We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



17

Toward Personalized RSS Retrieval Service: The Effect of Using User's Context

Haesung Lee[1] and Joonhee Kwon[2]
Dept. Computer Science, Kyonggi University
Republic of Korea

1. Introduction

Web 2.0 is a new way to find, save, and share information generated on the web. The goal of Web 2.0 is to facilitate forming network and sharing knowledge between users. Useful techniques that allow users easily to edit, store and publish contents on the web have been created. After that, multiple personal contents platform or personal media platform such as blog have developed. Using those platforms, a private individual publishes and shares his or her useful knowledge with others on the web very easily (Tim Oreilly. 2007). In addition, personalization is very critical concept to Web 2.0. Multiple personal contents platform or media platforms such as a blog or YouTube [www.youtube] allow users easily to edit, store and publish contents on the web. Moreover, those techniques help people publish and share his or her useful knowledge with others on the web very easily (Maged N Kamel Boulos et al. 2006). Web rapidly enlarges with explosively and newly increased information formed in UCC (User Created Contents). More that 513 million people around the world now have access to this global information resource (Lyman et al. 2000). However, how do all of these people find the information they are most interested in from among all of those terabytes? One among the most time-consuming human activities in modern times is keeping up updated and useful information with a huge amount of continuously generated information. The need of quickly obtaining and effectively sharing that information should be more concerned.

In Web 2.0, it is very important that useful information should be found, acquired and consumed efficiently. It is no longer the user who searches the information he or she is looking for, but it is the information that reaches directly its consumers. Also, each user is provided with the different information what he or she interests individually and shares his or her own worth knowledge with others (Tim Oreilly. 2007).

After appearance of the World Wide Web, initial search engines possessing only 1000 Web page indices, the method of acquiring information and ranking mechanisms have repeated a numerous change to provide more satisfactory services to users (Sergey Brin et al. 1999). In spite of dazzling technical development, the expansion of the web continually has demanded new and epoch-making techniques relevant the mechanism through which a search engine efficiently acquires information generated on the web and effectively provides each user with useful information.

The RSS is considered as the next generation information delivery technique, which deliveries newly created information to users simultaneously and frequently without asking them to visit the site every time when new content is published. Practically, many services in Web 2.0 provide useful information in RSS feed format and a great many RSS feeds has risen in quite short term (Don E. Descy. 2004).

Although there are actually millions of feeds available, finding those that are appealing and relevant to user's interests is not always easy. By rising RSS feed, consequently, many RSS feed search services have appeared. Those RSS feed search service provide useful information such as the best favorite feed or blog (A. Gulli. 2005). However, those RSS feed search services are all tiptoeing around RSS search, but none have yet to launch a full-blown personalized RSS search service.

As the scope of the internet gets larger and larger, the need for personalization to bring it within our scope becomes more and more important. Nevertheless, typically existing RSS retrieval systems offer user RSS feeds as results using mainly keyword based matching mechanism or statistical approach such as favorite RSS feeds with many persons (Chris Sherman. 2005). Therefore, existing RSS feed retrieval services have limitations to answer efficiently to user's need to obtain useful and appropriate RSS feeds. In other words, because of this limitation, those services are not useful to many users.

To provide more useful and personalized RSS feed retrieval service to each user, it is important to incorporate the feed characteristics and the user's context. We describe this corporation in this article later.

Contexts are any information that can be used to characterize the situation of an entity such as a person, place or object (Jong-yi Hong et al. 2009). To provide each user or searcher more personalized and useful information, number of applications or web services takes the advantage of various contexts such as user's preference (Haveliwala et al. 2003). For realizing completely personalization in Web 2.0, the user's preference is very efficiently used to find and to acquire useful information among overexposure of information (Diane Kelly et al. 2003). That is, user's context has a large influence on the interest and intent of one particular user.

Considering those described above, we introduce personalized RSS feed retrieval system in which two factors are mainly considered, which are user's context and existing information retrieval techniques based on RSS characteristics.

In detail this article is organized as follows: Section 2 explores conceptual personalization in Web 2.0 and important challenges to implementation of personalized application in Web 2.0; Section 3 introduces RSS's potential ability to deriver personalized information to each user. Also, in this section, we explore the existing RSS services or applications and their limitations and introduce some solutions; Section 4 focuses on the use of user's context in personalized RSS retrieval through the case of previous study; Section 5 outlines our design of personalized RSS feed retrieval. For new and efficeint personalized RSS feed retrieval, we describe some algorithms for incorporating information retrieval techniques and characteristic of RSS. And we introduces each module of designed personalized RSS retrieval system in detail and an equtation for ranking RSS feeds based on user's preference. Section 7 looks at future possibilitof our proposed personalized RSS feed service in ubiquitous computing environment and tha way of our research in the future.

2. Personalization in Web 2.0

Web personalization can be described as any action that can customize the content or structure of a web site to the user's taste or preferences. Personalization has widely been utilized by e- commerce organizations to better serve their customers (Bamshad Mobasher et al. 2000).



Fig. 1. A person who has attributes

For personalization of Web 2.0, there are two important things: people, and content objects. People are represented by identities (Ryan Turner, 2007). Most of people have many attributes. Identities have information attached to them. In Fig. 1, there is a person. The person has attributes.

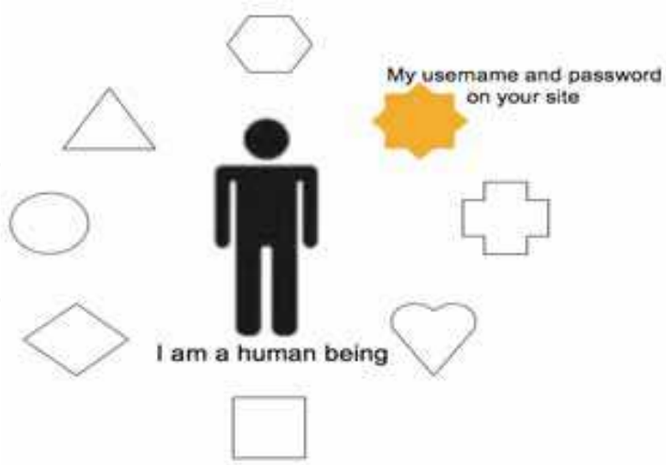


Fig. 2. The user who has some account attributes

When a person comes to your web site, you often know a few things about them, like their referring URL, their browser, etc. This isn't much. An early goal for conversion on Web 2.0 sites is account creation. In Figure 2, there is a person with a brand-new account. You know

their account information, but not much else. The person has many attributes, but you don't know what they are. A web site is made of contents objects, and the content objects have attributes. In Web 2.0, you know what some of the attributes are. Other attributes, like user-generated tags, might not yet be knowable. One of the most basic kinds of personalization is access control. Based on the person's account information, they can access some content objects and not others.

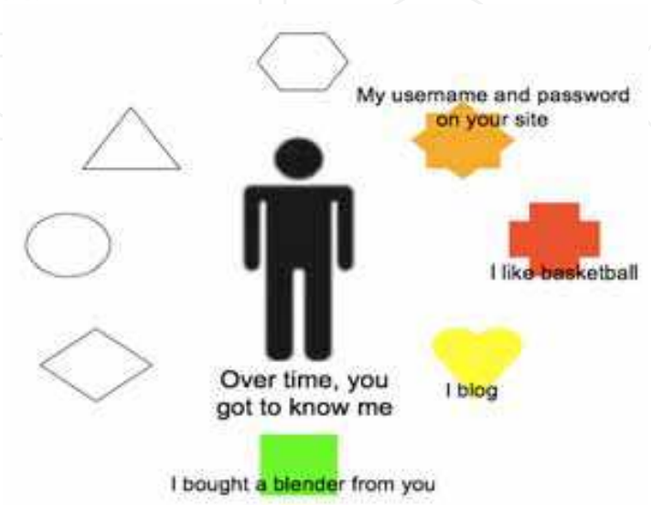


Fig. 3. A person frequently using your web site

Like the case saw in Figure 3, as a person uses your web site, you learn more about their attributes. Sometimes they tell you ("subscribe to newsletter") and other times you learn about people by knowing what they do ("I bought Scoble's book about blogging").
Different content objects have different attributes. Some, like content type (photo, video, downloadable document, product info) are structured, and others, like title, are unstructured. Some, like tags on a Flickr photo, are added by users.

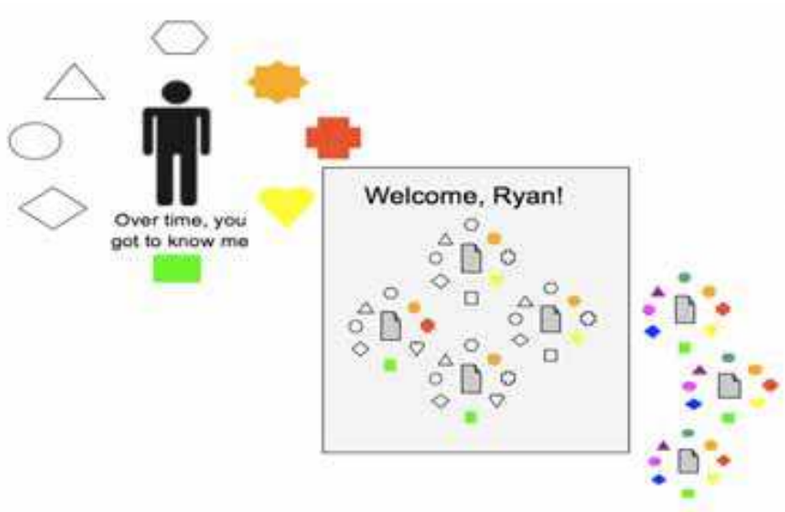


Fig. 4. Personalized information for an user

Like Figure 4, basic content personalization is a matter of matching up the attributes of a person with the attributes of a content object. You know the person likes basketball because they bought a Duke Basketball jersey, so you show them a content object that has to do with March Madness.
On the UK's Guardian newspaper site, writer Jemina Kiss suggested that web 3.0 will be about recommendation. "If web 2.0 could be summarized as interaction, web 3.0 must be about recommendation and personalization." she wrote (Jemima Kiss. 2008). We don't think whether this day is web 2.0 or web 3.0, but we know that the personalized information service is very important and needed service for information providers and information consumers in both two generations.



Fig. 5. Amazon's personalization service

Like figure 5, Amazon (http://www.amazon.com) has been the early adopter of personalization technology to recommend products to shoppers on its site, based upon their previous purchases (Linden, G. Et al. 2003). Amazon makes extensive use of collaborative filtering in its personalization technology.
Google provides a personalized service shown in figure 6, the gadget maker, which allows users to make their personalized web page by themselves reflecting their preference.



Fig. 6. Gadget Maker

Another example providing personalized service on the web is Pandora's music genome project offering users music streaming service based on their preference for music (Shane et al. 2007). This provides very useful service that allows users not to take the time of searching music related their preferences. With this service, user listens music related their preferences continually. The use of those contexts in user-centered service makes it possible to provide more useful and personalized information satisfying each user.

In Web 2.0, the concept of personalization is very important. As the scope of the Internet gets larger and larger, the need for personalization to bring it within our scope becomes more and more important (Udi Manber et al. 2000). But, any online examples where personalization was really well integrated with the user experience currently are very few. The best example we can come up with was Amazon's personalization engine. However this case is just scratching the surface of what's possible.

Most application or studies about personalized information service mainly focus on statistical or artificial intelligence techniques related with the semantic web. In this article, however, we focus on the use of user's contexts. Using user's context in information deliveries process is to satisfy user's need for information than other computational approaches.

There are currently many projects underway exploring methods of personalization for information seeking. The general starting point is to represent the interests of a user by means of one or more keyword profiles expressing aspects of the user's interest. In a very simple approach incoming information is compared with the profiles and is possessed to the user if there is a sufficiently high match (Stephen S. Yau et al. 2003). This case is very simple to standard information filtering techniques where the profiles represent topics of interest to the users. The most important issue in personalized information retrieval is how the interests of the user are captured. Various options are available; the user selects from a number of preset topics, the user enters sets of keywords which they believe represent topics that they are interested in, or the personalization system monitors the user's behavior and learns profiles from this.

It is sure that personalization in Web 2.0 and future Web is very important thing. So, we use user's context to provide personalized information to each user.

3. The next generation information delivery technique, RSS

The amount of the different information generated on the broad-scale web increases explosively. In Web 2.0, it is very important that useful information for each consumer should be found, acquired and consumed effectively. It is no longer the user that searches the information she is looking for, but it is the information she values that reaches directly its consumers. RSS (Really simple syndication) is a very useful technique in Web 2.0 providing users a efficient new way to share content on their website with other users and makes it available to offer them various information without asking them to visit the site every time when new content is published (S. Jeff Cold. 2006).

```
<rss version="2.0">
 − <channel>
    <title>Information Retrieval Gupf</title>
    <atom:link href="http://irgupf.com/feed/" rel="self" type="application/rss+xml"/>
    <link>http://irgupf.com</link>
   − <description>
      Information Retrieval Research, Issues, and Discussion
    </description>
    <pubDate>Thu, 16 Apr 2009 21:11:40 +0000</pubDate>
    <generator>http://wordpress.org/?v=2.7</generator>
    <language>en</language>
    <sy:updatePeriod>hourly</sy:updatePeriod>
    <sy:updateFrequency>1</sy:updateFrequency>
   − <item>
      <title> "Improving Findability" Falls Short of the Mark</title>
     − <link>
        http://irgupf.com/2009/04/16/improving-findability-falls-short-of-the-mark/
      </link>
     − <comments>
        http://irgupf.com/2009/04/16/improving-findability-falls-short-of-the-mark/#comments
      </comments>
      <pubDate>Thu, 16 Apr 2009 21:08:23 +0000</pubDate>
      <dc:creator>jeremy</dc:creator>
      <category>Exploratory Search</category>
      <category>Social Implications</category>
      <guid isPermaLink="false">http://irgupf.com/?p=502</guid>
     − <description>
        Via Tim O&#8217;Reilly on Twitter, I came across this article by Vanessa Fox on how go
        better informed and government to be more transparent. Fox writes: In an earlier post, I sa
      </description>
     − <content:encoded>
        <p>Via <a href="http://twitter.com/timoreilly">Tim O&#8217;Reilly on Twitter</a>, I came
```

Fig. 7. RSS Feed

RSS can literally be used with just about any kind of web-based content. RSS fundamentally is a simple specification that uses XML and a format for web-based content in a standard way. RSS is a form of XML. An example of RSS feed is shown in Figure 7. A big advantage of XML is that the data can be self-contained. Inside of XML, there's an opportunity to include a description of the methods required to use the data, along with the data itself. Because RSS uses XML to glean relevant information, RSS may well become the universal method people use to mine information from the Internet (Ronald J. Glotzbach et al. 2007). Figure 9 shows a RSS format with number of tags. Those tags are composed structurally and expressed in the tree as viewed at Figure 8.

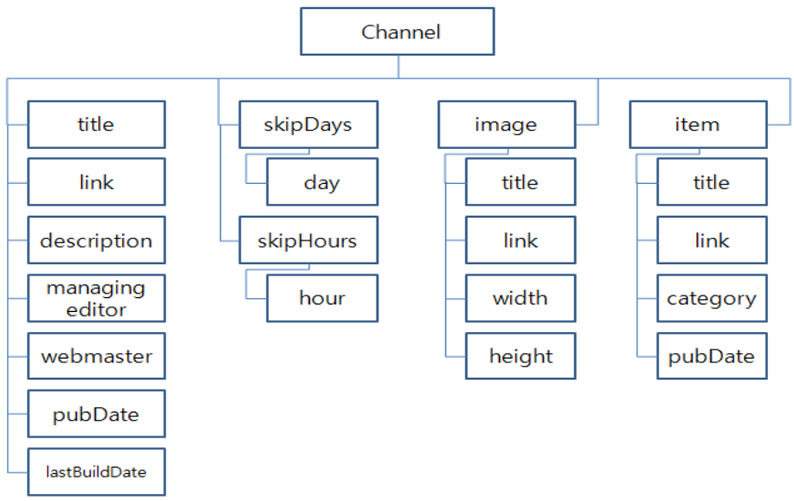


Fig. 8. RSS feed structure

RSS feeds are increasingly being used for other types of content. For example, you can get RSS feeds with weather forecasts, company news and financial information, package tracking and lots of others. RSS completely revolutionizes the paradigm according to which people collect information generated at number of information sources. A recent Pew Internet & American Life Project poll shows that only 9% of Americans understand what RSS feeds can do, and according to Forrester Research, only 2% of households with Internet access use RSS (Overhole et al. 2005)

Although there are actually millions of feeds available, finding those that are appealing and relevant to a user isn't always easy. While RSS is not a widely used internet tool today, a simple-to- use software client can make its use widespread in the near future. Email didn't go mainstream because people understood how data packets made their way around the Net. It happened when an email client made that technology invisible and easy to use. RSS will likely follow a similar path. Microsoft has already pledged to support all forms of RSS in its new browser: Internet Explorer 7 (Langley. 2005). RSS will be used increasingly as a way to cope with information overload.

News Aggregators (aggregators) or feed collectors download information from RSS feeds to which you subscribe based on key. Increasingly, publishers are providing RSS feeds for aggregation. This makes it easy to search their publications for research purposes. For example, the New York Times provides a free RSS feed (http://www.nytimes.com/services/xml/rss/). The range of current events and even topic-specific research that researchers can collect by subscribing to several RSS feeds is significant (Richardson. et al). Using a web-based aggregator, users can access their feeds from any place they have an internet connection. Once users create a free account on web-based aggregator, they can add the RSS URL of any information resource to begin the feed. But, this simple and efficient RSS aggregator (or reader) services have some limitations to satisfy user's underlying need for information. To provide each user RSS feeds automatically and periodically, these services request users to input RSS feed URLs themselves. However those RSS feed search engine annoy users with those requests. Because, it is very difficult to users to find which RSS channels publish useful RSS feeds satisfying them.

Although there are actually millions of feeds available, finding those that are appealing and relevant to a user isn't always easy. By rising necessity of RSS, consequently, RSS feed search engines mainly dealing with RSS feed appear.



Fig. 9. Blogpulse

Blogpulse (www.blogpulse.com) viewed at Figure 9 is known primarily as a tool for tracking trends and hot topics in the blogosphere. Also, it has the largest indexes of feed-based search engine. Blogpulse's search service provides phrase search, all the words or any of the words filters, and even allows a user to create her or his own free-form boolean queries.



Fig. 10. Bloglines

Figure 10 shows Bloglines(www.bloglines.com). It is both a feed search tool and a feed reader/aggregator. A drop-down menu next to its search form allows a user to search all of the indexed blogs or to add a feed to user's subscription.

They are all tiptoeing around RSS search, but none have yet to launch a full-blown RSS search service. While there are a number of smaller, specialized blog and feed search engines, their lack of resources and the problem of blog and feed spam mean their search results are often useless. So finding relevant feeds, at least for the time-being, often remains a hit-or-miss affair. Also, those RSS feed search engines cannot provide effectively and quickly useful results related each user's need for some information.

To provide more relevant information to the user in RSS retrieval service, it is important to incorporate the feed characteristics and the user's contextual information into the retrieval process. Because the previous retrieval services have not taken into account them, however, they provide no personalized information to each user.

Using RSS techniques, information delivery service to each user is very efficient way for implementing personalization in Web 2.0. We focus on this RSS potential. Also, we expand its techniques to integrate user's context providing each users with their underlying needing information. Later, we describe the integration of RSS and user's context in follow sections.

4. The potential ability of user's context

In this section, we describe user's contexts in ubiquitous computing area and introduce some applications to see potential ability of using user's context in our approach.

Contexts are any information that can be used to characterize the situation of an entity (Gregory D.Abowd et al. 1999). An entity is a person, place, or object that is considered

relevant to interaction between a user and an application. When humans talk with humans, they are able to use implicit situational information, or context, to improve more communicational functionalities. Unfortunately, this ability to convey ideas does not transfer well to humans interacting with computers. By improving the application's access to context, it is available to increase the richness of communication in human-computer interaction and make it possible to produce more useful computational services. There are various human factor related context such as knowledge of habits and emotional state of each user (Anind K. Dey. 2001).

That is an especially favorable case to use context to help select the information that is needed by a user. To provide each user or searcher more interesting and useful information, number of application or solution has been developed using many various contexts such as user's preference (Albrecht Schmidt et al. 1999). A user's context consists of the their present state, their previous history and their predicted future states; this can be enhanced by the contexts of other, similar or related, humans and other objects, or even by the context of information itself.

Currently, a considerable amount of research has been done on providing personalized information service. Existing personalization techniques on information retrieval can be either manual personalization or automatic personalization. Manual personalization, such as MyYahoo (http://my.yahoo.com), allows a user to select the user's own interests from a predefined list. In these services, an account is created on the web page, and the user selects checkboxes to determine what types of content are of interest to the user. There are some problems in manual personalization; (1) It is inconvenient for users to specify many options, (2) The provided options may be too coarse-grained to reflect users' preferences, (3) It is static and requires users to update their profile after their preferences change. Perkowitz, et al. demonstrated the feasibility of automatic personalization for desktop access to the Web (M. Perkowiz et al. 1997). Different from manual personalization, automatic personalization does not require explicit effort from users to customize their profiles. It automatically does the personalization based on access logs using some machine learning, such as Daily Learner, or data mining techniques, such as usage-based Web personalization (D. Billsus et al. 2000 and M. Perkowitz et al. 1999). Widely used data mining techniques include clustering, classification, association rules, etc. Clustering plays an important role in Web mining and has found numerous applications, such as Web search engines, Web directories, decision making assistance, and Web personalization. However, these techniques do not adapt quickly to changing user interests along with the changes in various useful contexts, such as time and location.

In conclusion, the existing systems that can provide personalized information either require users to provide a lot of personal information through questionnaires or to use data-mining techniques to gather and analyze information about users to generate user profiles. These systems cannot effectively reflect the actual needs of users because the actual needs of users may change according to current situation of users and many user actions (e.g. the most recently visited web page) are not considered by this type of systems. So, we seek to more useful user's context and different approach with existing studies to feed each users the information that she want, when she want it. We know that retrieval is much more effective if the context is richer than just location, and includes fields such as temperature, objects nearby, user's current interests, even her emotional state and etc. A context used to aid

retrieval can also usefully include fields that may be considered as aspects of the user model.
Context can also be associated with each of the documents that are candidates for retrieval. Thus a document may have contextual fields representing an associated preference or user's current location. Sometimes these contextual fields are part of the explicit mark-up of a document such as XML based RSS.
In later sections, we introduce related techniques for integrating information delivery technique and user's context to provide each user with personalized information.

5. Personalized RSS Feed Service Using User's Context

In the future, we will be provided information that we want, when we want it. But, there are no tries to integrate next generation syndication technique, RSS and user's context to provide personalized information to each user. In this section, we attempt to look beyond the needs of current personalization of Web 2.0.
In order to locate documents of interest users frequently make use of search engines such as Google. However, current retrieval engines take no account of the individual user and their personal interests or their physical context. We refer to extension of established information retrieval as personalized information retrieval.
IR (Information Retrieval) and the related technology of IF (Information Filtering) are concerned with the finding of information, often in the form of text documents, which are in some sense about a topic that a user is interested in. But, both are not concerned with satisfying the user's underlying information need. Typically, the user expresses their information need as a query, which is then matched against the available documents. Then, information is retrieved from a collection of discrete documents.
In a different way of typical RSS search engine, we subdivide each RSS feeds by feed tags into fields. So, our proposed RSS retrieval service can match user's query with not only contents but also RSS tags. This is useful for users, because they can limit search range to specific content of RSS feeds.

Name of Field	Description	Example
Title	The title of content	Information Retrieval Gupf
PubDate	The publishing date of content	Thu, 16 Apr 2009 21:11:40 +0000
Author	The author of content	jeremy
link	URL of content	http://www.irgupf.com/2009/04/16/improving-findability-falls-short-of-the-mark/
Description	The content of feed	I came across this article by Vanessa Fox on how government can improve the findability of their web pages, and thereby allow citizens to become better informed and government to be more transparent.

Table 1. Subdivided fields

These fields may be textual, such as title, author, keywords, and full text of content. Table 1 shows subdivided fields by RSS tags. The objective of incorporating contextual information into the retrieval process is to attempt to deliver information most relevant to the user within their current context. Use of context for retrieval within personalized information retrieval is to determine the manner and timing of any information passed to the user. An issue is that, since the information is based on the user's context, it should be delivered in a timely fashion. So, for this consideration we take advantage of simple RSS techniques to deliver information to each user in time when her or his interest changed.

We think to provide more relevant information to the user in RSS retrieval service, it is important to incorporate the feed characteristics and the user's contextual information into the search process. However, the previous retrieval services have not taken into account them. This seriously limits to offer users with useful information in RSS retrieval service.

To provide most relevant contents formed in RSS such as blog or news contents to the users within their context, we incorporate user's contextual information into retrieval process. We consider the following we use the RSS tag structure in query processing. It enables users or search engine to limit retrieval range in a specific RSS feed tag, not searching full RSS documents. Secondly, we consider both the query term frequency and the update frequency in RSS channel. It enables users to get more useful RSS channel by considering multiple RSS channel features. Thirdly, we use the user's context in retrieval service. To provide a personalized retrieval service, it is needed to understand the user's interest or preference. The context has a large influence on the interest and intent of one particular user.

The retrieval task is to deliver the RSS feed that best match the current user's context such as the latest topic which is subscribed by the user. As the user's preference or interest changes, new information may need to be retrieved.

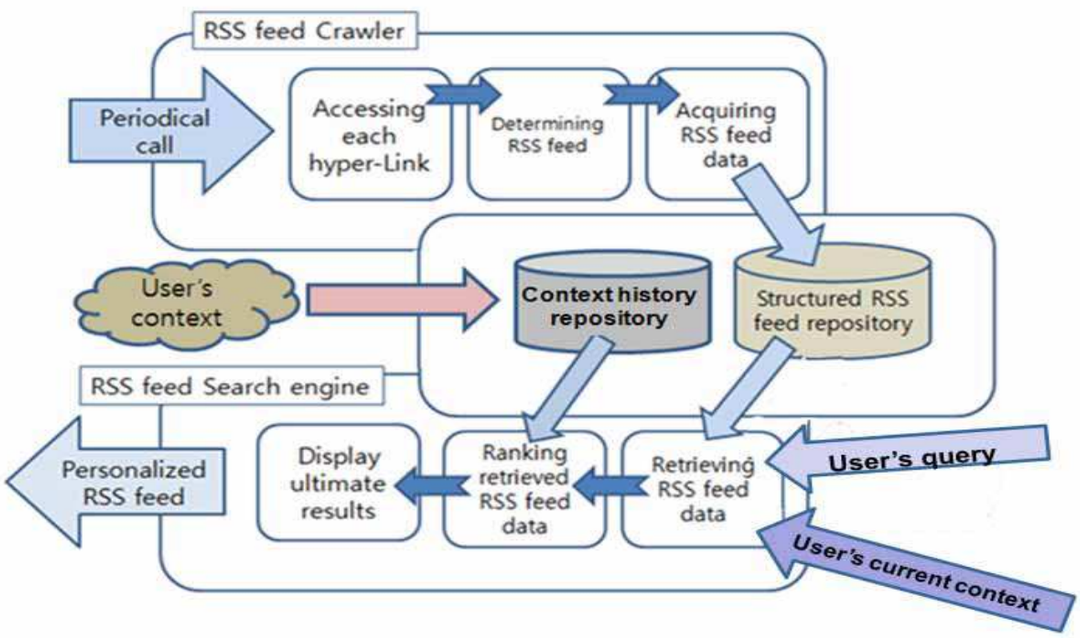


Fig. 11. System Architecture

Considering those described above, we introduce personalized RSS feed retrieval system in which two factors are mainly considered, characteristic of the RSS and user's context. It is a new RSS feed retrieval system that uses user's context for matching retrieval results to their interests.
Figure 11 shows the new architecture of the personalized RSS retrieval system. The architecture is composed of RSS feed crawler, RSS feed repository, and RSS feed retrieval engine. RSS feed crawlers visit RSS feed sites on the Web, down their RSS channels. After duplicated RSS feed elimination processing, captured RSS feeds are indexed in structured RSS feed repository. Historical repository is for storing user's previous context and current context from user's behavior such as web search. From this historical repository, RSS feed retrieval engine capture user's current interest or preference. Matching captured user's interest with indexed RSS contents deliver personalized RSS feed to each user. User's context history data can also associate with each of the RSS feeds in RSS feed repository. The RSS feeds in RSS feed repository are candidates for retrieval.

5.1 Collecting and Updating User's Context

Traditional information retrieval techniques usually require the users to spend much effort to repeatedly refine their request making their requests more specific to eliminate irrelevant results to obtain the result they want. However, those request cause annoyances to each user. And those techniques not provide related information with user's preference.
In order to get personalized RSS feeds based on user's preference, user's context data is needed. So, we store user's context data in a context history tuple. Historical repository showed in Figure 13 consists of a set of context history tuples. A context history tuple has the following structure.

Context History Schema		
Context Type	Context Value	RSS channel

Table 2. A context history schema

A tuple consists of context type, context value and context related data. The context type is to record context attributes such as time, location. Context value data can describe user's context such as time or location. RSS channel data express URL of RSS channel in which the user visit with some context value.

5.2 RSS feed crawler

Called periodically, RSS feed crawlers wander about on the web through the link to find RSS feed link. When accessing corresponded URL address, RSS feed crawlers judge whether the URL present RSS feed address or not. It is very important to acquire valid RSS feed address promptly. Our each RSS feed crawler include feed information cash with which RSS crawlers directly check whether acquired RSS feed URL is duplicated to elevate the whole system's performance. If acquired RSS feed URL is verified as new RSS feed URL, RSS feed crawlers get the RSS feed URL to structured RSS feed database. Otherwise, RSS feed crawlers extract new contents from the RSS feeds.

Figure 13 shows the operation flowchart of RSS feed crawler. By periodical calls which generate the crawler instance, the RSS feed crawler accesses site along hyperlink in the site and then downs web page to determine whether this page is RSS channel or not. If accessed page is RSS channel, RSS feed crawler checks duplicated accesses. RSS feed crawler acquire RSS feeds, and then insert it into structured RSS feed database. RSS feed crawlers use depth-first retrieval algorithm (DFS), which is used typically by many web crawlers (Allan Heydon et al. 1999). Proposed RSS feed crawlers input links acquired from web pages to queue and draw each links out orderly to judge whether the link is RSS channel link or not.

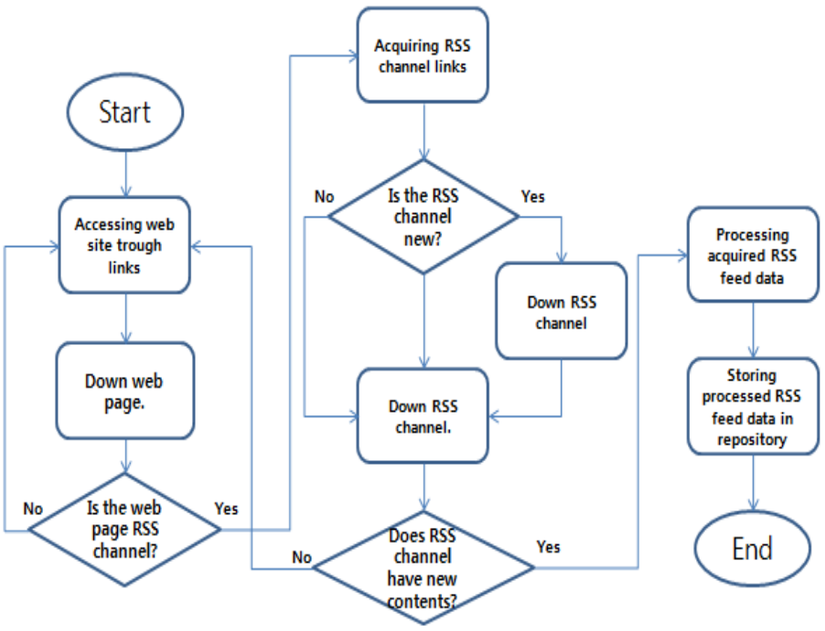


Fig. 12. Operation Flowchart of RSS Feed Crawler

The use of those structural characteristic of RSS not makes it very simple to express the web contents but also makes the RSS feed data to be accessed by crawlers and retrieval engine efficiently. The RSS feed data are structured by sub tags of item and stored in structured RSS feed database. Structured RSS feed enables users to limit retrieval in a specific feed tag, not searching all RSS feeds. This method can provide more useful RSS channel to users. For example, it allows retrieval engine to limit results to content published within a particular date range, and sort results by date reflecting user's context more efficiently.

5.3 RSS feed retrieval engine

In our proposed system, RSS feed retrieval engine takes important key role. The goal of proposed RSS feed retrieval engine is to offer users more useful information. We consider RSS feed characteristics and user's contextual information. The proposed retrieval method is comprised of four main tasks.

First, retrieval engine initiates the query and user's context from each user. Then, RSS feed retrieval engine accesses RSS feed repository and look up related contents using a Term

Frequency (TF) algorithm that retrieves RSS feeds on a similar topic (MJ Pazzani et al. 1996). It is commonly used method in traditional search engines. Second, we consider the publishing rate in RSS channel. The reason of considering the publishing frequency of RSS channel is to conclude whether those RSS channels publish useful contents continually. So, we use the pubDate tag of all feeds of each RSS channel to calculate update frequency. Calculating publishing frequency of each RSS channel can make it possible to filter out the spam RSS channel. Third, we conclude which content includes user's preference. By using user's context, it is possible to provide information with user's preference. Our RSS feed search engine takes the step in providing personalized search results based on user's preferences specifying what interests each users. By using a set of user's context history tuples in the historical repository, we can assume information that the user prefer. Finally, our method ranks RSS feeds using three tasks. Our ranking score is computed by (1) and (2). The RS_{uki} denotes the ranking score with respect to RSS feed i and term T_k of user u. It is possible to compute RS_{uki} with adjustable parameters α, β and γ ($\alpha + \beta + \gamma = 1$):

$$
\begin{aligned}
&RS_{uki} = \alpha \frac{tf_{ik}}{tf_{\max}} + \beta \frac{uf_i}{uf_{\max}} + \gamma pd_{ui} \\
&tf_{ik} = \text{frequency of term } T_k \text{ in RSS channel } D_i \\
&tf_{\max} = \text{maximum frequency of term} \\
&uf_i = \text{updatae frequency pf RSS channel } D_i \\
&uf_{\max} = \text{maximum feed update frequency of RSS channel} \\
&pd_{ui} = \text{preference degree of user } u \text{ in RSS channel } D_i
\end{aligned}
\tag{1}
$$

Equation (2) computes the preference degree of user u in RSS channel D_i. In equation (2), n is the number of context types and C_{uj} is the current context value with respect to context type j of user u.

$$
pd_{ui} = \sum_{j=1}^{j \le n} \delta_j \cdot C_{uj}, \; Where \left(\sum_{j=1}^{j \le n} \delta_j = 1 \right) \tag{2}
$$

After ranking RSS channels, RSS search engine provides result set of RSS channels in descending order by the computed ranking score.

6. Conclusion

Web 2.0 is a new way to find, save, and share information on the Web. The goal of Web 2.0 is to facilitate forming network and sharing knowledge between users. In Web 2.0, RSS is a one of the most important techniques and a new way to provide a simple way for user to share contents on their website and make it available to users without asking them to visit the site every time when new contents are added. Although there are a lot of RSS feeds available, finding which information are appealing and relevant to user's underlying information need is not easy. Therefore, it is important to incorporate the user's contextual

information and traditional information techniques based on RSS feed characteristics into the retrieval process to get relevant information to user. However, the previous search system have not taken into account them when provide retrieved information as a result to user. This seriously limits to offer users with reliable and useful information automatically.
Personalized RSS feed retrieval needs to bring together a number of disparate technologies. We have discussed related issues in personalized RSS feed retrieval. In this article, we proposes a new RSS feed retrieval system using user's context and traditional information techniques based on the characteristics of RSS feed. The proposed RSS feed search system efficiently collects the data from RSS service site by categorizing RSS feed structure and then rank RSS channel using user's context and relevant tag value.
Using user's context, it does provide personalized RSS feeds that automatically find RSS service site for user based on her or his interests and other RSS service site he or she read. That is, our proposed RSS feed search system uses personal preferences to deliver custom search results based on interests selected by users. Being different with the previous works, our proposed system acquires RSS feed on various web sites more efficiently using RSS feed structure and ranks information more reliably considering user's context such as preference, or predefined profile.
Consequently, RSS feed search engines can learn from what a user do to help him or her find what he or she need. With integration of processes for each factor, we can effective rank the retrieval set of searching result by reliable RSS channel, the source of RSS feed. It makes it possible to provide users with more personalized contents generated on reliable RSS feeds.

7. Future Work

Recent rapid advances in Internet-based information systems and handheld devices make it possible for users to retrieve information anytime and anywhere. Mobile and ubiquitous computing environments provide a challenging and exciting new domain for information retrieval. Some of the challenges are related with providing relevant and reliable information to users often engaged in other contexts of ubiquitous computing environment or to agents acting on behalf of the users.
Through our study described in this article, we know that identification of relevant information can be achieved by integration of existing methods from information retrieval and user's context. In the future work, we take advantage of contributions of human - computer interaction, mobile computing and context-awareness technologies to determine how and when to deliver the information to the user or how best to act on their user's behalf in ubiquitous computing environments. So, we will extend those techniques to meet the challenges and opportunity of ubiquitous context-aware environment. We will focus on context-awareness RSS feed retrieval mobile service in the ubiquitous computing environments in the future.

8. Acknowledgement

This work was supported by the Gyonggi Regional Research Center (GRRC) and Contents Convergence Software (CCS) Research Center.

9. References

A. Gulli.; (2005). The anatomy of a news search engine, *proceedings of international World Wide Web Conference,* pp.880 - 881, ISBN: 1-59593-051-5.

Albrecht Schmidt.; Michael Beigl.; Hans-W. Gellersen.; (1999), There is more to context than location, *Computers & Graphics,* pp. 893-901, Vol. 23, No. 6.

Allan Heydon.; Marc Najork.; (1999). Mercator: A scalable, extensible Web crawler, *World Wide Web*, pp. 219-229, Vol. 2, No. 4, ISBN: 1573-1413.

Anind K. Dey.; Understanding and Using Context, *Personal and Ubiquitous Computing, Springer-Verlag,* pp.4 – 7, Vol. 5, No. 1, ISSN:1617-4909

Bamshad Mobasher.; Robert Cooley.; Robert Cooley.; (2000). Automatic personalization based on Web usage mining, *Communications of the ACM,* pp. 142 – 151, Vol. 43, NO. 8.

Chris Sherman.; (2005), RSS Search Engines, *Search Engine Watch,* http://searchenginewatch.com/3 531191.

D. Billsus.; J. Pazzani.; J. Chen.; A learning agent for wireless news access, *Intelligent User Interfaces*, pp. 33-36, ISBN:1-58113-134-8

Diane Kelly.; Jaime Teevan.; (2003). Implicit feedback for inferring user preference: a bibliography, *ACM SIGIR Forum, pp. 18 – 28,* Vol. 37, No. 2, ISSN:0163-5840

Don E. Descy.; (2004). All aboard the internet: Introducing rss: your one stop for news and information!,*Jounal of TechTrends,* pp. 3-5, Vol.49, NO. 3.

Gregory D.Abowd.; Anind K. Dey.; Peter J. Brown.; Nigel Davies.; Mark Smith.; Pet Steggles.;(1999). Towards a better understanding of context and context- awareness, *Proceedings of the 1st international symposium on Handhel*d *and Ubiquito*us *Computing*, pp. 304-307, ISBN:3-540-66550-1.

Haveliwala.; T.H.; (2003). Topic-sensitive PageRank: a context-sensitive ranking algorithm for Web search, *Knowledge and Data Engineering, IEEE Transactions,* pp. 784- 796, Vol.15, No.4, SSN: 1041-4347.

Jemima Kiss.; (2008). Web 3.0 is all about rank and recommendation, *Article history,* http://www.guardian.co.uk/media/2008/feb/04/web20?gusrc=rss&feed=media

Jong-yi Hong.; Eui-ho Suh.; Sung-Jin Kim.; (2009). Context-aware systems: A literature review and classification, *Expert Systems with Applications: An International Journal,* pp. 8509-8522, Vol.36, No. 4, ISSN:0957-4174.

Langley.; Nick.; (2005). Explorer 7 and Window Vista set to drive demand for RSS skills. *Computer Weekly (Septemper 6, 2006),* pp. 40.

Linden, G.; Smith, B.; York, J.; (2003). Amazon.com recommendations item-to-item collaborative filtering, *Industry report of Amazon.com*, pp. 76-79.

Maged N Kamel Boulos.; Inocencio Maramba.; Steve Wheeler.; (2006). Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education, *BMC Medical Education 2006,* http://www.biomedcentral.com/1472-6920/6/41.

M. Perkowitz.; Etzioni.; (1997). Adaptive web sites: an AI challenge, *Proceedings of 15th International Joint Conference on Artificial Intelligence*, pp. 16-23, ISBN 1-55860-480-4

M. Perkowitz.; O. Etzioni.; (1999). Towards adaptive Web sites: conceptual framework and case study, *Computer Networks*, pp. 1245-1258, Vol. 31, No. 11-16.

MJ Pazzani.; J Muramatsu.; D Billsus.; (1996). Syskill & Webert: Identifying interesting web sites, Proceedings of the National Conference on Artificial Intelligence, pp. 69-77.

Overholt.; Alison.; (2005). Learning to Love RSS. *Fast Company*, 99 (October, 2005), pp. 43.

Peter Lyman.; Hal R. Varian.; (2000). How Much Information?, *Counting the Numbers,* Vol. 6, No. 2.

Richardson.;Will.; (2005). RSS: A Quick Start Guide for Educators, *http://www.weblogg- ed.com/rss_for_ed*, Updated march 29, 2005.

Ronald J. Glotzbach.; James L. Mohler.; Jaime E. Radwan.; (2007). RSS as a course information delivery method, *International Conference on Computer Graphics and Interactive Techniques,* No. 16, ACM

Ryan Turner.; (2007). The Basic of Web 2.0 Personalization, http://www.websocial architecture.com/community/2007/03/the_basics_of_w.html

Sergey Brin.; Lawrence Page.; (1999). *T*he anatomy of a large-scale hypertextual Web search engine*, Proceedings of the Seventh International World Wide Web Conference*, pp. 107-117,Vol. 30, No. 1-7.

S. Jeff Cold.; (2006). Using Really Simple Syndication (RSS) to enhance student research, *ACM SIGITE Newsletter,* pp. *6 – 9,* Vol. 3, No. 1, ISBN: 1550-1469.

Stephen S. Yau.; Huan Liu; Dazhi Huang,; Yisheng Yao.; (2003). Situation-Aware Personalized Information Retrieval for Mobile Internet, *Proceedings of the 27th Annual International Conference on Computer Software and Applications,* pp. 638, ISBN:0-7695-2020-0.

Tim Oreilly.; (2007). What is the Web 2.0, *Communications & Strategies*, pp. 17, No. 1, First Quarter 2007.

Tim Oreilly.; (2007). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, *Communications & Strategies*, pp. 17, No. 1, First Quarter 2007 collaborative filtering, *Internet Computing, IEEE*, *pp.* 76- 80, Vol. 7, Vo. 1

Udi Manber.; Ash Patel.; John Robison.; (2000). Experience with personalization of Yahoo!, *Communications of the ACM,* Vol.43, No. 8, ISBN: 0001-0782.



Engineering the Computer Science and IT

Edited by Safeeullah Soomro

ISBN 978-953-307-012-4

Hard cover, 506 pages

Publisher InTech

Published online 01, October, 2009

Published in print edition October, 2009

It has been many decades, since Computer Science has been able to achieve tremendous recognition and has been applied in various fields, mainly computer programming and software engineering. Many efforts have been taken to improve knowledge of researchers, educationists and others in the field of computer science and engineering. This book provides a further insight in this direction. It provides innovative ideas in the field of computer science and engineering with a view to face new challenges of the current and future centuries. This book comprises of 25 chapters focusing on the basic and applied research in the field of computer science and information technology. It increases knowledge in the topics such as web programming, logic programming, software debugging, real-time systems, statistical modeling, networking, program analysis, mathematical models and natural language processing.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Haesung Lee and Joonhee Kwon (2009). Toward Personalized RSS Retrieval Service: the Effect of Using User's Context, Engineering the Computer Science and IT, Safeeullah Soomro (Ed.), ISBN: 978-953-307-012-4, InTech, Available from: http://www.intechopen.com/books/engineering-the-computer-science-and-it/toward-personalized-rss-retrieval-service-the-effect-of-using-user-s-context

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen