Title:

The users' hierarchical mental model of Internet Search Engines.

Authors:

Sarah E. Crudge.

Department of Information & Communications, Manchester Metropolitan University, Geoffrey Manton Building, Rosamond Street West, Manchester, M15 6LL, UK. Tel: (+44)161 247 1745 Fax: (+44)161 247 6351 E-mail: s.crudge@mmu.ac.uk

Frances C. Johnson.

Department of Information & Communications, Manchester Metropolitan University, Geoffrey Manton Building, Rosamond Street West, Manchester, M15 6LL, UK. Tel: (+44)161 247 6156 Fax: (+44)161 247 6351 E-mail: f.johnson@mmu.ac.uk

Abstract:

Users' internal representations of their interactions with systems are often termed 'mental models', and for successful system use, the users' mental models and system designers' conceptual models of the tools should be congruent. This study explores a method for non-biased determination of the user's subconscious view of Internet search engines, in order to derive a mental model comprising those aspects of the systems of importance to the users. The investigation utilises a repertory grid approach in combination with laddering technique, the latter being based on the cause and effect style of mental model development. The detailed qualitative analysis of the data determined through use of laddering interviews is presented here in the development of a mental model comprising three strata. The main hierarchical stratum of the model conveys the interrelations between basic system description, evaluative description, and the key evaluations of ease, efficiency, effort and effectiveness. Two additional strata relating to the perceived process and the experience of emotion are also discussed. The conjunction of the procedural elements with the key evaluations is of particular significance, and further research proposes the extension of this to provide a framework for search engine evaluation.

Introduction

An investigation of usage of search engines indicates several market leaders. The Nielsen Net Ratings reported by Sullivan (2003) recorded search specific traffic at US search sites and found the top three sites in terms of audience reach (the percentage of US users who visited the site at least once during the month) to be Google, Yahoo, and MSN. Many of the less well used engines will either cease to operate, or more commonly be taken over by other services over time. Nevertheless, technology will always progress, and the market leaders are not themselves that old, as indeed the Internet is a relatively new technology. If an engine is to survive, it must suit its market and as engines become increasingly similar in effectiveness, they must look to other means to ensure their competitive edge.

Fundamental to the progression of Web retrieval system development is the need to understand the way the tools are perceived by the end-users. Unlike the target audience of systems such as DIALOG, web searchers form a large body of 'ordinary' users, with little or no formal IR training. Jansen, Spink and Saracevic (2000) report on an analysis of transactions at the Excite search engine, finding that "about two in three users submitted a single query". Although users do not search for long using the engines, they will nevertheless form an opinion of the tools they have used, based on these very brief interactions, and the opinions will inform any subsequent choice of search tools.

Users' internal representations of their interactions with systems are often termed 'mental models', and HCI researchers have proposed that for successful system use, the users' mental models and system designers' conceptual models of tools should be congruent. However, not a great deal is known about the mental models that users form for Internet search engines. Furthermore, an examination of the techniques used to determine such mental models reveals

that no one technique has proved favourable for use in previous studies. Many of the techniques involve generalisations, such as requiring a user to draw a picture of the technology, or to describe it in a few sentences. This type of mental model determination stems from the belief that users' mental models are formed on analogies to similar technologies (Staggers & Norcio, 1993). Whilst this may be true in part, Norman (1983) suggests that mental models are formed as a result of interactions with the systems, and that cause and effect chains would be an integral part of this (DeKleer and Brown, 1983).

This study investigates an approach to non-biased determination of the user's subconscious view of Internet search engines, in order to obtain a mental model for the search systems comprising those aspects of the systems that are important to the users. The chosen method utilises a repertory grid approach in conjunction with laddering technique. Grid technique originates in the field of clinical psychology (Kelly, 1955/1991) but has in more recent years been applied to the study of attitude to technology. The method exploits the human capacity for drawing comparison between items, thus providing an evaluative view, and it is expected that the resulting mental models will provide a framework for selection of search engine evaluation criteria.

The suitability of the repertory grid technique for eliciting a mental model of search engines was presented in Crudge and Johnson (2004) with a quantitative analysis of the user statements of system aspects with discriminating ability and which cluster around a central overall user rating. A key benefit of the method is that it minimises bias by requiring the user to define a set of ratings scales without influence from the researcher. Furthermore, the procedure allows the user to state as many or as few aspects of the system as they wish, thus resulting in a model that does not focus solely on one or two facets. Finally, the repertory grid approach especially when used in conjunction with laddering technique produces a large

quantity of data of considerable complexity, but at the same time only requires a small number of participants to determine a full set of system aspects. The detailed qualitative analysis of the statement explorations determined by the laddering process is presented here in the development of a hierarchical mental model. This paper describes in detail the implementation of the laddering technique based on the cause and effect aspect of mental model development. A detailed mental model is obtained and an analysis is presented of the extent to which it represents a complete, evaluative and explanatory model of users' perceptions of search engines.

Related Research

There are two main types of models that bear relation to Internet search engines, namely process models and conceptualisations. With reference to process models, Saracevic (1997) states that "the role of models is to depict the essential elements and relations of an object." Numerous IR process models have been proposed, many of which form constituents of the broader information seeking process models. Saracevic (1996) outlines the traditional model of IR, as query formulation, comparison searching, and retrieving of documents, with the inclusion of a simple feedback loop to allow reformulation of the query. However, few process models are proposed for search engines specifically. Holscher and Strube (2000) derived a global model of Internet searching as well as a close up model of direct interaction with a search engine. The models were derived from experts using mental walkthroughs and card sorting techniques, and then probable paths through the model were determined using a larger sample. The most common process was identified as launching the engine, selecting terms, formulating query, obtaining results, examining results, and selecting and examining individual documents. Reformulation will be likely; the probabilities presented indicate that reformulation is more likely than document selection after examination of results. This is in

contrast to the findings of the large scale transaction log analyses of Excite (Spink, Bateman, & Jansen, 1999; Jansen et al., 2000; Spink, Jansen, & Ozmultu, 2000; Spink, Wolfram, Jansen, & Saracevic, 2001) and AltaVista (Silverstein, Henzinger, Marais & Moricz, 1999). These studies suggest a low level of query reformulation and a reluctance to view beyond the first few pages of retrieved items.

In an information system context, the term mental model most frequently refers to conceptualisations of systems. Such mental models are "a psychological representation that aids in understanding, explaining, or predicting how a system works" (Slone, 2000). In contrast to Saracevic's comment on process models, Seadle (2003) states that "the point of examining mental models is not their accuracy, but their power to set expectations." The definition of the term 'mental model' varies across the literature, and has been the subject of much debate in the field of human computer interaction (Staggers & Norcio, 1993). A particular confusion lies with the interchangeable use of the terms mental model and conceptual model. Norman (1983) has a set of four entities to clarify the distinctions in terminology, and gives the four possible areas for consideration as,

- The target system,
- The conceptual model of the target system, this is essentially the system designers view of the system,
- The user's mental model of the system,
- The scientist's conceptualisation of the user's mental model, sometimes called the cognitive model.

Staggers and Norcio (1993) correspondingly define the mental model as the "users' own mental representations of their interactions with devices," the conceptual model as "the system designers', instructors' or scientists' invented model of a system created for design or

instruction purposes," and the cognitive model as "researchers' various conceptions about the structure, process and content of users' mental models."

Conceptual and mental models must be similar in order for the user/ system interaction to prove successful. Staggers and Norcio (1993) suggest that the conceptual model should facilitate the correct development of the corresponding mental model. It would clearly be unwise for a system designer to be unaware of the users mental picture of the system, but it would equally not be productive for the system to be designed entirely to match the mental model, which might be incomplete, unscientific, parsimonious and unstable (Norman, 1983).

Analogies and metaphors. Many researchers believe that mental models are formed through analogies and metaphors, and several have exploited and examined this. A study by Slone (2002) asked participants to explain how the Internet and on-line catalogues worked. The resulting Internet models were categorised as vague, satisfactory, technical, glowing or metaphorical, whilst the on-line catalogue models were classified as vague, satisfactory, technical or comparative (i.e. obtained through comparison of other system types). The Internet was often given 'magical' or human characteristics, felt by the researcher to be suggestive of fragmented or immature models. Ratzan (2000) reported a study of 350 participants who were surveyed to determine views of the Internet, reporting the frequent use of metaphors, of type varying according to skill level and gender. Here a view of the Internet as a disorganised library was common, but expert users suggested metaphysical metaphors such as 'fractal' and 'new dimension'.

Visual representations. A common method for the study of mental models requires users to produce representative drawings. Thatcher and Greyling (1998) determined mental models of the Internet, using protocol analysis, but also by obtaining drawings from participants,

required to represent how the Internet worked. The resulting drawings were classified into six categories according to complexity, and these were found to be related to the level of experience the participants had with the Internet.

It is common to take a navigational approach when determining mental models for systems. Navigational models include schematic drawings of the layout and interlinking within a site. The premise for such studies is that "if a user has a poor mental model of the hypertext system's structure, then it is likely that they will experience disorientation" (Otter & Johnson, 2000). The study by Otter and Johnson (2000) required participants to draw mental models as 'schema' of the layout of the sites. The results were found to suggest that the method had not been entirely successful, because there was no relation between the accuracy of the models drawn and the degree of 'lostness' as measured by other methods.

Modelling through use of queries. Muramatsu and Pratt (2001) investigated models by determining the users' understanding of the system interpretation of queries. The results indicated that the participants expected engines to combine search terms with 'OR' rather than 'AND', and were found to expect term suffix expansion. Little knowledge of stopwords was exhibited, and only slightly more understanding of term order variations was detected. The authors concluded that the participants' models were naïve and incorrect. Moukdad and Large (2001) described user mental models of the WebCrawler search engine through examination of a sample of the queries posed to the engine. The study speculates that users pose questions to the engine because they view it as they would a human respondent.

Mental models of Internet Search engines. Relatively few studies have specifically investigated users' mental models of Internet search engines, and those that have reported studies pertaining to mental models have rarely attempted to provide a complete model. The

study by Muramatsu and Pratt (2001) provides an indication of the differences between the users' mental and experts' conceptual model, but focuses only on the treatment of the query at a search engine. Similarly the transaction log analysis studies focus on one type of observed behaviour, and recognise the limitations of such an approach. Mental model studies often discuss the difference between the mental models of ordinary users and experts. Brandt and Uden (2003) present preliminary results of a study into users mental models, concentrating on the inaccuracies in the mental models. The results indicate that users expect semantic meaning to be derived from web sites by the engines, the difference between directory and search is not fully understood, there is little perseverance for scanning of result lists, and little understanding of the search index overlap with the index of other engines.

Research Objective

The aim of this study is to determine a representation for the users' mental model. The model of the ordinary user is not expected to be complete, or even accurate. However, it is proposed that a small set of individual models could be combined to determine an overall summary model, which would then define the complete general mental model. Each individual user's model would then be formed uniquely from some portion of this overall summary model.

Methods

One method suitable for determination of mental models stems from 'personal construct theory', as proposed by clinical psychologist George Kelly (1955/1991). Kelly suggested that our expectations of the world are governed by hypotheses which we derive from our experiences and develop from theories represented by constructs. Constructs are defined as "a way in which some things are construed as being alike and yet different from others" (p. 74). These evaluations are modified by experience, and will be unique to each individual but share a degree of commonality with others. The finite set of constructs will be interrelated to form a

system, or mental model, and Kelly proposed a method, termed repertory grid technique, to elicit such systems. Relatively few studies have employed grid technique in the field of IR, but the method has been used to model information space (McKnight, 2000), to determine mental models of IR (Zhang & Chignell, 2001), and in the classification of text types (Dillon, 1994; Dillon & McKnight, 1990) and digitised photographs (Burke, 2001).

The repertory grid technique is employed in this study to elicit a set of constructs, defined here as user statements relating to those system aspects of importance to the user. These are then further investigated by laddering, a process often employed in conjunction with repertory grid technique. The use of a finite number of probes during elicitation and laddering leads to a flexible approach with minimal bias, but provides data with a degree of inherent structure. Two pilot interviews were conducted to determine the best design for the study and the final methodology is presented here.

Ten first year undergraduates were recruited for the study during October 2002. These participants had basic levels of knowledge of search tools and techniques but had not received formal IR training. A small sample size is commonly used when implementing a repertory grid investigation (Dillon & McKnight, 1990; Hassenzhal & Trautmann, 2001; Moynihan, 1996; Dunn, 1986). For a given population, the use of ten participants will ensure determination of the complete set of important constructs. Data was collected on an individual basis, and involved three stages, introduction to a selection of search engines during a familiarisation session, a tape-recorded interview during which constructs were generated for inclusion in a ratings grid, and exploration of the constructs using probing questions. The process is outlined in more detail below and focuses on the generation of the qualitative data, which occurred mainly during the final laddering stage.

Familiarisation session. To identify the engines for use in study, a number were profiled to determine a small set representative of common search technologies. The engines chosen were AltaVista UK, Google UK, Lycos UK, and Wisenut, and these formed a set of 'elements' from which to elicit the constructs. Each participant searched using each engine for information to satisfy a chosen coursework assignment, thus ensuring sufficient motivation and realism of task. Time spent with each system was constant across the set, but the order of presentation of the systems varied across participants, to reduce learning effects.

Construct elicitation and grid completion. Participants gave an overall rating of success for each search engine, taken immediately after familiarisation. The method of dyadic elicitation was then used to generate constructs for use in the qualitative study. During this process, participants considered the search engines in pairs, and stated either a similarity or a difference between the members of each pair. The opposite of the stated similarity or difference was then obtained to form a construct, represented by a five-point scale along which all engines were rated. During elicitation, an additional engine, the participants' perceived 'ideal' search engine, was introduced; 'ideal' elements are commonly included in grid studies where element number is low (Whyte & Bytheway, 1996; Hunter, 1997). Pairs of engines were presented until no new constructs were elicited.

Laddering. The grid completion phase provided quantitative data and a great deal of qualitative data relating to more detailed exploration of the constructs was also obtained. Kelly (1955/1991) put forward a corollary to his theory of personal constructs, the organisation corollary, which indicated his belief that construct systems are hierarchically organised. The constructs are essentially interrelated by cause and effect. Some constructs are

central to a person's beliefs, and can be visualised as forming the topmost points of a pyramid. The lower positions can be filled with the system of constructs as they relate to each other. Thus, starting at any point within this organisation, termed a 'seed item', it would be possible for an interviewer to guide a participant up, down and across his construct system by using a series of probing questions (Rugg et al., 1999). This method is essentially a combination of the laddering technique used to move upwards within the hierarchy (Hinkle, 1965), with the pyramid technique used to move downwards in the hierarchy (Landfield, 1971). It has now become standard for the term 'laddering' to refer to the combined method.

Laddering has been used in the field of knowledge engineering, with particular success in the determination of the structure of knowledge in classificatory domains (Corbridge, Rugg, Major, Shadbolt, & Burton, 1994). Even where structure is only of minor interest, laddering will ensure full construct elicitation, by the decomposition of constructs to give more precise definitions, thereby ensuring that a construct represents only one facet. Tan and Hunter (2002) suggest laddering will clarify "underlying assumptions and interpretations of the label associated with the construct", and Hunter (1997) employed the technique within a repertory grid environment, which was felt to "…offer the research participant the fullest amount of freedom to comment upon a subject, yet still maintain a structured method to the data-gathering process" (Hunter, 1997).

Corbridge et al. (1994) emphasise that the probes used during the process should be standardised. The general rules given by Stewart and Stewart (1981) recommend use of 'why?' questions to take the participants higher up their pyramids, while 'how?' questions will move lower. A common strategy begins with the determination of a construct using an elicitation technique. The participant is then asked to identify which pole of that construct is

preferable and this is then taken as a seed item. The participant is then asked to state a reason for the expressed preference for the seed item, and the stated reason then becomes a new seed item and the process is repeated. Once the participant is no longer able to move upwards within their hierarchy of constructs, the interviewer returns to the original construct and begin a series of probes that will assist the move downwards, commonly by requiring the participant to state how the two poles of the construct are different from each other. An explanatory example of the laddering process, using the sample construct 'Interface simple / Interface cluttered', is given by Figure 1. The type of probe being used at each stage is indicated by the italicised comments, and a visual representation of the construct hierarchy is also provided beneath. This type of diagram was used during data collection for this study, to record the basic laddering information in note form.

Laddering Upwards

Interviewer: Considering the construct of interface simple / interface cluttered, which would you prefer? *[Determining the positive pole of the construct]*
Participant: A simple interface
Interviewer: Why would you prefer a simple interface? *[Probe to move upwards]*
Participant: Because it is easier for me to see where I have to type in the words
Interviewer: Why is that better for you? *[Probe to move upwards]*
Participant: Because then it's quicker to search.

Laddering Downwards

Interviewer: Thinking about the difference you just mentioned of a simple or cluttered interface. Can you think of any ways in which simple and cluttered interfaces are different? *[Probe to move downwards]*
Participant: A cluttered interface has lots of writing on it.
Interviewer: Can you explain what you mean by writing? *[Clarifying answer]*
Participant: Links to other things
Interviewer: Can you think of any other ways in which simple and cluttered interfaces differ? *[Probe to move sideways]*
Participant: Cluttered interfaces have lots of adverts.

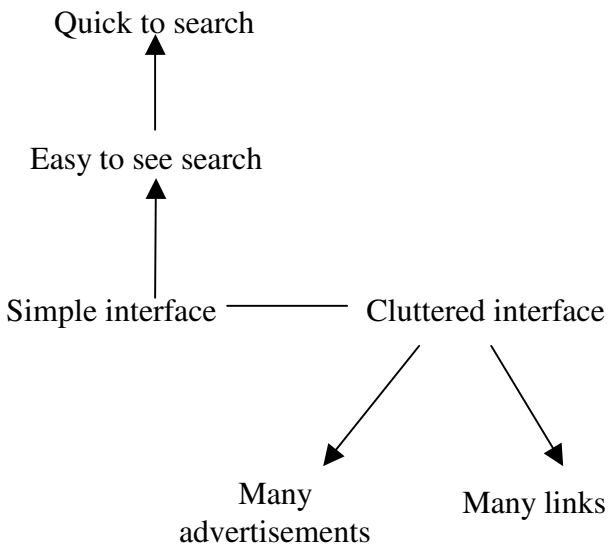

Figure 1: Elicitation of ladders using the construct 'Interface simple/ Interface cluttered', with suitable outline representation in diagrammatic form.

Data Analysis

Analysis of the quantitative data arising from the numerical grid completion is presented in Crudge and Johnson (2004), together with the complete set of raw constructs. This paper concentrates on the large quantity of qualitative data arising from the detailed exploration of the constructs. The analysis of the resulting qualitative data set was designed to exploit the hierarchical data structure obtained by the laddering method. The analysis was informed by the Grounded Theory approach of Strauss and Corbin, (1998), and the means-end chain analysis of Reynolds and Gutman (1988).

Following transcription of the tape recorded interviews, the raw construct set was used to provide a partial template to facilitate first level coding. The data was divided into 479 short segments, indexed by 65 different codes. Atlas/ti (Muhr, 1997) was used to enable grouping of the coded sections into themes, and the themed groupings were then divided into subsections following detailed examination. The hierarchical consequential relations between data segments were then determined using the probing questions of the laddering technique to facilitate identification. This stage was derived primarily from the means-end chain analysis method proposed by Reynolds and Gutman (1988) for the analysis of laddering data, but also corresponded to the axial coding phase of Grounded Theory. An example of a consequence chain derivation is provided by Figure 2. The direction of the arrows indicates the direction of the implication, with the left hand side corresponding to the lowest levels of the hierarchy, and the probe ‘Why is that important to you?’ being used to move across to the higher levels of the hierarchy on the right.

Participant: This is better, it's at the top [refers to directory link].
Interviewer: **Why** is it better at the top?
Participant: More aware of it at the top. When you see the page you see that first.
Interviewer: **Why** do you want to be more aware of it?
Participant: So I don't waste time looking to see… I want to know it knows my query.

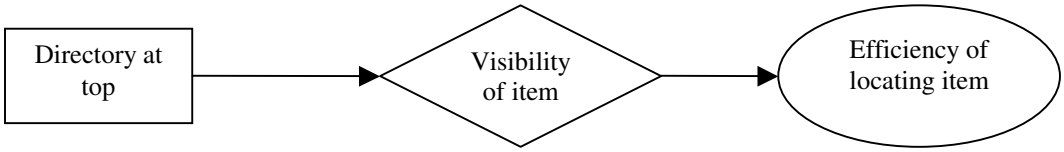

Figure 2: Derivation of a consequence chain from the raw data.

All the consequence chains were examined, together with the themed groupings already identified, and a generalised consequence chain was determined. This represented all the possible hierarchical interrelations, with a large proportion of the data appropriately assigned to one of three categories, ranging from the lower hierarchical levels of basic description, through the middle levels of evaluative description, to the highest levels termed key evaluations. A discussion of the types of data contained in each of these categories is provided subsequently. The generalised chain is included as Figure 3, with the causal relations indicated by the arrows, the thickest of these providing the main pathway through the hierarchy. The thinner solid lines indicate the possibility that statements from one data type could cause statements drawn from the same data type. Finally, the broken lines indicate the presence of affective statements within consequence chains. There was a substantial portion of data pertaining to emotional responses to the systems, and this was observed to occur at a variety of points in the hierarchy.

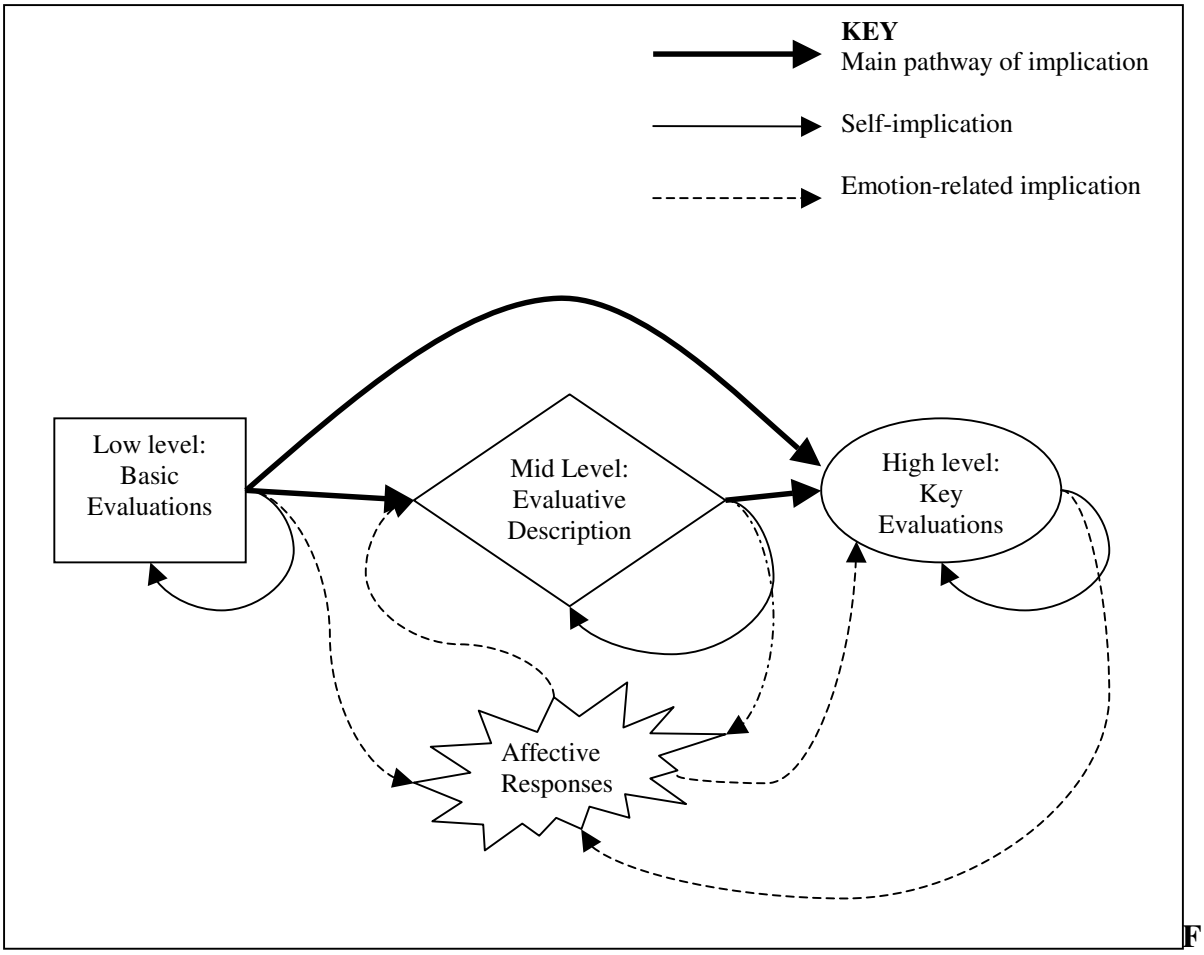

Figure 3: Generalised consequence chain demonstrating the relations between the main data types

Following the Grounded Theory approach, coding for process followed, during which the interview transcripts were recoded with the purpose of identification of evidence of the perceived process.

Thematic Discussion

The final code types identified serve to divide the data into three main hierarchical areas comprising basic description, evaluative description, and key evaluations. The process was also identified during the analysis process, as was a substantial amount of data pertaining to affective responses to the systems, and the data forming each main area is now considered in more detail. Frequency counts are also reported, although the aim of the study was not to obtain these, and conclusions regarding relative importance cannot be drawn based on

frequencies for the small sample size. Nevertheless, the frequency counts give an idea of the extent of overlap of the data between the respondents, and the proportion of the overall model that may be held by any one participant. For tables 1, 2 and 3, the frequency column refers to the number of participants presenting the item within their data at least once.

Basic description

The basic description forms the lowest levels of the consequence chains derived from the participants during laddering. The data from this section is characterised by its focus on description, and a lack of inclusion of more evaluative judgements. The section divides into description of the screens and the features of the systems. Table 1 provides the number of participants reporting aspects of screen layout or features.

Aspect	Frequency
Front page layout	**8**
Result page layout	**9**
Functionality	**10**
Presentation of features	9
Specific named features	9

Table 1: Frequency of reporting for the main areas within the basic description.

Screens. Three screens are described, the main entry page, commonly called the front page or interface, the result pages, and the advanced search page. The main issue emerging for the design of the front page related to the style, which usually reflected the streamlined or portal appearance. Participants gave description such as ‘busy’, ‘plain’, or ‘cluttered’, and discussed the presence of links, writing, adverts and ‘stuff’ on the page. Plain front pages were usually preferred, but one participant felt such pages could have too few colours. The issue of colour resulted in difference of opinion, with other participants preferring fewer colours. The colour was also referred to as mellow, garish or heavy. Five participants referred to the search box, and it was apparent that this featured strongly in the mental models of these participants.

All but one participant discussed the layout of the results page. In a parallel with the front page, participants distinguished between plain and busy pages, with only one preferring the busier variety. Definition of plain and busy varied, relating to font type and size, inclusion of URL and other information, use of numbering, and size of site descriptions. Other result page aspects included the need for a statement of the results quantity, use of colour to identify visited URLs, and a link to further pages of results.

Advertising. Advertising was discussed by eight participants, with varying reference to location on front or result pages. The participant set varied in attitude to advertising; although several participants felt that adverts should not be present there was some degree of acceptance, and one participant even felt advertising could be a positive issue if it took the form of a joke or cartoon. The quantity of advertisements was connected to the overall style of the interface. When considering advertising on the result pages, the two main issues were location and relevance. Positioning immediately prior to search results was a bad aspect; better presentation had advertisements grouped together and at the side of the page. While participants mentioned that relevant advertising might be acceptable, there was some disagreement over the definition of this, with one participant stating that a link to a store selling books on the search topic was relevant, and a second participant giving the same example as not relevant. Pop-ups and moving advertising were not favoured, and the colour, size, and ‘subliminal’ nature of the advertising were also mentioned.

Functionality. All participants discussed features, with specific features mentioned including categories, sneak-a-peek, directory, advanced search, image search, language facility, news, and e-mail. Several general issues pertaining to functionality were raised, including the quantity and variety of options available, and the relevance of features to searching. Only one

participant stated a preference for the inclusion of non-search related options. The presentation of the features was also referred to, and the tab-style was usually favoured, but one participant preferred a drop-down menu. More commonly, participants referred to the location of the features, or the location of the point of access to the features. Location at the top of the page or on the front page was often preferred.

Evaluative description

Table 2 provides the number of participants reporting each of the three main areas of evaluative description, namely readability of the screens, visibility of items including the search box, and the content of the results. These are areas that typically appear lower down the consequence chains, but are not purely descriptive.

Aspect	Frequency
Readability	**7**
Entry page	4
Result page	4
Features	3
Visibility	**7**
Items	7
Search Box	5
Content criteria	**10**
Relevance	9
Quantity	9
Precision/ ranking	7
New/ familiar results	4
Utility	4

Table 2: Frequency of reporting for the main areas within the evaluative description.

Readability of the screens was most affected by the choice of a streamlined or portal style, and the use of colour; plainer interfaces were more readable, while heavy or large amounts of colour were 'hard on the eyes'. Readability of features such as the tabs or pull-down menus was also discussed, with size and colour affecting this. Visibility was an issue, especially for the search box and access to features such as the directory. The location most affected the visibility of access to features, whilst the interface style was commonly stated to affect search

box visibility. Locating access to features at the top of the page or on the front page would increase visibility, whilst the adverts, writing and 'stuff' on an interface caused reduced visibility, especially noted for the search box. Several participants discussed the use of colour to highlight terms or the statement of number of 'hits', thereby rendering them more visible.

The content of the results was commonly discussed, but the criteria upon which results were assessed varied greatly. The table indicates areas where a degree of commonality occurred across the participant set, but a number of more individualised criteria were also specified. Some of the content criteria seemed heavily dependent on each other, and were not clearly delineated within the participant's mental models. There is interlinking of concepts such as quantity, precision, ranking, and relevance, which is also complicated by the inclusion of the process element of refining. A selection of participant comments illustrating the interlinking of these issues is provided as Figure 4. In addition to these main issues, raised by a high proportion of the participant set, smaller numbers of participants also raised a variety of other issues, including presence of familiar results, utility, and quality of retrieved sites. Only one participant discussed recall.

- "Because that's why you're visiting the search engine in the first…. You want relevant useful results, or at least results that are going to make you think about searching on a different term, or that are heading the right way towards finding the answer that you want."
- "I wouldn't mind how many results there were so long as they were all relevant to what I was looking for. But if they were totally unrelated or they were, they weren't what I was looking for, then obviously the less results the better really, less but pertinent results."
- "If I was left with, say, 12 results, I would expect that to be more in depth and detailed and more useful for what I was looking for."
- "You don't want to have, well the ideal thing is to possibly have fifty results or something, you don't want any more than that otherwise you'd be…. So of course if you don't have it, you don't want to have 300,000 results or something and if you've got no way of reducing them you're just going to be lost."

Figure 4: A selection of participant comments relating to the issues of relevance, refinement and quantity of results.

Key evaluations

Evaluations that occur at higher up the consequence chains, often resulting from laddering of constructs typically provided at lower levels, are core concepts and as such are termed here 'key evaluations'. These are the reasons why a system aspect was important to a participant, and are grouped here as ease of use, effort, efficiency, or effectiveness. The terminology chosen for the four sections reflects the ideas of literature and research, and although the terms efficiency and effort can have a more complex interpretation, for the purposes of this study they simply represent participant statements such as 'time taken' or the 'amount' a task must be performed. Effectiveness is taken to be the often highly individualised combination of the content criteria.

Ease and efficiency occur quite frequently in the data, with all participants referring to efficiency and nine referring to ease. Effort occurred less often, with only half the participant

set referring to this. Some examples of user statements relating to the key evaluations are provided in Figure 5. Identification of several co-occurrences of key evaluations within the data suggests that the concepts may be interlinked. However, there is inconsistency in reporting that makes it impossible to draw conclusions about a possible hierarchical order for the concepts. For the purposes of discussion, references in the data have been explicitly separated as far as possible.

Ease

- "Oh, I just found it user-friendly, I just found it nice. Because sometimes I must admit, I can just close a window and close the whole damn thing, you know, and I've got to go back again, whereas with that it's easier not to do that, isn't it." [sneak-a-peek]
- "I think AltaVista is near to my ideal engine, it's very good, easy to find the result. The way you search is very good."

Efficiency

- "When you're looking for something you don't want to spend hours and hours searching for it, you just want to find it and get on with what you're doing basically."
- "What I did like was on the Wisenut one you could take a preview of the actual site…if you're looking for something quickly, saves you having to like click forwards and backwards and that sort of stuff."

Effort

- "You have to work out a little bit more yourself more words to put in."

Effectiveness

- "I expect it to find relevant data, I expect it to find all the data, because it's supposed to be powerful, and I expect it not to give rubbish providing your search command is reasonably precise."

Figure 4: Selection of participant comments relating to key evaluations.

Ease of use. There are several types of ease of use; the ease of use of search functionality, ease as increased by search functionality, and ease as affected by the design of the screens. When discussing the ease of use of search functionality, one participant related the complexity of the advanced search to ease of use, and felt that a complicated advanced search would result in non-use. Wisenut's sneak-a-peek feature was stated to be easy to use, either navigationally or by reducing errors. Term suggestion features were also easy to use navigationally. Comments such as "you just click" were common explanations for ease of using categories. The location of categories at the top of the page increased their visibility and thus affected their ease of use. Finally, one participant discussed that the style of presentation of features would have an effect on the ease of use, preferring drop down menus to tabs, with the vertical list approach of the pull-down menu being more readable and so easier to use.

The ease as increased by search functionality was referred to by three participants as ease of use of the general search mechanism itself. Screen design issues such as quantity of information, colour, and advertisements impinged on the general ease of use, and one participant elaborated that for an interface with many links, it became more difficult to pick things out, and so was harder to use.

Efficiency. Efficiency was commonly reported during laddering, with all participants stating the time required to be a consequence of at least one lower level descriptive element. The length of time taken whilst using a search engine was always a negative aspect.

Eight participants gave time saving as a reason why the results content was important. The relevance, quantity and precision of the results were all stated to have an impact on the time required, and one participant mentioned scrolling through the results as the reason why the time was increased. Another participant expressed a dislike of the inclusion of PDF file types in the results content, and explained that these could take too long to load in.

The layout of the result pages was also stated to lead to extra time being required. Colours to indicate visited links would save time by reducing unnecessary revisiting of sites. A greater number of lines in the site descriptions would speed up the assessment process and the navigational aspect of clicking on titles to visit a site was also felt to be quick.

The evaluative descriptions of readability and visibility both affected the time required when using the front page. One participant who found the ‘busy’ interfaces harder to read felt that this impacted on time. The extra time required to locate the search box if it was surrounded by other information, was also highlighted. Advertising slowed down two participants, who

cited pop-ups and moving adverts as causes of this problem. The visibility of the adverts was another cause of time expenditure, for pop-ups this was explained as the time required to close them down.

Features were often stated as time-saving, with the reasons usually linked to the perceived use of the feature. Two participants stated that the cache feature saved time by ensuring access to sites even when 'down'. Half the participant set felt that sneak-a-peek would save them time; the reduce need to open a new page, reduction of navigational forward and backward clicking, and the use for relevance assessment were reasons provided. Categorised results/ term suggestions also saved time, either by allowing quick access to a subset or to quickly obtain more relevant information.

The presentation of the features was also linked to the time required. The 'quick launch' access to news at AltaVista, use of tabs or clicking to access things, and the location of functionality would all save time. The location of the directory was mentioned by three participants; placement at the top of the page increased visibility and reduced time. Finally, the location of pull-down menus or tabs at both top and bottom of page would reduce the need to scroll, thus saving time, and stated by one participant.

Effort. This key evaluation was the most difficult to identify from the data, and is taken to mean the 'amount' that a participant must do something. Participants variously discuss the amount of formulation required, the amount of navigation, the scrolling as linked to location of features and results precision, and the changing of pages, especially during relevance assessment, as linked to features that reduce it. Finally, the amount of refinement required as related to the content of the results was also discussed.

Effectiveness. Many participants were observed to combine various content criteria in an individual manner in order to define a 'good' result. The content criteria for effectiveness are more complicated than just having the information you want. There are simple methods the users employ to judge the relevance of a site at a glance. For example, several participants judge sites to be 'right' if they are the same as those retrieved from other sites. Others use term proximity as observed in site descriptions for relevance assessment. The presence of such shortcuts to relevance assessment perhaps stems from the users in the study providing general constructs relating to the overall effectiveness, rather than criteria for an individual site's relevance. However, the main issues for effectiveness were the quantity and relevance of the results.

The issue of results' quantity was closely linked to the precision, relevance and ability to refine. Participants were often unable to separate these concepts out in their discussion. Several participants equate the quantity of results with irrelevance, expecting a greater number of irrelevant results to be present in a larger retrieved set. To this extent, the quantity of results influences the perception of them by participants and the attitude to refining then becomes important, with some participants being more prepared than others to formulate or refine. The interrelations between these issues, and the combination of the criteria to produce an overall effective result, are complex and highly individualised, as would be expected.

Process.

The data relating to the process was grouped into three main phases, namely query input, results phase and refining phase. Further subdivisions are outlined in Table 3, together with the frequency of reporting. A set of action statements found to be common across process phases was also identified and is included in the table. These actions involved location of items such as the search box or features, reading the screens, typing in, and the navigational

actions of scrolling, activating items, and changing pages. The frequency of scrolling is noticeably lower than the other actions, perhaps because participants were less aware of it, or perhaps because they did not undertake it.

Aspect	Frequency
Query input phase	**8**
Formulation	8
Advanced search	4
Results phase	**10**
Assess sites	7
View sites	5
Manipulate results	4
Visualise query	2
Refining phase	**10**
Reduce quantity	6
Improve relevance	5
Actions	
Read screen	9
Locate item	8
Type in	7
Activate items	7
Change page	7
Scroll	3

Table 3: The frequency of reporting of the process and action statements.

Although the actual process statements were derived from the data, the groupings here are imposed by the researcher. Comparison of the process data determined during this study with the process models and research of the literature gives rise to several areas of consideration. The overall picture when compared with existing IR process models is suggestive of the traditional IR model, as outlined by Saracevic (1996). When compared to the transaction log studies of search engines, however, a higher concern for refinement is evident in this data set than might be expected. Finally, the delineation of the procedural data into process stages and action statements is suggestive of a micro level of perception on the part of users that is more commonly analysed in usability studies and navigational explorations than IR models.

Unfortunately, there is insufficient data and evidence in this study to formulate a detailed overall model of the procedural interaction with the system as perceived by the participant set. The data provides information as to the process stages and actions, but does not allow the patterns of process stages and actions to be identified. However, the absence of any unexpected process elements suggests that the perceived procedure is indeed in line with the user-side of Saracevic's traditional IR model (1996). This model moves from representation of the query, through formulation, comparison searching, and retrieving of documents. It also incorporates a simple feedback loop to allow reformulation of the query. All of these elements appear in the process data obtained through this study.

Affective Responses

There were five types of emotional statements elicited from the participant set during the interview process, namely frustration, confusion, overload, distraction and boredom. These emotional responses were identified as stemming from a variety of lower level data, and were usually the final members of a consequence chain. However, distraction also appeared as a cause of other aspects. Examination of the main causes of distraction identified a high proportion of presentational aspects, such as adverts, colour and amount of writing. Confusion similarly resulted from presentation aspects, such as the fonts and formatting, readability and clutter on interfaces. Advertising and visibility of features were the main causes of frustration, and the time taken was also a strong influence on an expression of frustration. Boredom related to having to perform a task repeatedly, such as refining or reading through imprecise results. The exploitation of the web-based medium by use of colour and layout was stated to increase interest. The causes of information overload were of two varieties, a profusion of retrieved sites, or an abundance of information provided by the search service itself.

Data relationships and models

The main data analysis has identified hierarchical delineations within the data, comprising basic description, evaluative description and key evaluations, related as consequence chains elicited during the laddering procedure. Although participants had not been required to make their process explicit at any point during the interview procedure, recoding of the data resulted in the addition of a surprisingly detailed outline of the process elements. Given the prevalence of procedural information within the data set, and thus the importance of the procedure to the users' mental model, the relationship among process elements with the other data groups was sought through an examination of data conjunctions. Examination of the location within the data where the process elements were identified produces two main conjunctions of data types, namely the process/ functionality conjunction and the process/ key evaluation conjunction.

Process/ functionality conjunction. The process/ functionality conjunction is evidenced by participants presenting information about the perceived uses they identify for search functionality. These might include, for example, use of categories to refine the search, or use of directory links to visualise the query. The focus of the study was to understand the perception of the systems overall, but the perceptions of individual features, although less useful because they are often specific to certain systems, are still of interest. Table 4 indicates five main features discussed by the participants, together with the process and action stages at which they were perceived to have use.

	Categories	**Directory**	**Sneak-a-peek**	**Cache**	**Advanced Search**
Assess relevance	-	-	✓	-	-
Manipulate results	✓	-	-	-	-
Refine general	✓	-	-	-	✓
Reduce quantity	✓	-	-	-	✓

Improve relevance	✓	✓	-	-	✓
Visualise query	-	✓	-	-	-
Navigate	-	-	✓	-	-
View sites	-	-	✓	✓	-

Table 4: Perceived use for the functionality available at Internet search engines.

Process/ key evaluation conjunction. The second point at which procedural elements were identified within the data set was as qualifying statements in conjunction with key evaluations. For example, a participant might discuss that it was 'easy to refine a search', thereby providing a conjunction of the key evaluation of 'ease' with the process phase of 'refinement'. The key evaluation of effectiveness, in the context of this study taken to be some combination of the content criteria, was not seen to occur in conjunction with the majority of the procedural elements. However, participants did refer to effectiveness without the use of refinement, and effectiveness after refinement had been carried out. The remainder of this section will focus on the other key evaluations of ease, effort and efficiency.

A full chart of possible conjunctions is provided as Table 5. Combinations occurring in the data are indicated by a tick, and the frequency is provided; combinations marked with a cross did not occur at all in the data set. For the three main divisions of query input, results and refinement phases, the frequency count indicated gives the number of participants who made reference to the category in general. This may have been in addition to one or more subcategories. The main phase division was still classed as present in the data if at least one subdivision was reported.

	Ease	**Efficiency**	**Effort**
Query input phase	✓ **(6)**	✓ **(2)**	✓
Formulation	✓ (1)	✓ (1)	✓ (2)
Advanced search	✓ (2)	×	×
Results phase	✓	✓	✓
Manipulate	✓ (1)	✓ (2)	×
Visualise	×	✓ (1)	×
View	✓ (2)	✓ (6)	×
Assess	✓ (3)	✓ (5)	✓ (3)

Refining phase	✓ **(3)**	✓ **(5)**	✓ **(2)**
Reducing Quantity	✓ (1)	×	✓ (1)
Improving Relevance	✓ (1)	✓ (3)	×

Table 5: Key evaluations as identified, with frequencies, at process points

From the total possible set of 24 specific process/ key evaluation conjunctions, two thirds were observed in the data. The highest percentage was observed for ease of use, with seven of the eight possible conjunctions occurring in the data. For efficiency, six of the eight conjunctions were reported, but the number observed for the effort evaluation was noticeably lower, with only three conjunctions reported. It is possible that the remaining conjunctions might have been observed had the sample size been larger. The continuation beyond ten participants in the study, whilst not expected to lead to the emergence of new facets in the data, might result in further shades of meaning. It is reasonable to conclude that some of the remaining key evaluation and process conjunctions would emerge in this way.

The Mental Model

The unification of the data into a summary model provides a suitable representation of the users' mental model of the systems. This model is presented as Figure 5, and shows the data as summarised by three strata. The main stratum contains the hierarchical data, and accounts for the majority of the data set. The other strata represent the affective data and the perceived process.

The main constituent of the mental model, the hierarchical evaluation stratum, comprises the basic description, evaluative description, and key evaluations. A pyramid is a common visual representation for laddered data, and in traditional terminology, the lowest levels are termed 'attributes', the middle levels are 'consequences', and the highest levels are 'values'. There are usually more attributes than consequences, and more consequences than values, hence the visual use of a pyramid representation. This is also the case for the data set here, with the key

evaluations providing the core concepts, positioned at the top of the pyramid and of fundamental importance to the participants. The pyramid is complicated by the interaction of the procedural elements with the key evaluations, and for this reason a suitable visualisation for the overall model is of overlapping strata with one of these strata taking the traditional pyramid form.

The affective stratum contains the set of emotional responses to the system. Nearly every participant was found to experience emotions as a result of the constituents of the hierarchical evaluation stratum. Most users would never verbalise their models as the participants in this study have done, and may thus have ceased their interactions with only emotional memories, and without themselves fully understanding the causes of the emotions at a conscious level. The occurrence of affective responses to the system requires interpretation informed by psychological theories of emotion, which is outside the scope of this paper.

The procedural stratum contains the data pertaining to process phases and actions, as derived from the participants' data as a by-product. The procedural data is important for correct interpretation of the hierarchical stratum; analysis indicated that the hierarchical data, in particular that drawn from the highest tier of the pyramid, occurred in conjunction with procedural data.

The model here is taken to be the compilation of the individual models of ten participants, and it is expected that this model is complete, in the sense that the addition of any further participants in a repertory grid study would not generate any new facets. Furthermore, previous research suggests that mental models in general will increase in accuracy and completeness as the experience level of an individual increases. The sample had moderate levels of experience, with 90% stating average or above average search engine experience,

and 80% stating average or above average Internet experience. It is thus expected that the participants in this study have presented reasonably complete models, with reasonable accuracy.

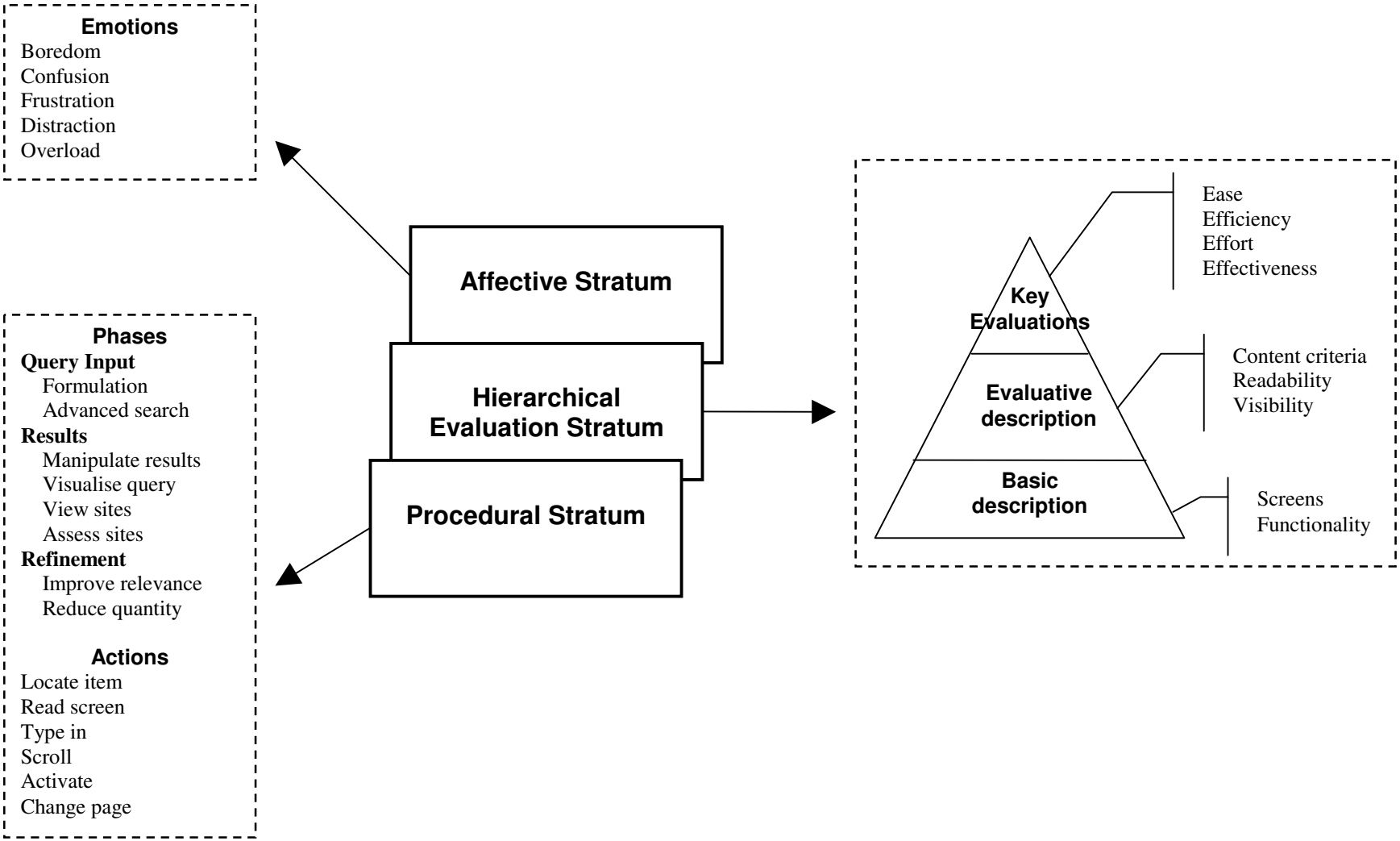

Figure 5: The users' hierarchical mental model

Discussion and conclusions

The mental model derived here from the qualitative laddering data is complete and hierarchical, with the importance of users' descriptions of system features suitably explained by the key evaluations. The value of such an explanatory mental model is explored here to identify avenues for further research.

Users' evaluative perceptions of search engines appear to be based on the key evaluations of ease, efficiency, effort and effectiveness. However, we can speculate that the models might not be as accurate and well developed as those elicited from an expert in the field. A survey of the literature pertaining to search engine evaluation, such as that provided by Su (2003), provides a suitable indication for the possible constituents of an IR expert's model. Traditional IR research often focuses on the importance of ranking and precision, in contrast to the users' mental model which indicates due concern for the search results but with a strong emphasis placed on the quantity of results retrieved. This perhaps highlights a naivety in the user model derived here – users are not so clearly able to interpret the quantity of results in terms of the success of the engine in dealing with them. Similarly, reduction of the quantity of results is commonly given by the users as a criterion for results refinement, where the IR expert might draw on their knowledge of IR and the engines and focus more strongly refining to improve the relevance.

Comparison of the user and IR expert mental model identifies further discrepancies. The indexing methods and database of the engine appear to be largely unrecognised by the user model. Although duplicates and dead links are mentioned singularly, it is not clear that the users understand that the search engine has a database and an index, only one user alludes to this through discussion of field searching. More typically the impact of the engine itself is

perceived through the screens and functionality, and not the index and database. The impact of screen design on the users' assessments of more 'obvious' measures such as ease of use, is evident from the hierarchy of the model.

From the user perspective these factors of search results and screen design have critical impact on the opinion formed. Users typically will not have an expert knowledge of the internal workings of a search engine and IR research, which has traditionally focused on the index and search, is now expanding to accommodate the user concerns for screen design. Yet whilst the user model naively does not include explicit perceptions of the index and database, the conjunction of process elements indicate a degree of sophistication in the user's perception of the engine as a tool to support the search process.

Both the user mental model and recent IR research (Johnson, Griffiths, & Hartley, 2001, 2003) show a consideration for conjunctions of evaluations taken at process stages. Although no single participant in this study identified every process/ key evaluation conjunction, the prevalence of process elements in the users' model is notable. Further research is needed to identify if the variety in the frequency of reporting of evaluations across the stages is significant. It is possible that the users' key evaluations may hold varying degrees of importance depending on the process stage in which the user is engaged.

Few IR studies have presented a methodology using a set of criteria such as 'ease', 'time' or 'effort' taken at identifiable process stages. Whether there is scope for the development of such a framework for user assessment of search engines remains to be seen. However, the current concern of IR evaluation studies to replace uni-dimensional measures such as user satisfaction with multidimensional constructs would suggest that the use of explicitly

delineated measures defined by the process/ key evaluation conjunctions identified in this study would result in an informative, meaningful evaluation of search engines.

In conclusion, the model determined here essentially represents the users' evaluative view of the search engines. The elicitation method made use of similarity and differences between systems to draw out comments defining the users' perceptions of the tools. As the derivation was based on comparisons, so the resulting model takes an evaluative context, and each individual interview essentially resulted in an evaluation of the system by that user. This study has not been concerned with the ratings given to individual engines, or to the association of the comments regarding good and bad design, results, ease of use, etc as related to individual search services. However, the model does provide a suitable framework for future development as a user-based system evaluation, in particular at the level of key evaluations and procedural conjunction.

References

Brandt, D. S., & Uden, L. (2003). Insight into mental models of novice Internet searchers. Communications of the ACM, 46(7), pp. 133-136.

Burke, M. (2001). The use of repertory grids to develop a user-driven classification of a collection of digitized photographs. Proceedings of the 64th ASIST annual meeting, Washington (pp. 76-92). Medford, NJ: Information Today, Inc.

Corbridge, C., Rugg, G., Major, N. P., Shadbolt, N. R., & Burton, A. M. (1994). Laddering: technique and tool use in knowledge acquisition. Knowledge Acquisition, 6, pp. 315-341.

Crudge, S. E., & Johnson, F. C. (2004). Using the information seeker to elicit construct models for search engine evaluation. Journal of the American Society for Information Science and Technology, 55(9), pp. 794-806.

DeKleer, J., & Brown, J. S. (1983). Assumptions and ambiguities in mechanistics mental models. In: Gentner, D., & Stevens, A. L. eds. Mental Models. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 15-34.

Dillon, A. (1994). Designing usable electronic text: ergonomic aspects of human information usage. London: Taylor and Francis.

Dillon, A., & McKnight, C. (1990). Towards a classification of text types: a repertory grid approach. International Journal of Man-Machine Studies, 33(6), pp. 623-636.

Dunn, W. N. (1986). The policy grid: a cognitive methodology for assessing policy dynamics. In: Dunn, W. N. (ed.) Policy analysis: perspectives, concepts and methods. Greenwich, USA: JAI Press, pp.355-375.

Hassenzahl, M., & Trautmann, T. (2001). Analysis of web sites with the repertory grid technique. Retrieved August 2003 from http://www.tu-darmstadt.de/fb/fb3/psy/soz/veroeffentlichungen_mh/ Chi01_hass_rgt.pdf

Hinkle, D. (1965). The change of personal constructs from the view point of a theory of construct implications. Unpublished Ph.D. thesis, Ohio State University.

Holscher, C., & Strube, G. (2000). Web search behaviour of Internet experts and newbies. Computer networks, 33, pp. 337-346.

Hunter, M. G. (1997). The use of RepGrids to gather interview data about information systems analysts. Information systems journal, 7, pp. 67-81.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real queries: a study and analysis of user queries on the Web. Information processing and management, 36(2), pp. 207-227.

Johnson, F. C., Griffiths, J. R., & Hartley, R. J. (2001). DEVISE: a framework for the evaluation of Internet search engines. Library and Information Commission Research Report 100.

Johnson, F. C., Griffiths, J. R., & Hartley, R. J. (2003). Task dimensions of user evaluations of information retrieval systems. Information research, 8(4), Retrieved July 9, 2004 from http://informationr.net/ir/8-4/paper157.html

Kelly, G.A. (1991). The psychology of personal constructs. London: Routledge (Original work published 1955).

Landfield, A. W. (1971). Personal construct systems in psychotherapy. Chicago: Rand McNally.

McKnight, C. (2000). The personal construction of information space. Journal of the American society for information science, 51(8), pp. 730-733.

Moukdad, H., & Large, A. (2001). Users' perceptions of the Web as revealed by transaction log analysis. Online information review, 25(6), pp. 349-358.

Moynihan, T. (1996). An inventory of personal constructs for information systems project risk researchers. Journal of information technology, 11, pp. 359-371.

Muhr, T. (1997) Atlas/ti: short user's manual. Berlin: Scientific Software Development.

Muramatsu, J., & Pratt, W. (2001). Transparent queries: investigation users' mental models of search engines. ACM SIGIR, New Orleans, Louisiana, USA, September 9-12, 2001.

Norman, D. A. (1983). Some observations on mental models. In: Gentner, D. & Stevens, A. L. eds. Mental Models. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 15-34.

Otter, M., & Johnson, H. (2000). Lost in hyperspace: metrics and mental models. Interacting with computers, 13, pp. 1-40.

Ratzan, L. (2000). Making sense of the Web: a metaphorical approach. Information research, 6(1). Retrieved July 9, 2004 from http://informationr.net/ir/6-1/paper85.html

Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis and interpretation. Journal of advertising research, 28, pp.11-31.

Rugg, B., Eva, M., Mahmood, A., Rehman, N., Andrews, S., & Davies, S. (1999). Eliciting information about organisational culture via laddering. Proceedings of the Enterprise Management and Resource Planning Studi, San Salvador, Venice, November 25-26. Retrieved August 2003 from http://leks.iasi.rm.cnr.it/emrps'99/papers/rugg_et_al.pdf

Saracevic, T. (1996). Modelling interaction in information retrieval: a review and proposal. Proceedings of the 59th annual ASIS meeting, Baltimore, October 21-24, (pp. 3-9). Medford, NJ: Information Today, Inc.

Saracevic, T. (1997). Extension and application of the stratified model of information retrieval interaction. Proceedings of the annual meeting of the American Society for Information Science, 34, pp. 3-9

Seadle, M. (2003). Editorial: mental models for personal digital assistants (PDAs). Library high tech, 21(4), pp. 390-392.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999). Analysis of a very large Web search engine query log. SIGIR Forum, 33(1), 6–12.

Slone, D. J. (2002). The influence of mental models and goals on search patterns during Web interaction. Journal of the American society for information science and technology, 53(13), pp.1152-1169.

Spink, A. (2002). A user-centred approach to evaluating human interaction with web search engines: an exploratory study. Information processing and management, 38(3), pp. 401-426.

Spink, A., Bateman, J., & Jansen, B. J. (1999). Searching the Web: a survey of Excite users. Internet research, 9(2), pp. 117-128.

Spink, A., Jansen, B. J., & Ozmultu, H. C. (2000). Use of query reformulation and relevance feedback by Excite users. Internet research, 10(4), pp. 317-328.

Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001). Searching the web: the public and their queries. Journal of the American society for information science and technology, 52(3), pp. 226-234.

Staggers, N. & Norcio, A. F. (1993). Mental models: concepts for human-computer interaction research. International journal of man-machine studies, 38, pp. 587-605.

Stewart, V. & Stewart, A. (1981). Business applications of repertory grid. London: McGraw-Hill.

Strauss, A. & Corbin, J. (1998). Basics of qualitative research: techniques and procedures for developing grounded theory (2nd ed.). London: Sage.

Su, L. T. (2003). A comprehensive and systematic model of user evaluation of Web search engines: I. Theory and background. Journal of the American society for information science and technology, 54(13), pp. 1175-1192.

Sullivan, D. (2003). Nielsen NetRatings Search Engine Ratings. Retrieved July 9, 2004 from http://www.searchenginewatch.com/reports/article.php/2156451

Tan, F. B. & Hunter, M. G. (2002). The repertory grid technique: a method for the study of cognition in information systems. MIS Quarterly, 26(1), pp. 39-57.

Thatcher, A. & Greyling, M. (1998). Mental models of the Internet. International journal of industrial ergonomics, 22, pp. 299-305.

Whyte, G. & Bytheway, A. (1996). Factors affecting information systems' success. International journal of service industry management, 7(1), pp. 74-93.

Zhang, X. & Chignell, M. (2001). Assessment of the effects of user characteristics on mental models of information retrieval systems. Journal of the American society for information science and technology, 52(6), pp. 445-459.