ABSTRACT

Classic IR (information retrieval) is predicated on the notion of users searching for information in order to satisfy a particular “information need”. However, it is now accepted that much of what we recognize as search behaviour is often not informational per se. Broder (2002) has shown that the need underlying a given web search could in fact be navigational (e.g. to find a particular site) or transactional (e.g. through online shopping, social media, etc.). Similarly, Rose & Levinson (2004) have identified the consumption of online resources as a further common category of search behaviour.

In this paper, we extend this work to the enterprise context, examining the needs and behaviours of individuals across a range of search and discovery scenarios within various types of enterprise. We present an initial taxonomy of “discovery modes”, and discuss some initial implications for the design of more effective search and discovery platforms and tools.

Categories and Subject Descriptors
H.3.3 [Information Search and Retrieval]: Search process;
H.3.5 [Online Information Services]: Web-based services

General Terms
Human Factors.

Keywords
Enterprise search, information seeking, user behaviour, knowledge workers, search modes, information discovery, user experience design.

1. INTRODUCTION

To design better search and discovery experiences we must understand the complexities of the human-information seeking process. Numerous theoretical frameworks have been proposed to characterize this complex process, notably the standard model (Sutcliffe & Ennis 1998), the cognitive model (Norman 1988) and the dynamic model (Bates, 1989). In addition, others have investigated search as a strategic process, examining the various strategies and tactics that information seekers employ over extended periods of time (e.g. Kuhlthau, 1991) and the effects of various levels of task context (e.g. Jarvelin and Ingwersen, 2004).

In this paper, we examine the needs and behaviours of varied individuals across a range of search and discovery scenarios within various types of enterprise. These are based on an analysis of scenarios derived from numerous customer engagements involving the development of search and business intelligence solutions based on the Endeca Latitude software platform. In so doing, we extend the classic IR concept of information-seeking to a broader notion of discovery-oriented problem solving, accommodating the much wider range of behaviours required to fulfil the typical goals and objectives of enterprise knowledge workers.

Our approach to enterprise discovery is an activity-centred model inspired by Don Norman’s Activity Centred Design (Norman 2006). This approach is an extension of previous activity-centred modelling efforts which focused on a “capturing [a systematic and holistic view of what users need to accomplish when undertaking information retrieval tasks more complex than searching]” (Lamantia 2006), employing Grounded Theory to provide methodological structure (Glaser 1967).

In this context, we present a model which has at its core an initial taxonomy of the “discovery modes” that knowledge workers employ to satisfy their information search and discovery goals. We then discuss some initial implications of this model for the design of more effective search and discovery platforms and tools.

2. INFORMATION RETRIEVAL MODELS

The classic model of IR assumes an interaction cycle consisting of four main activities: the identification an information need, the specification of an appropriate query, the examination of retrieval results, and reformulation (where necessary) of the original query. This cycle is then repeated until a suitable result set is found (Salton 1989).

In the above models, the user’s information need is assumed to be static. However, it is now acknowledged that information seekers’ needs often change as they interact with a search system. For example, Bates (1989) proposed the dynamic “berry-picking” model of information seeking, in which the information need (and consequently the query) changes throughout the search process. This model also recognizes that information needs are not satisfied by a single, final result set, but by the aggregation of results, insights and interactions along the way.

Bates’ work is particularly interesting as it explores the search strategies and tactics that professional information-seekers employ. In particular, Bates identifies a set of 29 individual tactics, organised into four broad categories (Bates, 1979).
Likewise, O’Day & Jeffries (1993) examined the use of information search results by clients of professional information intermediaries and identified three distinct categories of search behaviour: (1) Monitoring a known topic or set of variables over time; (2) Following a specific plan for information gathering; (3) Exploring a topic in an undirected fashion. O’Day and Jeffries also observed that a given search would often evolve over time into a series of interconnected searches, delimited by certain triggers and stop conditions that indicate the transitions between modes or individual searches executed as part of an overall enquiry or scenario.

More recently, Cool & Belkin (2002) proposed a faceted classification of interactions with information, in which their Information Behaviors facet contained nine disjunctive activity types (Create, Disseminate, Organize, Preserve, Access, Evaluate, Comprehend, Modify and Use). By contrast, Marchionini (2005) identifies three major categories of search activity (Lookup, Learn and Investigate) while Spencer (2006) suggests four modes of information seeking (Known-item, Exploratory, Don’t know what you need to know, and Re-finding).

### 3. A TAXONOMY OF ENTERPRISE SEARCH AND DISCOVERY

The primary source of data in this study is a set of 104 user scenarios captured during numerous customer engagements involving the development of search and business intelligence solutions based on the Endeca Latitude software platform. These scenarios were collected using a variety of methods, e.g. interviews, stakeholder workshops, direct observation, etc. They take the form of a simple narrative that illustrates the user’s end goal and the primary task or action they take to complete it, followed by a brief description of their job function or role, for example:

- “I need to understand a portfolio’s exposures to assess portfolio-level investment mix” (Portfolio Manager)
- “I need to understand the quality performance of a part and module set in manufacturing and the field so that I can determine if I should replace that part” (Engineering)

These scenarios were manually analyzed to identify themes or modes that appeared consistently throughout the set, using a number of iterations of a ‘propose-classify-refine’ cycle based on that of Rose & Levinson (2004). Inevitably, this process was somewhat subjective, echoing the observations made by Bates (1979) in her work on search tactics: “While our goal over the long term may be a parsimonious few, highly effective tactics, our goal in the short term should be to uncover as many as we can, as being of potential assistance. Then we can test the tactics and select the good ones. If we go for closure too soon, i.e., seek that parsimonious few prematurely, then we may miss some valuable tactics.”

There are however some guiding principles that we can apply to facilitate convergence on a stable set. For example, an ideal set of modes would exhibit properties such as:

- Consistency (they represent approximately the same level of abstraction)
- Orthogonality (they operate independently to each other)
- Comprehensiveness (they address the full range of discovery scenarios)

An initial set of nine discovery modes emerged from this analysis, which were subsequently grouped according to the three top-level categories proposed by Marchionini (2005). The nine modes are as listed below with a brief definition:

### 1. Lookup

1a. **Locating**: To find a specific (possibly known) item; 1b. **Verifying**: To confirm or substantiate that an item or set of items meets some specific criterion; 1c. **Monitoring**: To maintain awareness of the status of an item or data set for purposes of management or control.

### 2. Learn

2a. **Comparing**: To examine two or more items to identify similarities & differences; 2b. **Comprehending**: To generate insight by understanding the nature or meaning of an item or data set; 2c. **Exploring**: To proactively investigate or examine an item or data set for the purpose of serendipitous knowledge discovery.

### 3. Investigate

3a. **Analyzing**: To critically examine the detail of an item or data set to identify patterns & relationships; 3b. **Evaluating**: To use judgment to determine the significance or value of an item or data set with respect to a specific benchmark or model; 3c. **Synthesizing**: To generate or communicate insight by integrating diverse inputs to create a novel artefact or composite view.

Evidently, this taxonomy has been derived from a single data set and in that respect would benefit from further refinement. For example, Monitoring may be classified as a Lookup activity in the context of an engineer receiving a simple alert message, but it acts more as an Investigate activity when viewed in the context of an executive reviewing an organizational dashboard. Conversely, Exploring is a concept whose level of abstraction seems somewhat higher than the others, potentially compromising the consistency principle suggested above.

However, the true value of the modes will be realised not by their conceptual purity or elegance but by their utility as a design resource. In this respect, they should be judged by the extent to which they facilitate the design process in capturing important characteristics common to enterprise search and discovery experiences, whilst accommodating arbitrary variations in domain, information resources, etc.

### 4. MODE SEQUENCES AND PATTERNS

A further interesting observation arising from this analysis is that the mapping between scenarios and modes is not one-to-one. Instead, the modes tend to cluster, forming distinct chains or patterns analogous to higher-level syntactic units. More often than not, one particular mode will play a dominant role in the sequence. These patterns provide a framework for understanding the transitions between modes (echoing the triggers identified by O’Day & Jeffries), and can be used be used to provide further insight into enterprise search and discovery behaviour.

These mode chains echo the above-mentioned efforts to create goal-based information retrieval models, which yielded modes and a set of broadly applicable “information retrieval patterns that describe the ways users combine and switch modes to meet...
goals: Each pattern is assembled from combinations of the same [elemental] modes” (Lamantia 2006).

![Diagram of Discovery Mode Network]

**Figure 1. Discovery mode network**
The five most frequent mode patterns are listed below. These have been assigned descriptive (if somewhat informal) labels and an associated example scenario:

1. **Comparison-driven optimization**: (Analyze-Compare-Evaluate) e.g. “Replace a problematic part with an equivalent or better part without compromising quality and cost”

2. **Exploration-driven optimization**: (Explore-Analyze-Evaluate) e.g. “Identify opportunities to optimize use of tooling capacity for my commodity/parts”

3. **Strategic Insight**: (Analyze-Comprehend-Evaluate) e.g. “Understand a lead’s underlying positions so that I can assess the quality of the investment opportunity”

4. **Strategic Oversight**: (Monitor-Analyze-Evaluate) e.g. “Monitor & assess commodity status against strategy/plan/target”

5. **Comparison-driven Synthesis**: (Analyze-Compare-Synthesize) e.g. “Analyze and understand consumer-customer-market trends to inform brand strategy & communications plan”

Further insight may be derived by examining how the mode patterns combine across all the scenarios to form a “mode network”, as shown in Figure 1. Evidently, some modes act as “terminal” nodes, i.e. entry points or exit points to a discovery scenario. For example, **Monitor** and **Explore** feature only as entry points at the initiation of a scenario, whilst **Synthesize** and **Evaluate** feature only as exit points to a scenario.

**5. DESIGN PRINCIPLES FOR SEARCH AND DISCOVERY SOLUTIONS**
The modes establish a ‘taskonomy’ or collection of defined discovery activities which are structurally consistent, domain independent, orthogonal, semantically distinct, conceptually connected, and flexibly sequenceable. Such a profile – analogous to notes in the musical scale, or the words and phrases we assemble into sentences – could serve as a language for the design of variable scale discovery solutions through the use of common constructive mechanisms such as concatenation, combination and nesting. And if the modes do act as an elementary grammar for discovery, then sustained use as a functional and interaction design language should result in the creation of larger and more complex units of meaning which offer cumulative value.

Professional experience with employing the modes as both an analytical framework for understanding discovery needs and as a design grammar for the definition of discovery solutions suggests that both implications are valid. Further, our observations of using the modes suggest the existence of recognizable patterns in the design of discovery solutions. We will briefly discuss some of the patterns observed, doing so at three common levels of solution scale: on the level of a single functional or interface element, for whole screens or interfaces composed of multiple functional elements, and for applications comprising multiple screens.

**5.1 Single element patterns**

**5.1.1 Comparison Views**
One of the most common design patterns is to support the need for the Compare mode by creating A/B type comparison views that present two display panes - each containing data display charts or tables; or single items or groups of items - side by side to emphasize similarities and differences.

**5.1.2 Contextual Views**
Another common design pattern supports the Analysis mode by allowing a fore-grounded view of a single chart, table, item, or list, accompanied by its contextual ‘halo’ - the full body of information available about the element such as status, origin, format, relationships to other elements; annotations; etc.

**5.2 Whole screen patterns**

**5.2.1 Dashboard**
One of the most common screen-level design patterns is to support the Monitoring and Synthesis modes by presenting a collection of metrics which in aggregate provide the status of independent processes, groups, or progress versus goals in a ‘dashboard’ style screen.

**5.2.2 Visual Discovery Screen: 4-Dimensions**
A second common screen-level design pattern for discovery experiences is the visual discovery screen, which supports modes such Exploration, Evaluation, and Verification by layering views that present visualizations of several dimensions of a single axis of focus such as a core process, organizational unit, or KPI. When switching between layered views, the axis in focus remains the same, but the data and presentation in the dimensions adjusts to match the preferred discovery mode.

**5.3 Application-level patterns**

**5.3.1 Differentiated Application**
The ‘Differentiated Application’ pattern assembles a collection of individual screens whose distinct compositions and designs support individual discovery modes of Analysis, Comparison, Evaluation and Monitoring in aggregate to address the ‘Strategic Oversight’ mode sequence. Application-level patterns often address a spectrum of discovery needs for a group of users with differing organizational responsibilities, such as management vs. detailed analysis.

**6. DISCUSSION**
The above analysis is based on the assumption that the user scenarios provide a unique insight into the information needs of enterprise knowledge workers. However, a number of caveats apply to both the data and the approach.
Firstly, the scenarios were originally generated to support the development of specific customer solutions rather than for the analysis above. Therefore, the principles governing their acquisition may not faithfully reflect the true distribution or priority of information needs among the various end user populations. Secondly, the particular sample selected for this study was based on a number of pragmatic factors (including availability), which may also not faithfully represent the true distribution or priority among enterprise organizations. Thirdly, the data will inevitably contain some degree of subjectivity, particularly in cases where scenarios were generated by proxy rather than with direct end-user contact. Fourthly, the data will inevitably contain some degree of inconsistency in cases where scenarios were documented by different individuals.

We should also acknowledge a number of caveats concerning the process itself. In inductive work with foundations in qualitatively centered frameworks such as Grounded Theory, it is expected that a number of iterations of the “propose-classify-refine” cycle will be required for the process to converge on a stable output. In addition, those iterations should involve a variety of critical viewpoints, with the output tested and refined using a separate, independent sample on each iteration. Likewise, the process by which scenarios are classified would benefit from further rigour: this is a critical part of the process and relies on human judgement and inference. However, that judgement needs to go beyond simple word matching and be consistently applied to each scenario so that subtle distinctions in meaning and intent can be accurately identified and recorded.

That said, some interesting comparisons can already be made with the existing frameworks. For example, the first and third of the search modes suggested by O’Day and Jeffries have also been observed in our own study, and the second (arguably) aligns with one or more of the mode sequences identified above. Likewise, the Evaluate and Comprehend Information Behavior types identified by Cool & Belkin also appear as distinct search modes in our own taxonomy.

7. CONCLUSIONS AND FUTURE DIRECTIONS

To design better search and discovery experiences we must understand the complexities of the human-information seeking process. In this paper, we have examined the needs and behaviours of varied individuals across a range of search and discovery scenarios within various types of enterprise. In so doing, we have extended the classic IR concept of information-seeking to a broader notion of discovery-oriented problem solving, accommodating the much wider range of behaviours required to fulfil the typical goals and objectives of enterprise knowledge workers.

In addition, we have proposed a model which has at its core a taxonomy of “discovery modes” that knowledge workers employ to satisfy their information search and discovery goals. We have also examined some of the initial implications of this model for the design of more effective search and discovery platforms and tools.

Suggestions for future work include further iterations on the “propose-classify-refine” cycle using independent data. This data should ideally be acquired using a principled sampling strategy that attempts where possible to address any biases introduced in the creation of the original scenarios. In addition, this process should be complemented by empirical research and observation of knowledge workers in context to validate and refine the discovery modes and triggers that give rise to the observed patterns of usage.

8. REFERENCES


