disjointed Incrementalism (DI) strategy for this purpose. There are also other theories which similarly criticise rationality, but for our purpose DI is appropriate.

3.1 The strategy of disjointed incrementalism

Management Science has long studied decision-making. Early on, rationalism was the dominating school of thought. Roughly, the rational school prescribes the following way of solving problems or executing tasks:

1. The decision-maker first needs to collect all information relevant for solving the task.

2. Thereafter he/she shall create and compare all possible alternative solutions.

3. Next all possible consequences of each solution on all criteria need to be addressed.

4. Finally he/she shall choose the most cost-effective solution.

Disjointed incrementalism (DI) was proposed in the early sixties as a challenge to rationalism (Braybrooke & Lindblom, 1970). It suggests that, in social/political decision making, rationalism is generally not followed, for good reasons, and therefore rationalism is only a prescription of decision making, not a description. Science progresses by understanding how matters of interest really are, not by thinking how they should be (that may be the goal of engineering).

Since Web information searching (WIS) involves fairly complex decision-making by people who are not educated as professional searchers, but may nevertheless actively use the Web, we would like to suggest descriptive theories of decision-making as an approach to study WIS.

We shall start by introducing DI. It was presented as a radically opposing view to the once dominant rationalism (e.g., Bergson, 1938; Arrow, 1951). Since its introduction there has been a lot of discussion in Management Science on decision-making theories, involving rationalism, disjointed incrementalism, and restricted rationalism (e.g., Cohen, March & Olsen, 1972; Newell & Simon, 1972; Brunsson, 1985). However, the original DI proposal makes its point very clear and therefore serves well our wish to have another look at WIS. We shall here briefly present the focal features of DI (Braybrooke & Lindblom, 1970, pp. 83-106). The features clearly interrelate and need to be taken into consideration as a whole rather than be rejected based on any individual attribute. We use Braybrooke and Lindblom's own examples, from political science, to explain each of the features.

The core of the DI- approach is its focus on how policy decision makers focus on small steps, or *margin-dependent choices*, when they consider the possible options to take into consideration in order to reach a goal. A policy analyst evaluating possible ways to control government bureaucracy will only take into consideration those policies whose known or expected social states *and*

consequences differ incrementally from the situation at hand. A revolutionary policy will not be taken into consideration.

Related to the previous point is another central feature of DI - that only a *restricted variety of policy alternatives* are considered. This means that there are many alternatives that are never taken into consideration because they do not differ incrementally from the status quo. There may, of course, exist an infinite number of incremental alternatives but due to factors such as the policy analyst's lack of imagination, the number of relevant alternatives considered is extremely small compared to the theoretically possible.

Also the analyst will consider *only a restricted number of consequences for any given policy*. Thus he/she will take into consideration only consequences that are understandable for him/her. There may be important consequences of a policy that the analyst is aware of, but nevertheless does not take into account, due to, e.g., the consequences being uninteresting to or poorly understood by the analyst. One example would be in the analysis of consequences of building highways which might "disregard certain indirect consequences for social mobility or family solidarity" (Braybrooke & Lindblom, 1970, p. 90).

Adaptation of objectives to policies is the fourth focal feature of the DI strategy. In common language it means that the ends are sometimes adjusted to the means rather than the conventionally desired opposite. This can be explained by the notion that sometimes one finds that certain means emerge that can be used to reach an acceptable solution to a not clearly defined problem (cf. the so called garbage can model by Cohen, March & Olsen, 1972). There may be the active choice of the policy analyst to only take into consideration those objectives that he/she has the means to solve.

Reconstructive treatment of data is tightly knitted together with the adaptation of objectives (Braybrooke & Lindblom, 1970, p. 98). The strategy is dynamic in the sense that it adapts to whatever resources are available. For example, a policy analyst may learn through exploring new data that the problem he deals with is not relevant in its "original" form, but that it should be transformed according to the new knowledge.

Using the DI-strategy problems should be subject to *serial analysis and evaluation* related to the overall policy steps. In other words the individual problems that are handled using the strategy also should be viewed in the light of problems or ideas of a more general kind. Among the examples used by Braybrooke and Lindblom (1970, p.99-10) are the US Social Security Act, labour legislation, and legislation on public education which are all continually changed incrementally.

Remedial orientation of analysis and evaluation has to do with the tendency of decision makers following the strategy to focus on avoiding certain consequences rather than moving towards specific goals. For example, the policy analyst may focus on reducing governmental inefficiencies without having any knowledge about what is the maximum competence level to expect.

The last core feature of the DI strategy is *social fragmentation of analysis and evaluation*. This deals with the fact that there will typically be many different institutions that work on analysing the same problem area, but from different and uncoordinated perspectives. A country's security policy, e.g., will be based on analysis made in governmental offices, private organisations, research institutes, universities, and other institutions.

We shall, in our analysis of Web information searching behaviour, use the DI strategy to illustrate why seemingly "irrational" behaviour takes place during searching. We believe the features of the

DI strategy add explanatory power to understanding what takes place when Web searchers select, *and choose not to select*, certain sources for information and certain ways of interacting with those sources.

3.2 Analysis of real-life Web information searching

As part of a previous study (Pharo, 2002; Pharo & Järvelin, 2004) we have collected comprehensive data on task-based Web information searching.

The study subjects were third year students working on their theses in LIS at Oslo University College. All third year students (in total 110 persons) were asked to participate, of which 55 returned a questionnaire on search skills, general search strategies, work tasks, and demographic data. In all 13 students volunteered to participate in our study where they agreed to have Web search activities related to their theses recorded and observed. For the search sessions the searchers logged into their user accounts and had access to their familiar web browser and previously saved bookmarks. Thus the search tasks performed by the searchers were all real, as opposed to simulated, and related to a work task that they owned and that was familiar to them. Also the settings in which the searchers performed their search tasks were as natural as possible. The sessions were recorded in an undisturbed room and with the help of a GrandArt video converter, which is a device that converts computer screen signals to video format. A microphone is connected to the converter, and the users were asked to talk aloud (Ericsson & Simon 1996) during the sessions, to make it easier to identify their intentions and goals. Their utterances were recorded simultaneously on the videotape. This resulted in 9 (successfully recorded) search sessions, which in total lasted 12 ½ hours.

In the excerpts shown below we have used transcriptions of videotaped WIS-sessions containing information about the accessed Web page, actions performed, time spent, and comments made

by the searcher(s) during interaction. In the original study a multitude of different data were used, including questionnaires, pre- and post-session interviews, observation notes, and content analysis of final theses. The combination, or triangulation, of the different kinds of data were used to create stories that focused on the relationship and interplay between different factors related to the work task, search tasks, searcher, social/organisational environment, and search process.

The stories were analysed using the categories of the SST method schema (Pharo, 2002, p. 95-108; Pharo & Järvelin, 2004, p. 640-644) and content analysis. This schema was developed using methods from grounded theory (Strauss & Corbin, 1990). In our analysis we identified occurrences of SST schema categories and attributes. We then used content analysis of stories to reveal interplay between the categories and attributes at each step in the WIS-process. This made it possible for us to see whether the searchers followed rational strategies or preferred to take ad-hoc decisions about what sources to explore and interaction techniques to use. For more details about this study we refer to Pharo (2002) and Pharo and Järvelin (2004).

In relation to the three excerpts presented below we compare what they reveal about the search behaviour to the advice prescribed by the textbooks (Section 2.3) and the features of the DI strategy (Section 3.1).

We will now closely look at excerpts from one session where the searcher is looking for information about a specific Norwegian author called "Ronald Fangen" (Table 4). In the notation, we have used "." to denote a pause of up to 5 seconds, i.e. in row 9 in Table 4 ". It's a bit difficult etc" signifies that the searcher does not utter anything before 5-10 seconds of time has passed. Further the original utterances have been translated from Norwegian. We have used

brackets for adding context to incomplete sentences, additional comments are made in parentheses.

As a whole we observe that the searcher in total selects 6 different sources for information about this specific author. What sources to use she has decided beforehand as part of her search strategy. It seems reasonable to limit oneself to a selection of sources, hence restrict the number of alternatives to use (cf. the DI strategy). This might be particularly important when dealing with information sources as heterogeneous as the Web.

Table 4. Excerpts from Web search session 1

In Table 4 the searcher explores a newspaper (Dagbladet) in order to find articles about the author ("Fangen"). The searcher on Rows 2 and 6 explains her search strategy, which indicates that she has little background knowledge about the topic she is looking for (is the author alive or dead?). Also we notice on Row 8 that the searcher enters her query in inverted form ("fangen, ronald"); this would be a natural query formulation used in a bibliographic database. However, when used in this local (Category 1) search engine this has the unfortunate result that the query is treated as a Boolean OR-query. A third interesting finding is that the query results do not present themselves in such a format that it is possible to judge an article's relevance without looking at it in full text. Thus one might argue that the information environment does not provide such a format that the textbooks take for granted when prescribing their search strategies.

Viewed in the light of the search strategy prescription presented in Table 2 we see that the searcher behaves "irrationally" in various respects. She clearly does not follow Step 4

("determining search features") as she uses an inappropriate query formulation. Neither does she choose to use the search engine's help feature to learn about how to use it properly (Step 6). Thirdly she also experiences synonymy problems (Step 3) which could have been avoided had the query formulation been appropriate. These problems might have been reduced if the searcher had used stringent guidelines for developing her search strategy, but this might not necessarily be the searcher's priority!

We shall use another excerpt to illustrate how searchers change their relevance criteria. In this excerpt there are two searchers (indicated by (1) and (2) in Table 5) working together in order to find graphical elements to include in a web site. The search process itself lasts 2 hours and 4 minutes. The searchers adopt an exploratory search strategy and we have put together excerpts from different stages of the process thus including the Time-column to indicate the "progress" of the search process. We step into the process after approximately 10 minutes. The searchers try to find an image that can be used to represent an author, which they believe should be a picture of a human being.

Table 5. Excerpts from Web search session 2

In the excerpts in Table 5 we can observe an interesting phenomenon that can be called "dynamic relevance". In the data we found 15 occurrences of the searchers discussing how to find an icon that could be used to represent "authors" in the web site, Table 5 includes three of these occurrences. We see that the searchers in Row 9 and Row 13 access the same page, but whereas in the earlier situation they dismiss all icons available on this page they decide, in the latter situation, that there is at least one relevant icon there ("looks rather neutral though"). From

the data it appears that the searchers' relevance criteria change as the search process evolves; elements (in this case graphical icons; in our data we also find examples of similar treatment of the textual content of web pages) that at a former stage were considered non-relevant are re-evaluated and considered relevant at a later stage in the process.

The phenomenon captured in Table 5 can be compared to two central features of the DI-strategy; 1) the adjustments of objectives and 2) reconstructive treatment of data. The searchers decide that their requirements need to change in accordance with available data; the icons they originally thought they could find, which would have been the perfect match to their needs, did not exist, or, at least did not exist on any of the resources they were able to explore within a two-hour web session, so they needed to adjust their needs to match what was the best possible alternative.

The textbook strategy (Table 2) prescribes that each information need should be treated systematically by repetitive procedures of concept identification, keyword selecting, synonym determination etc. used in one search engine at a time. The Table 5 excerpts represent a fundamental breakdown of this strategy. Rather than describing a searcher following a stringent prescribed strategy the excerpts emphasise the point that searchers learn throughout the search process and that they may find that the resource/page initially not considered relevant actually was the most appropriate answer to their needs.

Our third excerpt is collected from a session where two searchers ("named" (1) and (2)) look for information on web usability and design. They are going to use the information in a study of web newspapers. We enter the session at an early stage.

Table 6. Excerpts from Web search session 3

The excerpts in Table 6 illustrate another common feature in our data; searchers' tendency to quickly pick out a very limited number of resources for further study. This is explicated on Row 7 where they say "we cannot look at everything". The searchers are aware of the existence of a large number of alternative sources (Row 3: 288 hits), but choose to invest their time in only a few of the existing sources. The selection of which sources to pick often seem very haphazard, in great contrast to the textbook prescriptions discussed above. This phenomenon exemplifies what Newell and Simon calls "satisficing", i.e. decision makers' choice to set "an acceptable level or aspiration level as a final criterion" (Newell & Simon, 1972, p. 681) and then select the first acceptable option that satisfies one's level of acceptance.

The data in Table 6 are comparable to the DI strategy's focus on how only a restricted number of alternatives are taken into consideration. The searchers "have to" make a selection between several information sources because they only wish to dedicate a certain amount of time on searching and would like to find sufficient relevant information. They consciously choose not to go through all potentially relevant alternatives, but rather pick out a representative sample.

In light of the textbook approach (Table 2) there are clear indications that the searchers start with a vague conceptualisation of their information need (Step 1 in Table 2). They are unaware of the meaning of the abbreviations used in the resource and their comments suggest that they intentionally do not want to perform a specific query – rather they want to pick out a small selection of resources to learn more about the topic. Rather than pointing out the searchers inability to formulate a "rational" search engine search strategy it exemplifies how unclear information needs facilitate browsing rather than querying.

4. Discussion

Research on information behaviour has shown that searchers' search strategies reflect what they learn during information searching and that searchers have subjective and dynamic information needs which change over time (e.g. Kuhlthau, 1993). In the preceding section we examined some empirical data supporting this. We compared the data to search strategies as prescribed by textbooks and saw that the searchers do not behave "rationally" in the sense of selecting optimal strategies. Rather they seem to make decisions that reflect an understanding of the Web as a complex information space requiring that searchers pick out a sample of possibly relevant sources, adjust their objectives during interaction and dynamically re-evaluate (or reconstruct) the sources throughout the search process. In our view this can be compared to decision makers' problem solving strategies, which are often characterised by lack of rationality (Braybrooke & Lindblom, 1970; Cohen, March & Olsen, 1972; Newell & Simon, 1972). More explicitly, we found that WIS behaviour strongly resembles the strategy of disjointed incrementalism proposed by Braybrooke and Lindblom in the early 1960s. We shall go into more detail on this below.

In the study described in (Pharo, 2002) we found that the interaction of many factors can be used to describe the seemingly haphazard behaviour of web searchers. It seems that the searchers prefer to select the information that coincidentally happens to cross their way, which can be compared to how decisions are often made coincidentally because the problems and means occur at the same time (Cohen, March & Olsen, 1972). This suggest that the searchers far from try to reach optimal matches to their information needs but rather that they look for sources that may help them constructing a satisfactory solution. This also relates to the theory of "satisficing" (Newell & Simon, 1972) recently discussed by Agosto (2002) in relation to young people's web site decision making.

The searcher whose WIS behaviour we have represented in Table 4 may seem haphazard, or even irrational by the (W)IR-textbook standards. However, the searcher is engaged in a real work task which she wants to complete. She also has some professional education in WIS, acquired through her education in LIS. Should one now conclude that the education has been poor? Or that the person in question is below average in her intellectual achievement – perhaps irrational? Or rather admit that WIS does not happen, in general, as guided by the rational textbooks?

The last alternative opens up an avenue to understand real-life WIS as it happens. Disjointed incrementalism can be used to suggest its features. Let us therefore relate some of the central features of DI to WIS processes as we have observed them in our data.

Restricted number of alternatives and consequences to consider

The searcher will, due to the limitations in human intellectual capacity, lack of imagination, the multitude of alternative resources available, and the limited time available restrict the number of resources she wishes to explore in the process. The searcher most often will focus on those resources most familiar to her and avoid extreme alternatives – which Braybrooke and Lindblom call "nonincremental changes". There may be a number of consequences that are not taken into consideration by the searcher during interaction. These may be important, but the searcher may be deliberate in not taking them into account. This happens for several reasons; the searcher may find them uninteresting, remote from her problem, imponderable, intangible or she may simply not understand their importance. Searchers use their pre-understanding when deciding what resources to explore. Although they may be aware of resources that have a reputation of treating a topic of interest well they still, in most cases, decide to use resources that they believe will give them a *satisfying* response. In other words resources that they can find quickly – as it happens, for many web searchers searching is a nuisance, not a priority.

Adjustment of objectives during interaction

During a search session the searcher may decide that it is not possible to find necessary resources to deal with the search task and/or work task. There are no resources that are usable for her initial purpose and instead she will make adjustments in her task(s). In other words we can describe it as an example of the ends adjusted to the means. She may collect material that was not what she expected to look for initially in the process, because ideal resources do not exist.

Reconstructive treatment of data

Another way of viewing the dynamics of WIS processes is to focus on searcher's reconstruction of the search task and/or work task during the process. The results obtained in the search process constitute active components in the reformulation of the search task. For example, the searcher may learn that she had an erroneous conception of an important factor in her initial search task formulation making it necessary to adapt her new knowledge into the formulation and possibly choose a new strategy to deal with it.

Serial analysis and evaluation

The work task constituting the point of departure for WIS processes may well be executed over a long time (see also Vakkari, 2001), for example in the case of a student working on her doctoral dissertation. The point of departure for the searcher at one stage of task processing may be different from what it was months and years ago. Still the searcher may use the same or similar Web resource types at different stages in the process, perhaps using different techniques, but all the same the goal may have changed incrementally and the searcher adjusts her interaction accordingly.

This kind of view is also hospitable to the critical views from IR research (Bates, 1989; Ingwersen, 1992; Borlund, 2003) and information seeking research (Kuhlthau, 1993; Fidel et al.,

1999) where vaguely formulated needs and learning during the process and multiple-goal information seeking are emphasised.

Rationalism seems well-founded in teaching WIR/WIS but unfounded as a description of what is actually happening.

Science asks "what is out there?" and we'd better learn that, no matter whether we want to see it or agree with that. Ordinary WIS users are better not seen as irrational – their interests and agendas are just different from those by IR specialists or researchers. And whom are the WIR systems for – for us, the IR-specialists – or the general public (or specialists in other areas) who outnumber IR specialists by several orders of magnitude?!

Usable systems are made to fit the users' needs (in a process where both sides adapt). Great numbers of IR-non-specialists will never take courses in IR or WIS and will use the systems in unexpected and ineffective ways. Therefore even IRtechnology, not just science, benefits from a descriptive approach to WIS behaviour instead of a rationalistic, prescriptive one.

5. Conclusion

In this article we have investigated the relationship between real-life Web information searching and textbook-prescribed Web search behaviour. Our study was based on a small selection of textbooks, which we believe represent a reasonably broad picture of such books, and an empirical study of web searchers that performed search tasks generated from their work on their bachelor theses.

From Management Science we introduced the strategy of disjointed incrementalism (DI) as a tool for understanding why searchers often behaved in a way that is not "optimal" or "rational" from the textbook-point of view.

It has not been our goal to despise the production of textbooks prescribing how to perform effective searching and develop rational search strategies. However, we think that the textbooks should provide a better combination of exemplary cases and advice for solving these through appropriate search tools. The cases provided by Sherman and Price (2002) are good examples of such an approach. By giving the readers examples that they can follow and relate to their own information needs it may be easier to illustrate that searching often may benefit from using creativity and heuristics in addition to rational search strategies.

6. References

Ackermann, E. & Hartman, K. (2003). Searching and researching on the Internet and the World Wide Web. 3rd Ed. Wilsonville, Or: Franklin Beedle & Assoc.

Agosto, D. E. (2002). Bounded rationality and satisficing in young people's Web-based decision making. Journal of the American Society for Information Science and Technology, 53(1), 16-27.

Arrow, K.J.. Social choice and individual values. New York: John Wiley and Sons, 1963.

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the on-line search interface. Online Review, 13(5), 407-431

Belkin, N.J. & Vickery, A. (1985). Interaction in information systems. London: British Library.

Bergson, A. (1938). A reformulation of certain aspects of welfare economics. Quarterly

Journal of Economics, 52, 310-334.

Borlund, P. (2003). The concept of relevance in IR. Journal of the American Society for

Information Science and Technology, 54(10), 913-925.

Braybrook, D & Lindblom, C. (1970). A strategy of decision: policy evaluation as a social process.

New York: Free Press.

Brunsson, N. (1985). The Irrational Organization. Chichester: John Wiley & Sons.

Chowdhury, G.G. & Chowdhury S. (2001). Information sources and searching on the World

Wide Web. London: Library Assoc.

Cohen, M.D., March, J.G. & Olsen, J.P. (1972) A garbage can model of organizational choice.

Administrative Science Quarterly, 17(1), 1-25.

Ericsson, K. A. & Simon, H. A. (1996). Protocol analysis: verbal reports as data. Cambridge,

Mass.: MIT Press,

Fidel, R et al. (1999). A visit to the information mall: Web searching behavior of high school

students. Journal of the American Society for Information Science, 50(1), 24-37.

Fidel, R. & Soergel, D. (1983). Factors affecting online bibliographic retrieval: a conceptual

framework for research. Journal of the American Society for Information Science, 34(3), 163-180.

Hartley, R.J., Keen, E.M., Large J.A. & Tedd, L.A. (1990). Online searching: principles and practice. London: Bowker-Saur.

Hock, R. (2001). The extreme searcher's guide to Web search engines. Medford, NJ: Cyberage Books.

Henry, W.M., Leigh, J.A., Tedd, L.A. & Williams, P.W. (1980). Online searching: an introduction. London: Butterworth.

Ingwersen, P. (1992). Information retrieval interaction. Taylor Graham, London.

Jansen, B. J. & Pooch, U. (2001). A review of Web searching studies and a framework for future research. Journal of the American Society for Information Science, 52(3), 235-246.

Kuhlthau, C. C. (1993). Seeking meaning: a process approach to library and information services. Norwood, N.J: Ablex.

Lancaster, F.W. (1979). Information retrieval systems: characteristics, testing and evaluation. New York: John Wiley & Sons.

Marchionini, G. (1995). Information seeking in electronic environments. Cambridge: Cambridge University Press.

Newell, A. & Simon, H. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

Pharo, N. (2002). The SST method schema: a tool for analysing Web information search processes. Tampere: University of Tampere [Doctoral dissertation]. Also available: http://acta.uta.fi/english/teos.phtml?6719.

Pharo, N. & Järvelin, K. (2004). The SST method: a tool for analysing Web information search processes. Information Processing & Management, 40 (4), 633-654.

Shepherd, M. & Watters, C. (1998). The evolution of cybergenres. In Proceedings of the 32nd Hawaii International Conference on System Sciences *(HICSS '98)*, Hawaii, vol. II, 97-109.

Sherman, C. & Price, G. (2001). The invisible Web. Medford, N.J. : Information Today.

Strauss, A. & Corbin, J. (1990). Basics of qualitative research: grounded theory, procedures and techniques. Newbury Park, CA: Sage.

Vakkari, P (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. Journal of Documentation, 7(1), 44-60.

Tables

Table 1

	Search engines		Subject indices	
	General	**Subject specific**	**General**	**Subject specific**
Local	**1**) E.g. American Society for Information Science and Technology's local search engine www.asis.org/search_site.html	**2**) E.g. Nasa's spacelink search engine: search.spacelink.nasa.gov	**3**) E.g. Human Rights Watch's sitemap: www.hrw.org/site-map.html	**4**) *
National /regional	**5**) E.g. the Europe oriented search engine Euroseek: www.euroseek.com	**6**) E.g. Autonius, the search engine on German medical resources: www.dr-antonius.de	**7**) E.g. Norwegian subject index Kvasir: www.kvasir.no	**8**) E.g. the subject index to Nordic libraries: www.lub.lu.se/resbyloc/Nordic_lib.html
Global	**9**) E.g. Google: www.google.com	**10**) E.g. Psychcrawler psychology search engine: www.psychcrawler.com	**11**) E.g. Yahoo: www.yahoo.com	**12**) E.g. the subject index to engineering Eels: eels.lub.lu.se/

Table 2

Step	Ackermann & Hartman (2003)	Lancaster (1979)
1	Identify concepts	Information need
2	Choose keywords	Stated request
3	Determine synonyms etc.	Search strategy
4	Determine search features (truncation, Boolean)	Search strategy
5	Choose search engine	Selection of database
6	Read help instructions	
7	Create search expression	Search in database
8	Evaluate results	Screening of output
9	Modify search	Stated request (2)
10	Choose another search engine	Selection of database (2)

Table 3

Feature	Ackermann & Hartman	Chowdhury & Chowdhury	Hock	Sherman & Price
Target group	General	Information science professionals, end-users	Users frequently searching for professional purposes	General
Level of experience	Novices to experts	All levels	Experienced searchers	Novices to experts
Web perspective	General search engines; general subject indices; special databases	General search engines and subject indices	General search engines, meta search engines	Various special databases
Level of advance	Descriptive and analytical	Mainly descriptive	Descriptive	Descriptive, analytical cases

Table 4

Row	Action	Time	Page	Comments
1.	Selects entry from bookmark	0.00	Department's main page	
2.	Looks at page		Page with pointers to different literature resources	Now I've found a page which gives a pretty good overview of literature on the net. This is the one we have used when we have decided who everyone in the group should look through. So I took a short look at it yesterday, but I found that it was a bit messy to return to like
3.	Selects link	0.43	Page with pointers to different literature resources	
4.	Scans page		Literature pointers from Dagbladet	
5.	Selects link		Literature pointers from Dagbladet	That one, dagbladet
6.	Scans page		Dagbladet - Book reviews	Now I'm going to search for an author, I don't know if he's alive or dead, so I'll just go into book reviews in Dagbladet to see if there might be something there, and I was in there yesterday
7.	Selects link		Dagbladet - Book reviews	to search – so I know that I need to go to that search button
8.	Enters query: fangen, ronald		Dagbladet query form	
9.	Looks at page		Dagbladet Query results	. it's a bit difficult when only such titles appear, or links without specific titles, only numbers, I don't understand what
10.	Selects link	1.15	Dagbladet Query results	that means, but I can just see
11.	Looks at page		Short article about prisoners [fanger] escaping from prison	if it works
12.	Selects back			
13.	Scans page		Dagbladet query results	
14.	Selects link		Dagbladet query results	
15.	Scans page		Article on Norwegian author Gunnar Reiss-Andersen	. yeah, here he is just mentioned in an article, and we had decided not to include, when he is just mentioned
16.	Selects back			in a sentence
17.	Scans page			
18.	Selects link		Dagbladet query results	
19.	Scans page			That's what I discovered when I searched in Dagbladet yesterday and
20.	Selects back			
21.	Scans page		Dagbladet query results	that I get a lot of hits which I don't understand how I've arrived at . no
22.	Selects back			I don't bother [to look at more of the result], (what she probably is unable to see is that a lot of the results refer to prisoners; in Norwegian "fangen" also means "the prisoner")
23.	Selects back	3.10		. (slow line)

Table 5

Row	Action	Time	Page	Comments
1.	Looks at page	10 min.	Realm Graphics – icons	
2.	Selects link		Realm Graphics – icons	Yeah, look at everything [link no. 1]
3.	Scans page		Realm Graphics – critics & games icons	We want a human being (1) No (2) Perhaps not that human perhaps (1) No, not that one (2) Wow, dices (1) What about that one, (laugher) (2) [points at a skull]
4.	Selects back			
5.	Selects link		Realm Graphics – icons	What about this one, info [link no. 2] [continues on another track; looking for "help"-icons]
6.	…	…	…	…
7.	Looks at page	21 min	Icon Bazaar - main page	But, what is it we haven't got yet, a human being perhaps? (2) yeah, a human being, we haven't seen any of …and, well we have found a (1) there are some very nice globes – it (2) [points at link] Is that it? But here are "people", "people", and then maybe, eh, we've briefly seen the magnifying glass (1) yeah, but it was in a frame (2) yeah, and then house, home (1) House, yea. And a light bulb
8.	Selects link			Or such a light (2)
9.	Scans page		Icon Bazaar – people symbols	I think it's the same all over (1) hey, that's the brain they've used at that Norwegian page (2) really? [the searchers continue discussing the design of an icon, the background colours they would need to use etc, this takes a few minutes, concluding that they might have to draw the icon themselves]
10.	Selects back			Yeah [laugher] with the drawing program
11.				[throughout the search session they return to discuss the usability of different icons representing "authors" several times [in total 10 more times], before they return to the current page in Icon Bazaar, 1 hour and 10 minutes later
12.	…	1h 10 min	…	…
13.	Scans page		Icon Bazaar – People symbols	..[giggling] shall we use one of those – author (1) well [quiet] (2) looks rather neutral though (1) yeah, it does doesn't it? (2) can take it (1) yeah do that (2)
14.	Right-clicks on icon,			
15.	Saves image			Can use it
16.	Scans page		Icon Bazaar – People symbols	At least for now, let's see was it only, yeah, should we take any others or [quiet] (1) don't think they were very (2) that one? (1) [laughter]

Table 6

Row	Action	Time	Page	Comments
1.	Scans page	40 secs	Infoseek : Internet	… here there are web sites, "about the internet", blah blah blah, "internet fun"? (2) I think it's all about the same, go a little bit further down to see, I think I've been there (1) web publishing? (2) ehe (1) [incomprehensible] (2) .
2.	Selects link			I've seen a little, but I haven't been looking at, read, if it's in full text maybe, it probably is
3.	Scans page		Infoseek : Internet : Web publishing	. yeah it hasn't appeared entirely yet, yeah except for HTML I don't understand what all these [abbreviations] stands for (2) here there are different things (1) yeah, here's something (2) "web usability", "design" (1) we'll have to look at that (14) look
4.	Selects link	2 min.	Infoseek : Internet : Web publishing	at it , 288 it said (1) yeah it was 288 [hits] you know (2)
5.	Scans page		J. Nielsen's Web Alert box, index to articles on web usability	[searchers explores web site looking for relevant articles on web usability, before they return to the Infoseek page almost 4 minutes later]
6.	…	…	…	…
7.	Scans page	5 min 50 secs	Infoseek : Internet : Web publishing	. "communication, publishing" … butt his is, well, what is this – 10 out of 288? (1) we cannot look at everything (2)