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 [27]. Other times, users lack well-articulated information needs, so they engage in opportunistic browsing [21]. 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 management and curation activities when designing tools that support information discovery.

To close their knowledge gaps, people turn to various Web technologies ranging from specialized search tools to visual discovery applications. Several studies have been directed at exploring high-level Web tasks, including information seeking tasks [17, 18, 26, 29], deriving models of information seeking behaviors [5, 8, 9, 10, 3, 2], and looking at methods of information curation [4, 31]. 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) using the framework to establish a set of questions that can be used when evaluating and designing new applications, and (3) evaluating the framework by using it to study and describe currently existing Web applications, which in turn helped us refine the framework of factors and questions. In summary, the framework addresses our research goal which is to gain an understanding of how existing tools support digital information curation and discovery.

The remainder of this paper is organized as follows. Section 2 highlights some of the studies and technologies related to information seeking and curation tasks. Section 3 outlines a conceptual framework of factors and provides 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 the limitations of the study, followed by future work and conclusions, 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, collaboration in information seeking, and modes of Web use.

2.1 Web Tasks

Kellar et al. [17] 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, 31]. In their later work, Kellar et al. [18] added communication and maintenance as additional Web tasks.

Similarly to Kellar, Sellen et al. [29] identified six tasks that are commonly performed by Web users: browsing, finding, housekeeping, information gathering, communicating, and transacting. Therefore, Kellar et al. and Sellen et al. both identified browsing, fact finding, and information gathering as information-related tasks that users perform online.

People often engage in information seeking activities to close some knowledge gap that occurred as a result of not having enough information to perform task [27]. 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. [26] 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 mechanisms.

Morrison also draws distinction between finding or looking up information and exploratory search. Whereas information lookup involves tasks such as fact retrieval, navigation, and verification, exploration is more cognitively demanding and involves learning and investigation [22]. Learning and investigation can be performed over multiple iterations, and they can involve learning though various media, "social searching", and serendipitous browsing performed with goals of knowledge acquisition, socialization, forecasting, and planning.

2.2 Information Behavior Models

A number of researchers have proposed models of information-seeking and information behavior. Wilson [32] summarized some of the key work [8, 7, 19, 33, 34] on information behavior along with proposing a new model. According to Wilson's original model of information behavior [34], information-seeking behavior results form the user

trying to satisfy her perceived information need. Consequently, the user makes demands on information systems. Success or failure of such demands dictates whether the process is repeated or, if the information need is satisfied, used or communicated with other people. In addition, Wilson defined possible barriers that can impede information seeking behaviors, as well as context that influences formation of the information need. These underlying ideas remained in the revision of Wilson's model [33]. Finally, Wilson proposed a "problem solving model" of information-seeking behavior. The model reflects on the idea that people engage in information seeking and searching in order to resolve some uncertainty that stands in a way of solving, defining, or identifying a problem.

Ellis et al. [8, 10, 9], proposed a model of information seeking characterized by six different patterns: starting, chaining, browsing, extracting, monitoring, and differentiating. Subsequently, 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, extraction can occur when the user systematically searches site to extract information of interest. Ellis' model is also complemented Kuhlthau's [19] work, who corresponded stages of information seeking with feelings, thoughts, actions, as well as anticipated information tasks.

Information retrieval behaviors are further studied by Saracevic [28] and Ingwersen [16]

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 de-

fined and the search is directed. Finally, monitoring and being aware are passive modes of information seeking although monitoring is directed and being aware is undirected.

2.3 Digital Curation

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 [31] believes that in terms of Web use, a significant shift is happening from information consumption to information curation, which means that people no longer just use the web to find and consume the information that they are interested in, but they also try to save and manage that information so that it can be reaccessed and exploited later.

2.4 Collaboration and Information Seeking

By surveying 204 Web users, Morris found that people often have desire to or do collaborate on information seeking tasks [25]. To collaborate on information seeking, people often use instant messaging, email, and create documents and Webpages to share information. Occationally collaborative information seeking occurs when collaborators work side by side, and they 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 [13].

2.5 Modes of Web Use

Categorizing web usage into information seeking, digital curation, and other web tasks does not necessarily give full insight about how information-related tasks are performed. Lindley et al. [21] 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.

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

factors and questions that facilitate information discovery and curation.

3 A conceptual framework for information discovery and curation on the Web

Although information discovery and curation tasks are commonly performed Web tasks today, as we mentioned above, there is a 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 an extensive literature review. From this review we derived design factors. Through a careful analysis of twenty information discovery applications, we iteratively expanded the framework, added concepts, and established relations between those concepts. The framework can be expanded further; however, we selected the most popular information discovery applications in use today and considered the full range of features in those tools (both by referring to the literature and documentation on those tools, as well as exploring the features). For presentation purposes, we present the final version of the framework we developed as the framework was refined iteratively as explored the literature and available tools.

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. The discovery category consists of serendipitous discovery, fact discovery, rediscovery, and channel-based discovery.

---

[1]Most applications were selected based on Website popularity rank provided by Alexa, commercial Web traffic data provider, available at www.alexa.com

Table 1: Conceptual Framework.

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

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 [17, 18]. The following are 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 [11].

Search-based navigation. Search-based navigation often serves as an entry point of information seeking [20]. 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 [20].

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

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 infor-

mation need and is easier to perform within systems that provide access to homogeneous types of information [17, 21]. Below is a list of factors that influence fact discovery:

Search-based navigation. With fact discovery, an information need is known [17, 18]. 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 [20]. In the case of fact discovery, such navigation should narrow the results to a specific type of resource so that further fact discovery is bounded by that type.

Uniform representation. Uniform representation is a method of displaying diverse resources in a common way, with each resource having the same set of components [14]. Such a representation assures that each resource has the same set of facts associated with it, and therefore, the user can afford to have expectations about information that can be found when looking for a specific fact.

Integration. Similarly 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. Similarly to serendipitous information discovery, spatial 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 [30]. The following is a list of factors that enable rediscovery:

History-based rediscovery. A Web application needs to automatically record browsing history in order to enable history-based rediscovery [30].

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 [26]. 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 [15]. 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 [23]. 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 [23].

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

Content news stream. Similarly to displaying the subscription news stream, displaying the content news stream 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, 31]. 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, 31]. In the context of Web information management, the following factors play a major role:

List-based categorization. Resource categorization helps establish relationships between various resources [4, 31]. Allowing to sort information into custom categories can aid rediscovery, discovery in a socially curated space, as well as add more value to resources.

Tag-based categorization. Similarly to list-based categorization, tagging aids rediscovery, adds value to resources, and aids discovery, especially in a socially curated space [12].

3.2.2 Information Preservation

Information preservation is a common Web task that is usually performed with the intent to revisit information [1, 31]. 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 [21]. Bookmarking is a traditional way of preserving information; many Web applications provide diverse bookmarking mechanisms.

Internal preservation of internal resources. 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. 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, 31]. 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 not only facilitates 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 (and consequently refining) the conceptual framework and understanding how to address the elements of the framework when designing real world applications.

4 Evaluating and Refining the Conceptual Framework

The study presented in this paper is motivated by two primary goals. The first goal is to evaluate 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, our questions are:

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 [35] 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 [35], 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 [35]. 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 methodology was an iterative process of selecting cases, analyzing each case, and determining whether or not it can be described and evaluated using our famework. 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 that and other applications. 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 that were selected among the most popular information discovery tools (see Table 2), we were able to build and validate the conceptual framework of factors and questions that facilitate

Table 2: Web-based information discovery and curation 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 infomration discovery, and information curation. It lacks support for fact finding.
Wikipedia	Free content Internet encyclopedia. Availabe 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. I supports certain aspects of information curation, sush 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.

digital information discovery and curation. In this section, we further expand our findings and answer research questions, RQ1 and RQ2.

RQ1: How do existing Web applications support information discovery?

Current Web-based information discovery applications can support one or more types of information discovery: **serendipitous discovery**, **fact discovery**, **rediscovery**, or **channel-based discovery**.

The main challenge with serendipitous discovery is facilitating browsing with upderdefined or absent information need. Many applications , such as Tumblr and StumbleUpon, support **arbitrary navigation** to allow for opportunistic jumping from one resource to another. Alternatively, applications use **category-guided navigation**. For example, Google Images, with every search query, suggests related categories of images to help users with defining an information need. **Search-based navigation** is also common with serendipitous browsing, where a search query returns many diverse resources, many of which can be valuable to the user. For instance, searching for a location within Pinterest returns numerous images of that location that link to (or **integrate** with) other resources, blogs, and Web pages, whereas searching for the same place on Google Maps usually returns a small set of possible locations with precise addresses of those places.

Applications that facilitate serendipitous information discovery often employ elaborate resource representation techniques. Many social-bookmarking systems, such as Scoop.it! and StumbleUpon, support **visual previews** of bookmarked pages. Delicious is a social-bookmarking application that lacks this type of link representation support, which makes it harder to determine if the link will lead to a relevant resource. In addition, information discovery applications that support serendipitous discovery often have a special way of **spatially arranging** resources to make it easier to browse through large amounts of information. For example, similarly to Pinterest, many tools use a 'pinboard' layout of resources.

The main challenge for designing applications for fact discovery is to facilitate finding of a specific piece information, leaving little room for uncertainly in the search results. Therefore, instead of suggesting related search topics, **category-guided**

navigation within fact discovery supporting applications is often designed to suggest types of resources to narrow search results to a specific type. For example, TripAdvisor lets the user choose among flights, hotels, vacation rentals, restaurants, and destinations. On contrary to **search-based navigation** for serendipitous browsing, for fact discovery, the search engine returns a small set of results, among which only one is usually of interest to the user. **Integration** for fact finding is important when it gives an opportunity to display specific information about resources that otherwise would not be accessible. For example, Google Maps displays business ratings as a result of its integration with Google+.

Having uniform resources helps with fact discovery since they can be **represented in a uniform way**. For example, Yelp displays rating, price range, and address for all restaurants, so not only it is easy to find specific information, but also the user can have expectations about the content of resources within the application. On the contrary, searching Tumblr for a restaurant will return a chaotic collection of information about the place. Applications that support fact discovery often use **visual link preview**, similarly to applications that support serendipitous browsing. However, the motivation behind having a link preview for fact finding is to make is possible to identify if the resource is indeed what the user is looking for. For example, searching for an actor in IMDb will return a list of actors with their photographs, so that the user can pick the one she is interested in. Finally, fact discovery applications often displays resources with a semantic to their **spatial arrangement**.

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 specific information, such as Wikipedia and Rotten Tomatoes, search can usually directly lead to a specific resource. However, within Web applications such as We Heart It or Pinterest, search-based rediscovery is often unreliable. To solve this problem, a lot of the information discovery tools support various forms of **bookmarking**. **History-based rediscovery** is less common; however, it can still be found in some Web applications, such as Google Maps.

Channel-based information discovery is usually enabled at sites that have regularly updated content, such as Pinterest and YoutTube. In some ap-

plications, 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 providers, for example blog moderators, so that users can **subscribe to site updates**. For example, Rotten Tomatoes allows subscriptions to newsletters; however, it does not allow subscriptions to movie critics. Different applications provide **notification mechanisms** that sometimes take place within applications, and sometimes they have informative emails and smart phone notifications. Alternatively, **user or content updates** are displayed within applications without the use of notifications.

Information discovery tools can have different implementations depending on the purpose of discovery. Using information discovery factors of the 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 **tag-** and **list-based categorizations**, where lists are represented as 'pinboards'. Tumblr, on the other hand, only supports tag-based categorization. In addition, Pinterest allows for private information categorization.

To **preserve information** for future exploitation, many applications support various forms of bookmarking. On We Heart It, the users can preserve **internal information** using **internal collections** and they can add information from **external** Websites. However, there are no means for bookmarking **internal content** using other bookmarking systems.

Many Web applications allow users to add value to the resources they curate. Descriptions, and comments are some types of **annotations** that are commonly seen within application supporting information discovery. In addition, many applica-

tions allow users to **evaluate** resources by rating them or recording other forms of approval or disapproval. Some sites, such as Wikipedia, do not allow any evaluations.

Socially curated information discovery tools usually support **information sharing** capabilities. First, they let users **add content** to the application. The content can be either created by users themselves or taken from some other source online, or both. For example, YoutTube 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. Finally, information discovery application commonly allow for sharing information on popular networking sites **outside of the application**.

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 Framework Application

Google Maps Evaluation
   Place discovery application

7 Research and design implications

8 Limitations

The case study we conducted has a number of limitations, including lack of reliable data, imprecise 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.

9 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 evaluated and iteratively refined 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 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] Dervin, Brenda. "Sense-making theory and practice: an overview of user interests in knowledge seeking and use." Journal of knowledge management 2.2 (1998): 36-46.

[8] Ellis, David. "A behavioural model for information retrieval system design." *Journal of information science* 15.4-5 (1989): 237-247.

[9] 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.

[10] 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.

[11] Foster, Allen, and Nigel Ford. "Serendipity and information seeking: an empirical study." *Journal of Documentation* 59.3 (2003): 321-340.

[12] 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.

[13] Golder, Scott A., and Bernardo A. Huberman. "Usage patterns of collaborative tagging systems." Journal of information science 32.2 (2006): 198-208.

[14] Herrera, Francisco, L. Martınez, and Pedro J. Sánchez. "Managing non-homogeneous information in group decision making." *European Journal of Operational Research* 166.1 (2005): 115-132.

[15] Java, Akshay, et al. "Feeds That Matter: A Study of Bloglines Subscriptions." *ICWSM.* 2007.

[16] Ingwersen, Peter. "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory." Journal of documentation 52.1 (1996): 3-50.

[17] 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.

[18] 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.

[19] Kuhlthau, Carol C. "Inside the search process: Information seeking from the user's perspective." JASIS 42.5 (1991): 361-371.

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

[21] 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.

[22] Marchionini, Gary. "Exploratory search: from finding to understanding."Communications of the ACM 49.4 (2006): 41-46.

[23] 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.

[24] Mishne, Gilad, and Maarten De Rijke. "A study of blog search." *Advances in information retrieval.* Springer Berlin Heidelberg, 2006. 289-301.

[25] Morris, Meredith Ringel. "A survey of collaborative web search practices." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2008.

[26] 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.

[27] 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.

[28] Saracevic, T. Modeling interaction in information retrieval (IR): a review and proposal. In: Hardin, S., ed. 59th Annual Meeting of the American Society for Information Science. Silver Spring, MD: American Society for Information Science, 1996, 3–9.

[29] 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.

[30] 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.

[31] Whittaker, Steve. "Personal information management: from information consumption to curation." *Annual review of information science and technology* 45.1 (2011): 1-62.

[32] Wilson, Tom D. "Models in information behaviour research." Journal of documentation 55.3 (1999): 249-270.

[33] Wilson, Thomas Daniel. "Information behaviour: an interdisciplinary perspective." Information processing & management 33.4 (1997): 551-572.

[34] Wilson, Tom D. "On user studies and information needs." Journal of documentation 37.1 (1981): 3-15.

[35] Yin, R. K. 2009. *Case study research*, 4th, Thousand Oaks, CA: Sage.