

Towards Understanding Digital Information Discovery and Curation

Elena Voyloshnikova
elenavoy@uvic.ca

Dr. Margaret-Anne Storey
mstorey@uvic.ca

University of Victoria
Victoria, BC, Canada

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 modelled information seeking strategies and characterized information seeking and curation activities on the Web. However, there is a lack of research on how existing Web applications support the discovery and management of information, especially concerning the motivations behind them and how different approaches can be compared.

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

1 Introduction

Today, people use Web technologies to satisfy their information needs. People research their interests and hobbies using various online resources, shoppers search online stores for product characteristics to make purchasing decisions, and travelers

visit online booking sites to find information about flights and hotels. In order to accommodate diverse and evolving user needs, Web applications continuously introduce new features and services empowering information discovery and curation.

Sometimes, Web users hope to find particular pieces of information, such as showtimes and phone numbers, to satisfy specific information needs [21]. Other times, users lack well-articulated information needs, so they engage in opportunistic browsing [17]. Often, people discover information online without even looking for it [3]. The nature of information discovery can vary, and therefore, require elaborate tool support.

In addition, people perform information curation tasks, such as management and preservation, to maintain and add value to collections of information [4]. With a rapidly increasing popularity of socially curated information spaces, it is important to understand how to enable curation activities when designing tools that support information discovery.

To close their knowledge gaps, people turn to various Web technologies ranging from specialized search tools to visual discovery applications. Several studies have been directed at exploring high-level Web tasks, including information seeking tasks [14, 15, 20, 22], deriving models of information seeking behaviors [5, 7, 8, 9, 3, 2], and looking at methods of information curation [4, 24]. However, more research is necessary to determine how different tools and their features provide fundamental support for information discovery and curation.

To enhance information seeking and curating experiences and support users' interactions, we extend existing research by (1) deriving factors that enable information discovery and curation and relating them within a framework, (2) validating the framework by studying and describing currently existing Web applications, and (3) providing questions to ask, to address concepts of the framework when evaluating and designing new applications. Specifically, our research goal is to gain understanding of how existing tools support digital information curation and discovery.

The remainder of this paper is organized as follows. Section 2 highlights some of the studies and technologies related to information seeking and curation tasks. Section 3 outlines a conceptual framework of factors and questions that enable digital information discovery and support curation. In Section 4, we describe the methodology for validating the framework, and in Section 5, we report our findings and demonstrate how the framework can be used to reveal missing features. In Section 6, we describe limitations of the study, followed by future work and conclusion, in section 7.

2 Web-based Information Discovery and Curation

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

Kellar et al. [14] separated Web tasks into five categories: transactions, browsing, fact finding, information gathering, and other uncategorized tasks, with information seeking being composed of browsing, fact finding, and information gathering. Although the authors categorized information gathering as part of information seeking, it is in fact more closely related to digital curation [4, 24]. In their later work, Kellar et al. [15] added communication and maintenance as additional Web tasks.

Similarly to Kellar, Sellen et al. [22] identified six tasks that are commonly performed by Web users: browsing, finding, housekeeping, infor-

mation gathering, communicating, and transacting. Therefore, Kellar et al. and Sellen et al. both identified browsing, fact finding, and information gathering as information-related tasks that users perform online.

People often engage in information seeking activities to close some knowledge gap that occurred as a result of not having enough information to perform task [21]. Therefore, when providing tool support for various information discovery tasks, it is useful to consider the motivation behind these tasks, as it may differ. Morrison et al. [20] make a distinction between methods of Web use and purposes. The authors derived a purposes taxonomy of Web use, including three purposes or motivations: finding information, comparing or choosing to make a decision, and using the Web to find relevant information to gain understanding of some subject. Consequently, methods of finding information identified by Morrison et al. are collecting, finding, exploring, and monitoring. The differences between the two taxonomies suggest that different information seeking tasks may be performed to satisfy more than one information seeking purpose. Therefore, each purpose may require more than one task-supporting mechanism.

A number of researchers have studied information seeking behavior [2, 3, 5, 7, 9, 8]. Ellis et al [7, 9, 8], proposed a model of information seeking characterized by six different patterns: starting, chaining, browsing, extracting, monitoring, and differentiating. Consequently, Choo et al. [5] derived anticipated Web tasks that correspond to these patterns. According to the authors, when users identify sources of interest, they usually identify which Websites can point to that information of interest. Chaining occurs when users navigate through links on those initial pages. When people browse, they scan top-level pages, headings, lists, and site maps. Differentiating takes place when people bookmark, print, copy and paste information, or choose an earlier selected sites. Monitoring occurs when users revisit web pages or receive updates from some earlier visited sites. Finally, extracting can occur when the user systematically searches site to extract information of interest.

Bates [3] proposed a model of four information seeking modes being aware, monitoring, browsing, and searching. Bates differentiated the modes based on the levels 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.

In 2002, Bates [2] extended her research with the notion of information farming. Information farming involves people collecting and organizing information for future use and revisitation. More commonly, it is referred to as digital curation, which is the notion of collecting and managing digital information for the purpose of adding value to the collection, and revisitation [4]. Wittaker [24] 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.

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. [17] conducted a qualitative study involving 24 participants tracking their daily web usages 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 web browse opportunistically when they look for information related to some personal interest, long-term goal, or future ambition. Purposeful use occurs when the user knows what information she needs to acquire or what online action she needs to perform in order to continue or finish some other activity. Respite mode usually occurs when users are in the process of waiting for something or taking a break, and it serves as a means for people to temporarily occupy themselves when high engagement with the content is not a requirement. Orienting mode usually occurs when people want to be updated on what has been happening in their environment. Examples of this mode are checking email at work or looking at the news and updates on a social networking site. Finally, lean-back mode of web use

can be thought of as listening to the radio or watching TV. It usually involves watching videos online or browsing through other types of entertainment content.

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

A number of researchers have studied how people search for and curate information, including information seeking tasks, information seeking behaviors, and information curation. However, there is a lack of research on how current technologies assist people in performing these tasks. With this overview of related work, we present a framework of Web application design factors and questions that facilitate information discovery and curation.

3 Conceptual Framework

Although information discovery and curation tasks are commonly performed Web tasks today, there is lack of literature on how to support them when building applications. We reduce this gap by presenting a framework of design factors facilitating digital information discovery and curation (see Table 1).

Development of the framework began with deriving design factors based on Web tasks described in existing literature. Through analyzing twenty information discovery applications, we iteratively expanded the framework, added concepts, and established relations. Perhaps, the framework can be expanded further; however, having most essential

Table 1: Conceptual Framework.

| Factors | Questions |
|--|---|
| Discovery <i>Serendipitous discovery</i> Arbitrary navigation Search-based navigation Category-guided navigation Integration Visual link preview Spatial arrangement <i>Fact discovery</i> Search-based navigation Category-guided navigation Integration Uniform representation Visual link preview Spatial arrangement <i>Rediscovery</i> History-based rediscovery Bookmark-based rediscovery Search-based rediscovery <i>Channel-based discovery</i> Site subscription User subscription Notifications Subscription news stream Content news stream | Does the application provide a means for arbitrary navigation among resources? Does the search engine help retrieve diverse resources related to the topic of interest? Do categories suggest and help with navigating to resources related to the topic of interest? If resources originate from a different site, do they link to their original sources? If resources are delivered as links, do they have visual previews? Is there a semantic to the spatial arrangement of resources? Does the search feature help discover the specific resource of interest? Do categories help narrow results to specific types of resources? If resources originate from a different site, do they link to their original sources? If resources are uniform, are they presented in a uniform way? If resources are delivered as links, do they have visual previews? Is there a semantic to the spatial arrangement of resources? Does the application save and provide access to browsing history? Does the application support bookmark-based resource revisitation? Is the search a reliable method for resource revisitation? Does the application allow subscriptions to news and updates? Does the application allow subscriptions to other users' activities? Does the application have one or more notification mechanisms? Can subscription updates be visible within the application? Can content updates be visible within the application? |
| Curation <i>Management</i> List-based categorization Tag-based categorization <i>Preservation</i> Internal preservation of internal resources Internal preservation of external resources External preservation of internal resources <i>Augmentation</i> Evaluation Annotation <i>Sharing</i> Adding resources Internal sharing External sharing | Does the application support information information sorting into list-like structures privately or publicly? Does the application support tagging privately or publicly? Does the application support bookmarking mechanism(s) for preserving internal information within the application? Does the application support bookmarking mechanism(s) for preserving external information within the application? Does the application support bookmarking mechanism(s) for preserving internal information outside of the application? Can the resource evaluations be recorded privately or publicly? Can resources be annotated privately or publicly? Can resources be publicly added to the collection of information within the application from other Web pages? Can internal resources be publicly reshared within the application? Can internal resources be publicly reshared outside of the application? |

factors makes the framework more usable.

The framework consists of two main categories, discovery and curation, that are consequently decomposed into subcategories. Each subcategory contains factors that determine use case enablers and corresponding questions that can help application design and evaluation. This section outlines the main components of the framework.

3.1 Information Discovery

In our framework, we built on existing classifications of information seeking tasks and methods (see Section 2) to derive corresponding design factors. Thus, the discovery category consists of serendipitous discovery, fact discovery, rediscovery, and channel-based discovery.

3.1.1 Serendipitous discovery

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 [14, 15]. The following key criteria that influence serendipitous information discovery:

Arbitrary navigation. In order to browse diverse information, an information discovery tool needs to provide a way to arbitrarily navigate among resources to support purely serendipitous information discovery [10].

Search-based navigation. Search-based navigation often serves as an entry point of information seeking [16]. In case of serendipitous discovery, since the information need is not well articulated, the search engine should retrieve diverse resources related to a topic.

Category-guided navigation. Similar to search-based navigation, category-guided navigation should provide a way to narrow the results to those related to one topic. In addition, categories can help the user formulate an information need by suggesting topics [16].

Integration. To users with ambiguous information needs, one information portal might not provide access to all information of interest. If an information discovery application gives access to resources from various sources, the user should be able to navigate to those sources.

Visual link preview. Abrams et al. [1] identi-

fied link representation as one of the problems with traditional bookmarking. Analogous with browsing through a bookmark manager, identifying relevant information when browsing through links to diverse resources can be a challenging task. A visual preview should make it easier to evaluate the relevance of resources.

Spatial arrangement. Similarly to link representation, spatial visualization of numerous links is another problem that occurs when browsing through diverse content [1]. Therefore, a semantic to the spatial arrangement of resources is of major importance.

3.1.2 Fact discovery

Fact discovery refers to information discovery resulting from looking 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 [14, 17]. Below is a list of factors that influence fact discovery:

Search-based navigation. With fact discovery, an information need is known [14, ?]. Therefore, the goal of a search-based navigation for fact discovery is to directly navigate to the resource of interest, as opposed to retrieving diverse information, as in serendipitous discovery.

Category-guided navigation. Category-guided navigation is used to direct the user to relevant resources [16]. In case of fact discovery, such navigation should narrow the results to a specific type of resources so that further fact discovery is bounded by that type.

Uniform representation. Uniform representation is a method of displaying diverse resources in a common way, with each resource having the same set of components [12]. Such representation assures that each resource has the same set of facts associated with it, and therefore, the user can afford to have expectations about information that can be found when looking for a specific fact.

Integration. Similar to serendipitous discovery, if an information discovery application gives access to resources from various sources, the user should be able to navigate to those sources since they may contain the facts of interest.

Visual link preview. If an application provided links to resources, a visual preview makes it easier to recognize the relevance of the resource [1].

Spatial arrangement. Similar to serendipitous

information discovery, apatial arrangement of resources is important [1], as a poor semantic to the arrangement can make it difficult to visually navigate to the fact of interest.

3.1.3 Rediscovery

Rediscovery refers to information discovery resulting from revisiting previously discovered resources [23]. The following is a list of factors that enable rediscovery:

History-based rediscovery. A Web application needs to automatically back up browsing history in order to enable history-based rediscovery [23].

Bookmark-based rediscovery. Bookmark-based revisitation is one of the most common ways of information rediscovery [1]. The majority of Web browsers are equipped with bookmarking features. However, some modern Web applications provide integrated mechanisms for bookmarking and bookmark-based information rediscovery.

Search-based rediscovery. Search-based rediscovery is not always a reliable way of refinding information [6]. In information portals that provide access to fairly ambiguous information and that have information regularly repopulated and updated, the search feature is usually designed around retrieving information related to some topic but not very specific. In order to revisit a resource, search must provide consistent results.

3.1.4 Channel-based discovery

Channel-based discovery can incorporate two different information seeking tasks, monitoring and awareness. It occurs when information is suggested to users based on the content that they are subscribed to. If users can actively look for updates, then an application affords monitoring [20]. If users can receive notifications about updates, then an application facilitates awareness [2, 3].

Site subscription. Subscriptions to updates from a site help users follow the news [13]. In order to support subscription-based discovery, an application must provide a subscription mechanism.

User subscription. Similar to site subscriptions, user subscriptions help networking and following users' activities [18]. Such subscriptions help to further filter new content delivered to the user.

Notifications. Notification mechanisms enable user awareness about the new content on the subscribed channel [18].

Subscription news stream. Displaying the

news stream within application further promotes awareness and can serve as a monitoring mechanism.

Content news stream. Displaying the news stream within application further promotes awareness and can serve as a monitoring mechanism.

3.2 Information Curation

By definition, digital information curation is the notion of managing, preserving, and adding value to collections of information [4, 24]. Thus, the curation category consists of information management, preservation, information enhancement, and sharing.

3.2.1 Management

Information management is one of the key elements of information curation [4, 24]. In the context of Web information management, the following factors plays the major role:

List-based categorization. Resource categorization helps establish relationships between various resources [4, 24]. Allowing to sort information into custom categories can aid rediscovery, discovery in a socially curated space, as well add more value to resources.

Tag-based categorization. Similar to list-based categorization, tagging aids rediscovery, adds value to resources, and aids discovery, especially in a socially curated spaces [11].

3.2.2 Information Preservation

Information preservation is a common Web task that is usually performed with intent to revisit information [1, 24]. However, in the case of opportunistic Web use, information gathering is sometimes performed with just the goal of collecting information rather than revisiting it in the future [17]. Bookmarking is a traditional way of preserving information; many Web applications provide diverse bookmarking mechanisms.

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

External preservation of internal resources. External preservation means bookmarking resources so that they become through other bookmarking systems. An application must facilitate integration with other applications in order to enable external preservation [1].

3.2.3 Augmentation

One of the most important elements of digital curation is augmentation, adding value to information [4, 24]. It is often performed within social bookmarking systems.

Evaluation. 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.

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

3.2.4 Sharing

Sharing information is key to empowering social information curation [4]. Therefore, the main components that facilitate sharing are adding resources, and external and internal information sharing.

Adding resources. Adding resources does not only facilitate global Web information curation, but it also scales the information available through the system providing more opportunities for information discovery.

External sharing. Sharing resources through different media supports channel-based information discovery within the media channels.

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

The following section describes the methodology for evaluating the conceptual framework and understanding how to address the elements of the framework when designing real world applications.

4 Methodology

The study presented in this paper is motivated by two primary goals. The first goal is to validate the conceptual framework of factors that enable digital information discovery and curation. The second goal of the study is to gain perspectives on how different elements of our framework support

real-world applications. Therefore,

RQ1: How do existing Web applications support information discovery?

RQ2: How do existing Web applications support information curation?

We used Yin's strategies for designing a case study [25] for guidance. The motivation behind choosing a case study over other methods of qualitative research was based on our choice of research questions (which have an explanatory nature), lack of control over existing applications and their development, and having to focus on contemporary use of real-life Web applications. According to Yin [25], a case study would be an optimal research strategy given the above characteristics.

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

To increase external validity of our study, we chose the cases based on replication logic [25]. Using replication logic in a case study design means carefully selecting each case so that it either predicts analogous results or predicts contrasting results but for anticipated reasons. Therefore, we used our conceptual framework (see Section 3) to predict if an application supported each of the information discovery and curation tasks based on the features that the application provided.

Consequently, our case study was an iterative process of selecting cases, analyzing each case, and determining whether or not it can be described and evaluated using our framework. If we found a feature that cannot be described, we adapted the framework according to the finding. We repeated the process of case selection and evaluation until the framework was usable for all cases. We then grouped the elements of the framework into categories, and recorded corresponding questions to ask in order to evaluate application. The resulting framework is presented in Table 1. Limitations of

our study are outlined in Section 6.

5 Trends in Digital Information Discovery and Curation

Through examining twenty different Web applications whose main function is to help people discover information (see Table 2), we were able to build and validate the conceptual framework of factors and questions that facilitate digital information discovery and curation. In this section, we further expand our findings and answer the research questions, RQ1 and RQ2.

RQ1: How do existing Web applications support information discovery?

Design of information discovery applications heavily depends on the motivations behind information seeking for each individual tool. Depending on the purposes of information discovery, an application may or may not support each of the factors outlined in our conceptual framework (see Table 1). Hence, there exists a large spectrum of information discovery applications.

When people try to gain knowledge or find inspiration related to some interest or hobby, they often have underdefined or hidden information needs [17]. Applications that supply diverse resources with an intent to satisfy underdefined information needs or that try to help users formulate their interest and uncover hidden information needs, tend to have strong support for **serendipitous discovery**. Serendipitous information discovery often requires support for **arbitrary navigation** so that users can freely move from one resource to another. Some applications utilize **category-guided navigation** in order to guide the user. For example, 500px, a professional photography hosting platform, provides categories of photographs. Categories can direct user's interest towards information needs that otherwise would be harder to formulate. **Search-based navigation** is another common navigation mechanism among tools that support serendipitous information discovery. In this case, the purpose of the search feature is usually to retrieve numerous resources related to some topic, as can be seen in applications such as Flickr, Tumblr, and We Heart It.

Applications that facilitate serendipitous infor-

mation discovery often employ elaborate resource representation techniques. Social-bookmarking systems, such as Scoop.it! and StumbleUpon, had indeed tried solving the two link representation problems indicated by Abrams et al.[1], visualization and representation. Most resources displayed within these applications have **visual previews** in a form of an image from the page where the resource is coming from, page title, and description. In addition, resources are **arranged** in a 'pinboard' fashion, similarly to Pinterest, visual information discovery tool.

When users have specific information need, which can occur, for example, as a result of lacking information to complete a real-world task, they often engage in fact finding activities online [14, 15, 22]. Fact discovery is a common task within applications that are tailored to support one type of resource. For example, it is easy to find directions through Google Maps because they help retrieve location data for places. Applications that support multiple types but need to provide factual information often use **category-guided navigation** [16] to filter resources into a particular type. For example, Expedia lets the user pick between hotels and flights.

Since when users browse within applications that are tailored towards fact discovery, they have well-defined information needs, the main goal of the search feature in such applications is to navigate to desired fact as directly as possible instead of returning related resources. In addition, to make it easier for users to find what they are looking for, many applications ensure that they display the same set of properties about the resource in a **uniform way**. For example, Yelp displays rating, address, contact number, and other properties for every business they list.

Depending on the nature of resources provided by an application, **search-based rediscovery** can be a challenging task. In information discovery applications that provide access to fairly specific information, such as Wikipedia and IMDb, search can usually directly lead to a specific resource. However, within Web applications such as We Heart It or Pinterest, that support image discovery, search-based revisitation is often unreliable. To solve this problem, a lot of the information discovery tools support various forms of **bookmarking**. **History-based rediscovery** is a rare feature, presumably because for many applications, the his-

Table 2: Web-based Information Discovery Tools

| Application | Description | Summary of Findings |
|-----------------|--|--|
| Pinterest | Visual discovery tool. Available at www.pinterest.com | The application supports serendipitous browsing, bookmark-based rediscovery, channel-based information discovery, and information curation. It lacks support for search- and history-based rediscovery and fact finding. |
| Delicious | Social bookmarking service. Available at delicious.com | The application supports channel-based discovery, bookmark-based rediscovery, and supports social curation. It lacks support for visual link preview and list-based categorization. |
| Tumblr | Microblogging platform. Available at www.tumblr.com | The application supports serendipitous browsing, bookmark-based rediscovery, channel-based information discovery. It lacks support for fact finding and list-based categorization. |
| StumbleUpon | Web page discovery tool. Available at www.stumbleupon.com | The application supports serendipitous browsing, bookmark- and history-based information rediscovery, channel-based information discovery, and information curation. It lacks support for fact finding. |
| Wikipedia | Free content Internet encyclopedia. Available at en.wikipedia.org | The application supports serendipitous discovery, fact finding, search-based rediscovery. It lacks support for history-based and bookmark-based rediscovery, personal preservation and resource evaluation. |
| Google Maps | Web mapping service. Available at www.google.ca/maps | The application supports fact finding and rediscovery. It lacks support for serendipitous browsing and curation mechanisms, except for personal information preservation. |
| Rotten Tomatoes | Movie and TV database. Available at www.rottentomatoes.com | The application supports fact discovery, serendipitous browsing, and search-based rediscovery. It lacks support for history-based and bookmark-based rediscovery, information preservation, and management. |
| 500px | Photography site. Available at 500px.com | The application supports serendipitous browsing, channel-based discovery, and social curation. It lacks support for fact discovery and list-based categorization. |
| BucketList | Goal tracking and discovery service. Available at bucketlist.org | The application supports serendipitous discovery, bookmark-based rediscovery, and channel-based discovery. It lacks support for fact discovery, search- and history-based rediscovery. |
| We Heart It | Visual discovery tool. Available at weheartit.com | The application supports serendipitous browsing, bookmark-based rediscovery, channel-based information discovery, and information curation. It lacks support for fact finding. |
| Scoop.it! | Online publishing platform. Available at www.scoop.it | The application supports serendipitous browsing, bookmark-based information rediscovery, channel-based information discovery, and information curation. It lacks support for fact finding. |
| Google Images | Image discovery service. Available at images.google.com | The application supports serendipitous browsing. It lacks support for rediscovery, channel-based discovery, fact finding, or information curation. |
| Vimeo | Video sharing Website. Available at vimeo.com | The application supports serendipitous discovery, bookmark-based rediscovery, and channel-based discovery, and information curation. It lacks support for fact discovery and list-based categorization. |
| LifeHacker | Daily weblog. Available at lifehacker.com | The application supports serendipitous discovery, but lacks support for channel-based discovery and information curation. |
| YouTube | Video hosting platform. Available at www.youtube.com | The application allows for serendipitous discovery, channel-based discovery, history- and bookmark-based revisitation, and information curation. It lacks support for internal sharing. |
| Yelp | Business review site. Available at www.yelp.ca | The application supports fact finding serendipitous browsing, search-based rediscovery, but not channel-based discovery. It supports certain aspects of information curation, such as evaluation and annotation. |
| IMDb | Movie database. Available at www.imdb.com | The application supports fact discovery, serendipitous discovery, and rediscovery. It lacks support for channel-based discovery. |
| Trip Adviser | Travel site. Available at www.tripadvisor.ca | The application supports serendipitous discovery, fact finding, and personal information curation. It lacks support for history-based rediscovery. |
| Urban Spoon | Online bar and restaurant guide. Available at www.urbanspoon.com | The application supports serendipitous browsing, fact finding, evaluation and annotations. It lacks support for channel-based discovery. |
| Thesaurus | Online thesaurus. Available at thesaurus.com | The application supports serendipitous browsing and fact discovery. It lacks support for information curation. |

tory would be too large to use. However, **history-based rediscovery** mechanism can still be found in some Web applications, such as Google Maps.

Subscription-based information discovery is usually enabled at sites that have regularly updated content, such as Pinterest, PolyVore, and various blogs. In some applications, the content is updated and curated by users, and therefore, users can subscribe to other users. In others, the content is updated by alternative content provider, for example blog moderators, so that users can subscribe to the site updates. Different application provide **notification mechanisms** that sometimes take place within applications, and sometimes they have a form of informative emails and smart phone notifications. Alternatively, news and updates are displayed within applications without use of notifications.

Information discovery tools can have different implementations depending on the purpose of discovery. Using information discovery factors in our framework (see Table 1), we described and evaluated currently existing tools. Similarly, the framework can be used for identifying gaps in information discovery support and developing new technologies.

RQ2: How do existing Web applications support information curation?

Information discovery applications vary from being fully socially curated and populated by users to those that lack any curation mechanisms.

Information categorization mechanisms are prevalent in applications that have a lot of information that is hard to categorize automatically or that can mean something different for each user. For example Pinterest supports categorization into customary-named boards. PolyVore allows categorization into collections. Custom information categorization helps information exploitation and rediscovery.

To preserve information for future exploitation, some Web applications use generic lists, such as Google Maps. Other application use pinboards and other mechanisms that interweave categorization with information preservation.

Many Web applications allow users to add value to the resources they curate. Tags, descriptions, and comments are some types of annotations that are commonly seen within application supporting

information discovery. In addition, many applications allow users to evaluate resources by rating them or recording other forms of likes and dislikes.

Socially curated information discovery tools usually have two features in common. First, they let users add content to the common information pool. The content can be either created by users themselves or taken from some other source online, or both. For example, YouTube allows for uploading your own videos, whereas Pinterest permits adding images from other sites in addition to users' personal images. Second, they allow for resource resharing within the application. Thus, this element provides support for subscription-based discovery.

Information curation is a common activity within many information discovery applications. By asking questions about application design in regards to information curation as in Table 1 of the conceptual framework, designers can find ways of adding value to information and enable information exploitation overtime.

6 Limitations

The case study we conducted has a number of limitations, including lack of reliable data, measures used to collect and interpret the data, and lack of prior research studies on the topic.

Lack of documentation, literature, and formal descriptions of available features for some applications introduces a threat to construct validity of the study. Therefore, the use of some features within information discovery applications was recorded based on the researchers' interpretations.

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

7 Future Work and Conclusion

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

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

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

References

- [1] Abrams, David, Ron Baecker, and Mark Chignell. "Information archiving with bookmarks: personal Web space construction and organization." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 1998.
- [2] Bates, Marcia J. "Toward an integrated model of information seeking and searching." *The New Review of Information Behaviour Research* 3 (2002): 1-15. APA
- [3] Bates, Marcia J. "An exploratory paradigm for online information retrieval." *Intelligent Information Systems for the Information Society*. Amsterdam: North-Holland (1986): 91-99.
- [4] Beagrie, Neil. "Digital curation for science, digital libraries, and individuals." *International Journal of Digital Curation* 1.1 (2008): 3-16.
- [5] Choo, C. W., Detlor, B., and Tunbull, D. (2000). Information seeking on the web: An integrated model of browsing and searching. *FirstMonday*, 5(2). Available from http://firstmonday.org/issues/issue5_2/choo/index.html.
- [6] Cockburn, Andy, et al. Improving Web page revisitation: Analysis, design, and evaluation. Department of Computer Science & Software Engineering, *University of Canterbury*, 2002.
- [7] Ellis, David. "A behavioural model for information retrieval system design." *Journal of information science* 15.4-5 (1989): 237-247.
- [8] Ellis, David, Deborah Cox, and Katherine Hall. "A comparison of the information seeking patterns of researchers in the physical and social sciences." *Journal of documentation* 49.4 (1993): 356-369.
- [9] Ellis, David, and Merete Haugan. "Modelling the information seeking patterns of engineers and research scientists in an industrial environment." *Journal of documentation* 53.4 (1997): 384-403.
- [10] Foster, Allen, and Nigel Ford. "Serendipity and information seeking: an empirical study." *Journal of Documentation* 59.3 (2003): 321-340.
- [11] Gruber, Thomas. "Ontology of folksonomy: A mash-up of apples and oranges." *emphInternational Journal on Semantic Web and Information Systems (IJSWIS)* 3.1 (2007): 1-11.
- [12] Herrera, Francisco, L. Martinez, and Pedro J. Sánchez. "Managing non-homogeneous information in group decision making." *European Journal of Operational Research* 166.1 (2005): 115-132.
- [13] Java, Akshay, et al. "Feeds That Matter: A Study of Bloglines Subscriptions." *ICWSM*. 2007.
- [14] Kellar, Melanie, Carolyn Watters, and Michael Shepherd. "A Goal-based Classification of Web Information Tasks." *Proceedings of the American Society for Information Science and Technology* 43.1 (2006): 1-22.

- [15] Kellar, Melanie, Carolyn Watters, and Michael Shepherd. "A field study characterizing Web-based information-seeking tasks." *Journal of the American Society for Information Science and Technology* 58.7 (2007): 999-1018.
- [16] Levene, Mark. *An introduction to search engines and web navigation*. John Wiley & Sons, 2011.
- [17] Lindley, Siân E., et al. "It's simply integral to what I do: enquiries into how the web is weaved into everyday life." *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012.
- [18] Millen, David, Jonathan Feinberg, and Bernard Kerr. "Social bookmarking in the enterprise." *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM 2006.
- [19] Mishne, Gilad, and Maarten De Rijke. "A study of blog search." *Advances in information retrieval*. Springer Berlin Heidelberg, 2006. 289-301.
- [20] Morrison, Julie B., Peter Pirolli, and Stuart K. Card. "A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions." *CHI'01 extended abstracts on Human factors in computing systems*. ACM, 2001.
- [21] Proper, Henderik Alex, and P. D. Bruza. "What is information discovery about?" *Journal of the American Society for Information Science* 50.9 (1999): 737-750.
- [22] Sellen, Abigail J., Rachel Murphy, and Kate L. Shaw. "How knowledge workers use the web." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2002.
- [23] Tauscher, Linda, and Saul Greenberg. "How people revisit web pages: Empirical findings and implications for the design of history systems." *International Journal of Human-Computer Studies* 47.1 (1997): 97-137.
- [24] Whittaker, Steve. "Personal information management: from information consumption to curation." *Annual review of information science and technology* 45.1 (2011): 1-62.
- [25] Yin, R. K. 2009. *Case study research*, 4th, Thousand Oaks, CA: Sage.