TOWARDS UNDERSTANDING DIGITAL INFORMATION DISCOVERY AND CURATION

by

ELENA VOYLOSHNIKOVA
B.Sc., University of Victoria, 2012

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. Margaret-Anne Storey, Supervisor, Dept. of Computer Science

---

Dr. Melanie Tory, 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 curation of information, especially concerning the motivations behind them and how different approaches can be compared.

In this thesis, I present a study of information discovery tools and how they relate to the nature of information seeking. I propose a conceptual framework that deals with Web application design elements that support different aspects of information discovery and curation. This framework can be used when designing, evaluating or updating Web applications.

Examiners:

Dr. Margaret-Anne Storey, Supervisor, Dept. of Computer Science

---

Dr. Melanie Tory, Member, Dept. of Computer Science

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

Table of Contents

List of Figures

List of Tables

Acknowledgement

Dedication

Chapter 1 — Introduction

Web technologies help people 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.

The term "information discovery" has been used by many researchers to define or explain various information behaviour paradigms, such as information exploration [1] and serendipitous information seeking [2]. However, the definition of information discovery itself is difficult to articulate.

Lynch describes resource discovery as a complex collection of activities ranging from locating a well-specified information to iterative research activities, that can involve the identification of potentially relevant resources, organization and ranking of resources, and resource exploration [3]. Proper and Bruza apply the term "information discovery" in the context of the identification and retrieval of relevant information from electronic sources [4].

In the field of cognitive psychology, Jerome S. Bruner [5] defines information discovery as "all forms of obtaining knowledge for oneself by the use of one's own mind." I build on Bruners's definition to underline the importance of the cognitive processes that govern information discovery. Therefore, I consider *information discovery* as a process of obtaining knowledge from digital sources that can involve complex mental tasks and information behavior.

Information behavior refers to the totality of ways in which humans interact with information [6]. It can enable and support information discovery when targeted at information maintenance and augmentation. This type of information behavior is also known as *digital curation*.

Similar to the term "information discovery", the term "digital curation" is per-

ceived differently across disciplines and among researchers. In this thesis, I use the definition proposed by Giaretta [7] and adopted by the Digital Curation Centre[1] which states that digital curation is a process of maintaining and adding value to an existing body of information to improve its future use and retrieval.

Information discovery can take on many forms. Web users might be hoping to find particular pieces of information, such as show times and phone numbers, to satisfy specific information needs [4]. Alternatively, they might be lacking well-articulated information needs, so they engage in opportunistic browsing [8]. Sometimes people discover information online without even looking for it [9]. The nature of information discovery can vary, and therefore, it requires elaborate tool support. The functionality required for information discovery and curation can also be distributed among multiple applications, which often leads to tools that provide integrated solutions. With people having such diverse information needs and methods of looking for information, designing for information discovery is a challenging task [10, 11].

My research goal is to gain an understanding of how existing tools support digital information discovery and curation addressing the problem of designing Web applications for information discovery. While several researchers propose frameworks targeted at designing information discovery systems [4, 12], the importance of information curation in the realm of information discovery has been largely overlooked despite the rapidly increasing popularity of socially-curated information spaces. Moreover, much of the existing work that focuses on how people look for and discover information online [9, 13, 14, 15, 8, 16, 17] fails to examine concrete features of existing Web-based information discovery applications that empower real-world users. 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' interactions, I 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, (3) iteratively evaluating the framework by using it to study and

[1]The Digital Curation Centre is a UK-based organization established to support expertise and practice in digital curation and preservation across communities of practice.

describe current Web applications as well as to design a new application, which in turn helped refine the framework of factors and questions, and (4) relating the framework to user information discovery and curation goals that drive the underlying usage of many Web-based applications.

This thesis is organized as follows. My methodology and the process of building and refining a conceptual framework is documented in Chapter 2. Chapter 3 highlights some of the studies and technologies related to information discovery and curation tasks. Chapter 4 describes preliminary attempts at building the conceptual framework and outlines their shortcomings. Chapter 5 outlines the conceptual framework and questions that enable digital information discovery and support curation, including specific examples from real-world Web applications. In Chapter 6, I illustrate the framework validation process, demonstrate how the framework can be used to reveal missing features in tools, and propose new directions for development with relation to user goals. I then showcase how the framework can be used for Web application design in Chapter 7. Chapter 8 summarizes the implications for research and practice. This is followed by future work and conclusions in Chapter 9.

Chapter 2 — Methodology

The methodology used for the study presented in this thesis consisted of five major steps. To gain a deeper understanding of the problem of information discovery and curation, (1) I conducted a systematic literature review. Based on the literature review, (2) I derived a preliminary set of information discovery and curation design factors and related them within a framework. (3) The framework was then applied for the evaluation of 20 different information discovery applications and iteratively refined after every evaluation. (4) The resulting framework was used to develop a novel place photo discovery application,revealing unforeseen gaps that were consequently addressed. Lastly, (5) the framework was applied to a reevaluation of some of the previously evaluated tools with the purpose of validating its effectiveness. A summary of the methodology is presented in Figure 2.1.

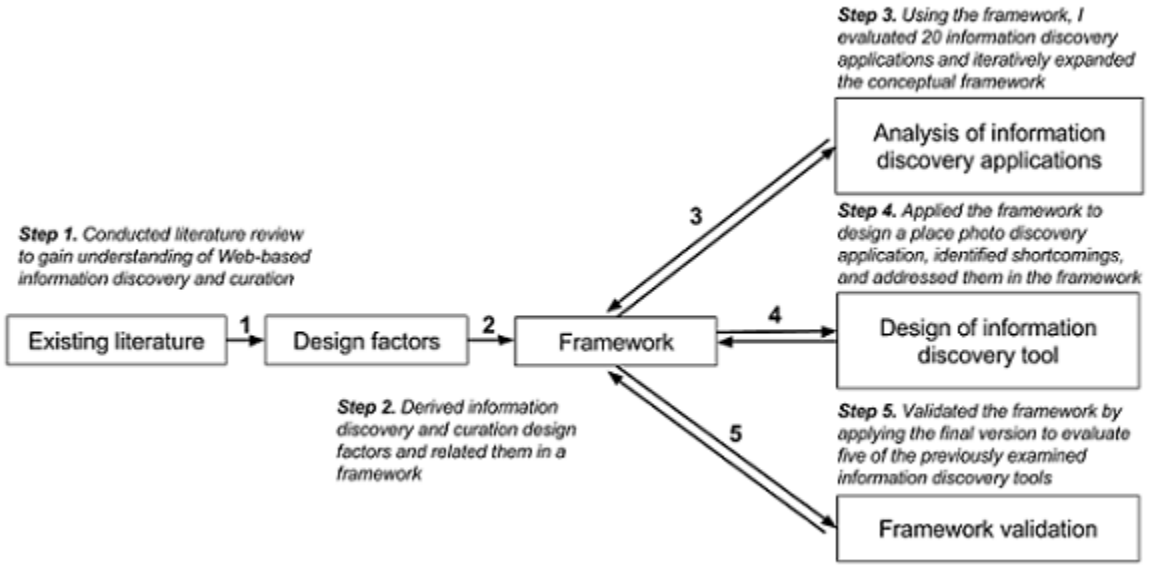

Figure 2.1. *Methodology Overview*

2.1 Research Questions

This study was designed to address the problem of designing Web applications for information discovery and 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?

RQ3: How can one conduct formative and summative evaluation of Web applications for information discovery and curation?

To address RQ1 and RQ2, I conducted a systematic literature review (see Section 2.2) and a case study of 20 information disocvery tools (see Section 2.3). I then developed and validated a conceptual framework to address RQ3 (see Sections 2.3, 2.4, 2.5).

2.2 Literature Review

The development of the framework began with an extensive literature review. A diverse set of topics contributed to forming an understanding of information discovery and curation, including information behaviour and information seeking models, high-level Web tasks and modes of Web use, exploration-based models of discovery, and methods of personal and social curation. From this review, the preliminary design factors for the framework were derived. Key findings in the current literature are presented in Chapter 3.

2.3 Building and Refining the Conceptual Framework

Through a careful analysis of 20 information discovery applications (see Table 2.3), the framework was iteratively expanded by adding new concepts and establishing relations between those concepts. The framework was refined as I explored the literature and available tools, and for presentation purposes in this thesis, I present only two versions of the framework. The preliminary framework was a result of this tool analysis and

depicted in Chapter 4. The final version of the framework (see Chapter 5) was a result of developing an information discovery application based on the preliminary work.

For my case study, I selected some of the most used information discovery applications 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.

I used Yin's strategies for designing a case study [18] for guidance. The motivation behind choosing a case study over other methods of qualitative research was based on my choice of research questions, 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 [18], a case study would be an optimal research strategy given the above characteristics.

My study consisted of 20 cases, whereby each case is a Web application that focuses on the support of information discovery. I examined the overall purpose of each application, its description as defined within the application, as well as 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, I checked if it was indeed intended to be used for information preservation.

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

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

A list of the tools that were used in this study are presented in Table 2.3. Other tools were considered throughout the study, however, only the 20 applications presented underwent systematic examination.

Table 2.1. *Web-based Information Discovery and Curation Tools as of May 15, 2014*

Application	**Address**	**Description**
Pinterest	www.pinterest.com	Visual discovery tool
Delicious	delicious.com	Social bookmarking service
Tumblr	www.tumblr.com	Microblogging platform
StumbleUpon	www.stumbleupon.com	Web page discovery tool
Wikipedia	en.wikipedia.org	Free content Internet encyclopedia
Google Maps	www.google.ca/maps	Web mapping service
Rotten Tomatoes	www.rottentomatoes.com	Movie and TV database
500px	500px.com	Photography site
BucketList	bucketlist.org	Goal tracking and discovery service
We Heart It	weheartit.com	Visual discovery tool
Scoop.it!	www.scoop.it	Online publishing platform
Google Images	images.google.com	Image discovery service
Vimeo	vimeo.com	Video sharing Website
LifeHacker	lifehacker.com	Daily blog
YouTube	www.youtube.com	Video hosting platform
Yelp	www.yelp.ca	Business review site
IMDb	www.imdb.com	Movie database
Trip Adviser	www.tripadvisor.ca	Travel site
Urban Spoon	www.urbanspoon.com	Online bar and restaurant guide
Thesaurus	thesaurus.com	Online thesaurus

2.4 Applying the Framework to the Design of an Information Discovery and Curation Application

In order to analyze the framework's capabilities when designing for information discovery and curation, I used the framework as a guide for developing a place photo discovery application. The motivation for choosing a place photo discovery application was based on the gaps that were exposed during analysis of some of the applications, such as Google Maps and Pinterest. Applying the framework to designing an application has triggered more changes within the framework, its further extension and refinement. The resulting application is discussed in Chapter 7.

2.5 Framework Validation

In order to further validate the framework, it was applied to the reevaluation of five of the previously examined tools (see Chapter 7). For each tool, I identified gaps and proposed directions for future development.

2.6 Limitations

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

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

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

Another limitation was the lack of prior research 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 3 — Web-based Information Discovery and Curation

Given the complexity of Web-based information discovery and curation tasks, a variety of topics was examined to gain an understanding of how current Web tools support these tasks, including known characteristics of information-related Web usage, currently existing information behavior models, and other aspects of information discovery and curation. This chapter outlines the key background literature that contributed to the development of the conceptual framework and helped answer the research questions.

3.1 Information Behavior

As defined previously, information behavior refers to the totality of ways in which humans behave in relation to information [6]. A number of models and frameworks were proposed to represent human information behaviour in its entirety or to represent some of its components, such as information seeking and searching, information retrieval, information discovery, and information curation.

3.1.1 Information Behavior Model

One of the early information behavior models was proposed by Wilson [19] in 1981. According to the original model, 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.

These underlying ideas remained in the revision of Wilson's model [20]. In the new model however, Wilson defined possible barriers (psychological, environments,

demographic, etc.) that can impede information seeking. Additionally, the model recognizes that information seeking behaviour can take on many forms and is not limited to active search. Saracevic [21] and Ingwersen [22] derived resembling models that focus on human behaviour when interacting with information retrieval systems.

3.1.2 Information Seeking Models

Information seeking refers to "the purposive seeking for information as a consequence of a need to satisfy some goal [6]." A number of researchers have tried identifying what modes of seeking information seeking behaviour may entail.

According to Kellar et al. [15], information seeking is composed of browsing, fact finding, and information gathering. Although the authors categorized information gathering as part of information seeking, it appears to be more closely related to digital curation [23, 24].

Bates [9, 25] 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.

Ellis et al. [14, 26, 27] proposed a model of information seeking characterized by six different patterns: starting, chaining, browsing, extracting, monitoring, and differentiating. Ellis' model complemented Kuhlthau's work, in which the researcher correlated stages of information seeking with feelings, thoughts, actions, as well as anticipated information tasks [28].

Finally, Wilson proposed a "problem solving model" of information seeking behavior [29]. 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.

3.1.3 Information Exploration

Information exploration, or exploratory search, does not have a single definition in realm of information behavior. Waterworth highlights that exploration is a "broad" activity and identifies browsing as an example of exploration [1]. According to Marchionini [11], exploratory search involves learning(knowledge acquisition, comparison, comprehension, etc.) and investigating(analysis, synthesis, evaluation, discovery, etc.) Similar to Janiszewski [30], my focus is on the visual aspects of information exploration, specifically visual and spatial data representations.

3.1.4 Information Foraging

Information foraging theory is another approach towards understanding how people adapt their strategies of interacting with technology when seeking, gathering, or consuming information depending on the environment [31]. The theory resonates with explanations of human behavior in the context of food foraging.

The underlying assumption of the information foraging theory is that people, similarly to when they forage for food, adopt their foraging strategies to the environment in order to gain the maximum amount of valuable information. The theory states that "natural information systems evolve towards stable states that maximize gains of valuable information per unit cost."

The theory introduces three key concepts to formulate an understanding of information foraging: information scent, information diet, and information patch. An *information scent* refers to proximal cues (often visual or linguistic) that people use to identify the value of information. An *information diet* deals with user preferences when it comes to information. At last, *information patches* are clusters of information that a an information system presents before the user. The theory with these concepts lay foundations for existing information foraging models [32, 33] as well as social information foraging models [34, 35].

3.1.5 Information Discovery

Kerne and Smith proposed an information discovery framework [12] that connects human cognitive processes or states to those of an information system. The frame-

work represents a continuum of information flowing through different system and cognitive states as a result of an iterative reformulation process. The framework consists of five mental states: formulating a problem, evaluating results, updating and forming mental models, running mental models, and discovering solutions. Each mental state have a corresponding interaction with the system. For example, browsing resources (human-system interaction) facilitates evaluation or immediate results (cognitive state). The framework helps understanding how to support the user's cognitive processes and provide affordances that facilitate information discovery.

3.1.6 Digital Curation

In 2002, Bates extended her research on the topic of information behaviour with the notion of *information farming*, which involves people collecting and organizing information for future use and revisitation [25]. More commonly, information farming is referred to as digital curation.

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 [24].

3.2 Web Tasks and Modes of Web Use

Outside of the realm of cognitive models and frameworks for information behavior, there exists a body of research that examines information discovery, curation, and other Web information behaviours in terms of Web use and corresponding tasks, methods, and modes.

Kellar et al. [15] separated Web tasks into five categories: transactions, browsing, fact finding, information gathering, and other uncategorized tasks. In their later work, Kellar et al. [36] added communication and maintenance as additional Web tasks. Similarly to Kellar et al., Sellen et al. [17] identified six tasks that are commonly performed by Web users: browsing, finding, housekeeping, information gathering, communicating, and transacting. Using different terms, Kellar et al. and

Sellen et al. identified highly comparable tasks, such as fact finding and finding, housekeeping and maintenance, etc.

Building on Ellis' model of information seeking [14, 26, 27], Choo et al. [13] derived anticipated Web tasks that correspond to the information seeking patterns in the model. According to the authors, when users *identify* sources of interest, they usually identify which Websites can point to that information of interest. *Chaining* corresponds to users navigating 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.

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 [4]. 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. [16] make a distinction between methods of Web use and purposes. 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 purposes. Therefore, each purpose may require more than one task-supporting mechanism.

Morrison et al. also draw 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 [11]. 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.

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. [8] 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 television, 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 unarticulated or 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 [8]. 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.

3.3 Collaborative Information Discovery and Curation

By surveying 204 Web users, Morris found that people often desire to or do collaborate on information seeking tasks [37]. 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 [38].

Today, there are a multitude of tools that support different aspects of information discovery 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 discovery and curation.

Chapter 4 — A Preliminary Framework for Information Discovery and Curation

A preliminary framework for information discovery and curation (see Tables 4.1 and 4.2) was designed in hopes of merging the gap between existing Web tools and high-level information behaviour models [39]. It was constructed by identifying important elements in current Web applications (see Table 2.3) and relating them among themselves with the help of background research (see Chapter 3). In this chapter, I describe the preliminary version of the framework to illustrate its evolution and outline some of the challenges with its construction. The final framework is discussed in Chapter 5.

4.1 Preliminary Framework Composition

The two main parts of the framework (discovery and curation) encapsulate other categories of design factors for Web applications. Serendipitous discovery, fact discovery, rediscovery, and channel-based discovery are the main types of information discovery. Curation consists of common curation tasks: information management, preservation, augmentation, and sharing. This section provides brief summaries of each part of the framework.

Serendipitous discovery refers to information discovery resulting from serendipitous browsing. Such discovery is characterized by underdefined, absent, or hidden information needs, and it usually involves browsing through diverse resources with varying content types [15, 36]. Here, a 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. Serendipitous discovery can be supported using arbitrary, search-based, and category-based navigation mechanisms, integration, visual link preview, and spatial arrangement of

resources.

Fact discovery is defined as 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 [15, 8]. Fact discovery can be supported using category-based and search-based navigation mechanisms, integration, uniform representation, visual link preview, and spatial arrangement of resources, as with serendipitous discovery.

Rediscovery refers to information discovery resulting from revisiting previously discovered resources [40]. Rediscovery can be enabled using search, history, or bookmarking.

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 they are subscribed to. If users can actively look for updates, then an application affords monitoring [16]. If users can receive notifications about updates, then an application facilitates awareness [9, 25]. Channel-based information discovery is usually enabled on sites that have regularly updated content, such as Pinterest and YouTube. Channel-based discovery can be supported using site, user, and news feed subscriptions, notifications, and by displaying the news feed.

Management of information can be performed through organizing information into lists (or collections) or tagging publicly or privately.

To *preserve* information, people use diverse bookmarking mechanisms. Information can be preserved within the application where it was found or in a different application. As another form of preservation, internal preservation of external resources, new information can be added to the Web application in question.

Information *augmentation* is the notion of adding value to existing digital assets [23]. Augmentation can be accomplished through activities such as rating, commenting, describing, and upvoting. In other words, by augmenting or evaluating information.

Information *sharing* is commonly performed within information discovery and curation applications. It is a way to communicate information to other individuals or groups of people though various Web channels. Information communication is an important aspect of Wilson's information behaviour model [19]. In information

discovery and curation tools, information sharing can be enabled by providing mechanisms for publicly adding resources, resharing resources within the same application or outside of it, in other applications.

The preliminary framework aims to help with the evaluation and design of currently existing tools, however, it has certain shortcomings, which are outlined in the next section.

4.2 Limitations of the Preliminary Framework

Although the preliminary framework can be applied to evaluate some aspects of information discovery and curation for Web applications, some of its characteristics make it difficult to use.

In the preliminary framework, there is a clear distinction between the types of curation and discovery subcategories. Discovery subcategories represent **types** of information discovery (serendipitous discovery, fact discovery, etc), whereas curation subcategories represent curation **tasks**. For comparison, information discovery tasks can include *navigating* to a target resource or *exploring* a resource in order to extract necessary information, whereas curation tasks would be to *preserve* a resource or to *manage* a collection of resources.

Furthermore, the types of information discovery in the framework are mutually independent. Serendipitous and fact discoveries are defined using specificity of the user's information need. Defined information needs result in fact discovery, and undefined information needs result in serendipitous discovery. However, rediscovery and channel-based discovery are defined mostly by how the information in question is related to the user, whether or not it has been discovered before, or if the user chose to receive it. Therefore, there can be serendipitous rediscovery, channel-based fact discovery, etc.

Another aspect of information discovery and curation support that the framework fails to thoroughly address is the ways in which the system provides cognitive support to the user and reduces the amount of effort the user needs to put in to perform a task. Examples of such support are automatic sharing of curated content and suggestion of search terms to the user. The framework has to be extended beyond just the factors that **enable** information discovery and curation and showcase strategies that can help **improve** these enabling mechanisms.

In the next chapter, I present the final version of the framework that addresses some of the major drawbacks of the preliminary framework.

Table 4.1. *Preliminary Framework - Discovery*

Design factors	**Questions to be posed during the design or evaluation of Web-based information discovery tools**
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 tools**
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, there is a lack of literature on how to enable and support them when building Web applications. I reduce this gap by presenting a framework of design factors facilitating digital information discovery and curation.

In my framework, I build on existing models and frameworks of information discovery and curation and analysis of existing Web tools to derive corresponding design factors for Web design. The first part of the framework deals with the *motives* behind information discovery and curation (see Section 5.1). These motives often define use cases for Web application design and help set initial assumptions about required functionality.

The second part of the framework defines *actions* that compose discovery and curation activities and design factors that enable those actions (see Section 5.2). Some examples of actions include managing and preserving information. In order to enable these actions, a Web-based application must provide corresponding mechanisms, such as bookmarking and tagging capabilities.

Actions can be further decomposed into *operations* performed using mechanisms that enable these actions. For example, information preservation (action) can be enabled using bookmaking feature (enabling mechanism), so that users can bookmark information using this feature (operation). Similar terminology is used in Activity Theory [41]. The remaining part of the framework deals with improving operations that are involved in information discovery and curation (see Section 5.3).

Figure 5.1 illustrates how different components of the framework are connected and gives high-level overview of the framework.

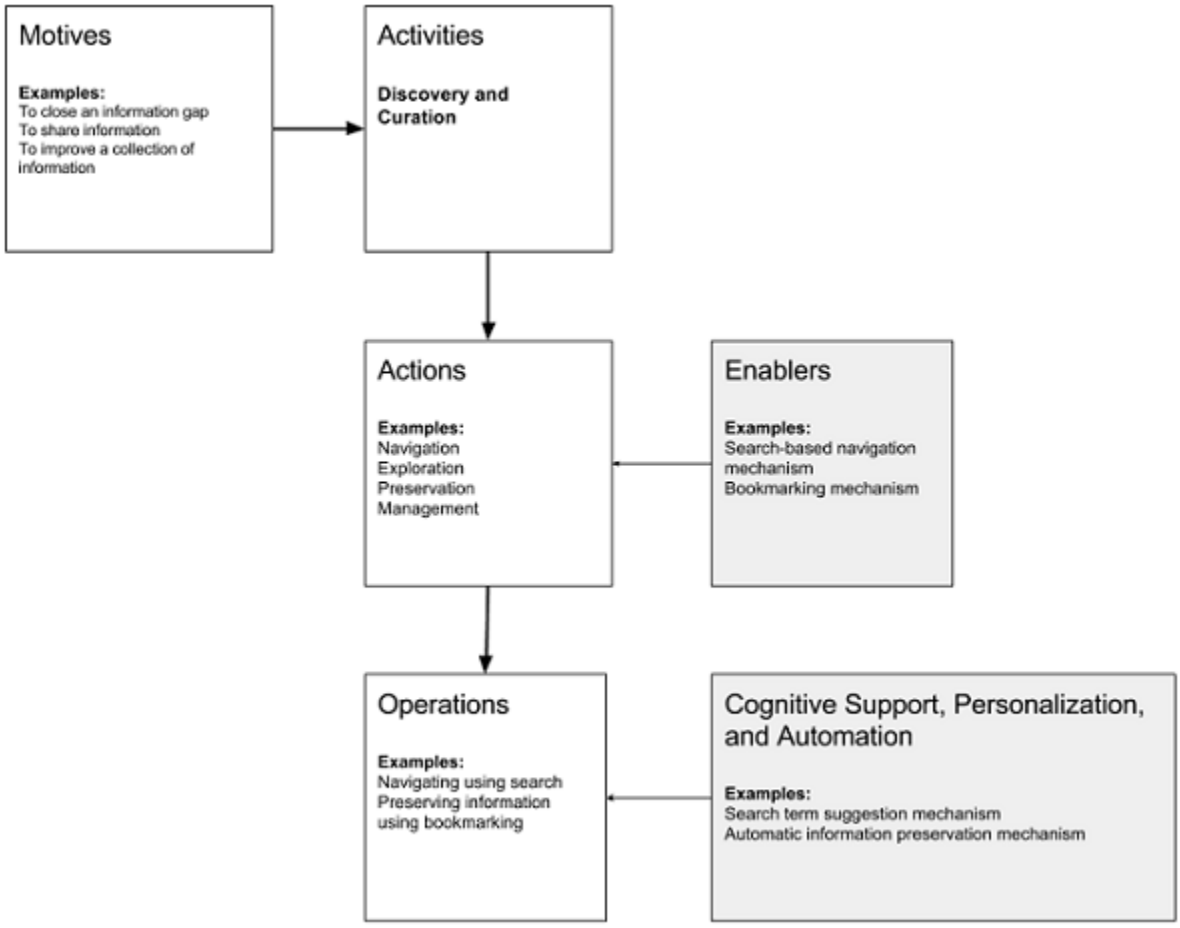

Figure 5.1. *Framework Composition*

5.1 Motives Behind Information Discovery and Curation

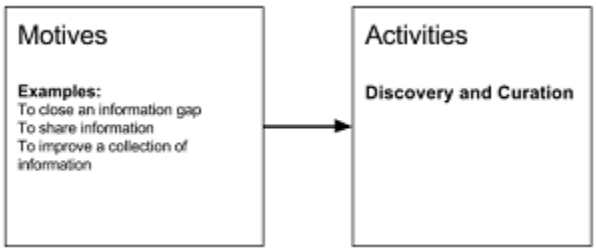

Figure 5.2. *Section Overview: Motives Behind Information Discovery and Curation*

Understanding the motives behind user activities can help form a conceptual model of a needed Web application and its features. There exists a wide variety of user motives behind information discovery and curation, and certain aspects of these motives can significantly impact the design of an application. The following generalizations of motives and their properties can help define conceptual models and identify primary use cases of Web tools for information discovery and curation. Figure 5.2 illustrates the part of the framework discussed in this section.

5.1.1 Closing a Knowledge Gap

The primary motive for information discovery is usually to close a knowledge gap that occurs when the user tries to accomplish a task and lacks information to do so. This motive can take up various forms. Its form commonly depends on the nature of an information need and other conditions surrounding the given motive.

Depending on the context in which it arose, an information need can have various degrees of specificity. For example, if the motive is to find inspiration for a project, an information need is only vaguely defined. However, if the motive is to find a phone number of a specific business, an information need is well-defined. In some cases, the information need may be hidden, and the user might not be aware of the existing knowledge gap. The specificity of an information need determines important properties of information discovery mechanisms, such as whether users can benefit more

from mechanisms that allow to specify an information need, help form an information need, or allow to randomly retrieve information. This property has to be taken into consideration when evaluating or designing a Web application.

The nature of an information need predetermines whether discovery is respectively serendipitous or oriented towards fact finding. Therefore, depending on the user needs, an application can be designed to increase serendipity and opportunistic discovery or to improve purposeful fact finding. On the one hand, displaying featured content can improve serendipitous discovery because of its unexpected nature and novelty. On the other hand, using context, such as location and date, to tailor search results to the user can improve fact finding.

The motive behind information rediscovery involves finding previously discovered information and closing the previously closed knowledge gap in case if the information was forgotten. It usually results in the user looking for previously found resources and Web pages. In fact, Web-page revisitation is one of the most commonly performed web-browsing activities [42, 43]. The percentage of revisited web pages involved in web browsing can range from 58% [40] to 81% [44]. Some of the reasons for revisiting pages include shopping, communication, entertainment, education, activity planning, and hobby-related information retrieval [42]. Some Web pages and resources can be rediscovered using navigation while others need to be previously preserved (bookmarked) in order to allow rediscovery. Rediscovery is one of the many ways in which information discovery and curation interweave.

Another type of motive for information discovery relates to the two qualities of the Web defined by Lindley et al. - temporality and persistence [8]. *Persistence* refers to the quality of the Web that allows people to habitually revisit Web pages and continue on-going Web projects. *Temporality* refers to the quality of the Web that allows the content of websites to be continuously updated to provide users with new information. Persistence alone usually facilitates information rediscovery. However, if persisitence is combined with temporality, they can facilitate discovery of new information within the same application or channel. I refer to this type of discovery as *channel-based discovery*. Some of the common motives for channel-based discovery include orienting (or monitoring for updates) and opportunistic information discovery [8].

5.1.2 Supporting Future Use and Reaccess

The main motive behind information curation is to make it possible to retrieve information and to consequently use it. In order to facilitate easy information retrieval, many Web applications employ various forms of bookmarking systems. Traditionally, bookmarks are manually organized by users into folders. However, this method of organization has been found inefficient because folders with bookmarks become easily cluttered [45]. Therefore, in order to efficiently support information rediscovery, Web tools need to provide mechanisms for information preservation along with information management.

5.1.3 Improving Collections

Reportedly, people gather information to improve existing collections [8]. Although some deeper motives may include self reflection or the possibility of future use, collecting information is a motive in itself. In general, information gathering may be stretched over a period of time [15], resulting in repeating page visitation. Although information gathering composes only 13.4% of Web usage, it highly contributes to various goal-supporting activities, such as decision making and planning [15].

5.1.4 Facilitating Communication

As part of his information behavior model, Wilson identified communication of information as an outcome of information seeking. Communication can also be thought of as a motive for information discovery and curation. To support communication of information, Web tools have to provide mechanisms that allow various users to share information among themselves.

In recent years, social bookmarking, a way to preserve and share information within various communities, has gained popularity as an effective way of communicating with other users [46]. One of the first visions of social bookmarking was associated with Web blogging. Oravec [47] believes that web blogs help users annotate or bookmark important information and build a “map” of the Internet. Evolution of social bookmarking led to advanced bookmarking technologies and provided a means for collaborative information discovery and curation.

5.1.5 Summary

Even though it was not feasible to list all of the possible motives for information discovery and curation, in this section I outlined some of the important motives and their aspects that can aid in developing use cases and formalizing conceptual models for Web applications. These motives also make it easier to showcase how mechanisms for discovery and curation presented in the next section complement each other.

5.2 Discovery and Curation Activities, Actions, and Their Enablers

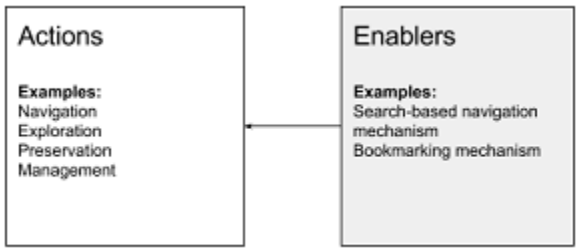

Figure 5.3. *Section Overview: Discovery and Curation Activities, Actions, and Their Enablers*

The next part of the framework (see Figure 5.3) deals with actions that compose information discovery and curation and their enablers. A more detailed overview is depicted in Figure 5.4, where the two main activities (discovery and curation) are decomposed into actions, and each of the actions is supported by a group of features or mechanisms that enable given aspect of discovery or curation in a Web application. The following subsections describe each of the feature groups and outline the corresponding questions a designer could ask to improve application design and evaluation.

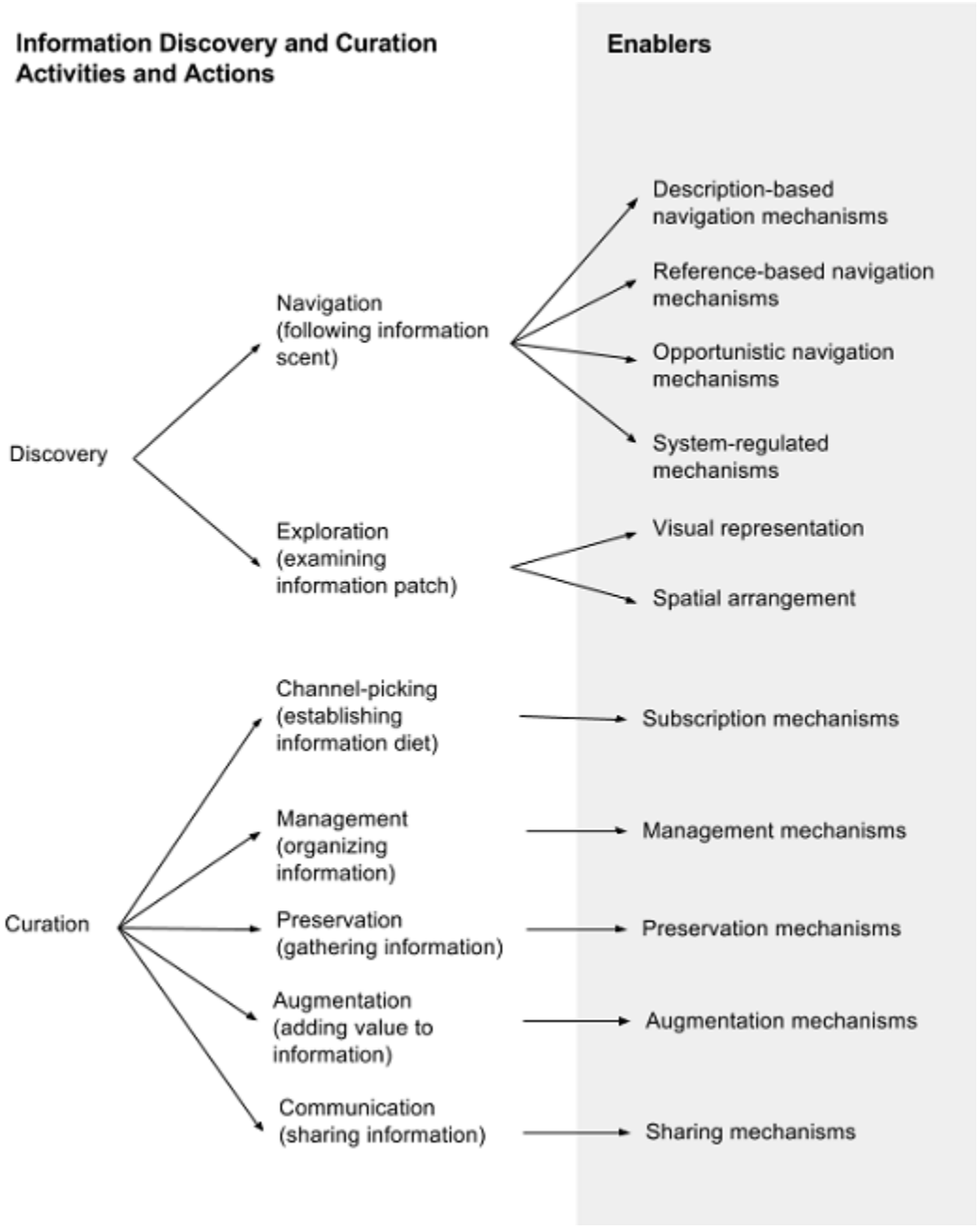

Figure 5.4. *Information Discovery and Curation Activities, Actions, and Corresponding Enablers*

5.2.1 Navigation in Discovery: Following Information Scent

In order to discover information, a user needs to have a way of navigating to it. Navigation in information discovery can be thought of as following an information scent. In general, information scent models deal with how the users identify value, cost, or access path of information sources based on proximal cues, such as links, icons, categories, etc. [31]. Common methods of navigation that facilitate information discovery include descriptional, referential, opportunistic, and system-regulated (see Table 5.1).

5.2.1.1 Descriptional Navigation

A navigation is descriptional when the user has a means of describing their information need. Most commonly it is implemented as *search-based navigation* since it allows users to enter the search query and describe what they want to find. Some of the modern descriptional navigation systems can be activated using voice.

Almost every present-day Web application has the search feature implemented, with rare exceptions of applications that utilize other methods of navigation, such as StumbleUpon and certain shopping websites. Some Web applications are *integrated* with others enabling users to search multiple websites at once.

There exist numerous ways in which descriptional navigation supports information discovery. Search-based navigation often serves as an entry point for information seeking [48]. When the motive behind information discovery has a well-articulated information need, then the user can express their information need by entering a search query.

Descriptional navigation can also help to rediscover information. However, it is not always a reliable way of refinding information [43]. In information portals that provide access to fairly ambiguous information and that have regularly updated information flow, the search feature is usually designed around retrieving information related to some general topic. In order to make search-based navigation a reliable way to rediscover information, it must return consistent results.

Table 5.1. *Navigation Mechanisms*

Navigation mechanisms	Questions to be posed during the design or evaluation of discovery and curation tools
Descriptional	
Search-based navigation	Is it possible to navigate within the application using a search mechanism?
Integrated search	Is it possible to retrieve information from other applications using a search mechanism?
Referential	
Categories	Is it possible to navigate using categories?
Facets	Is it possible to navigate using facets?
Filters	Is it possible to navigate using filters?
Tags	Is it possible to navigate using filters?
Search by item or resource	Is it possible to search by item or resource?
Integrated reference	Is it possible to retrieve information from other applications using any of the referential mechanisms?
Opportunistic	
Opportunistic navigation	Is it possible to opportunistically navigate through information within the application?
Integrated opportunistic navigation	Is it possible to opportunistically retrieve information from other application?
System-regulated	
Static direct display	Is it possible to view static information directly without active search?
Integrated static display	Is it possible to view static information from other applications without active search?
Featured content	Is it possible to view featured content?
Integrated featured content	Is it possible to view featured content from other applications?
News feed	Is it possible to view news feed?
Integrated news feed	Is it possible to view news feed from other applications?

5.2.1.2 Referential Navigation

A navigation is referential when the user finds a reference, such as link or icon, to the term that they are looking for. This reference represents an information scent. The underlying assumption of this method of navigation is that the user can recognize the needed information or a reference to it as they see it [1].

Referential navigation mechanisms can take many different forms. Some common types are **categories**, **facets**, **filters**, and **tags**. In some applications, users can search by a given **resource**. For example, YouTube provides a playlist with music related to the currently playing song. Information scent representatives may also reference sources outside of the given system. This enables another type of **integration** of Web applications.

Referential navigation can help the user identify their information needs by suggesting terms, topics, or categories to use, and therefore, directing the user to relevant resources [48]. It can also help narrow the results to a specific type of resource so that further discovery is bounded by that type. For example, TripAdvisor allows the user to choose among hotels, flights, vacation rentals, restaurants, and destinations, so the application helps to narrow search results.

5.2.1.3 Opportunistic Navigation

Opportunistic navigation is a method of navigating 'randomly' through resources and Web pages. As expected, this type of navigation is not authentically random; therefore, I apply the term 'opportunistic' to describe this type of navigation. However to the user, this navigation type appears to be randomized, and it promotes serendipitous discovery. This method is especially useful when the information need is fully undefined.

Many applications, such as StumbleUpon and Wikipedia, support **opportunistic navigation** to allow for opportunistic jumping from one resource to another. StumbleUpon makes it possible to explore the Web in general - other websites and Web applications, allowing for **integrated** navigation, whereas Wikipedia provides opportunistic access to its own articles.

5.2.1.4 System-regulated Navigation

Often, Web applications display or update information without the user's active participation. This information can be a **news feed**, **featured** deals or articles, **static information**, or other types of content. In my thesis, I refer to this type of navigation as system-regulated because it occurs when the application brings the content to the user instead of the user applying any effort to find it. It differs from opportunistic navigation because the the user cannot choose when to observe new information; instead, all updates are regulated by the application.

One example of an application that utilizes system-regulated navigation is Yelp. This tool displays featured restaurants as well as its user's recent activities as soon as the user enters the site. As any other navigation method, system regulated navigation can ensure cross-application **integration** by displaying content from other Web applications.

5.2.2 Exploration in Discovery: Examining Information Patches

Exploration of resources is another action that facilitates information discovery. Visual and spatial cues, which help representing single or multiple resources, enable this action by allowing the users to examine information patches conveniently(see Table 5.2).

Abrams et al. [45] 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 in a Web application can be a challenging task. **Visual** and **textual previews** make it easier to evaluate the relevance of resources by providing the user with more information scent. Many social bookmarking systems, such as Scoop.it! and Pinterest, support visual previews of bookmarked pages. Delicious is a social bookmarking application that lacks this type of link representation support, so that it is harder to determine if the link will lead to a relevant resource.

Visual and **textual** information cues and representations are also important for a single resource exploration. Not only they help navigating within the resource or the Web page, but they can also contribute to learning experience. For example, in case if the user would like to know what something looks like, they can learn it from

Table 5.2. *Visual and Spatial Exploration Mechanisms*

Exploration mechanisms	Questions to be posed during the design or evaluation of discovery and curation tools
Visual and textual cues of multiple resources	
Visual preview	Are there visual previews of resources to help identify resources of value?
Textual preview	Are there textual previews of resources to help identify resources of value?
Visual and textual cues of a single resource	
Visual cues	Are there visual cues to help identify the value of information within a resource?
Textual cues	Are there textual cues to help identify the value of information within a resource?
Spatial proximal cues of multiple resources	
List	Are resources presented in a list?
Grid	Are resources presented in a grid?
Gallery	Are resources presented in a gallery layout?
Spatial semantic	Is there a semantic to the spatial arrangement of multiple resources?
Consistency	Are resources presented in a consistent way?
Spatial proximal cues of a single resource	
Spatial semantic	Is there a semantic to the spatial arrangement of information within a resource?
Consistency	Are same types of resources presented in a consistent way?

the representation in question.

Similar to link representation, spatial visualization of numerous links is another

problem that occurs when browsing through diverse content [45]. Therefore, a semantic to the **spatial arrangement** of information (single and multiple resources) is of major importance. Information discovery applications often employ a sophisticated 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. Common ways of arranging multiple resources include list, grid, and gallery layouts.

Additionally, **consistency** in the way multiple and single resources are represented helps forming a conceptual model of how the application can be used and provides some degree of predictability [49].

5.2.3 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 5.3 of the conceptual framework, designers can find ways to add value to information and enable information discovery 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 [23, 24]. Thus, the curation activity consists of actions such as information management, preservation, information augmentation, sharing, and channel-picking.

5.2.3.1 Management

Information management is one of the key elements of information curation [23, 24]. Information management mechanisms are prevalent in applications that have a lot of information that is hard to categorize automatically or can mean something different for each user. In the context of Web information management, **tagging** and **collection-based** information categorization play major roles.

Resource categorization helps establish relationships between various resources [23, 24]. Allowing people to tag can aid rediscovery and discovery in a socially curated space, as well as add more value to resources [50]. Sample applications that facilitate

Table 5.3. *Curation Mechanisms*

Curation mechanisms	Questions to be posed during the design or evaluation of discovery and curation tools
Management	
Collection-based categorization	Is it possible to sort information into collection-like structures, either privately or publicly?
Tag-based categorization	Is it possible to tag information, either privately or publicly?
Preservation	
Internal preservation of internal resources	Is it possible to preserve internal information within the application?
Internal preservation of external resources	Is it possible to preserve external information within the application?
External preservation of internal resources	Is it possible to preserve internal information outside of the application?
Augmentation	
Annotation	Is it possible to annotate resources, either privately or publicly?
Evaluation	Is it possible to evaluate resources, either privately or publicly?
Sharing	
Adding resources	Is it possible to add resources to the collection of information within the application from other websites?
Internal sharing	Is it possible to publicly reshare internal resources within the application?
External sharing	Is it possible to publicly reshare internal resources outside of the application?
Channel-picking	
User subscription	Is it possible to subscribe to activities of other users?
Site subscription	Is it possible to subscribe to site updates?
Artifact subscription	Is it possible to subscribe to artifact updates?

information management are Pinterest, a tool that supports tagging and collection-based categorization, and Tumblr, a tool that only supports tagging.

5.2.3.2 Preservation

Information preservation is a common Web task that is usually performed with the intent of revisiting information [45, 24]. However, information gathering is sometimes performed with just the goal of collecting information rather than revisiting it in the future [8].

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 [45].

For example, in one of the image discovery application, 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.

5.2.3.3 Augmentation

One of the most important elements of digital curation is augmentation: adding value to information [23, 24]. It is often performed within social bookmarking systems. Many Web applications allow users to add value to the resources they curate.

One way to augment information is by **annotating** it. Annotations are metadata attached to a resource, such as comments and descriptions. Annotations make it easier to search for and interpret information. For example, Yelp and TripAdvisor largely rely on reviews written by their users.

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. Many applications allow users to evaluate resources by rating them or recording other forms of approval or disapproval, such as "I like this" and "I dislike this" buttons on YouTube.

5.2.3.4 Sharing

Sharing information is key to empowering social information curation [23]. 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 and resharing them within the Web application supports channel-based information discovery within the media channels. Information discovery applications commonly allow for sharing information on popular networking sites outside the application.

5.2.3.5 Channel-picking

Channel-picking is an action of selecting information sources. A common enabler for this action is subscriptions. Subscriptions to updates from a site help users follow the news [51]. In order to support channel-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, such as the one in Pinterest, would allow.

In some applications, the content is updated and curated by users, and users can **subscribe** to other **users** or **artifacts**. Similar to site subscriptions, user and artifact subscriptions are subscriptions to activity updates. These subscription mechanisms help with networking and provide awareness about other users' activities [52]. Such subscriptions also help to further filter new content delivered to the user.

5.2.4 Summary

Information discovery and curation tools can have different implementations depending on the motives behind these activities. The design factors presented in this chapter can help enable different actions associated with information discovery and curation. However, the activities can be significantly improved by additional support and automation, as described in the next section.

5.3 Enhancing Information Discovery and Curation Experience

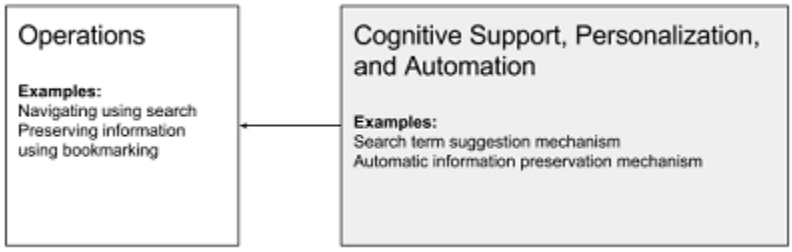

Figure 5.5. *Section Overview: Enhancing Information Discovery and Curation Experience*

Information discovery and curation enablers presented in the previous section are design elements that afford various operations. For example, the search feature affords typing in a query and searching. These operations can be further aided by another set of design elements. A high-level overview of this part of the framework is illustrated in Figure 5.5. The primary goal of this part is to highlight opportunities for improvement over various information discovery and curation enabling mechanisms.

Strategies for improvement include providing additional cognitive support for a given operation, personalizing user experience, and automating an operation. Not all of the strategies are feasible for every single operation. Some operations can be supported in multiple ways. The following sections outline some of the possibilities for advancing information discovery and curation features.

5.3.1 Enhancing Navigation

There are two common methods of enhancing information discovery when search-based navigation is used (see Table 5.4). The first method entails returning personalized results when the user enters a search query. **Personalization** can be accomplished using a variety of techniques, including predefined user preferences, social interactions, context, browsing history, etc. The second method is to **suggest search terms** to make it easier for the user to formulate the information need. For example, Yelp suggests search terms as the user enters their query.

To further support referential navigation, applications can **personalize** reference suggestions, such as **categories**, **tags**, and **topics** of interest. They can also suggest relevant resources based on the one that the user already picked. As an example, Pinterest showcases similar 'pins' once the user clicks on any of them.

For opportunistic navigation, Web tools sometimes allow users to **personalize** types or categories of information that they the users would like to discover. StambleUpon allows to not only choose topics of interests, but it can also help discover new topics that might be promising for the user.

Featured content can also be **personalized** to improve information discovery with system-regulated navigation. For example, Yelp showcases restaurants from a predefined area, such as the city where the user is from.

Finally, to make better use of subscribed content and reduce human efforts in information search, an application can support various **notification mechanisms**. These mechanisms can advise the user about updates on the **Website content**, various **artifacts**, and activities of other **users**.

5.3.2 Enhancing Exploration

Personalization of the **spatial** information representation usually has limited support in Web applications. Presumably, it is because consistency is more welcomed within information discovery applications than spatial personalization. However, it is still a possibility to personalize how multiple resources or information within a single resource are arranged (see Table 5.5).

Visual and **textual personalizations** are more common, especially when the content within the application is curated by its users. For example, a Web application

Table 5.4. *Cognitive Support, Automation, and Personalization for Navigation*

Support, automation, and personalization elements	Questions to be posed during the design or evaluation of discovery and curation tools
Descriptional	
Personalized results	Does the search mechanism return personalized results?
Guided search	Does the system suggest search terms to the user?
Referential	
Suggesting categories	Does the system suggest categories of interest?
Suggesting topics of interest	Does the system suggest topics of interest?
Suggesting tags	Does the system suggest similar tags?
Suggesting similar resources	Does the system suggest similar resources?
Opportunistic	
Personalized opportunistic navigation	Is it possible to personalize opportunistic navigation?
System-regulated	
Personalized featured content	Is featured content personalized to the user?
User activity update notification	Is it possible to receive notifications about other users' activities?
Application activity update notification	Is it possible to receive notifications about website content updates?
Artifact update notification	Is it possible to receive notifications about artifact-related updates?

for tracking personal goals, BucketList, allows its users to personalize how their goals look, so that it is easier to rediscover information. Similarly, 'pinboards' on Pinterest can have personalized cover images.

5.3.3 Enhancing Curation

Management of information can be improved if the system helps the user make decisions about information categorization or tagging (see Table 5.6). Alternatively, information can be **categorized** or **tagged automatically**. For example, when

Table 5.5. *Visual and Spatial Exploration Cognitive Support and Personalization*

Cognitive support and personalization design elements	Questions to be posed during the design or evaluation of discovery and curation tools
Visual and textual cues of multiple resources	
Personalized visual preview	Is it possible to personalize visual previews of resources?
Personalized textual preview	Is it possible to personalize textual previews of resources?
Visual and textual cues of a single resource	
Personalized visual cues	Is it possible to personalize the visual cues within a resource?
Personalized textual cues	Is it possible to personalize the textual cues within a resource?
Spatial proximal cues of multiple resources	
Personalized arrangement of multiple resources	Is it possible to personalize the arrangement of resources?
Spatial proximal cues of a single resource	
Personalized arrangement of information within a resource	Is it possible to personalize the arrangement of information within a resource?

the user bookmarks a restaurant on Yelp, it is automatically categorized. The user can filter bookmarks by category whenever they go into the embedded bookmark manager.

Preservation operations can also be automated. An example of the most common automatic preservation mechanism is **history**. Many applications, such as YouTube and Google Maps, preserve users' browsing history so that they can review it later. Additionally, preservation mechanisms can be **suggested** to the user.

Table 5.6. *Cognitive Support, Personalization, and Automation for Curation*

Cognitive support, personalization, and automation elements	Questions to be posed during the design or evaluation of information discovery and curation tools
Management	
Suggesting collections	Does the system suggest relevant collections?
Suggesting tags	Does the system suggest relevant tags?
Automated classification into collections	Does the system automatically sort resources into collections?
Automated tagging	Does the system automatically tag resources?
Preservation	
History	Does the system automatically preserve information found by the user?
Suggested preservation	Does the system suggest preservation channels to the user?
Augmentation	
Automated augmentation	Does the system automatically annotate resources?
Suggested augmentation	Does the system suggest annotation options to the user?
Sharing	
Automated sharing	Does the system support automatic sharing?
Suggested sharing	Does the system suggest sharing channels to the user?
Subscription	
Suggesting users for subscription	Does the system suggest which users to subscribe to?
Suggesting artifacts for subscription	Does the system suggest which artifacts to subscribe to?
Automated subscription	Can the system subscribe the user automatically to the website activity?

YouTube allows users to **automatically share** information about their activities, such as comments, added videos, liked or disliked videos, and created playlists. In general, socially curated spaces **offer sharing channels** to support convenient information communication.

Augmentation is another aspect of information curation that can be either **automated** or **suggested** to the user. For example, Yelp asks to rate places which the application identifies as visited by the user.

Notification mechanisms enable user awareness about new content on the subscribed channel [52]. Web applications, that facilitate rapidly updating content, support various notification mechanisms, such as messages within the application, informative emails, and smartphone notifications. Many types of notifications include suggestions of users or artifacts to follow. Some Web tools subscribe users to their notifications automatically, usually along with registration.

5.3.4 Summary

Providing cognitive support, personalization, and automation dramatically improves user experience when they interact with information discovery and curation systems. Using the information discovery factors in my framework, I described and evaluated currently existing tools. Similarly, the framework can be used for identifying gaps in information discovery support and developing new technologies (see Chapters 7 and 6).

Chapter 6 — Framework Validation

In order to validate the conceptual framework (see Chapter 5) and verify its stability, I applied it to the evaluation of five of the applications that were used in the construction of the preliminary framework (see Chapter 4): Pinterest, Google Maps, Wikipedia, Delicious, and Yelp. For each of the Web applications, I first summarize my observations resulting from asking the questions from the framework in a systematic manner. Based on my assumptions and judgment and with help of the framework, I propose directions for future development and reflect on the necessity of having some of the mechanisms, as not all mechanisms are always needed.

6.1 Pinterest

Pinterest is a Web application designed for image discovery and curation, which is oriented towards finding inspiration and collecting knowledge about hobbies and interests [53, 54, 55]. Users of Pinterest are commonly referred to as 'pinners'. Resources on Pinterest are called 'pins', and each 'pin' consists of an image, short description, user's name, and name of a collection that the pin belongs to. More information is available once the user clicks on one of those 'pins'.

Motivated by the desire to gain inspiration and knowledge, as noted above, users of Pinterest have either underdefined or absent information needs. Other motives for using Pinterest, that follow from the main purpose of the tool, could be to rediscover previously found information (and possibly use it), to be oriented about new 'pins' that emerge from subscribed channels, and to collect information for the purposes of future rediscovery and collection itself.

Navigation in Pinterest is mostly supported by descriptional, referential, or system-regulated mechanisms. Although an explicit **opportunistic navigation mechanism** is absent, both descriptional and referential mechanisms usually return novel and serendipitous results to facilitate opportunistic browsing. Descriptional naviga-

tion is enabled with a **guided search** mechanism, which suggests search terms to the user.

Referential navigation is enabled in Pinterest using a range of techniques. To support articulation of an information need, **category-based** navigation mechanism makes further suggestions on subcategories, or interests. Through clicking on any of the 'pins', the user can see related resources, which enables **resource-based** referential navigation. Most of the images in Pinterest are 'pinned' from other Websites, and users are provided with links to their original sources. Therefore, Pinterest supports **integrated** referential navigation.

System-regulated navigation within Pinterest is highly personalized. When the user enters the site, they see the history of their own information gathering activities and updates from the people they are subscribed to. Additionally, the application suggests **featured** 'pins' selected based on the user's personal interests.

To reinforce discovery of visual data, Pinterest extensively supports various exploration mechanisms. Multiple resources are represented in a **gallery layout** which is often referred to as a 'pinboard'. Such layout provides good spatial support for exploration and makes it easier to build a mental model of the tool by drawing analogies with a real pinboard. Users can create multiple 'pinboards' (also known as 'boards'), and they can **personalize** covers of all 'boards' to enhance future exploration and rediscovery.

A single resource does not have a lot of distinct spatial arrangements; however, visually it provides a glimpse into what can be found on the Website that the image came from, with **textual preview** being limited to the address of the Website.

Information management is accomplished through sorting 'pins' into different collections ('pinboards') thus enabling **collection-based classification** and **internal preservation of internal and external resources**. All of user information collecting actions are **automatically** preserved and displayed. Users can augment the information pool by uploading new 'pins', commenting on existing 'pins', or adding descriptions. Users can also **internally share** 'pins' among themselves. Channel-picking actions are carried out through following or **subscribing** to users or individual 'pinboards'. The system also 'automatically' sends 'notifications' though emails and 'suggests' new 'boards' to follow.

Applying the framework to Pinterest revealed that the tool employs a variety of techniques to facilitate information discovery and curation. However, individual mechanisms could be further improved. For example, **textual previews** of multiple and individual resources is rather limited, and provides little insight to what information source Websites actually contain. As another example, Pinterest could benefit from **automatically classifying** 'pins' into 'boards' because finding an appropriate 'board' for a 'pin' can be difficult with a large number of existing 'boards'. Overall, Pinterest provides rich support for information disocvery and curation, and in some ways, enable each of the discovery or curation actions of the conceptual framework.

6.2 Google Maps

Google maps is a Web application oriented towards navigation and place discovery [56]. It provides services for finding directions to places, their addresses, and other information. By answering the questions from the conceptual framework for information discovery and curation and analyzing the application, I arrive at the following description of Google Maps.

The primary motive behind using Google Maps is usually to find specific information about a place, most commonly directions to that place. The information need can be either very precise, such as looking for an address of a particular place, or it can be slightly more ambiguous, such as looking for a coffee shop within a certain area. Sometimes users also return to the site to rediscover previously found directions or addresses.

Information discovery in Google Maps is usually initiated by a **search**, and thus, the user needs to formulate their information need—the application lacks some of the internal referential navigation mechanisms, so there is nothing that aids users in this task. The one type of referential navigation that Google Maps does support is **resource-based**. For example, the user can click on the "Search nearby" suggested link to find places near another place. Google Maps is conveniently **integrated** with Google+, allowing access to relevant information, such as reviews, images, and hours of operation, and thus, enabling resource-based integrated referential navigation. Search-based navigation within Google Maps is usually precise and returns

accurate search results for specific places, making it easy to rediscover information using search.

Google Maps lacks **opportunistic navigation mechanisms**, and it provides limited support for system-regulated navigation by displaying **personalized featured content** in a form of a map of the user's location.

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**. **Consistency** in how resources are represented makes it easy to find information, such as addresses and contacts. Furthermore, multiple and individual resources provide **visual previews** that show photographs added by users or street views respectively.

Google Maps supports curation mainly through personal preservation. Users can only bookmark places to the "My Places" list — by adding **internal content to internal storage**. Other types of place preservation are possible through Google+, however, not within Google Maps. Users can also **evaluate** and **annotate** places through Google+, and aggregated reviews and ratings are visible on Google Maps. Sharing is enabled through providing functionality to add new locations to Google Maps and supplying links and code for embedding.

Channel-picking actions are usually enabled within applications with frequently updating content. Content provided by Google Maps is fairly stable, and therefore, there are no channel-picking mechanisms used by the application.

Evaluating Google Maps using my conceptual framework also exposes some gaps in its design. 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 as well as adding functionality for channel-picking introduces the possibility of channel-based 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 channel-based discovery, thereby complimenting the Google Maps application.

6.3 Wikipedia

Wikipedia is an open encyclopedia containing millions of articles contributed by people from all over the world [57, 58]. The primary motive for discovering information on Wikipedia is to gain knowledge to either answer a specific question or to learn more about a general topic, such as art or history.

Wikipedia supports a wide range of navigational mechanisms. Descriptional navigation on Wikipedia is accomplished through **search**. Results, however, are not **personalized to the user**, and the search mechanism is not **guided** by suggestions of what search terms to use, which could help the user to formulate their information need. Referential mechanisms include a **categories** which consist of broad topics and return **featured articles**. Not all Wikipedia articles are **integrated** with other Websites and Web applications; however occasionally, articles contain "External Links" section that provides links to external resources.

Opportunistic navigation facilitates serendipitous discovery of new articles, and it is enabled though "Random article" link located on the navigation sidebar. Wikipedia regularly updates **featured content** that can be viewed on the front page and when navigated to using categories. Users can also see the history of recently updates articles.

Exploration of multiple resources is limited to when the target of the search query is unclear. Then, search results are presented in a **list** layout, where links are represented as **text**. Single resource exploration mechanisms include a table of contents in large articles, which can serve as a referential navigation mechanism as well as **textual cue** of what the article consists of, and occasional **visual** material that aids in communicating ideas of the article.

Information on Wikipedia is publicly curated by thousands of users, who improve existing articles and **add** new. Although it is not possible to **subscribe** to any particular channel, moderators of Wikipedia regularly update featured articles and information. Augmentation is possible through personal contribution to the content of articles. However, there are no private curation mechanisms that could be used for personal benefit.

The application of the framework to Wikipedia revealed that major gaps in its discovery and curation support are related to personal curation and visual explo-

ration of multiple resources. For example, since there are no mechanisms for personal preservation and management, users cannot build their own knowledge maps and continuously engage with the content. Adding more cognitive support or personalization, such as suggesting search terms or topics of interest could perhaps also improve user experience.

6.4 Delicious

Delicious is a Web application designed for social bookmarking and information discovery [59, 60]. The primary motive for using Delicious is to preserve articles found on other Websites for future access and to discover new articles or blog posts. When used for discovery, the information need is usually underdefined or absent unless the user's motive is to rediscover previously found information.

In Delicious, users can navigate using **search** (descriptional navigation). Referential navigation within the application is accomplished through **resource-based** search, which returns related links. Delicious is also integrated with many other Websites through linking, which enables **integrated referential navigation**. The application does not support opportunistic browsing, but it does provide an option for system-regulated browsing through the "Trending" section of the Website which displays **featured content** based on article popularity. The "Trending" section displays results of the social curation, and therefore, enables channel-based discovery. Since the "Trending" section displays results of social curation, it enables channel-based discovery.

Exploration of a single resource is not enabled in Delicious, and the mechanisms for exploring multiple resources, for the most part, are limited to **textual previews** and a **list** layout. However, the "Trending" section of the system does provide **visual previews** and arrange resources in a **grid** layout. In addition, it provides extended **textual previews** or snippets of corresponding articles, making it easier to follow the information scent when navigating across various Websites. Although these mechanisms help with visual and spatial exploration, having them applied to only one section of Delicious simultaneously undermines the **consistency** of multiple resource representation.

Since the primary motive for using Delicious is to preserve and share information, support for curation is the core feature of this Web application. Management can be performed through **tagging**, and the system **suggests** tags based on the tag cloud. Delicious supports **internal preservation of external and internal resources**, **external sharing of internal resources**, as well as **adding** new resources. Information augmentation is possible through commenting on (or **annotating**) added links. Channel-picking is performed thorough **subscription mechanisms** — users can follow other users and build their networks.

According to the framework, Delicious lacks extensive visual exploration mechanisms in most of its sections. Since Delicious is designed to facilitate article discovery, and Web page titles often provide limited insight about an article, the tool could benefit from providing **textual** and **visual** article previews in all of its sections to help the user follow the information scent.

6.5 Yelp

Yelp is a Web application used to discover local businesses, such as restaurants, beauty salons, and shops [61]. The primary motive for discovering information on Yelp is to evaluate and compare businesses in certain domains and geographical locations. Therefore, users either have defined information needs, such as rating of a specific business, or underdefined information needs, such as a good restaurant in a certain area. Most of the content on Yelp consists of user reviews and business evaluations or ratings.

Descriptional navigation in Yelp is once again supported using the **search feature**, which does not only **suggest** search terms to the user, but also allows to further specify the proximate location of a business. Referential navigation is enabled using **category-based** navigation and **filtering**. **Integrated references** are provided to navigate to Google Maps and to business Websites. On Yelp, the user can see a **news feed** of activities of other users, as well as **featured** businesses tailored to the user based on their location.

Both visual and spatial exploration mechanisms are enabled on Yelp. Visual explorations of multiple and single resources are facilitated by numerous photographs,

icons, maps, and visuals depicting ratings. Spacial representation of information on Yelp consists of a blend of **list**, **grid**, and **gallery** layouts and other spatial arrangements.

In addition to discovery, Yelp supports various curation actions. Users can bookmark businesses they like within the system, thus performing **internal preservation of internal resources**. They can further augment information by either writing reviews or performing an **evaluation** of businesses or other reviews. Evaluation of businesses is enabled using a five-star rating system, and reviews can be evaluated by choosing between 'Useful', 'Funny', and 'Cool' metrics.

Identified gaps include lack of management mechanisms when businesses are bookmarked and lack of channel-picking mechanisms. A lot of information on Yelp is continuously updated, so channel-picking could help filter updated information. Furthermore, adding mechanisms for opportunistic navigation could make it possible to discover new restaurants every time and help the user when their information need is undefined.

6.6 Discussion

The evaluations of existing Web tools, provided in this chapter, using the conceptual framework for information discovery and curation demonstrate applicability of the framework. A set of questions, provided by the framework, can help in the process of tool evaluation, and they can also be applied in drawing distinctions between different tools.

The framework associates different information discovery and curation actions with concrete mechanisms. However, it is not always clear which actions in the framework the tool needs to support. With help of the framework, some of the requirements (but not all) can be derived from the analysis of other applications, which might be in a similar domain or possess desired properties.

Another source that can hint on which actions need support is the motive for discovery and curation activities in an application. For example, if the motive is to discover information with an undefined information need, the application can either be tailored to support serendipitous discovery by providing opportunistic navigation

mechanisms, or it can help the user to formulate the information need by suggesting search terms and categories.

Finally, the need for a given discovery or curation-supporting mechanism can be evaluated using intuition and experience of the designer. In some cases, it is especially difficult to estimate the importance of a mechanism in a specific application using known characteristics of the tool; however, as with many other decisions when it comes to developing or improving an application, the designer can use their judgment along with subsequent studies and evaluation.

Evaluation and comparison of different Web applications can reveal useful insights about the mechanics of how the system induces user experience, and they can expose certain unadressed needs. The next chapter illustrates the design process of a Web application that emerged from the evaluations of other tools using the conceptual framework.

Chapter 7 — Framework Application for Design

To verify that the conceptual framework is effective, I applied it to design a place photo discovery application. This chapter outlines the role the framework played in the design process of the Web application, the resulting application and its features, and some prospects for future application development.

7.1 Applying the Conceptual Framework to Design an Application

A need for a place photo discovery application was revealed during the construction phase of the framework. Asking questions from the preliminary framework (see Chapter 4) about existing applications, such as Pinterest and Google Maps, helped expose the need for discovery and curation of place photos with additional access to place location data and other details. It also helped gather some of the requirements for a photo discovery application. Once user needs and motives for information discovery and curation of place photographs where established, I repeatedly consulted the framework throughout the development process of the application in order to systematically select the next feature to be implemented.

In general, Web applications that are tailored towards image discovery, such as Pinterest and We Heart It, support user's motive to close a knowledge gap that is characterized by underdefined information needs. To deal with the issue of having an underdefined information need, an application has to help the user to formulate their information need as well as support serendipitous discovery of information. In order to enable serendipitous discovery, Web applications regularly update the content they provide by allowing users to add new resources and curate information.

The task of image seeking for the purpose of finding inspiration (as it is the

case for the majority of Pinterest users) can stretch out to multiple sessions over an undetermined period of time. Curation mechanisms, such as preservation and management, help the user to rediscover information that allows them to reflect on the previous findings and continue image seeking.

It is common for users to discover place photographs on Pinterest, and Pinterest does display a map when a location of a place is known within the system. However, this feature only applies to a relatively small fraction of existing 'pins'. In addition, Pinterest facilitates discovery of images related to diverse topics and interests, and not only places, which makes it harder to tailor the user experience to facilitate discovery and curation based on their desired motives.

When it comes to place discovery, the Google Maps application provides the ultimate support for finding place and business locations. It is also possible to see what a place looks like based on an associated image. However, since the application is oriented towards finding specific information, visual and spatial photo exploration mechanisms are not well-developed. The user can preserve a given place but cannot preserve or organize photographs of places. Google Maps also lacks category-based navigation mechanisms which can help the user identifying their information needs.

The findings above helped me define a motive for a place photo discovery application, which is *to find inspirational (underdefined) place photographs, to collect and manage found information for future use and retrieval, as well as to provide access to more defined information about the place*, such as its location. After formalizing the motive for the application use, I refer back to the framework to choose options for supporting various aspects of information discovery and curation while developing the application.

7.2 KeePlaces Features and Future Prospects

The resulting Web application, KeePlaces[1] (see Figure 7.1), supports discovery and curation of place photographs, and it is integrated with Google Maps. This section outlines the main features of KeePlaces in accordance with the conceptual framework.

KeePlaces supports descriptional, referential, and partially system-regulated nav-

[1] A prototype of KeePlaces is available at www.keeplaces.com

igation methods. It is possible to perform descriptional navigation using **integrated search** that in turn utilizes Google Maps APIs to search for photographs of different places. The search feature is not **guided**, and at this time, results are not **personalized**. Descriptional navigation could be improved by suggesting search terms to the user once they start typing. However, personalizing the results of searching might not be a good strategy because users might want to explore photographs that they have not seen before or that are of places not related to them.

Users can navigate using **categories** which enable referential navigation. Currently, categories that users might be interested in are only approximately estimated, and no other traditional referential mechanism is employed for navigating within the application. However, the users can navigate to Google Maps by clicking "View Google Maps" link beside every photograph to see where the place is located and perform any other actions within the Google Maps application. This feature enables **integrated referential navigation**.

With the preliminary prototype, as the user first visits KeePlaces, the system displays predefined tourist attraction photographs, and therefore, it supports system regulated navigation by displaying **featured content**. However, this solution is temporary since system-regulated navigation could be further improved by **personalizing featured content** and delivering **notifications** about content updates to the user.

Currently, **opportunistic navigation** is not enabled in KeePlaces although users with undefined information needs could benefit from this method. Alternatively, other navigation methods could return serendipitous results.

Spatial exploration of multiple resources is enabled using **gallery layout**. A **grid layout** could be an alternative way to present information within the application; however, **list** is not always an optimal solution to presenting visual data. Resources are represented as photographs, and these photographs serve as **visual cues** to what the places they represent look like. In addition to visual cues, **textual cues** provide names of different places delivering additional exploration support.

KeePlaces does not support exploration of individual resources as of today; however, enabling it could improve future information discovery. Furthermore, allowing users to **personalizate** of visual or textual cues can help them rediscover place pho-

tographs and collections.

Curation in KeePlaces is supported through management and preservation. Management is implemented using **collection-based classification**. Every photo discovered on the site can be bookmarked by clicking the the 'Keep' button and choosing a collection. This bookmarking mechanism enables **internal preservation of internal resources** since it allows to save information found within KeePlaces.

Some aspects of curation, such as information sharing, augmentation, and channel-picking have not been enabled yet. These activities are important because they contribute to collaborative and creative environments as well as help building community around the Web application. In KeePlaces, having the users add new photographs and share them among themselves could scale the application usage up and enrich the quality of the content provided to the users.

In order to support channel-picking, a Web application must regularly update its content, which can be done by either moderators or general users. Then, adding **subscription mechanisms** and **notifications** can further empower channel-based discovery. For place photo discovery, a tool, such as KeePlaces, can provide updates about photographs preserved by other users, new photographs added to the pool of information, spatial featured photographs, etc.

Although KeePlaces has not been released as a stable Web application yet, it supports the discovery and curation of place photographs from all over the world. Applying the framework above as presented, can guide its future development and evaluation.

7.3 Discussion

The conceptual framework for information discovery and curation guided the design of the place photo discovery application, KeePlaces. The framework assisted in identifying the need for a Web application that facilitates the discovery of place photographs in particular, and it highlighted which design elements are important in this specific case. Similarly, the framework can aid in the design process of other applications.

When the motive for an application use is known, one can evaluate Web applications from similar domains to identify gaps in the provided features. Finding

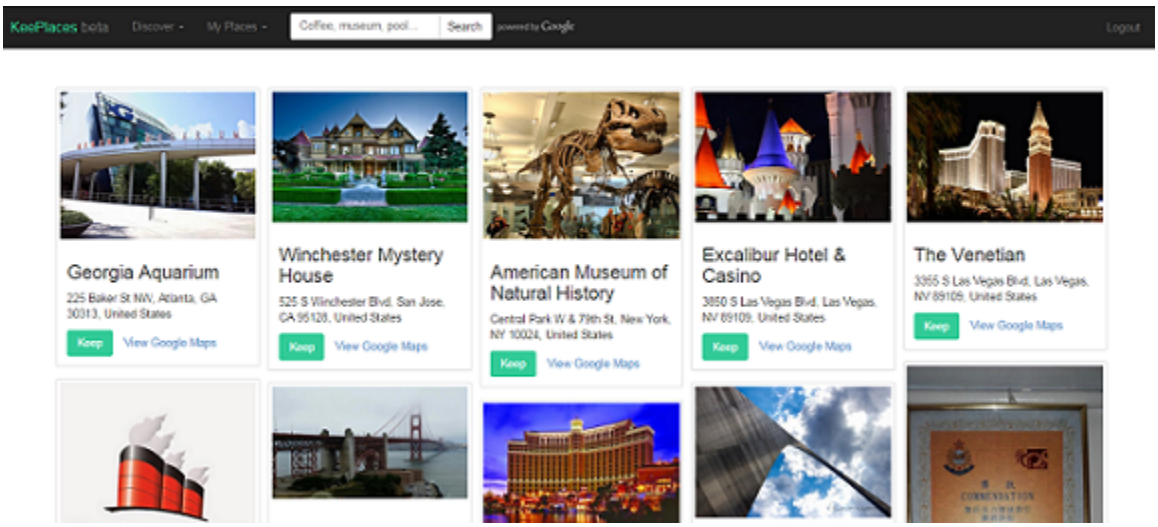

Figure 7.1. *KeePlaces Interface*

feature gaps is a challenging task, but the framework can assist by making easier to relate relevant information behaviour with concrete mechanisms and features. Ongoing reevaluation of the tool and its competitors using the framework can help with continuous development process and improve user experience when they interact with the system.

Chapter 8 — Research and Design Implications

The conceptual framework for information discovery and curation is designed to perform formative and summative evaluation of existing Web applications, with a goal to reveal how these tools support information-related activities in question. The framework, as a tool, and its capability to guide the process of analyzing Web applications makes it broadly applicable in research and Web design.

In Chapter 6, I demonstrate how the framework can be used to reveal missing features in tools. Using similar methods, the framework can also be applied to compare different Web applications. When used for evaluation, the framework helps to identify which areas of a given tool require further attention. Therefore, the framework can be helpful for designers who wish to improve existing tools or get ideas for new information discovery and curation applications.

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

User interface designers face certain trade-offs when developing applications. Therefore, it is not always advantageous to implement all of the missing features. For example, providing the support to customize spatial arrangement of multiple resources can undermine the consistency of their representation.

In the research domain, the framework can serve as a guide for drawing distinctions between different Web-based information discovery and curation applications, finding gaps in tools that can be studied, and selecting cases for studies based of required functionality. Hence, both researchers and developers can benefit from the systematic tool examination guided by the framework.

Even though applying the framework requires initial expertise and critical reasoning, it opens up opportunities for research and practice. Systematic evaluation of Web tools for information discovery and curation helps improve user experience and gain better understanding of information behaviour within a given system.

Chapter 9 — Future Work and Conclusions

In my study, I 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 and curation tools. I then evaluated and iteratively refined the framework by analyzing twenty different information discovery applications and provided concrete examples of tool support addressing various concepts of the framework. Finally, I designed a Web-based application for place photo discovery and curation using the conceptual framework, and validated the framework by reevaluating five of the previously examined tools.

The current version of the framework is generalized to be applicable to most information discovery applications. Finding ways to instantiate the framework and extend it to be used in domain-specific practices could serve as a potential research goal in the future. For example, video discovery and curation activities have unique properties related to the type of data to be discovered — information is mostly found in the video itself, and it cannot be viewed all at the same time. Hence, the framework could be extended to address domain-specific challenges.

Another potential research direction would be to expand my investigation to include factors that influence the need for one information discovery type over another and further deepen an understanding of the relationship between the motives for information discovery and curation activities and information discovery types.

Additionally, one could investigate how collaboration in information discovery and curation relates to the conceptual framework. Generally, collaboration mechanisms in most Web information discovery applications are limited to information sharing, public information augmentation and tagging. However, collaboration often involves other activities, such as communication, coordination, and other domain-specific shared activities.

My framework opens up opportunities for structured information discovery and curation 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.

References

[1] J. A. Waterworth and M. H. Chignell, "A model of information exploration," *Hypermedia*, vol. 3, no. 1, pp. 35–58, 1991.

[2] A. Foster and N. Ford, "Serendipity and information seeking: an empirical study," *Journal of Documentation*, vol. 59, no. 3, pp. 321–340, 2003.

[3] C. A. Lynch, "Networked information resource discovery: an overview of current issues," *Selected Areas in Communications, IEEE Journal on*, vol. 13, no. 8, pp. 1505–1522, 1995.

[4] H. A. Proper and P. Bruza, "What is information discovery about?," *Journal of the American Society for Information Science*, vol. 50, no. 9, pp. 737–750, 1999.

[5] J. S. Bruner, "The act of discovery.," *Harvard educational review*, 1961.

[6] T. D. Wilson, "Human information behavior," *Informing science*, vol. 3, no. 2, pp. 49–56, 2000.

[7] D. Giaretta, "Dcc approach to digital curation," 2006.

[8] S. E. Lindley, S. Meek, A. Sellen, and R. Harper, "It's simply integral to what i do: enquiries into how the web is weaved into everyday life," in *Proceedings of the 21st international conference on World Wide Web*, pp. 1067–1076, ACM, 2012.

[9] M. J. Bates, "An exploratory paradigm for online information retrieval," *Intelligent Information Systems for the Information Society. Amsterdam: North-Holland*, pp. 91–99, 1986.

[10] A. Conaway, C. Pikas, U. McLean, S. Morris, L. Palmer, L. Rosman, S. Sears, E. Uzelac, and S. Woodson, "Designing for information discovery: User needs analysis," *Johns Hopkins Applied Technical Digest*, vol. 28, no. 3, pp. 290–291, 2010.

[11] G. Marchionini, "Exploratory search: from finding to understanding," *Communications of the ACM*, vol. 49, no. 4, pp. 41–46, 2006.

[12] A. Kerne and S. M. Smith, "The information discovery framework," in *Proceed-*

ings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques, pp. 357–360, ACM, 2004.

[13] C. W. Choo, B. Detlor, and D. Turnbull, "Information seeking on the web: An integrated model of browsing and searching," *first monday*, vol. 5, no. 2, 2000.

[14] D. Ellis, "A behavioural model for information retrieval system design," *Journal of information science*, vol. 15, no. 4-5, pp. 237–247, 1989.

[15] M. Kellar, C. Watters, and M. Shepherd, "A goal-based classification of web information tasks," *Proceedings of the American Society for Information Science and Technology*, vol. 43, no. 1, pp. 1–22, 2006.

[16] J. B. Morrison, P. Pirolli, and S. K. Card, "A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions," in *CHI'01 extended abstracts on Human factors in computing systems*, pp. 163–164, ACM, 2001.

[17] A. J. Sellen, R. Murphy, and K. L. Shaw, "How knowledge workers use the web," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 227–234, ACM, 2002.

[18] R. K. Yin, *Case study research: Design and methods*. Sage publications, 2014.

[19] T. D. Wilson, "On user studies and information needs," *Journal of documentation*, vol. 37, no. 1, pp. 3–15, 1981.

[20] T. D. Wilson, "Information behaviour: an interdisciplinary perspective," *Information processing & management*, vol. 33, no. 4, pp. 551–572, 1997.

[21] T. Saracevic, "Modeling interaction in information retrieval (ir): A review and proposal.," in *Proceedings of the ASIS annual meeting*, vol. 33, pp. 3–9, ERIC, 1996.

[22] P. Ingwersen, "Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory," *Journal of documentation*, vol. 52, no. 1, pp. 3–50, 1996.

[23] N. Beagrie, "Digital curation for science, digital libraries, and individuals," *International Journal of Digital Curation*, vol. 1, no. 1, pp. 3–16, 2008.

[24] S. Whittaker, "Personal information management: from information consump-

tion to curation," *Annual review of information science and technology*, vol. 45, no. 1, pp. 1–62, 2011.

[25] M. J. Bates, "Toward an integrated model of information seeking and searching," *The New Review of Information Behaviour Research*, vol. 3, pp. 1–15, 2002.

[26] D. Ellis, D. Cox, and K. Hall, "A comparison of the information seeking patterns of researchers in the physical and social sciences," *Journal of documentation*, vol. 49, no. 4, pp. 356–369, 1993.

[27] D. Ellis and M. Haugan, "Modelling the information seeking patterns of engineers and research scientists in an industrial environment," *Journal of documentation*, vol. 53, no. 4, pp. 384–403, 1997.

[28] C. C. Kuhlthau, "Inside the search process: Information seeking from the user's perspective," *JASIS*, vol. 42, no. 5, pp. 361–371, 1991.

[29] T. D. Wilson, "Models in information behaviour research," *Journal of documentation*, vol. 55, no. 3, pp. 249–270, 1999.

[30] C. Janiszewski, "The influence of display characteristics on visual exploratory search behavior," *Journal of Consumer Research*, vol. 25, no. 3, pp. 290–301, 1998.

[31] P. Pirolli and S. Card, "Information foraging.," *Psychological review*, vol. 106, no. 4, p. 643, 1999.

[32] W.-T. Fu and P. Pirolli, "Snif-act: A cognitive model of user navigation on the world wide web," *Human–Computer Interaction*, vol. 22, no. 4, pp. 355–412, 2007.

[33] M. Kitajima, M. H. Blackmon, and P. G. Polson, "A comprehension-based model of web navigation and its application to web usability analysis," in *People and Computers XIVUsability or Else!*, pp. 357–373, Springer, 2000.

[34] P. Pirolli, "An elementary social information foraging model," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 605–614, ACM, 2009.

[35] W.-T. Fu, "The microstructures of social tagging: a rational model," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pp. 229–238, ACM, 2008.

[36] M. Kellar, C. Watters, and M. Shepherd, "A field study characterizing web-based information-seeking tasks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 999–1018, 2007.

[37] M. R. Morris, "A survey of collaborative web search practices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1657–1660, ACM, 2008.

[38] S. A. Golder and B. A. Huberman, "Usage patterns of collaborative tagging systems," *Journal of information science*, vol. 32, no. 2, pp. 198–208, 2006.

[39] E. Voyloshnikova and M.-A. Storey, "Towards understanding digital information discovery and curation," in *Proceedings of the 2014 Conference of the Center for Advanced Studies on Collaborative Research*, pp. 247–261, ACM, 2014.

[40] L. Tauscher and S. Greenberg, "How people revisit web pages: Empirical findings and implications for the design of history systems," *International Journal of Human-Computer Studies*, vol. 47, no. 1, pp. 97–137, 1997.

[41] K. Kuutti, "Activity theory as a potential framework for human-computer interaction research," *Context and consciousness: Activity theory and human-computer interaction*, pp. 17–44, 1996.

[42] E. Adar, J. Teevan, and S. T. Dumais, "Large scale analysis of web revisitation patterns," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 1197–1206, ACM, 2008.

[43] A. Cockburn, S. Greenberg, S. Jones, B. McKenzie, and M. Moyle, "Improving web page revisitation: Analysis, design, and evaluation," 2003.

[44] A. Cockburn and B. McKenzie, "What do web users do? an empirical analysis of web use," *International Journal of human-computer studies*, vol. 54, no. 6, pp. 903–922, 2001.

[45] D. Abrams, R. Baecker, and M. Chignell, "Information archiving with bookmarks: personal web space construction and organization," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 41–48, ACM Press/Addison-Wesley Publishing Co., 1998.

[46] E. Estellés, E. Del Moral, and F. González, "Social bookmarking tools as facilitators of learning and research collaborative processes: The diigo case," *Interdis-*

ciplinary Journal of E-Learning and Learning Objects, vol. 6, no. 1, pp. 175–191, 2010.

[47] J. A. Oravec, "Bookmarking the world: Weblog applications in education," *Journal of Adolescent & Adult Literacy*, pp. 616–621, 2002.

[48] M. Levene, *An introduction to search engines and web navigation*. John Wiley & Sons, 2011.

[49] D. A. Norman, *The design of everyday things*. Basic books, 2002.

[50] T. Gruber, "Ontology of folksonomy: A mash-up of apples and oranges," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 3, no. 1, pp. 1–11, 2007.

[51] A. Java, P. Kolari, T. Finin, A. Joshi, and T. Oates, "Feeds that matter: A study of bloglines subscriptions.," in *ICWSM*, 2007.

[52] D. Millen, J. Feinberg, and B. Kerr, "Social bookmarking in the enterprise," *Queue*, vol. 3, no. 9, pp. 28–35, 2005.

[53] E. Gilbert, S. Bakhshi, S. Chang, and L. Terveen, "I need to try this?: a statistical overview of pinterest," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2427–2436, ACM, 2013.

[54] M. Zarro and C. Hall, "Pinterest: Social collecting for# linking# using# sharing," in *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, pp. 417–418, ACM, 2012.

[55] R. Ottoni, J. P. Pesce, D. B. Las Casas, G. Franciscani Jr, W. Meira Jr, P. Kumaraguru, and V. Almeida, "Ladies first: Analyzing gender roles and behaviors in pinterest.," in *ICWSM*, 2013.

[56] R. Gibson and S. Erle, *Google maps hacks*. " O'Reilly Media, Inc.", 2006.

[57] A. Kittur, E. Chi, B. Pendleton, B. Suh, and T. Mytkowicz, "Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie," *World wide web*, vol. 1, no. 2, p. 19, 2007.

[58] M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer, "Semantic wikipedia," in *Proceedings of the 15th international conference on World Wide Web*, pp. 585–594, ACM, 2006.

[59] E. Rader and R. Wash, "Influences on tag choices in del. icio. us," in *Proceedings*

of the 2008 ACM conference on Computer supported cooperative work, pp. 239–248, ACM, 2008.

[60] M. Tesconi, F. Ronzano, A. Marchetti, and S. Minutoli, "Semantify del. icio. us: Automatically turn your tags into senses," in *The 7th International Semantic Web Conference*, p. 67, Citeseer, 2008.

[61] M. Luca, "Reviews, reputation, and revenue: The case of yelp. com," *Com (September 16, 2011). Harvard Business School NOM Unit Working Paper*, no. 12-016, 2011.

VITA

Surname: Voyloshnikova ***Given Names:*** Elena
Place of Birth: Nikolaev, Ukraine

Educational Institutions Attended

University of Victoria 2008 to 2012

Degrees Awarded

B.Sc. University of Victoria 2012

Honors and Awards

award date

PARTIAL COPYRIGHT LICENSE

Title of Thesis: TOWARDS UNDERSTANDING DIGITAL INFORMATION DISCOVERY AND CURATION.

Author: _____

ELENA VOYLOSHNIKOVA

21 April 2015