TOWARDS UNDERSTANDING DIGITAL INFORMATION DISCOVERY

by

ELENA VOYLOSHNIKOVA
B.Sc., University of Victoria, 2002

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

We accept this thesis as conforming
to the required standard

Dr. ******, Supervisor, Dept. of Computer Science

Dr. ******, Member, Dept. of Computer Science

Dr. ******, Member, Dept. of Computer Science

Dr. ********, External Examiner

Supervisor: Dr. Margaret-Anne Storey

ABSTRACT

Everyday life revolves around the discovery and curation of digital information. People search the Web continuously, from quickly looking up the information needed to complete a task, to endlessly searching for inspiration and knowledge. A variety of studies have modeled information seeking strategies and characterized information seeking and curation activities on the Web. However, there is a lack of research on how existing Web applications support the discovery and management of information, especially concerning the motivations behind them and how different approaches can be compared.

In this paper, we present a study of information discovery tools and how they relate to the nature of information seeking. We propose a conceptual framework that deals with the opportunistic and purposeful aspects of how people discover and manage digital information. This framework can be used when designing, evaluating or updating Web applications.

Examiners:

Dr. ******, Supervisor,	Dept. of Computer Science
Dr. ******, Member,	Dept. of Computer Science
Dr. ******, Member,	Dept. of Computer Science

Dr. ********, External Examiner

Table of Contents

List of Figures

List of Tables

Acknowledgement

Dedication

Chapter 1 — Introduction

Today, people use Web technologies to satisfy their information needs. People research their interests and hobbies using various online resources, shoppers search online stores for product characteristics to make purchasing decisions, and travelers visit online booking sites to find information about flights and hotels. In order to accommodate diverse and evolving user needs, Web applications continuously introduce new features and services, empowering information discovery and curation.

Sometimes, Web users hope to find particular pieces of information, such as show times and phone numbers, to satisfy specific information needs [**?**]. Other times, users lack well-articulated information needs so they engage in opportunistic browsing [**?**]. Often, people discover information online without even looking for it [**?**]. The nature of information discovery can vary, and therefore, it requires elaborate tool support. Required functionality for information discovery and curation can also be distributed among two or more applications, which often leads to tools providing integrated solutions.

In addition, people perform information curation tasks, such as management and preservation, to maintain and add value to collections of information [**?**]. With the rapidly increasing popularity of socially-curated information spaces, it is important to understand how to enable management and curation activities when designing tools that support information discovery.

To close their knowledge gaps, people turn to various Web technologies ranging from specialized search tools to visual discovery applications. Several studies have been directed at exploring high-level Web tasks, including information seeking tasks [**?**, **?**, **?**, **?**], deriving models of information seeking behaviors [**?**, **?**, **?**, **?**, **?**, **?**], and looking at methods of information curation [**?**, **?**]. However, more research is necessary to determine how different tools and their features provide fundamental support for information discovery and curation.

To enhance information seeking and curating experiences and support users' in-

teractions, we extend existing research by (1) deriving factors that enable information discovery and curation and relating them within a framework, (2) using the framework to establish a set of questions that can be used when evaluating and designing new applications, and (3) iteratively evaluating the framework by using it to study and describe current Web applications, which in turn helped us refine the framework of factors and questions. In summary, the framework addresses our research goal which is to gain an understanding of how existing tools support digital information curation and discovery.

The remainder of this paper is organized as follows. Section 2 highlights some of the studies and technologies related to information seeking and curation tasks. The process of building and refining a conceptual framework of factors is documented in Section 3. Section 4 outlines the conceptual framework and provides questions that enable digital information discovery and support curation, as well as gives specific examples from real-world Web applications. In Section 5, we demonstrate how the framework can be used to reveal missing features and propose new directions for development. Section 6 summarizes implications for research and practice. In Section 7, we describe the limitations of the study, followed by future work and conclusions in Section 8.

Chapter 2 — Related Work

2.1 Web-based Information Discovery and Curation

Several researchers have studied various aspects of Web-based information discovery. To gain an understanding of how current Web tools support information discovery and curation, we first studied known characteristics of information-related Web usage, including high-level Web tasks, information seeking behavior, information curation, collaboration in information seeking, and modes of Web use.

2.1.1 Web Tasks

Kellar et al. [?] separated Web tasks into five categories: transactions, browsing, fact finding, information gathering, and other uncategorized tasks, with information seeking being composed of browsing, fact finding, and information gathering. Although the authors categorized information gathering as part of information seeking, it is in fact more closely related to digital curation [?, ?]. In their later work, Kellar et al. [?] added communication and maintenance as additional Web tasks.

Similarly to Kellar et al., Sellen et al. [?] identified six tasks that are commonly performed by Web users: browsing, finding, housekeeping, information gathering, communicating, and transacting. Therefore, Kellar et al. and Sellen et al. both identified browsing, fact finding, and information gathering as information-related tasks that users perform online.

People often engage in information seeking activities to close some knowledge gap that occurred as a result of not having enough information to perform a task [?]. Therefore, when providing tool support for various information discovery tasks, it is useful to consider the motivation behind these tasks as it can be different for each task. Morrison et al. [?] make a distinction between methods of Web use and pur-

poses. The authors derived a purpose-based taxonomy of Web use, including three purposes or motivations: finding information, comparing pieces of information or choosing products to make a decision, and using the Web to find relevant information to gain an understanding of some subject. Consequently, methods of finding information identified by Morrison et al. are collecting, finding, exploring, and monitoring. The differences between the two taxonomies suggest that different information seeking tasks may be performed to satisfy more than one information seeking purpose. Therefore, each purpose may require more than one task-supporting mechanism.

Morrison et al. also draws distinction between finding or looking up information and exploratory search. Whereas information lookup involves tasks such as fact retrieval, navigation, and verification, exploration is more cognitively demanding and involves learning and investigation [**?**]. Learning and investigation can be performed over multiple iterations, and can involve learning though various media, "social searching", and serendipitous browsing performed with the goals of knowledge acquisition, socialization, forecasting, and planning.

2.1.2 Information Behavior Models

A number of researchers have proposed models of information seeking and information behavior. Wilson [**?**] summarized some of the key work [**?**, **?**, **?**, **?**, **?**] on information behavior and proposed a new model. According to Wilson's original model of information behavior [**?**], information seeking behavior results form the user trying to satisfy their perceived information need. Consequently, the user makes demands on information systems. Success or failure of such demands dictates whether the process is repeated or, if the information need is satisfied, used or communicated with other people. In addition, Wilson defined possible barriers that can impede information seeking behaviors, as well as context that influences formation of the information need. These underlying ideas remained in the revision of Wilson's model [**?**]. Finally, Wilson proposed a "problem solving model" of information seeking behavior. The model reflects on the idea that people engage in information seeking and searching in order to resolve some uncertainty that stands in the way of solving, defining, or identifying a problem.

Ellis et al. [**?**, **?**, **?**] proposed a model of information seeking characterized by six

different patterns: starting, chaining, browsing, extracting, monitoring, and differentiating. Subsequently, Choo et al. [**?**] derived anticipated Web tasks that correspond to these patterns. According to the authors, when users identify sources of interest, they usually identify which Websites can point to that information of interest. Chaining occurs when users navigate through links on those initial pages. When people browse, they scan top-level pages, headings, lists, and site maps. Differentiating takes place when people bookmark, print, copy and paste information, or choose an earlier selected site. Monitoring occurs when users revisit Web pages or receive updates from previously visited sites. Finally, extraction can occur when the user systematically searches sites to extract information of interest. Ellis' model also complemented Kuhlthau's [**?**] work which corresponded stages of information seeking with feelings, thoughts, actions, as well as anticipated information tasks.

Information retrieval behaviors are further studied by Saracevic [**?**] and Ingwersen [**?**] who derived models concentrating on cognitive processes of information seeking.

Bates [**?**] proposed a model of four information seeking modes: being aware, monitoring, browsing, and searching. Bates differentiated the modes based on the user's level of attention being active or passive, and information needs being directed or undirected. Thus, browsing can be characterized as undirected active information seeking because users do not know directly what information they are looking for, but they are actively looking. Searching falls under active directed information seeking because the information need is clearly defined and the search is directed. Finally, monitoring and being aware are passive modes of information seeking although monitoring is directed and being aware is undirected.

2.1.3 Digital Curation

In 2002, Bates [**?**] extended her research with the notion of information farming, which involves people collecting and organizing information for future use and revisitation. More commonly, information farming is referred to as digital curation, which is the notion of collecting and managing digital information for the purpose of adding value to the collection and revisitation [**?**]. Wittaker [**?**] believes that in terms of Web use, a significant shift is happening from information consumption to information curation, which means that people no longer just use the Web to find and consume

the information that they are interested in, but they also try to save and manage that information so that it can be reaccessed and exploited later.

2.1.4 Collaboration and Information Seeking

By surveying 204 Web users, Morris found that people often desire to or do collaborate on information seeking tasks [?]. To collaborate on information seeking, people often use instant messaging, email, and create documents and Webpages to share information. Occasionally, collaborative information seeking occurs when collaborators work side by side and share search results in person.

Collaborative information-related activities on the Web are not limited to information seeking. Collaborative information tagging is a way of organizing content for future search and navigation. Although it is usually performed for personal reasons, tagging greatly enhances information retrieval [?].

2.1.5 Modes of Web Use

Categorizing Web usage into information seeking, digital curation, and other Web tasks does not necessarily give full insight about how information-related tasks are performed. Lindley et al. [?] conducted a qualitative study involving 24 participants, tracking their daily Web usage in the form of a diary. As a result of this study, the researchers identified five distinct modes of Web use: respite, orienting, opportunistic, purposeful, and lean-back. According to the authors, people browse the Web *opportunistically* when they look for information related to some personal interest, long-term goal, or future ambition. *Purposeful use* occurs when the users know what information they need to acquire or what online action they need to perform in order to continue or finish some other activity. *Respite* mode usually occurs when users are in the process of waiting for something or taking a break, and it serves as a means for people to temporarily occupy themselves when high engagement with the content is not a requirement. *Orienting* mode usually occurs when people want to be updated on what has been happening in their environment. Examples of this mode are checking email at work or looking at the news and updates on a social networking site. Finally, *lean-back* mode of Web use can be thought of as listening to the radio or

watching TV, and usually involves watching videos online or browsing through other types of entertainment content.

Lindley et al.'s primary motivations behind looking at use modes that occur when people browse the Internet was that traditional Web use studies and Web tasks discovered by other researchers cannot reflect the depth of user's intentions online. Understanding the characteristics of different modes guides the design of Web interaction. For example, opportunistic use can have blurry and continuously changing information needs. People often cannot indicate the completion of Web tasks, and they finish whenever they have been browsing the Internet for too long, or whenever they need to complete some other task of higher priority. Then, they will often resume their opportunistic information seeking. Finally, opportunistic use is 'grasshopper-like', which means that users jump from one resource to another [**?**]. From these factors, we can assume that to support such Web usage, we would need to consider mechanisms for supporting users' information needs and support revisitation and arbitrary navigation.

Today, there are a multitude of tools that support different aspects of information exploration and curation, but understanding how these tools are similar (or differ) is difficult. Moreover, the existing research is not useful at helping identify gaps in current tools or ways that current tools may be improved to support information exploration and curation. Thus, we present a framework of Web application design factors and questions that facilitate information discovery and curation (see Sec. 4).

Chapter 3 — Methodology

3.1 Building and Refining the Conceptual Framework

Development of our framework began with an extensive literature review. Although the previous section outlines only the key research that was considered, it illustrates the diversity of topics that contributed to forming an understanding of information seeking. From this review, we derived preliminary design factors.

Through a careful analysis of 20 information discovery applications (see Table 1), we iteratively expanded the framework, added concepts, and established relations between those concepts. The framework can be expanded further, however, we selected the most popular information discovery applications in use today and considered the full range of features in those tools (both by referring to the literature and documentation on those tools, as well as exploring the features). The popularity of information discovery applications was determined using Website popularity ranks provided by Alexa[1], a commercial Web traffic data provider. The focus was on applications that had strong information discovery components and lesser priority was given to applications whose purpose revolved only around curation. The framework was refined iteratively as we explored the literature and available tools, and for presentation purposes, we present the final version of the framework.

The exploration of information discovery tools was motivated by the following research questions:

RQ1: How do existing Web applications support information discovery?

RQ2: How do existing information discovery applications support information curation?

[1] Alexa is available at www.alexa.com

Table 3.1. *Web-based Information Discovery and Curation Tools*

Application	Description	Summary of findings
Pinterest	Visual discovery tool, available at www.pinterest.com	- Supports serendipitous browsing, bookmark-based rediscovery, channel-based information discovery, and information curation. -Lacks support for search- and history-based rediscovery and fact finding.
Delicious	Social bookmarking service, available at delicious.com	- Supports channel-based discovery, bookmark-based rediscovery, and supports social curation. - Lacks support for visual link preview and list-based categorization.
Tumblr	Microblogging platform, available at www.tumblr.com	- Supports serendipitous browsing, bookmark-based rediscovery, channel-based information discovery. - Lacks support for fact finding and list-based categorization.
StumbleUpon	Web page discovery tool, available at www.stumbleupon.com	- Supports serendipitous browsing, bookmark- and history-based information rediscovery, channel-based information discovery, and information curation. - Lacks support for fact finding.
Wikipedia	Free content Internet encyclopedia, available at en.wikipedia.org	- Supports serendipitous discovery, fact finding, search-based rediscovery. - Lacks support for history-based and bookmark-based rediscovery, personal preservation and resource evaluation.
Google Maps	Web mapping service, available at www.google.ca/maps	- Supports fact finding and rediscovery. - Lacks support for curation mechanisms, except for personal information preservation.
Rotten Tomatoes	Movie and TV database, available at www.rottentomatoes.com	- Supports fact discovery, serendipitous browsing, and search-based rediscovery. -Lacks support for history-based and bookmark-based rediscovery, information preservation, and management.
500px	Photography site, available at 500px.com	- Supports serendipitous browsing, channel-based discovery, and social curation. - Lacks support for fact discovery and list-based categorization.
BucketList	Goal tracking and discovery service, available at bucketlist.org	- Supports serendipitous discovery, bookmark-based rediscovery, and channel-based discovery. - Lacks support for fact discovery, search- and history-based rediscovery.

We used Yins strategies for designing a case study [**?**] for guidance. The motivation behind choosing a case study over other methods of qualitative research was based on our choice of research questions (which have an explanatory nature), the lack of control over existing applications and their development, and having to focus on contemporary use of real-life Web applications. According to Yin [**?**], a case study would be an optimal research strategy given the above characteristics.

For each case of our study, we chose a Web application whose primary purpose is to support information discovery. We examined the overall purpose of each application, its description as defined within the application, and literature and documentation related to the application (if they were available) against the features that the application provided. For example, if an application provided bookmarking features, we checked if it was indeed intended to be used for information preservation.

To increase external validity of our study, we chose cases based on replication logic [**?**]. Using replication logic in case study design means carefully selecting each case so that it either predicts analogous results or predicts contrasting results but for anticipated reasons. Therefore, we used our preliminary conceptual framework to predict if an application supported each of the information discovery and curation tasks based on the features that the application provided. If our predictions were inaccurate, we would modify the framework accordingly and move onto the next case.

Consequently, our methodology was an iterative process of selecting cases, analyzing them, and determining whether they could be described and evaluated using our framework. If we found a key feature that could not be described, we adapted the framework according to the findings. We repeated the process of case selection and evaluation until the framework was usable for all cases. We then grouped the elements of the framework into categories, recording corresponding questions to ask in order to evaluate applications.

A list of the tools that were used as cases as well as brief summaries of our findings for each tool are presented in Table 1. Summaries are limited and provide a general idea of the results of examining the tools using the framework. Other tools were considered throughout the study, however, only the 20 applications presented

underwent systematic examination. The framework itself is covered in the next section and presented in Table 2. Limitations of our study are outlined in Section 7.

Chapter 4 — Preliminary Framework

Serendipitous discovery refers to information discovery resulting from serendipitous browsing. Such discovery is characterized by under-defined, absent, or hidden information needs, and it usually involves browsing through diverse resources with varying content types [?, ?]. Here, resource is defined as a collection of information about a single unit of inquiry, usually bundled together for presentation purposes. Some examples of resources are places, images, blog posts, and Web pages.

Fact discovery refers to information discovery resulting from the search for a specific piece of information. It is characterized by a well-defined information need and is easier to perform within systems that provide access to homogeneous types of information [?, ?].

Rediscovery refers to information discovery resulting from revisiting previously discovered resources [?]. The following is a list of factors that enable rediscovery.

Table 4.1. *Preliminary Framework - Discovery*

Design factors	Questions to be posed during the design or evaluation of Web-based information discovery or curation tools
Discovery	
Serendipitous discovery	
Arbitrary navigation	Does the application provide a means for arbitrary navigation among resources?
Search-based navigation	Does the search engine help retrieve diverse resources related to the topic of interest?
Category-guided navigation	Do categories suggest and help with navigating to resources related to the topic of interest?
Integration	If resources originate from a different site, do they link to their original sources?
Visual link preview	If resources are delivered as links, do they have visual previews?
Spatial arrangement	Is there a semantic to the spatial arrangement of resources?
Fact discovery	
Search-based navigation	Does the search feature help discover the specific resource of interest?
Category-guided navigation	Do categories help narrow results to specific types of resources?
Integration	If resources originate from a different site, do they link to their original sources?
Uniform representation	If resources are uniform, are they presented in a uniform way?
Visual link preview	If resources are delivered as links, do they have visual previews?
Spatial arrangement	Is there a semantic to the spatial arrangement of resources?
Rediscovery	
History-based rediscovery	Does the application save and provide access to browsing history?
Bookmark-based rediscovery	Does the application support bookmark-based resource revisitation?
Search-based rediscovery	Is the search a reliable method for resource revisitation?
Channel-based discovery	
Site subscription	Does the application allow subscriptions to news and updates?
User subscription	Does the application allow subscriptions to other users' activities?
Notifications	Does the application have one or more notification mechanisms?
Subscription to news feed	Can subscription updates be visible within the application?
Content news feed	Can content updates be visible within the application?

Table 4.2. *Preliminary Framework - Curation*

Design factors	**Questions to be posed during the design or evaluation of Web-based information discovery or curation tools**
Curation	
Management	
List-based categorization	Does the application support sorting information into list-like structures, either privately or publicly?
Tag-based categorization	Does the application support tagging, either privately or publicly?
Preservation	
Internal preservation of internal resources	Does the application support bookmarking mechanism(s) for preserving internal information within the application?
Internal preservation of external resources	Does the application support bookmarking mechanism(s) for preserving external information within the application?
External preservation of internal resources	Does the application support bookmarking mechanism(s) for preserving internal information outside of the application?
Augmentation	
Evaluation	Can the resource evaluations be recorded privately or publicly?
Annotation	Can resources be annotated privately or publicly?
Sharing	
Adding resources	Can resources be publicly added to the collection of information within the application from other Web pages?
Internal sharing	Can internal resources be publicly reshared within the application?
External sharing	Can internal resources be publicly reshared outside of the application?

Chapter 5 — A Conceptual Framework for Information Discovery and Curation on the Web

Although Web-based information discovery and curation tasks are commonly performed today, as we mentioned above, there is a lack of literature on how to support them when building Web applications. I reduce this gap by presenting a framework of design factors facilitating digital information discovery and curation (see Figure 5.1).

In my framework, I built on existing classifications of information seeking tasks and methods and existing Web tools to derive corresponding design factors. The framework consists of two main categories (discovery and curation) that are consequently decomposed into subcategories. Each of the lower subcategories contains mechanisms that enable given aspect of discovery or curation and corresponding questions that can help application design and evaluation. Every component in the framework has corresponding automation and support elements that can improve user experience. This chapter outlines the main components of the framework.

5.1 Discovery and Navigation

In order to discover information, a user needs to have a way of navigating to it. Common methods of navigation that facilitate information discovery include descriptional, referential, random, and direct (see Table 5.1).

5.1.1 Descriptional Navigation

A navigation is descriptional when the user describes the information need. Most commonly is implemented as search-based navigation since it allows users to enter the search query and describe their information need.

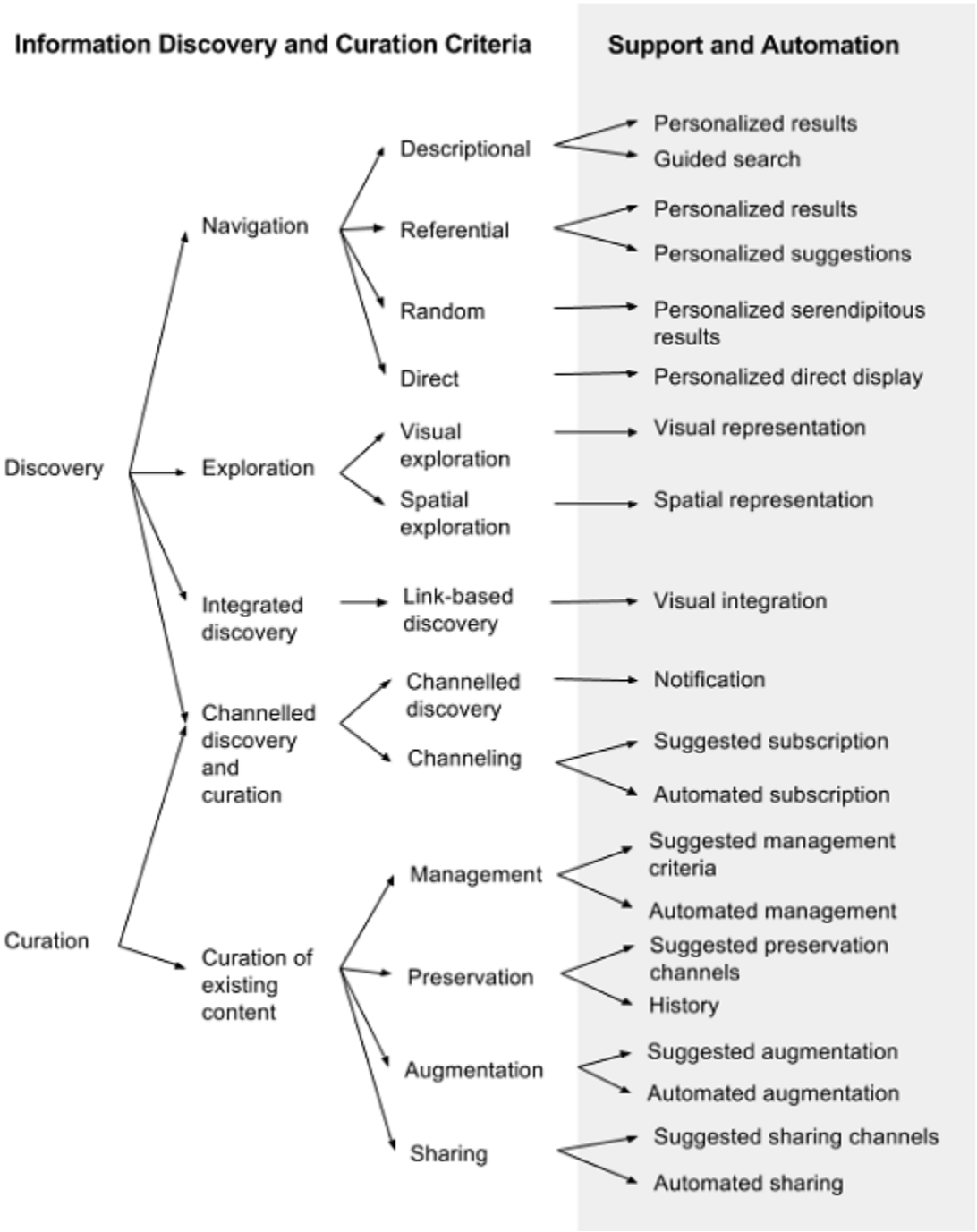

Figure 5.1. *Conceptual Framework Overview*

Table 5.1. *Navigation Mechanisms*

Navigation mechanisms	Questions to be posed during the design or evaluation of tools
Descriptional	
Search-based navigation	Can users navigate the site using search mechanism?
Referential	
Category-guided navigation	Can users navigate using categories?
Facet-guided navigation	Can users navigate using facets?
Filters-guided navigation	Can users navigate using filters?
Tag-guided navigation	Can users navigate using filters?
Search by resource	Can the user search by resource?
Random	
Random navigation	Is it possible to randomly navigate through resources?
Direct	
Direct display	Is any information displayed directly without active search?

There are two common ways of aiding information discovery when search-based navigation is used. The first method entails returning personalized results when the user enters a search query. There are a number of ways to accomplish personalization, but this chapter only focuses on which features can can support information discovery and curation have rather than their implementations. The second method is to suggest search terms to make it easier for the user to formulate the information need (see Table 5.2).

There exist numerous ways in which descriptional navigation supports information discovery. With fact discovery, an information need is known [**?**, **?**]. Therefore, descriptional navigation provides a way of for the user to express her information need.

Search-based navigation often serves as an entry point for information seeking [**?**]. In case of serendipitous discovery, since the information need is not well articulated,

descriptional navigation can be used to express a topic that could potentially relate to the information need. For instance, searching for a location within Pinterest returns numerous images of the location that link to (or integrate with) other resources, blogs, and Web pages, whereas searching for the same place on Google Maps usually returns a small set of possible locations with limited information about those places.

Descriptional navigation can help rediscovering information. However, search-based rediscovery is not always a reliable way of refinding information [**?**]. In information portals that provide access to fairly ambiguous information and that have information regularly repopulated and updated, the search feature is usually designed around retrieving information related to some topic, but is not very specific. In order to revisit a resource, search must provide consistent results. In information discovery applications that provide access to specific information, such as Wikipedia and Rotten Tomatoes, search can usually lead directly to a specific resource. However, within Web applications such as We Heart It or Pinterest, search-based rediscovery is often unreliable.

5.1.2 Referential Navigation

A navigation is referential when the user finds a reference to the term that she is looking for. The underlying assumption of this method of navigation is that the user can recognize needed information as she sees it.

Referential navigation methods can take many different forms. Some common ones are searching categories, facets, and filters. Often, Web applications implement tag-based navigation. In some applications, users can search by a given resource (see Table 5.1).

To further support referential navigation, applications can either personalize search results (similarly to descriptional navigation) or personalize reference suggestions (see Table 5.2).

rentialRefe navigation is used to direct the user to relevant resources [**?**]. In the case of fact discovery, such navigation should narrow the results to a specific type of resource so that further fact discovery is bounded by that type. For example, TripAdvisor lets the user choose among flights, hotels, vacation rentals, restaurants, and destinations.

For serendipitous discovery, referential navigation should provide a way to narrow the results to those related to one topic. In addition, categories, facets, filters, and tags can help the user formulate an information need by suggesting topics [?]. For example, when using Google Images, every search query suggests related categories of images to help users define an information need.

5.1.3 Random Navigation

In order to browse diverse information, an information discovery tool needs to provide a way to randomly navigate among resources, thereby supporting serendipitous information discovery [?]. Many applications, such as StumbleUpon, support random navigation to allow for opportunistic jumping from one resource to another. This method is useful when the information need is undefined.

To further enhance random navigation, Web tools sometimes allow users to personalize this type of navigation, which makes it less 'random'. However, this way the user can discover new information within a specific category, for example.

5.1.4 Direct Navigation

In a broad sense of Web-based navigation, direct navigation is associated with entering an address of a site and being redirected directly to it. In the context of Web applications for information discovery, direct navigation means displaying certain content to the user without user's active participation.

Often, applications display certain information as soon the user visits the site. It can be news feed, featured content, context-dependent information, or other types of information. Displayed content can also be personalized to improve information discovery with direct navigation.

5.2 Exploration and Discovery

Exploration of resources is another factor that enables information discovery. In particular, visual and spatial explorations of single or multiple resources allow for rapid information searching (see Table 5.3).

Table 5.2. *Automation and Support for Navigation*

Automation	Questions to be posed during the design or evaluation of tools
Descriptional	
Personalized results	Do the descriptional mechanisms return personalized results?
Guided search	Is the descriptional mechanism guided by suggested search terms?
Referential	
Suggesting topics of interest	Does the application suggest topics of interest?
Suggesting similar resources	Does the application suggest similar resources?
Suggesting tags	Does the application suggest similar tags?
Random	
Personalized destination	Is random navigation personalized to the user?
Direct	
Personalized display	Is direct display personalized to the user?

Abrams et al. [**?**] identified link representation as one of the problems with traditional bookmarking. Analogous with browsing through a bookmark manager, identifying relevant information when browsing through links to diverse resources can be a challenging task. A visual preview should make it easier to evaluate the relevance of resources. Applications that facilitate serendipitous information discovery often employ elaborate resource representation techniques. Many social bookmarking systems, such as Scoop.it! and StumbleUpon, support visual previews of bookmarked pages. Delicious is a social bookmarking application that lacks this type of link representation support, which makes it harder to determine if the link will lead to a relevant resource.

Similar to link representation, spatial visualization of numerous links is another problem that occurs when browsing through diverse content [**?**]. Therefore, a semantic to the spatial arrangement of resources is of major importance. Information discovery applications that support serendipitous discovery often have a special way of spatially arranging resources to make it easier to browse through large amounts of

information. For example, many tools use a 'pinboard' layout of resources similar to Pinterest.

Uniform representation. Uniform representation is a method of displaying diverse resources in a common way, with each resource having the same set of components [**?**]. Such a representation assures that each resource has the same set of facts associated with it, and therefore, the user can afford to have expectations about information that can be found when looking for a specific fact. For example, Yelp displays rating, price range, and address for all restaurants, so not only is it easy to find specific information, but the user can have expectations about the content of resources within the application. On the contrary, searching Tumblr for a restaurant will return a chaotic collection of information about the place.

Visual link preview. If an application provides links to resources, a visual preview makes it easier to recognize the relevance of the resource [**?**]. Applications that support fact discovery often use visual link preview, similar to applications that support serendipitous browsing. However, the motivation behind having a link preview for fact finding is to make it possible to identify if the resource is indeed what the user is looking for. For example, searching for an actor in IMDb will return a list of actors and their photographs, so that the user can pick the one they are interested in.

Spatial arrangement. Similar to serendipitous information discovery, spatial arrangement of resources is important [**?**] as a poor semantic to the arrangement can make it difficult to visually navigate to the facts of interest.

5.3 Integration

Similar to serendipitous discovery, if an information discovery application provides access to resources from other Websites, the user should be able to navigate to those sites as they may contain the facts of interest. Integration for fact finding is especially important when it gives an opportunity to display specific information about resources that otherwise would not be accessible. For example, Google Maps displays business ratings as a result of its integration with Google+.

To users with ambiguous information needs, one information portal might not

Table 5.3. *Exploration Mechanisms*

Method of exploration	Questions to be posed during the design or evaluation of tools
Visual exploration	
Visual exploration of a single resource	Does the system visual exploration of a single resource?
Visual exploration of multiple resources	Does the system allow visual exploration of multiple resources?
Spatial exploration	
Spatial exploration of a single resource	Does the system provide means for spatial resource exploration?
Spatial exploration of a multiple resources	Does the system provide means for spatial exploration for multiple resources?

Table 5.4. *Support for Exploration of Multiple Resources*

Method of exploration of multiple resources	Questions to be posed during the design or evaluation of tools
Visual exploration	
Visual preview	Are there visual previews of resources?
Textual preview	Are there textual previews of resources?
Spatial exploration	
List	Are resources presented in a list?
Grid	Are resources presented in a list?
Gallery	Are resources presented in a gallery layout?
Consistent representation	Are resources presented in a consistent way?

provide access to all information of interest. If an information discovery application gives access to resources from various sources, such as other Websites, the user should be able to navigate back to those sources.

Table 5.5. *Support for Exploration of a Single Resource*

Method of exploration of single resource	Questions to be posed during the design or evaluation of tools
Visual exploration	
Visual cues	Are there visual cues?
Textual cues	Are there textual cues?
Spatial exploration	
Spatial semantic	Is there a semantic to the spatial arrangement of resources?
Consistent representation	Are resources presented in a consistent way?

Table 5.6. *Integration*

Integration mechanism	Questions to be posed during the design or evaluation of tools
Integration	
Linking	Is application linked to another application?

Table 5.7. *Support for Integration*

Integration support	Questions to be posed during the design or evaluation of tools
Integration	
Visual integration	Is another application's data visually integrated?

5.4 Curation

Information curation is a common activity within many information discovery applications. By asking questions about application design with regards to information curation as in Tables 5.8 and 5.9 of the conceptual framework, designers can find ways to add value to information and enable information exploitation over time.

Information discovery applications vary from being completely socially curated and populated by users, to those that lack any curation mechanisms. By definition, digital information curation is the notion of managing, preserving, and adding value to collections of information [**?**, **?**]. Thus, the curation category consists of information

management, preservation, information enhancement, and sharing.

Table 5.8. *Curation Mechanisms*

Curation support	Questions to be posed during the design or evaluation of tools
Management	
Collection-based categorization	Does the application support sorting information into collection-like structures, either privately or publicly?
Tag-based categorization	Does the application support tagging, either privately or publicly?
Preservation	
Internal preservation of internal resources	Does the application support mechanism(s) for preserving internal information within the application?
Internal preservation of external resources	Does the application support mechanism(s) for preserving external information within the application?
External preservation of internal resources	Does the application support mechanism(s) for preserving internal information outside of the application?
Augmentation	
Evaluation	Can the resource evaluations be recorded privately or publicly?
Annotation	Can resources be annotated privately or publicly?
Sharing	
Adding resources	Can resources be publicly added to the collection of information within the application from other Web pages?
Internal sharing	Can internal resources be publicly reshared within the application?
External sharing	Can internal resources be publicly reshared outside of the application?

5.4.1 Management

Information management is one of the key elements of information curation [**?**, **?**]. Information categorization mechanisms are prevalent in applications that have a lot of

Table 5.9. *Curation Support and Automation*

Curation support	Questions to be posed during the design or evaluation of tools
Management	
Suggesting collections	Does the application suggest relevant collections?
Suggesting tags	Does the application suggests relevant tags?
Preservation	
History	Does the application automatically preserve found information?
Augmentation	
Automatic augmentation	Does the application automatically annotates resources?
Sharing	
Automatic sharing	Are resources shared automatically?

information that is hard to categorize automatically or can mean something different for each user. In the context of Web information management, the following factors play a major role.

Resource categorization helps establish relationships between various resources [**?**, **?**]. Allowing people to sort information using custom categories can aid rediscovery, discovery in a socially curated space, as well as add more value to resources.

Similar to list-based categorization, tagging aids rediscovery, adds value to resources, and aids discovery, especially in a socially curated space [**?**]. For example, Pinterest supports tag- and list-based categorizations, where lists are represented as 'pinboards'. Tumblr, on the other hand, only supports tag-based categorization. In addition, Pinterest allows for private information categorization.

5.4.2 Information Preservation

Information preservation is a common Web task that is usually performed with the intent of revisiting information [**?**, **?**]. However, in the case of opportunistic Web use, information gathering is sometimes performed with just the goal of collecting information rather than revisiting it in the future [**?**]. Bookmarking is a traditional way

of preserving information and many Web applications provide diverse bookmarking mechanisms.

Internal preservation of internal resources means bookmarking resources to be reaccessed within the same application. Such bookmarking facilitates information curation within the system.

Internal preservation of external resources signifies bookmarking other Web pages within an application.

External preservation means bookmarking resources so that they become available through other bookmarking systems. An application must facilitate integration with other applications in order to enable external preservation [**?**].

On We Heart It, users can preserve *internal information* using *internal collections* and they can add information from *external* Websites. However, there are no integrated means for bookmarking *internal content* using other bookmarking systems.

Bookmark-based revisitation is one of the most common ways of information rediscovery [**?**]. The majority of Web browsers are equipped with bookmarking features. However, some modern Web applications, such as YouTube and Pinterest, provide integrated mechanisms for bookmarking and bookmark-based information rediscovery.

A Web application needs to automatically record browsing history in order to enable history-based rediscovery [**?**]. History-based rediscovery appears to be the least common rediscovery mechanism, however, it can still be found in some Web applications, such as Google Maps.

5.4.3 Augmentation

One of the most important elements of digital curation is augmentation: adding value to information [**?**, **?**]. It is often performed within social bookmarking systems. Many Web applications allow users to add value to the resources they curate.

Evaluation methods can have various forms. They usually take place in socially curated information systems. However, evaluation can also contribute to personal reflection and information preservation. In addition, many applications allow users to evaluate resources by rating them or recording other forms of approval or disapproval. Some sites, such as Wikipedia, do not allow any evaluation.

Annotations are metadata attached to a resource, such as comments and descriptions. Annotations make it easier to search for and interpret information.

5.4.4 Sharing

Sharing information is key to empowering social information curation [?]. Therefore, the main components that facilitate sharing are adding resources, and external and internal information sharing.

Adding resources not only facilitates global Web information curation, but it also scales the information available through the system, providing more opportunities for information discovery. Resources can be created by users themselves, taken from some other sources online, or both. For example, YouTube allows users to upload their own videos, whereas Pinterest permits adding images from other sites in addition to users' personal images.

Sharing resources through different media supports channel-based information discovery within the media channels. Information discovery applications commonly allow for sharing information on popular networking sites outside the application.

Resharing resources within the system supports channel-based information discovery.

5.5 Channelled Curation and Discovery

5.5.0.1 Channel-based Discovery

Channel-based discovery can incorporate two different information seeking tasks, monitoring and awareness. It occurs when information is suggested to users based on the content that they are subscribed to. If users can actively look for updates, then an application affords monitoring [?]. If users can receive notifications about updates, then an application facilitates awareness [?, ?]. Channel-based information discovery is usually enabled at sites that have regularly updated content, such as Pinterest and YouTube.

Subscriptions to updates from a site help users follow the news [?]. In order to support subscription-based discovery, an application must provide a subscription

mechanism. For example, Rotten Tomatoes allows subscriptions to newsletters; however, it does not allow subscriptions to movie critics, as a user-based subscription mechanism would allow.

In some applications, the content is updated and curated by users, and users can subscribe to other users. Similar to site subscriptions, user subscriptions are subscriptions to activity updates from individual users rather than all content updates, and help with networking and following users' activities [**?**]. Such subscriptions help to further filter new content delivered to the user.

Notification mechanisms enable user awareness about new content on the subscribed channel [**?**]. Different applications provide various notification mechanisms including messages within the application, informative emails, and smartphone notifications.

Displaying a news feed within the application further promotes awareness and can serve as a monitoring mechanism. For such.

Similar to displaying a subscription news feed, displaying a content news feed promotes awareness and can serve as a monitoring mechanism.

Information discovery tools can have different implementations depending on the purpose of discovery. Using the information discovery factors in our framework (see Table 2), we described and evaluated currently existing tools. Similarly, the framework can be used for identifying gaps in information discovery support and developing new technologies (see Sec. 5).

Table 5.10. *Chenneling Mechanisms*

Channelling mechanisms	Questions to be posed during the design or evaluation of tools
Subscriptions	
User subscription	Can the user subscribe to activities of other users?
Site subscription	Can the user subscribe to site updates?
Artifact subscription	Can the user subscribe to artifact updates?
Notifications	
User activity overview	Does the application display activities of other users?
Site activity overview	Does the application display site updates?
Artifact activity overview	Does the application display activities of other users?

Table 5.11. *Channeling Support*

Channeling support	Questions to be posed during the design or evaluation of tools
Subscriptions	
Suggesting users	Are users suggested to the user?
Automatic subscription	Can the system subscribe the user automatically?
Suggesting artifacts	Are artifacts suggested to the user?
Notifications	
User activity update notification	Does the application display activities of other users?
Site activity update notification	Does the application notify the user about site updates?
Artifact update notification	Does the application notify the user about updates on an artifact?

Chapter 6 — Framework Evaluation and Application

Evaluation of the framework was performed in three different stages. At first

6.1 Preliminary Evaluation

6.2 Application

6.3 Final Evaluation

To verify that the final version of the framework was effective (see Chapter 5), I applied it to evaluate five more applications. Three of them were from the previous list of applications and two more new applications.

6.3.1 Pinterest

6.3.2 Google Maps

To illustrate how the framework can be applied to evaluate current Web applications and suggest new tooling, we use it to examine one of the cases of the study, Google Maps. By answering the questions from the framework, we get the following description of Google Maps.

Serendipitous discovery. Although there are some possibilities for serendipitous discovery within Google Maps, it is limited by a few factors. Arbitrary navigation is only possible when the user looks at the map itself or browses through the images of nearby places. Any other information discovery must be initiated by search, and thus, the user needs to formulate their information need—the application lacks category-based navigation, so there is nothing that aids users in the formulation of

an information need. Once the application returns search results, the possibility for serendipitous information discovery increases. Some interesting information can be discovered on the business' official Website or integrated Google+ page that the user can navigate to by clicking on 'reviews'. However, the 'reviews' link doesn't have a visual preview to indicate that there are more than just reviews on the linked page. Considering the nature of Google Maps, the semantic of the spatial arrangement of resources is defined by the locations of actual places on the map. More information is presented as a list.

Fact finding. Fact finding is well supported in Google Maps. Since the application provides access to only one type of resource (places), there is no need for category-based navigation. Direct navigation is not always possible, but some places are visible on the map so the user can click on a place and the application will display relevant information. Search-based navigation within Google Maps is usually precise and returns accurate search results for specific places. The application is conveniently integrated with Google+, allowing access to relevant information, such as reviews, images, and hours of operation. Resources are displaced in a uniform fashion making it easy to find information such as addresses and contacts.

Rediscovery. There are a few ways to rediscover information through Google Maps. Google Maps employs a history-based revisitation mechanism, so users can see the last few places they searched for when opening the page. Users can bookmark a place on a list called "My Places" by clicking on the 'star' icon. Lastly, it is easy to rediscover information about a place by simply searching for it. Returned results are usually both accurate and reliable.

Channel-based discovery. Channel-based rediscovery is common among applications with content that is frequently updated. Content provided by Google Maps is fairly stable, and therefore, there are no channel-based discovery mechanisms used by the application.

Management. Google Maps does not allow the creation of custom lists nor does it allow tagging. Users can only bookmark places to the "My Places" list.

Preservation. Personal preservation in Google Maps is possible through adding the place to the "My Places" list as mentioned above—by adding internal content to internal storage. Other types of place preservation are possible through Google+,

however, not within Google Maps.

Augmentation. Users can evaluate and annotate places through Google+. However, aggregated reviews and ratings are visible on Google Maps.

Sharing. It is possible to add a new location to Google Maps. Sharing functionality is limited to the tool providing links and code for embedding.

Evaluating Google Maps using our conceptual framework helped expose some gaps in its design, so we propose directions for future development. From the description above, it can be estimated that Google Maps' curation mechanisms lack some functionality for public and private curation. Improving public curation mechanisms introduces the possibility of channel-based discovery. Furthermore, adding category-guided navigation mechanisms can help with serendipitous discovery. By no means should an application like that be a replacement to Google Maps. However, it could be oriented towards social discovery and curation, as well as opportunistic place exploration, thereby complimenting the Google Maps application.

6.3.3 Wikipedia

6.3.4 The deviantArt

6.3.5 Google Search

Chapter 7 — Implications

7.1 Research and Design Implications

In the previous section, we demonstrated how the framework can be used to reveal missing features in tools. We also showed how the framework can be helpful for designers who wish to improve existing tools or get ideas for new information discovery applications.

Factors and questions of the framework are there to guide the developer, but they do not dictate which features should be in the application. In other words, the framework helps expose gaps, but it is up to designers to decide whether those gaps need to be closed—some gaps cannot be closed because of certain constraints, such as data type and system design.

As with the Google Maps example, designers face certain trade-offs when developing applications with the help of the framework. For example, high precision with navigation mechanisms can potentially eliminate some opportunities for serendipitous discovery.

In the research domain, the framework can serve as a guide for selecting cases for studies and drawing distinctions between different Web-based information seeking applications. Hence, both researchers and developers can benefit from the systematic tool exploration guided by the framework.

Chapter 8 — Limitations

8.1 Limitations

The case study we conducted has a number of limitations: a lack of documentation, literature, and formal descriptions of available features for some applications introduces a threat to construct validity of the study. In addition, information discovery tools and features can be used in manners unintended or unforeseen by designers and developers. Therefore, the use of some features within information discovery applications was recorded based on the researchers' interpretations. To compensate for such limitations, the researchers employed the tools for personal use over an extended period of time to gain a deeper understanding of their use. In addition, the researchers considered some cases with repeating functionality and design to be able to validate or clarify prior findings.

Many Web applications rapidly evolve. Therefore, our tool analysis only applies to tools at the moment of our study.

Only Web applications running in browsers on a desktop computer were considered in this study. Our study can be extended with use of various devices, such as smartphones and tablets, as information discovery patterns and mechanisms may very for different platforms.

Another limitation was the lack of prior research studies on the subject matter. Some researchers have studied information seeking models and high-level Web tasks, but there is a lack of literature on how to enable and support different Web tasks. This opens up opportunities for future research to analyze methods of developing and building frameworks for facilitating and evaluating tools that support other Web tasks, such as communication, transactions, and goal realization.

Chapter 9 — Future Work

9.1 Future Work and Conclusions

In our study, we analyzed information curation and seeking tasks and developed a conceptual framework of factors and questions that are important when building and evaluating Web information discovery tools. We then evaluated and iteratively refined the framework by analyzing 20 different information discovery applications and provided concrete examples of tool support addressing various concepts of our framework.

One of the possible future research objectives would be to apply the framework to identify a gap in available information discovery tools, and then further use the framework to design an application that would close that gap. Another potential research question would be to expand our investigation to include the factors that influence the need for one information discovery type over another.

Our framework opens up opportunities for structured information discovery tool evaluation and design. As more tools are being developed within the social space of information discovery and curation, understanding how these tasks can be supported promises advancements in how Web applications are designed.

Chapter 10 — Conclusion

VITA

Surname: Mouse ***Given Names:*** Mickey

Place of Birth: California

Educational Institutions Attended

University of Victoria ... to ...

Degrees Awarded

B.Sc. University of Somewhere ...

Honors and Awards

award date

PARTIAL COPYRIGHT LICENSE

Title of Thesis: TOWARDS UNDERSTANDING DIGITAL INFORMATION DISCOVERY.

Author: _____

ELENA VOYLOSHNIKOVA

30 Dec 2002