International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 6 (2014) pp. 279 - 293
© MIR Labs, www.mirlabs.net/ijcisim/index.html

Semantic User Profiling for News Domain

Shikha Agarwal[1] and Archana Singhal[2]

[1]Department of Computer Science
University of Delhi, Delhi
Shikha_8june@rediffmail.com

[2]Department of Computer Science
University of Delhi, Delhi
Singhal_archana@yahoo.com

***Abstract*-Online news reading is becoming an important part of the daily routine of online web users. The amount of information available online is increasing exponentially. Although this information is a valuable resource but lots of scattered, unstructured, irrelevant data is increasing on the web every moment which limits its value. Many research projects and companies are exploring the use of personalized applications that manage this deluge by tailoring the information presented to individual users. These applications all need to gather, and exploit, some information about individuals in order to improve the relevance of news recommended to the end user.**

User behavior analysis is very important step of news recommendation. As a very first step we have identified different problems faced by online news readers and problems faced in user profiling of online news readers. Then in our proposed approach we are trying to curb these problems to make the system more efficient for the end user. For semantic user profiling we have designed ontologies for various categories of news items. Knowledge of news domain captured in ontologies, act as a classifier for news items and help in semantic user profiling as well. We are incorporating International Press Telecommunication Council (IPTC) standards in our design since IPTC has proposed various standards to make the system more interoperable. We have classified RSS (Really Simple Syndication) feed news items into categories specified in our designed ontologies. Characteristics and preferences of online news readers have been analyzed semantically. In user profiling for news domain we are considering properties of news items also, along with user information. Information has been collected explicitly, through direct user intervention, as well as implicitly, through agents that monitor user activity.

This user profiling helps us to better understand our end user. Ontological inference in making user profiling also helps to find the user interests in the area which could not be identified from the user browsing history directly. In contrast to the previous work done in the area, our system can handle issues mentioned in the paper in a better way.

Keywords:User behavior analysis; web server log file; Semantic classification of news items; News Domain Ontology; RSS news feeds.

I. Introduction

The expansion of the World Wide Web (Web for short) has resulted in a large amount of data that is now in general freely available for user access. Sifting through all the digital information available poses a significant challenge to users. One gets lost in loads of documents, unlinked data and scattered knowledge.

The recent decade has witnessed an explosive growth of online news. These numbers have been steadily increasing over the past years and show the growing appeal of reading news online. Also, with advent of new technologies, online news reading is becoming popular. One important advantage of online news over traditional newspapers is that the former can be augmented with hyperlinks to other related news.

The different types of data have to be managed and organized in such a way that they can be accessed by different users efficiently. A concept of personalized information filtering or search for personalization or recommendation was introduced as a key technology to overcome the problem of information overload, but we are all aware that technologies such as user profile acquisition should be studied in detail before the information filtering technologies are put into any practical use. User profiling on the web consists of studying important characteristics of the web visitors. To offer user with more precise and relevant result we have analyzed user behavior semantically. We make use of knowledge captured in our designed news domain ontologies [13]. Aim of this study is to understand user preference and the historical browsing interests. Taking the multiple and diverse information needs of the user into consideration, the proposed mechanism constructs clusters of ontological user profiles that adapts according to the changing user requirements.

We have identified problems faced by online news readers. As mentioned earlier, Web consists of loads of information and sifting through all the available information is time consuming and a tedious task. Moreover, news items are provided by various news sources. This also creates confusion for end user to decide which source of news to prefer. News items itself is very dynamic in nature and has time value. So notification of novel news items should reach the end user in real time basis. End user itself has very dynamic interest regarding topic of news.

We have identified various problems in making user profiling. Cold start problem is that limited information is available about news user. Problem of sparsity arises due to lack of information about user interest or disinterest, in majority of news areas. Problem of scalability arises as the number of online news readers and news items and news sources are increasing exponentially. In proposed system we have tried to overcome these issues.

Information about online news readers [16] can be gathered explicitly, implicitly or using both (mixed approach). We have adopted mixed approach. We are gathering user's personal information explicitly by asking questions. We are also gathering user information implicitly, by analyzing log files and user behavior, while reading online news. We have semantically analyzed user behavior to gather information about user interest in those areas also which could not be identified directly. We are analyzing user behavior online after each session.

We also make clusters in user profile based on the similarity in user interest areas.

We have classified RSS (really simple syndication) feed new items semantically [14]. In future this preclassification of news items will help to match user clusters in a better way, to recommend news in real time.

In this paper we have expanded user profile semantically using ontologies. It helps to capture those user interests, which can not be captured directly by analyzing user behavior.

The paper is organized as follows. Section II gives the insight of background of the research area. Section III summarizes related work done in the area. Section IV focuses on our approach. Section V is about experimental study. Section VI concludes the work done and paves the path for future work. It is followed by list of references at the end.

II. Background

Personalized Recommendation

In order to personalize the recommendations, the system should be able to distinguish between different users or groups of users. This process is called user profiling and its objective is the creation of an information base that contains the preferences, characteristics, and activities of the users. In the web domain, user profiling has been developed significantly because Internet technologies provide easier means of collecting information about the users of web site.

In literature user profiling is typically either knowledge based (static user model) or behavior based (logging). In previous approach knowledge is obtained based on interview or questionnaire. In later approach patterns are extracted from user's browsing history using various techniques.

RSS feed

In making of user profile and in classification of news items both, properties of news items in RSS feed need to be analyzed. Online News items exist in different formats. RSS [19] is a simple and popular way of syndicating information over the internet. RSS feeds news item from different news sources consists of fields like title, link, and description. An example of RSS feed has been shown below. It has a channel element and item element(s).

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<rss version="2.0">
 <channel>
<title>Landofcode.com web tutorials updates</title>
<link>http://www.landofcode.com</link>
<description>Web development tutorials</description>
          <item>
<title>Online code editor</title>
<link>http://www.landofcode.com/online-code-
editor.php</link>
 <description>New functionality added to the code
editor. Check it out!</description>
</item>
 <item>
<title>Errors fixed</title>
<link>http://www.landofcode.com</link>
<description>Fixed a bunch of errors. Site runs much
better now</description>
</item>
</channel>
</rss>
```

This example consists of two item tags in a channel tag. Channel has properties title, link and description which are common for all the item tags contained in channel tag.

Format of Server Log File: Web servers register a Web log entry for every single access they get, in which important pieces of information about accessing are recorded, including the URL requested, the IP address from which the request originated, and a timestamp. A log file can be located in three different places-Web Servers, Web proxy Servers, and Client browsers [15]. Web server stores log files in different formats [11]. The most popular log file format is the Common Log Format (CLF)- <ip><baseurl><date><method><file><protocol><code> <bytes> <referrer><user agent>. In apache log files are stored in combined log format. Our system has been implemented on Apache tomcat server. Each entry on the server in combined log format has been described below by taking an example of a record:

192.2.79.245 - - [15/Mar/2012:08:11:59 -0500] "GET /dogs/george HTTP/1.1" 200 0 "http://www.photo.net/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows NT; DigExt)"

In above example, a user on a computer at the IP address 192.2.79.245, who is not telling us his login name on that computer nor supplying an HTTP authentication login name to the Web server (- -), on March 15, 2012 at 8 hours 11 minutes 59 seconds past midnight in a timezone 5 hours behind Greenwich Mean Time (15/Mar/2012:09:11:59 -0500), requested the file /dogs/george using the GET method of the HTTP/1.1 protocol. The file was found by the server and returned normally (status code of 200) but it was returned by an ill-behaved script that did not give the server information about how many bytes were written, hence the 0 after the status code. This user followed a link to this URL from http://www.photo.net/ (the referer header) and is using a browser that first falsely identifies itself as Netscape 4.0 (Mozilla 4.0), but then explains that it is actually merely compatible with Netscape and is really Microsoft Internet Explorer 5.0 on Windows NT (MSIE 5.0; Windows NT).

It is apparent that Web server logs do not contain information in the form which can be used directly to infer the user behavior. Cleaned and preprocessed data is required to create user interest profile, for analysis.

Preprocessing of log files: The main objectives of preprocessing are to reduce the quantity of data being analyzed while, at the same time, to enhance its quality. The inputs to the pre-processing phase are the log and site files. The outputs are the user session files and transaction files.

Pre-processing contains following sub steps[15]: Merging of Log files from Different Web Servers (if required), Data Cleaning, User Identification, Session Identification and Formatting.

As mentioned earlier, the data in the log files (on server) can not be used for analysis in the form it is stored. Firstly, the collected data must be cleaned. Since all the log entries are not valid, we need to eliminate the irrelevant entries. Usually, this process removes requests concerning non-analyzed resources such as images, multimedia files, and page style files. Moreover, in most cases, the log file provides only the computer address (name or IP) and the user agent. For Web sites requiring user registration, the log file also contains the user login (as the third record in a log entry) that can be used for the user identification.

Secondly, the different sessions belonging to different users should be identified. A session is understood as a group of activities performed by a user when he is navigating through a given site. To identify the sessions from the raw data is a complex step, because the server logs do not always contain all the information needed. There are Web server logs that do not contain enough information to reconstruct the user session. A user session is a directed list of page accesses performed by an individual user during a visit in a Web site A user may have a single (or multiple) session(s) during a period of time. The session identification problem is formulated as “Given the Web log file, capture the Web users’ navigation trends, typically expressed in the form of Web users’ sessions”. After identifying the sessions, the Web page sequences are generated which task belongs to the first step of the preprocessing.

The third step is to convert the data into the format needed by the algorithms. This is the last step of data preprocessing. Here, the structured file containing sessions and visits are transformed to a relational database model.

Ontology

The subject of ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called ontology, is a catalog of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D. The types in the ontology represent the predicates, word senses, or concept and relation types of the language L when used to discuss topics in the domain D.

TF-IDF method to assign weights to terms

This is a statistical method used to determine the relative importance of word within a document, in corpus of documents [20]. A TF value depends on the number of occurrences of a term, so the TF value of a general word is high. This method alone can’t feature documents clearly because it is not able to distinguish important words with trivial words. Inverse Document frequency (IDF) weakens importance of general words. Thus tf and idf are used in unison.

$id\ f\ (T) = \log 2\ (\ dnum\ /\ df\ req(T)\ )+ 1$

where dnum is the total number of documents. Likewise, dfreq(T) is the number of documents that have term T.

Then TFIDF is calculated as:

$t\ f\ id\ f\ (T, D) = t\ f\ (T, D)\ X\ id\ f\ (T, D)$

When t f (T, D) is high and id f (T) is low, t f id f (T, D) is very high. If a term T appears many times in a document D and it is used by few documents, the term T is a feature word of the document D.

Instead of considering all the terms we are using concepts of the news domain. We call this method as CF-IDF (Concept frequency-Inverse document Frequency). The domain knowledge has already been stored in ontologies.

III. Related work

Recently, Personalization for recommending physical or digital products has attracted researcher's attention. Personalization can be improved if while making user profile, we can gather more specific information about end user, both explicitly as well as implicitly. We now brief the related work done in past in the area, to illustrate several techniques that have been applied to build and analyze user profile.

In [1] authors have made use of ontological user profiling. Limitation of their approach is that ontology is based on just is_a relation and they are not considering short term and long term interests. We have overcome these limitations. Our ontology consists of different relations. We are also considering both long term and short term user interest as it is very essential in case of news domain to handle dynamic nature of users and dynamic nature of news. In [2] authors have applied agent based user modeling (both explicit and implicit). They have improved the trust in recommendation by using argumentation. In contrast we are analyzing user behavior semantically. In [3] concepts have been identified in user profile using cf-idf. In contrast for identification of concepts of user interests we are using knowledge captured in our designed news domain ontology, designed based on IPTC standards, which makes the system widely acceptable and interoperable. In [4], authors have analyzed users reading history to identify various attributes. Jaccard similarity had been used to find the similarity in access patters (offline calculation). We are analyzing user's reading history by keeping in mind the various properties of news items mentioned in he paper. In [5] user model for smartTV is based on category frequency- inverse user frequency. Users with similar preference patterns are grouped using ISODATA algorithm to reduce SSE. We have designed user model for online news considering news domain specific features. In [6] description of user preference is in terms of concepts appearing in set of domain ontology. They have considered both long term and short term interest (current semantic interest) of user. The model is based on single user. Including this we are updating user profiles in real time to give more relevant news in real time. In [7] authors have used defeasible logic programming to draw credibility of news reports based on the opinions of set of users. System maintains a pool of viewers. In order to improve user's trust in the system our main focus is on better technique for collecting user interests. [8] Allows the end user to view and edit their interest profile (with cautions) to make the system of recommendation more transparent. We are giving user the option to make changes in the personal information only up to a certain extent as sometimes it can be misused and will give unexpected results.. In [9] survey shows that less focus has been given to recommendation of news. That's why we are working in this direction. In [10] authors have made changes in ontologies by swapping different properties and their values. This may lead to very unexpected results sometimes. In [12] authors have made use of wikepedia as a source of knowledge. In contrast we are using news domain specific ontology as a knowledge source and it serves the purpose of semantic inferencing also. In [22] authors have shown use of domain ontology as knowledge source in web dialogue system as a basis to define semantics of information to be exchanged among the components of web dialogue system. Their purpose is to make system adaptable for different web services. In [23] a preliminary work has been presented for making use of ontology based semantic user profiling for personalization of recommendation. Taxonomy of ontology used is based on Amazon taxonomy having IS-A relations. We in our work have used IPTC based news domain ontologies for interoperability and acceptability. Moreover ontology has many other realtions and axioms which results in better inferencing results which will improve recommendations. In [24] authors have designed a multi agent system to bring the interoperabilty to retrieval of heterogenous learning object repositories. For recommendation of educational resources it uses case based reasoning. Our approach for news recommendation is semantically motivated and based on news industry standards also. We have modelled semantically enriched user behavior for personalization of RSS feed news item recommendations. We have handled the various issues identifid in news personalization. In [25] authors have described various levels of ontology modelling for paperless office in university. They have shown the importance of ontology to increase the efficiency of workflow in office using resoning. We have also relaized the importance of ontology to provide background knowledge of news domain. As mentioned earlier also, our work is based on IPTC based news domain ontology.

User's web site accesses are recorded in web logs. Web log data is usually noisy and ambiguous and preprocessing is an important process before analyzing. Here we give insight of the recent work done in the area.

G.Castellano et al. [21] present LODAP tool (log data pre-processor) which design and implemented in order to perform preprocessing of log file. LODAP takes input log file related to website and output a database containing pages visited by user and identified user sessions. LODAP tool reduces size of web log file and groups web request into a number of user sessions.

We also need to do the same in early sages but our focus is in semantic analysis of user interests to expand the user preferences identified directly in first stage.

Our approach in next section shows, that ontological inference in profiling process results in better analysis of user interest.

IV. Our Approach

We have created Ontological user profile based on knowledge and behavior both. We have analyzed user profile semantically by considering different relationships specified in our designed Ontology of news domain. It takes into consideration the current semantic context of interest of the user. Various problems faced during the study of finding user preferences are discussed below.

A. Cold start problem: It is about lacking the information about user preferences, at the very beginning. To handle this problem we will offer end users the pre classified news items. We will give users the choice to select different categories of news.

B. Problem of Sparse data: In news domain there is constant stream of new news item and each user only rates a few, leads to sparse data. In user behavior analysis only few interest areas can be identified. This leads to lots of missing values in the records. To handle this problem we are extending the user interest by semantically relating concepts of user interest with other concepts specified in our designed ontology.

C. Scalability issue: It is the ease with which a system or component can be modified, added, or removed, to accommodate changing load. To handle the increasing load we are clustering (making groups) end users based on the similarity in user preferences. The clustering process is an important step in establishing user profiles.

We have also noticed that in user study for news domain we have to take care of few domain specific problems:

A. Novelty control: this is a new area in news recommendation having less focus. If two news articles are too similar to each other it is clear that user may not be interested in one given that they already read the one, since there will be no novel information in the other. After user profile analysis, every time diverse news items need to be recommended to the user. To handle this issue we have classified news items. We have analyzed news items using concept frequency-inverse document frequency to identify their classes. We have referred ontologies to consider the semantics of identified class. While classifying, news items with similar context will be stored in the database, only once.

B. Dynamic nature of news: We are analyzing user profile after every user session in real time. This will help to recommend ever changing dynamic news. In user profiling we are considering time value of news data in each session.

C. Dynamic nature of a user: Choice of user changes with time in case of news, so we are considering both long term and short term interest of user. Short-term profiles represent the user’s current interests whereas long-term profiles indicate interests that are not subject to frequent changes over time. To consider long term interest we give high rank to the concepts which are identified in previous user sessions also. To consider short term interest we consider the concepts identified in current session only.

D. Diverse end users: We have to cater to diverse end users in news domain. User may be a student, a homemaker, a businessman or a researcher. We will make clusters of users based on features considering both short term and long term interests.

To consider the semantics of the identified concepts in user profile making, we have made news domain ontologies. It consists of classification structure, relationships and instances of knowledge base. For powerful inferencing our designed Ontology includes relationships like is-a, type of, held_in. We have developed IPTC based ontologies for 17 news domains as IPTC has divided all the news items into 17 major categories [18]. Table 1 shows ontologies of all the 17 categories and number of classes in each ontology.

S. No.	Ontologies of various Categories of News Domain	Classes
1.	Human Interests	6
2.	Labour	6
3.	Weather	9
4.	Disasters, accidents	16
5.	Education	20
6.	Crime, law, justice	22
7.	Unrests, conflicts, wars	23
8.	Health	26
9.	Lifestyle, leisure	29
10.	Religion, belief	31
11.	Social issues	39
12.	Environmental issues	41
13.	Science, technology	50
14.	Politics	54
15.	Arts, culture, entertainment	87
16.	Sports	124
17.	Economy, business, finance	161

Table 1: Ontologies with number of classes

Each class (concept) has various properties. Ontology consists of various relationships among the classes and the properties. For lexical representation of concepts we have enriched ontology with semantic lexicon Wordnet [17].

These ontologies are referred in making News item profile as well as user profile. Both the profiles are explained below.

News Item Profile: Different kind of items or products has different features. Product we are offering in our designed system is online news item from various sources (like Times of India, Economic Times, and Hindustan Times). In our designed system user can select the link to news source of their own choice, from the given list of choices, as shown in figure 4. News items are internally represented using concepts and relationships captured in ontologies. As specified above Ontologies acts as source of information in identifying concepts, sub concepts and relationships. News had been classified based on frequency of concepts [14] in title and description features of news items. News had also been classified based on identified named entities (like person, location, event, organization) in the news items. Before classification we have tokenize news item. Then stop word have been removed, and then stemming took place. Removal of stop words enhances precision of retrieval by removing semantically meaningless but otherwise most frequent words from text. Stemmer reduces variant form of words to a common form.

We have classified news items so that it can be easily offered to the end user based on user likings and disliking. This initiates the need to capture user interests

User Profile: We make user profile by gathering information about end user both explicitly as well as implicitly. In registration process new users are asked to provide personal information explicitly to be stored in user profile database. Existing users need not to give this information again. User profiling algorithm has been activated for logged in registered users. The algorithm analyzes user behavior and implicitly gathers user information. We measure the frequency with which the concepts and entities appear in the news items red by user in a session(s).

When a user interacts with the system, web server log records and accumulates data about user interactions, implicitly. Web servers register a Web log entry for every single access they get, in which important pieces of information about accessing are recorded, including the URL requested, the IP address from which the request originated, and a timestamp. The data in the log files (on server) can not be used for analysis in the form it is stored. It needs to be preprocessed first. We identify user and session from the server log data. Both the registration data as well as the preprocessed server log data are stored in a database (MYSQL) for further analysis.

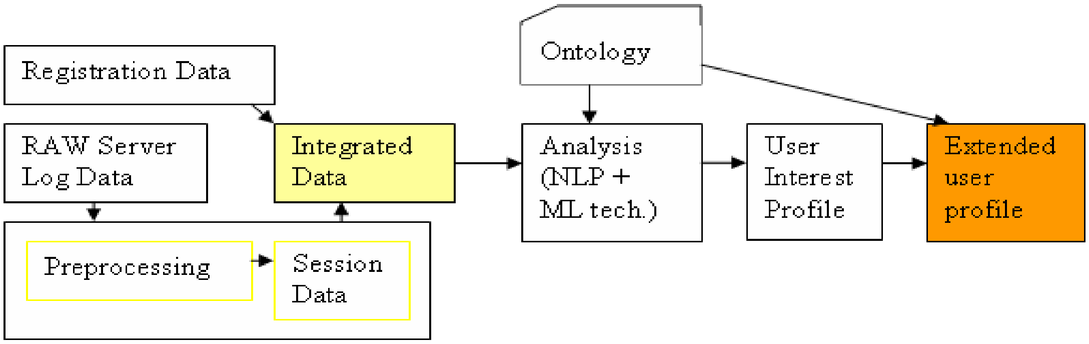


Figure 1. Framework for creating user profile semantically

IP	USER NAME	COUNTRY	REGION	CITY	CLIENTS ID	REMOTE USER NAME	DATE TIME	CONTENT	RESPONSE CODE	BYTES TRANSFERRED	REFERER	BROWSER

Figure 2. Fields of user profile

Figure 1, illustrates the framework for creating user profile semantically.

Fields in integrated database file that we have created is shown in figure 2. We also find out and store the current geographic location of the end user in the database.

In making user interest profile we have to keep track of the news items read by user. Features of news items (title and description) will provide us with information about user's interest. In addition to log file entries, we will incorporate features of news like topic, subtopic, person, location, event, organization in making user profile. For each day we will compute interest values in these topics. NLP tool will be used to extract named entities from user reading behavior. Using ML technique cf-idf (concepts of our designed ontology) and ef-idf (entities of our designed ontology) we are giving weight to the identified concepts and entities. Documents, here, are news items read in a session. Concepts and entities which are related to the concepts and entities found in the user profile analysis will also be considered relevant for the end user. This inference process using ontology will help us to discover interests which are not directly observed in user's behavior. This gives us extended user interest profile.

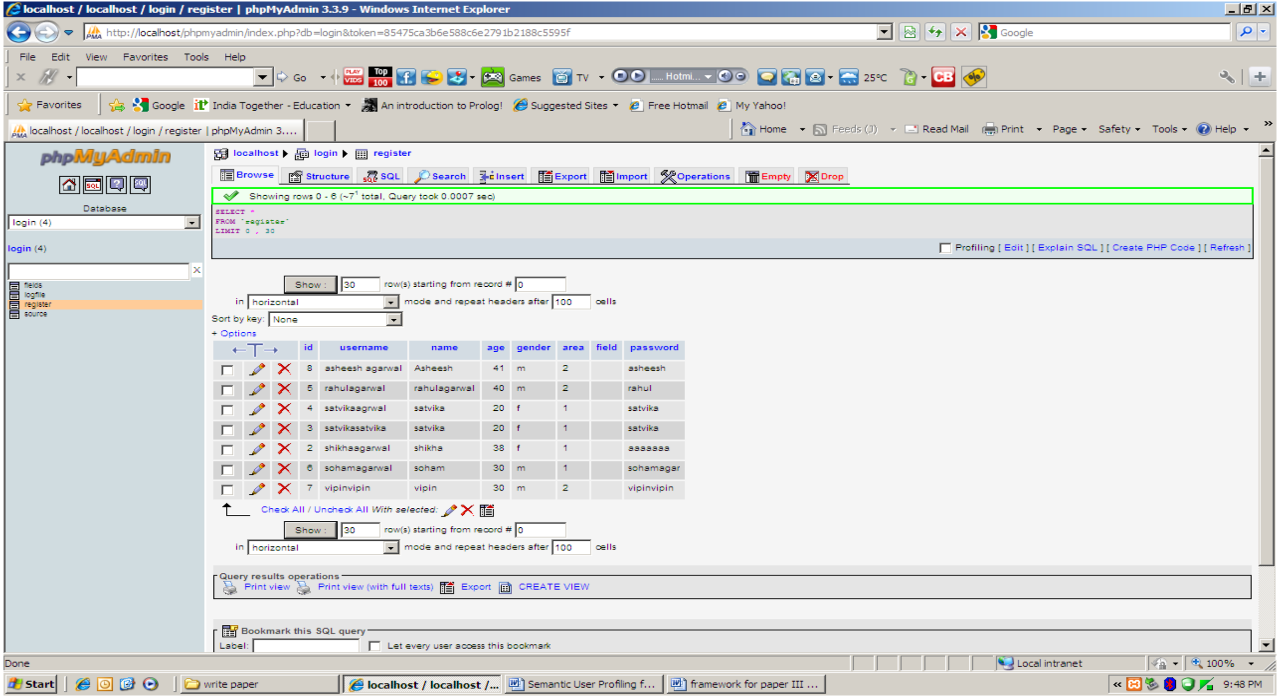
To make the system scalable we need to cluster diverse end users, by identifying similarity in the user interests.

Clustering of users:

Various criterion functions and methodologies have been developed which may be used in clustering systems. They can be grouped into three main categories: (i) proximity based methods, (ii) feature based methods, and (iii) model-based methods. Proximity based methods compute dissimilarity between two data objects in terms of distance computed from their feature values. Feature based methods extract a set of "characteristic" features from individual data objects that capture temporal information in individual sequences. In model-based approach, an explicit model is learned from one or more temporal data, the similarity of two data objects is computed based on the likelihood of one data object given the model derived from the other data object.

V. Experimental Study

For experimental study a news website has been designed on XAMPP server running on local host. User profiling and news profiling has been implemented using open source tools and software components. The OS for development is windows XP. For server side coding of user registration and news website, we are using PHP. Databases are stored in MYSQL. Server is Apache Tomcat. For development of news domain ontologies, WEB


localhost / localhost / login / register | phpMyAdmin 3.3.9 - Windows Internet Explorer
http://localhost/phpmyadmin/index.php?db=login&token=85475ca3b6e588c6e2791b2188c5595f
File Edit View Favorites Tools Help
phpMyAdmin
Database
login (4)
fields
logfile
register
source
localhost ▶ login ▶ register
Browse Structure SQL Search Insert Export Import Operations Empty Drop
Profiling [Edit] [Explain SQL] [Create PHP Code] [Refresh]
Check All / Uncheck All With selected:
Query results operations
Print view Print view (with full texts) Export CREATE VIEW
Bookmark this SQL query
Done
Local intranet
100%

Showing rows 0 - 6 (~7¹ total, Query took 0.0007 sec)

```
SELECT *
FROM `register`
LIMIT 0 , 30
```

Show : 30 row(s) starting from record # 0
in horizontal mode and repeat headers after 100 cells
Sort by key: None

id	username	name	age	gender	area	field	password
8	asheesh agarwal	Asheesh	41	m	2		asheesh
5	rahulagarwal	rahulagarwal	40	m	2		rahul
4	satvikaagrwal	satvika	20	f	1		satvika
3	satvikasatvika	satvika	20	f	1		satvika
2	shikhaagarwal	shikha	38	f	1		aaaaaaa
6	sohamagarwal	soham	30	m	1		sohamagar
7	vipinvipin	vipin	30	m	2		vipinvipin

Figure 3. Personal information of registered user gathered explicitly

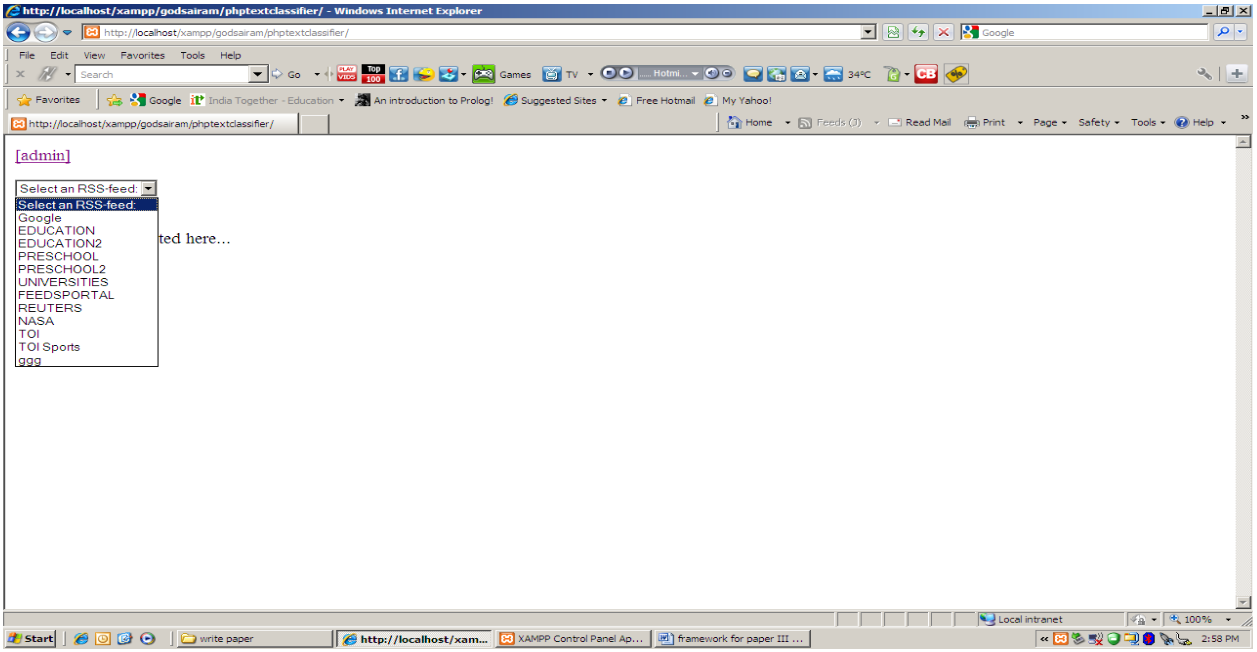
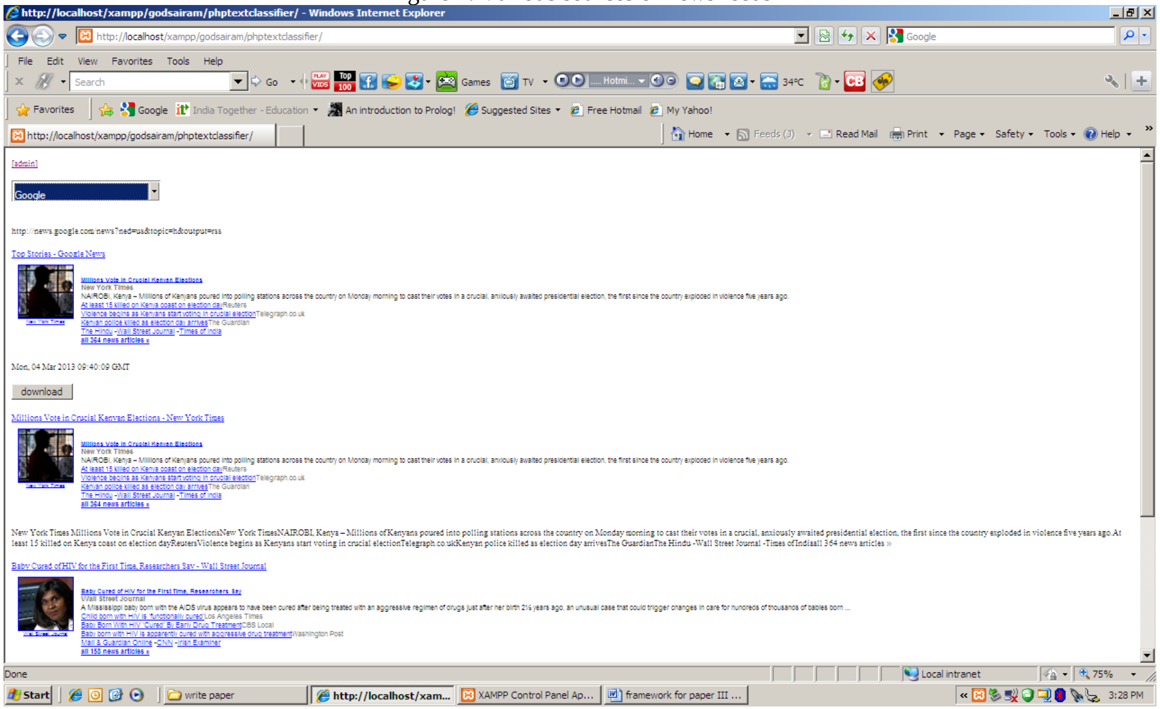


Figure 4. Various sources of news feeds


http://localhost/xampp/godsairam/phptextclassifier/ - Windows Internet Explorer
File Edit View Favorites Tools Help
[admin]
Google
download
Done
Local intranet
75%
Start
write paper
XAMPP Control Panel Ap…
framework for paper III …
3:28 PM

http://news.google.com/news?ned=us&topic=h&output=rss

Top Stories - Google News

Millions Vote in Crucial Kenyan Elections
New York Times
NAIROBI, Kenya – Millions of Kenyans poured into polling stations across the country on Monday morning to cast their votes in a crucial, anxiously awaited presidential election, the first since the country exploded in violence five years ago.
At least 15 killed on Kenya coast on election dayReuters
Violence begins as Kenyans start voting in crucial electionTelegraph.co.uk
Kenyan police killed as election day arrivesThe Guardian
The Hindu -Wall Street Journal -Times of India
all 364 news articles »

Mon, 04 Mar 2013 09:40:09 GMT

Millions Vote in Crucial Kenyan Elections - New York Times

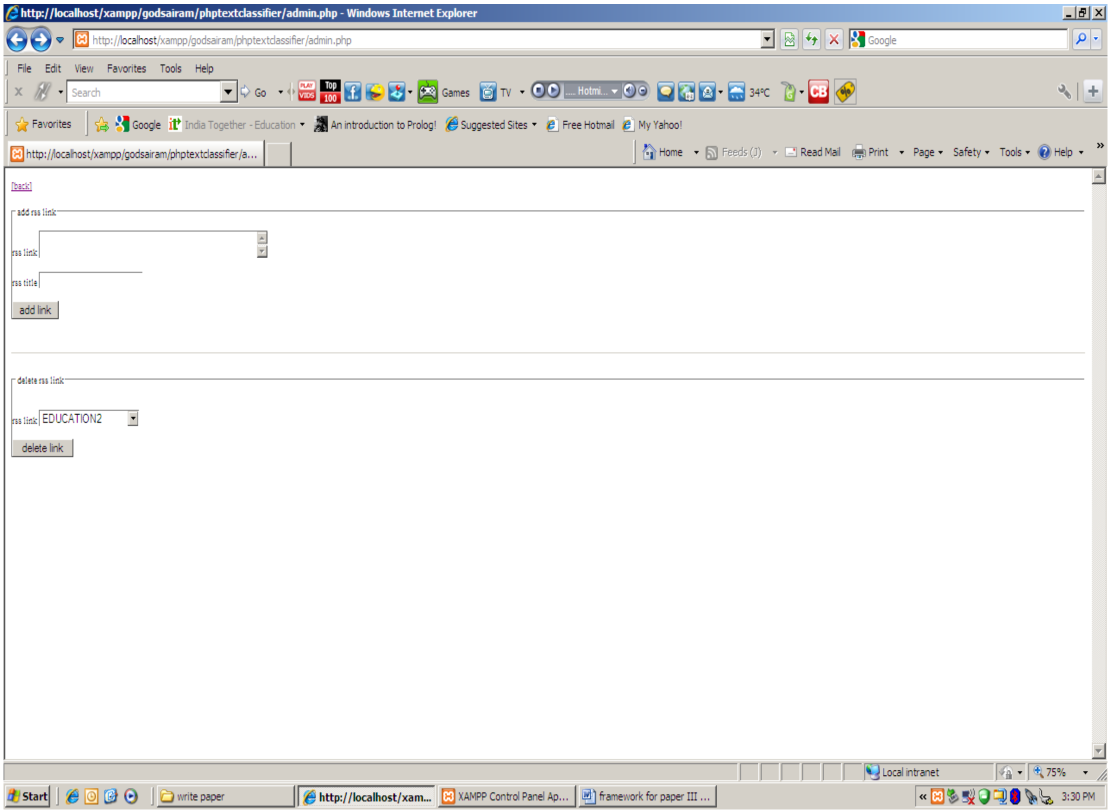
Millions Vote in Crucial Kenyan Elections
New York Times
NAIROBI, Kenya – Millions of Kenyans poured into polling stations across the country on Monday morning to cast their votes in a crucial, anxiously awaited presidential election, the first since the country exploded in violence five years ago.
At least 15 killed on Kenya coast on election dayReuters
Violence begins as Kenyans start voting in crucial electionTelegraph.co.uk
Kenyan police killed as election day arrivesThe Guardian
The Hindu -Wall Street Journal -Times of India
all 364 news articles »

New York Times Millions Vote in Crucial Kenyan ElectionsNew York TimesNAIROBI, Kenya – Millions of Kenyans poured into polling stations across the country on Monday morning to cast their votes in a crucial, anxiously awaited presidential election, the first since the country exploded in violence five years ago.At least 15 killed on Kenya coast on election dayReutersViolence begins as Kenyans start voting in crucial electionTelegraph.co.ukKenyan police killed as election day arrivesThe GuardianThe Hindu -Wall Street Journal -Times of Indiaall 364 news articles »

Baby Cured of HIV for the First Time, Researchers Say - Wall Street Journal

Baby Cured of HIV for the First Time, Researchers Say
Wall Street Journal
A Mississippi baby born with the AIDS virus appears to have been cured after being treated with an aggressive regimen of drugs just after her birth 2½ years ago, an unusual case that could trigger changes in care for hundreds of thousands of babies born ...
Child born with HIV is 'functionally cured'Los Angeles Times
Baby Born With HIV 'Cured' By Early Drug TreatmentCBS Local
Baby born with HIV is apparently cured with aggressive drug treatmentWashington Post
Mail & Guardian Online -CNN -Irish Examiner
all 153 news articles »

Figure 5. News items from news source of user’s choice


http://localhost/xampp/godsairam/phptextclassifier/admin.php - Windows Internet Explorer
File Edit View Favorites Tools Help
Favorites
Google
India Together - Education
An introduction to Prolog!
Suggested Sites
Free Hotmail
My Yahoo!
Home
Feeds (J)
Read Mail
Print
Page
Safety
Tools
Help
Local intranet
75%
Start
write paper
http://localhost/xam...
XAMPP Control Panel Ap...
framework for paper III ...
3:30 PM

[back]

add rss link

rss link

rss title

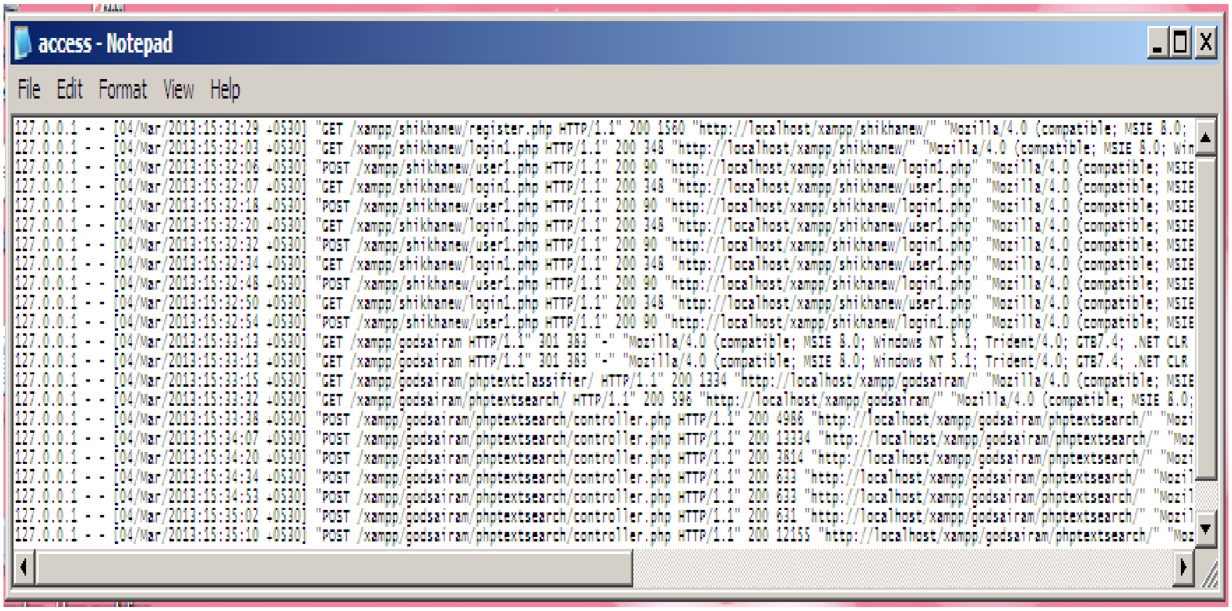
add link

delete rss link

rss link EDUCATION2

delete link

Figure 6. Option to add or remove RSS news source from the list



```
127.0.0.1 - - [04/Mar/2013:15:31:29 +0530] "GET /xampp/shikhanew/register.php HTTP/1.1" 200 1560 "http://localhost/xampp/shikhanew/" "Mozilla/4.0 (compatible; MSIE 8.0;
127.0.0.1 - - [04/Mar/2013:15:32:03 +0530] "GET /xampp/shikhanew/login1.php HTTP/1.1" 200 348 "http://localhost/xampp/shikhanew/" "Mozilla/4.0 (compatible; MSIE 8.0; Win
127.0.0.1 - - [04/Mar/2013:15:32:06 +0530] "POST /xampp/shikhanew/user1.php HTTP/1.1" 200 90 "http://localhost/xampp/shikhanew/login1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:32:07 +0530] "GET /xampp/shikhanew/login1.php HTTP/1.1" 200 348 "http://localhost/xampp/shikhanew/user1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:32:18 +0530] "POST /xampp/shikhanew/user1.php HTTP/1.1" 200 90 "http://localhost/xampp/shikhanew/login1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:32:20 +0530] "GET /xampp/shikhanew/login1.php HTTP/1.1" 200 348 "http://localhost/xampp/shikhanew/user1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:32:32 +0530] "POST /xampp/shikhanew/user1.php HTTP/1.1" 200 90 "http://localhost/xampp/shikhanew/login1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:32:34 +0530] "GET /xampp/shikhanew/login1.php HTTP/1.1" 200 348 "http://localhost/xampp/shikhanew/user1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:32:48 +0530] "POST /xampp/shikhanew/user1.php HTTP/1.1" 200 90 "http://localhost/xampp/shikhanew/login1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:32:50 +0530] "GET /xampp/shikhanew/login1.php HTTP/1.1" 200 348 "http://localhost/xampp/shikhanew/user1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:32:54 +0530] "POST /xampp/shikhanew/user1.php HTTP/1.1" 200 90 "http://localhost/xampp/shikhanew/login1.php" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:33:13 +0530] "GET /xampp/godsairam HTTP/1.1" 301 383 "-" "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; GTB7.4; .NET CLR
127.0.0.1 - - [04/Mar/2013:15:33:13 +0530] "GET /xampp/godsairam HTTP/1.1" 301 383 "-" "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; GTB7.4; .NET CLR
127.0.0.1 - - [04/Mar/2013:15:33:15 +0530] "GET /xampp/godsairam/phptextclassifier/ HTTP/1.1" 200 1334 "http://localhost/xampp/godsairam/" "Mozilla/4.0 (compatible; MSIE
127.0.0.1 - - [04/Mar/2013:15:33:32 +0530] "GET /xampp/godsairam/phptextsearch/ HTTP/1.1" 200 596 "http://localhost/xampp/godsairam/" "Mozilla/4.0 (compatible; MSIE 8.0;
127.0.0.1 - - [04/Mar/2013:15:33:38 +0530] "POST /xampp/godsairam/phptextsearch/controller.php HTTP/1.1" 200 4986 "http://localhost/xampp/godsairam/phptextsearch/" "Mozi
127.0.0.1 - - [04/Mar/2013:15:34:07 +0530] "POST /xampp/godsairam/phptextsearch/controller.php HTTP/1.1" 200 13334 "http://localhost/xampp/godsairam/phptextsearch/" "Moz
127.0.0.1 - - [04/Mar/2013:15:34:20 +0530] "POST /xampp/godsairam/phptextsearch/controller.php HTTP/1.1" 200 3814 "http://localhost/xampp/godsairam/phptextsearch/" "Mozi
127.0.0.1 - - [04/Mar/2013:15:34:34 +0530] "POST /xampp/godsairam/phptextsearch/controller.php HTTP/1.1" 200 633 "http://localhost/xampp/godsairam/phptextsearch/" "Mozil
127.0.0.1 - - [04/Mar/2013:15:34:53 +0530] "POST /xampp/godsairam/phptextsearch/controller.php HTTP/1.1" 200 633 "http://localhost/xampp/godsairam/phptextsearch/" "Mozil
127.0.0.1 - - [04/Mar/2013:15:35:02 +0530] "POST /xampp/godsairam/phptextsearch/controller.php HTTP/1.1" 200 631 "http://localhost/xampp/godsairam/phptextsearch/" "Mozil
127.0.0.1 - - [04/Mar/2013:15:35:10 +0530] "POST /xampp/godsairam/phptextsearch/controller.php HTTP/1.1" 200 12155 "http://localhost/xampp/godsairam/phptextsearch/" "Moz
```

Figure 7. Raw log file stored on the server

ontology language (OWL) in Protégé IDE has been used.

This news website offers RSS feed news items to two types of users: Registered users or non registered ordinary user. In registration process user is asked to provide some personal data and also some preferences about source of news. The personal information will be stored in the database as shown in figure 3. This database stores id, username, name, age, gender, area (like student or professional) and also stores the password.

Website offers various sources of news feeds as shown in figure 4. User has been provided news items from the source of user's choice as shown in figure 5. Links to new source of news items can be added to the list or removed from the list as shown in figure 6. This authority has been given to the authentic registered users only.

As mentioned above log file contains various fields which need to be separate out for the processing. An unprocessed raw log file has been shown in figure 7.

The process of separating field from the single line of the log file is known as field extraction. The server used different characters as delimiters. The most commonly used delimiter is 'comma' or 'white space' character. Here space is the delimiter.

The Field Extraction algorithm has been given below:

```
Input: Log File on Apache server
Output: A table (LD) in Data Base(DB) on MySQL
Begin
1. Open a DB connection
2. Create a table (LD) to store log data
3. Open Log File
4. Read all fields contained in Log File
5. Separate out the Attribute in the string Log
6. Extract all fields and Add into the LD
7. Close a DB connection and LD
End
```

An algorithm for cleaning the entries of server logs is presented next.

```
Input: Log data Table (LD)
Output: Summarized Log data Table (LDT)
‘E’ = access pages consist of embedded objects (i.e .jpg, .gif, etc)
‘S’ =successful status codes and requested methods (i.e 200, GET etc)
Begin
1. Read records in LD
2. For each record in LD
3. Read fields (Status code, method)
4. If Status code=‘E’ and method= ‘S’
Then
5. Get IP_address and URL_link
6. If suffix.URL_Link= {*.gif,*.jpg,*.css}
Then
7. Remove suffix.URL_link
8. Save IP_sddress and URL_Link
End if
Else
9. Next record
End if
End
```

Output of the algorithms has been shown in figure 8. It contains fields extracted from raw log file.

Users will be offered pre classified news items. Figure 9 shows classified news items for educational category.

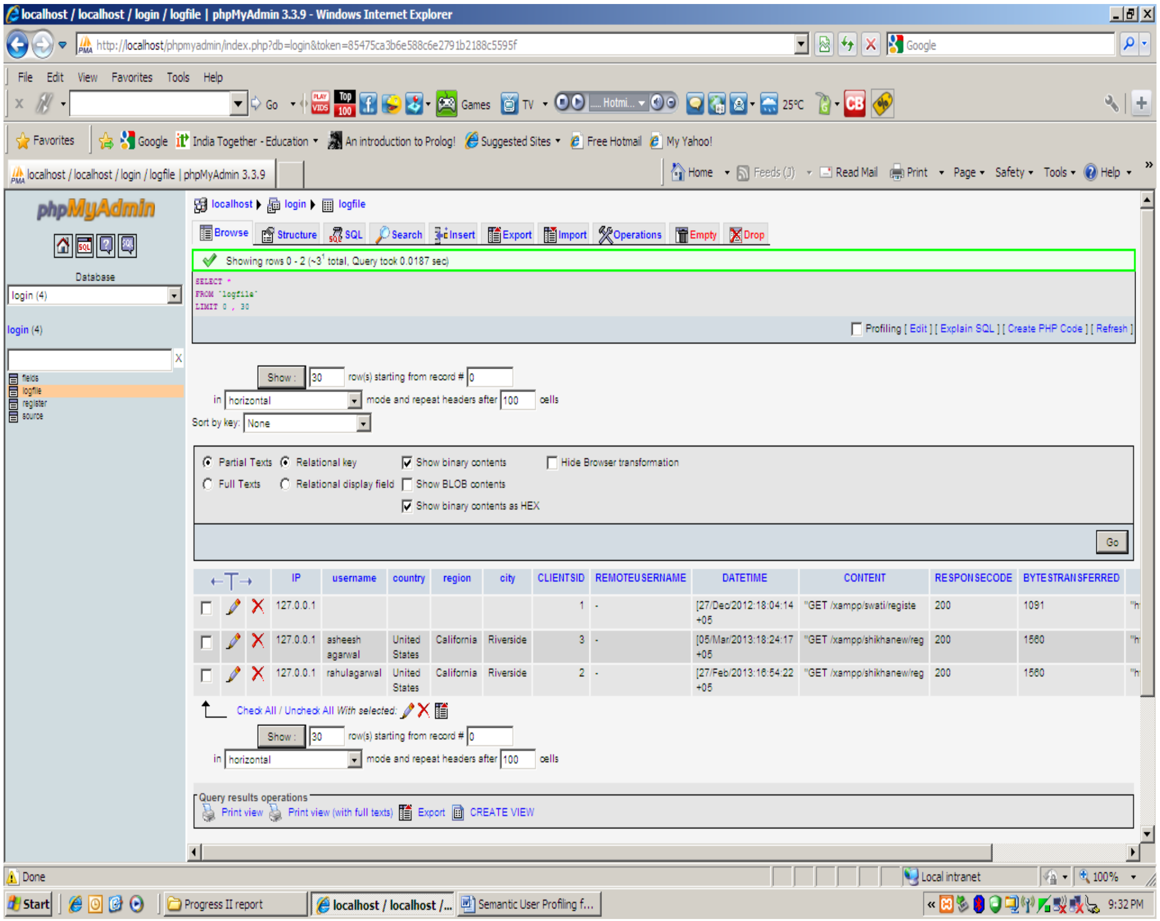
News items have been classified into 17 broad categories mentioned above. For each category ontology has been created using web ontology language OWL in Protégé. One is shown in the figure 10.

With the help of the example given in table 3 we are giving the steps of making user profile. We have identified news items given in table 2, read by a registered user (user ‘A’) in a session:

In news items given in table 3 we have identified following concepts, entities and relations, making use of ontologies.

1. CBSE is an organization.(c1 related to c2 by property ‘is an’)
2. School sports meet is a sport event.
3. Delhi is a state.
4. Delhi is located in India.
5. Private school is a type of school.
6. BHU is a university.
7. BHU is located in Varanasi. (Varanasi has been given a label Banaras also.)
8. Varanasi is located in state UP.
9. UP is a state.
10. UP is located in India.
11. Entertainment is an event held in University.

Underlined terms are relations. Here, ‘School’, ‘university’ and ‘Organization’ are classes. ‘School’ has subclass ‘Private School’.. ‘Geographic location’ is an entity having subdivision ‘continent’, ‘country’, ‘state’, and ‘city’. Organization has value ‘*CBSE’*. Country has value ‘*India*’. State has value ‘*Delhi’*, *‘UP’*. University has value *‘BHU’*. *Entertainment* is an event.


localhost / localhost / login / logfile | phpMyAdmin 3.3.9 - Windows Internet Explorer
http://localhost/phpmyadmin/index.php?db=login&token=85475ca3b6e588c6e2791b2188c5595f
File
Edit
View
Favorites
Tools
Help
phpMyAdmin
Database
login (4)
fields
logfile
register
source
localhost ▸ login ▸ logfile
Browse
Structure
SQL
Search
Insert
Export
Import
Operations
Empty
Drop
Done
Local intranet
100%
Start
Progress II report
localhost / localhost /...
Semantic User Profiling f...
9:32 PM

Showing rows 0 - 2 (~3¹ total, Query took 0.0187 sec)

```
SELECT *
FROM `logfile`
LIMIT 0 , 30
```

□ Profiling [Edit] [Explain SQL] [Create PHP Code] [Refresh]

Show : 30 row(s) starting from record # 0

in horizontal mode and repeat headers after 100 cells

Sort by key: None

● Partial Texts ● Relational key ☑ Show binary contents □ Hide Browser transformation

○ Full Texts ○ Relational display field □ Show BLOB contents

☑ Show binary contents as HEX

Go

	IP	username	country	region	city	CLIENTSID	REMOTEUSERNAME	DATETIME	CONTENT	RESPONSECODE	BYTESTRANSFERRED	
□	127.0.0.1					1	-	[27/Dec/2012:18:04:14 +05	"GET /xampp/swati/registe	200	1091	"h
□	127.0.0.1	asheesh agarwal	United States	California	Riverside	3	-	[05/Mar/2013:18:24:17 +05	"GET /xampp/shikhanew/reg	200	1560	"h
□	127.0.0.1	rahulagarwal	United States	California	Riverside	2	-	[27/Feb/2013:16:54:22 +05	"GET /xampp/shikhanew/reg	200	1560	"h

Check All / Uncheck All *With selected:*

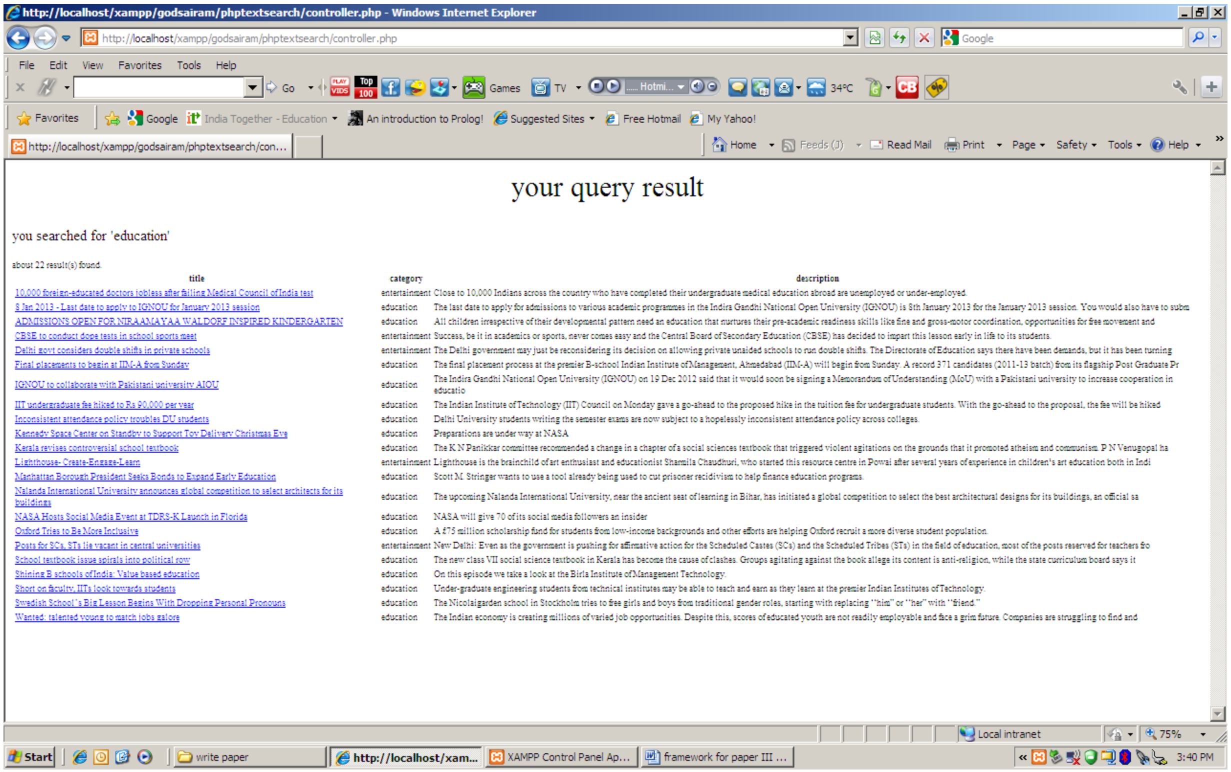
Show : 30 row(s) starting from record # 0

in horizontal mode and repeat headers after 100 cells

Query results operations

Print view Print view (with full texts) Export CREATE VIEW

Figure 8. Fields extracted from raw log file


http://localhost/xampp/godsairam/phptextsearch/controller.php - Windows Internet Explorer
http://localhost/xampp/godsairam/phptextsearch/controller.php
File Edit View Favorites Tools Help
Favorites
Google
India Together - Education
An introduction to Prolog!
Suggested Sites
Free Hotmail
My Yahoo!
Home
Feeds (J)
Read Mail
Print
Page
Safety
Tools
Help
Local intranet
75%
Start
write paper
http://localhost/xam...
XAMPP Control Panel Ap...
framework for paper III ...
3:40 PM

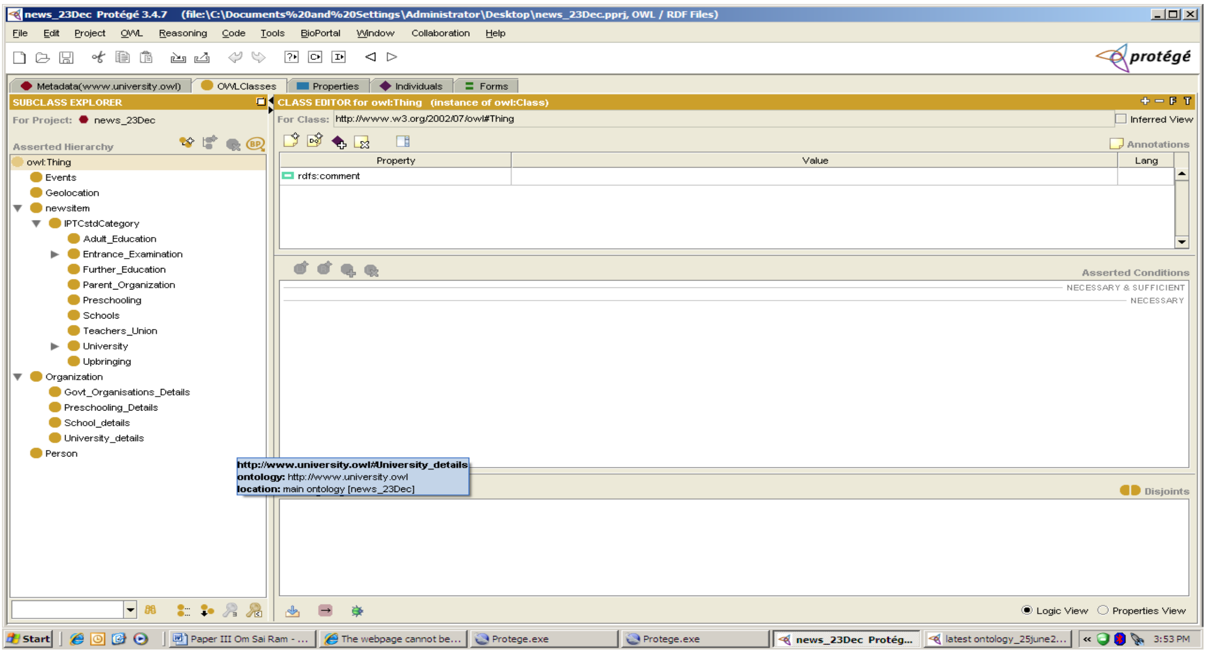
your query result

you searched for 'education'

about 22 result(s) found.

title	category	description
10,000 foreign-educated doctors jobless after failing Medical Council of India test	entertainment	Close to 10,000 Indians across the country who have completed their undergraduate medical education abroad are unemployed or under-employed.
8 Jan 2013 - Last date to apply to IGNOU for January 2013 session	education	The last date to apply for admissions to various academic programmes in the Indira Gandhi National Open University (IGNOU) is 8th January 2013 for the January 2013 session. You would also have to subm
ADMISSIONS OPEN FOR NIRAAMAYAA WALDORF INSPIRED KINDERGARTEN	education	All children irrespective of their developmental pattern need an education that nurtures their pre-academic readiness skills like fine and gross-motor coordination, opportunities for free movement and
CBSE to conduct dope tests in school sports meet	entertainment	Success, be it in academics or sports, never comes easy and the Central Board of Secondary Education (CBSE) has decided to impart this lesson early in life to its students.
Delhi govt considers double shifts in private schools	entertainment	The Delhi government may just be reconsidering its decision on allowing private unaided schools to run double shifts. The Directorate of Education says there have been demands, but it has been turning
Final placements to begin at IIM-A from Sunday	education	The final placement process at the premier B-school Indian Institute of Management, Ahmedabad (IIM-A) will begin from Sunday. A record 371 candidates (2011-13 batch) from its flagship Post Graduate Pr
IGNOU to collaborate with Pakistani university AIOU	education	The Indira Gandhi National Open University (IGNOU) on 19 Dec 2012 said that it would soon be signing a Memorandum of Understanding (MoU) with a Pakistani university to increase cooperation in educatio
IIT undergraduate fee hiked to Rs 90,000 per year	education	The Indian Institute of Technology (IIT) Council on Monday gave a go-ahead to the proposed hike in the tuition fee for undergraduate students. With the go-ahead to the proposal, the fee will be hiked
Inconsistent attendance policy troubles DU students	education	Delhi University students writing the semester exams are now subject to a hopelessly inconsistent attendance policy across colleges.
Kennedy Space Center on Standby to Support Toy Delivery Christmas Eve	education	Preparations are under way at NASA
Kerala revises controversial school textbook	education	The K N Panikkar committee recommended a change in a chapter of a social sciences textbook that triggered violent agitations on the grounds that it promoted atheism and communism. P N Venugopal ha
Lighthouse- Create-Engage-Learn	entertainment	Lighthouse is the brainchild of art enthusiast and educationist Sharmila Chaudhuri, who started this resource centre in Powai after several years of experience in children's art education both in Indi
Manhattan Borough President Seeks Bonds to Expand Early Education	education	Scott M. Stringer wants to use a tool already being used to cut prisoner recidivism to help finance education programs.
Nalanda International University announces global competition to select architects for its buildings	education	The upcoming Nalanda International University, near the ancient seat of learning in Bihar, has initiated a global competition to select the best architectural designs for its buildings, an official sa
NASA Hosts Social Media Event at TDRS-K Launch in Florida	education	NASA will give 70 of its social media followers an insider
Oxford Tries to Be More Inclusive	education	A £75 million scholarship fund for students from low-income backgrounds and other efforts are helping Oxford recruit a more diverse student population.
Posts for SCs, STs lie vacant in central universities	entertainment	New Delhi: Even as the government is pushing for affirmative action for the Scheduled Castes (SCs) and the Scheduled Tribes (STs) in the field of education, most of the posts reserved for teachers fro
School textbook issue spirals into political row	education	The new class VII social science textbook in Kerala has become the cause of clashes. Groups agitating against the book allege its content is anti-religion, while the state curriculum board says it
Shining B schools of India: Value based education	education	On this episode we take a look at the Birla Institute of Management Technology.
Short on faculty, IITs look towards students	education	Under-graduate engineering students from technical institutes may be able to teach and earn as they learn at the premier Indian Institutes of Technology.
Swedish School''s Big Lesson Begins With Dropping Personal Pronouns	education	The Nicolaigarden school in Stockholm tries to free girls and boys from traditional gender roles, starting with replacing ''him'' or ''her'' with ''friend.''
Wanted: talented young to match jobs galore	education	The Indian economy is creating millions of varied job opportunities. Despite this, scores of educated youth are not readily employable and face a grim future. Companies are struggling to find and

Figure 9. Classified news items for news category: Education


news_23Dec Protégé 3.4.7 (file:\C:\Documents%20and%20Settings\Administrator\Desktop\news_23Dec.pprj, OWL / RDF Files)
File Edit Project OWL Reasoning Code Tools BioPortal Window Collaboration Help
protégé
Metadata(www.university.owl)
OWLClasses
Properties
Individuals
Forms
SUBCLASS EXPLORER
CLASS EDITOR for owl:Thing (instance of owl:Class)
Inferred View
Annotations
Property
Value
Lang
Asserted Conditions
NECESSARY & SUFFICIENT
NECESSARY
Disjoints
Logic View
Properties View
Start
Paper III Om Sai Ram - ...
The webpage cannot be...
Protege.exe
Protege.exe
news_23Dec Protég...
latest ontology_25june2...
3:53 PM

For Project: news_23Dec

For Class: http://www.w3.org/2002/07/owl#Thing

Asserted Hierarchy

- owl:Thing
  - Events
  - Geolocation
  - newsitem
    - IPTCstdCategory
      - Adult_Education
      - Entrance_Examination
      - Further_Education
      - Parent_Organization
      - Preschooling
      - Schools
      - Teachers_Union
      - University
      - Upbringing
  - Organization
    - Govt_Organisations_Details
    - Preschooling_Details
    - School_details
    - University_details
  - Person

Property	Value	Lang
rdfs:comment		

http://www.university.owl#University_details
ontology: http://www.university.owl
location: main ontology [news_23Dec]

Figure 10. OWL ontology created in Protege

Concepts, session->	C1, tn1(s1)	C2, tn1(s1)	C3, tn1(s1)	C4, tn2(s2)	C2, tn2(s2)	e1 tn2(s2)	C5 (semantically related to c4)	C6 (related to c4)	e2 (related to c2)
CF-IDF(s) ->	0.4	0.3	0.2	0.6	0.2	0.5	-	-	-
Weight->	2	5	1	4	5	3	3.5	3.5	4.5
Interest type	Short Term	Long Term	Short Term	Short Term current	Long Term	Short Term current	expanded	expanded	expanded

Table 2. Identified concepts/entities and expanded concepts/entities

	News Item 1	News Item 2	News Item 3
User A (session 1)	"CBSE to conduct dope tests in school sports meet"	"Delhi govt. considers double shift in private schools"	"BHU goes hi-tech to provide better entertainment facilities"
User A			

Table 3. News items read by a user in a session

	News item1
User A (session 2)	"BHU goes hi-tech to provide better entertainment facilities

Table 4. News read by same user in next session

Table 4 presents news items read by the same user in session 2. It shows user interest in news about 'BHU'.

This shows that user A has more interest in news about concept 'university' having value 'BHU'. Gathering information from previous sessions gives long term interest of user.

In table 2, we have given weighs to the identified concepts (cn) and entities (en) in profile of registered user A, based on below mentioned certain factors. Table shows the concepts identified in two sessions (s1 and s2) of user 1 in time period tn1 and tn2. Concepts c1, c2, c3 are identified in session s1.

Concepts c2, c4 and entity e1 are identified in session s2.

Weight will be assigned based on both the time value and concept frequency. Maximum weight will be given to the concepts which are common in both the sessions. We will assign a time value to the concepts based on the timing of session. To give more importance to the currently read news items than older ones, time decay function will be used.

Interest value in topic (a) = ∑ [Interest Value in session n in topic (a) / x]

Where, x=positive integer depending on the time value of the session. More current is the session i.e., larger is the value of tn lesser is the value of the x.

User Profile(u1)=((t1, IntVal (i1)), (t2, IntVal (i2)),……)If tn1<tn2, then s2 session gives us information about latest interest of the user. We will give more weight to concepts identified in session s2. Common concepts in both the sessions will be given highest weight. This tells us about long term interest of the user. In above table concept c2 has been given highest weight(highest value is 5) as it is common in both the sessions. New concepts and entities in a session are ranked based on frequency calculation in a session. In s2, Concept c4 has been given rank 4 and e1 has been ranked 3 based on the values in second row of the table. Then in s1 remaining concepts are c1 and c3. Similarly in s1, c1 will be given rank 2 and c3 will be given rank 1.

Expansion of User Profile

We are expanding user interests by identifying the related concepts and entities in the ontologies. If a user is interested in concept c1, it may be interested in all the subconcepts of c1 also. If a user is interested in entity e1, it may be interested in entities closely related to e1 found in the ontology. This will give us the extended user interest profile. We will consider concepts of current session for expansion because the concepts which do not exist in current session will not be stored in user profile database. Concept/Entity (e2) semantically related to common concept c2 has been given more weight (4.5) than the concepts/entities (c5, c6) related to rest of the concepts(c4), weight is(3.5).

Clustering of user profiles

Clusters will be made, based on the similarities found in the user interests. User interest vectors contain ranked concepts and entities. Users having similar ranking for same concepts and entities are considered similar in nature.

VI. Conclusions and Future Work

Our main focus in the mentioned work is about creating and analyzing user profile. In general, the goal of user profiling is to collect information about the subjects in which a user is interested, and the length of time over which they have exhibited this interest. In our work we have handled various issues faced in making user profile. In order to improve the quality of information access and infer user's intentions, we are collecting user information both explicitly as well as implicitly. For semantic analysis of user interest, ontologies for various news item categories have been designed. IPTC standards have also been incorporated in ontology design process to make the system more interoperable. Making use of knowledge captured in ontologies in the form of hierarchical relations helps us to extend the user profile. We have also classified RSS feed news items into various categories of news domain.

We will now focus on recommending news items based on the profiles constructed and will also evaluate the user's trust in the system. Trust will help us to know user satisfaction level with the quality of recommendations and their acceptance ratio of our designed system. This will help to further improve our system.

References

[1] Stuart E. Middleton, Nigel R. Shadbolt and David C. De Roure, "Ontological user profiling in recommender systems" in ACM 1073-0516/01/0300-0034 2001.

[2] Pooja, Deepak, Bipin, Punam Bedi, "Trust enabled Argumentation based Recommender Systems", in International Conference on Intelligent Systems Design And Applications(ISDA) 978-1-4673-5118-8 PP 137-142, IEEE 2012.

[3] Wouter Ijntema, Frank Goossen et.al., "Ontology-Based News Recommendation" in EDBT 2010, Copyright 2010 ACM 978-1-60558-945-9/10/0003 ...$10.00, March 22-26 2010, Laussane, Switzerland ACM 2010.

[4] Lei Li, Dingding Wang, Daniel Knox e.al., "SCENE: A Scalable Two-Stage Personalized News Recommendation System" in ACM 978-1-4503-0757-4/11/07 Beijing, Chine 2011.

[5] Myung-Won Kim, Eun-Ju Kim et.al., "Efficient Recommendation for Smart TV Contents" in BDA 2012, LNCS 7678, pp. 158-167, Springer-Verlag Berlin Heidelberg 2012.

[6] Ivan Cantador, Pablo Castells, Alejandro Bellogin, "An enhanced semantic layer for hybrid recommender systems: Application to News Recommendation". Publication unknown

[7] Cristian E. Brigcuz, Fernando M. Sagui et.al. "System Architecture for Trust Based News Recommenders on the web", in Conngreso Argentino De Ciencias De La Computacion, pg. 211-220, CACIC 2011.

[8] Peter Brusilovsky, Jonathan Grady et.al., "Open User Profiles for Adaptive News Systems: Help or Harm?", in ACM 978-1-59593-654-7/07/005 Banff, Alberta, Canada 2007.

[9] Deuk Hee Park, Hyea Kyeong Kim et. Al., "A Review and Classification of Recommender Systems Research", in IPEDR vol. 5 2011, International conference on Social science and humanity, Singapore 2011.

[10] Elmar P. Wach "A heuristic as basis for an adaptive e-commerce recommender system" in 978-1-4673-5118-8, 12th ISDA IEEE 2012.

[11] Grace, Maheshwari et.al, "Analysis of web logs and web user in web mining" in IJNSA, Vol 3, No. 1, Jan 2011.

[12] Jinming Min, Gareth J.F.Jones, "Building User Interest profiles from Wikipedia Clusters" in SIGIR 2011 Workshop on Enriching Information Retrieval (ENIR 2011), beijing, China, july 28, 2011.

[13] Shikha Agarwal, Archana Singhal and Punam Bedi, "IPTC based Ontological Representation of Educational News RSS Feeds" in ITC 2012,LNEE, pp. 277-282, 2012. © Springer-Verlag Berlin Heidelberg 2012.

[14] Shikha Agarwal, Archana Singhal and Punam Bedi, "Classification of RSS Feed News Items Using Ontology" 978-1-4673-5118-8, PP 491-496, 2012 IEEE Presented in ISDA 2012, Kochi.

[15] K. R. Suneetha and Dr. R. Krishnamoorthi, "Identifying User Behavior by Analyzing Web Server Access Log File",IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009, pp. 327-332.

[16] P. Brusilovsky, A. Kobsa, and W. Nejdl (Eds.): The Adaptive Web, LNCS 4321, 2007, pp. 90–135

[17] http://wordnet.princeton.edu

[18] www.iptc.org/

[19] http://www.whatisrss.com/

[20] G. Salton, and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval" in 'Information Processing & Management 24 (5) , 513-523 (1988).

[21] G.Castellano et al.," LODAP: a log data preprocessor for mining web browsing patterns " in AIKED'07 Proceedings of the 6th

Conference on 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases - Volume 6 pp 12-17. (2007).

[22] Marta Gatius and Meritxell González, “ The use of Domain Ontologies for Improving the Adaptability and Collaborative Ability of a Web Dialogue System” in International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 5 (2012) pp. 185-194 © MIR Labs, www.mirlabs.net/ijcisim/index.html

[23] Hubert Kadima, Maria Malek, “Toward ontology-based personalization of a Recommender System in social network” in International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 5 (2013) pp.499508©MIRLabs,www.mirlabs.net/ijcisim/index.html

[24] A.B. Gil1, S. Rodríguez1, F. de la Prieta1 and De Paz J.F., “Personalization on E-Content Retrieval Based on Semantic Web Services” in International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 5 (2012) pp. 243-251© MIR Labs, www.mirlabs.net/ijcisim/index.html

[25] Sanjay K. Dwivedi1 and Anand Kumar2, “Ontology Exemplification and Modeling for aSPOCMS in the Semantic Web” in International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 5 (2013) pp. 542-549 © MIR Labs, www.mirlabs.net/ijcisim/index.html

Author Biographies

Shikha Agarwal is a researcher in Departmnt of Computer Science at University of Delhi, Delhi. Her research work focuses on personalization of new items recommendation using AI techniques and semantic web technologies. She has publeshed papers on ontological knowledge representation and news items classification using natural language processing and machine learning techniques.

Archana Singhal is working as Head of the Deptt. and as Associate prof. in Indraprastha College For Women, Delhi University, Delhi. She has recived her Ph.D. from Jawaharlal Nehru University, Delhi. Her research interests includes Natural language processing, Semantic Web, Multi-agent Systems, Agile development, Intelligent Software Engineering, Requirement Engineering, Information Retrieval and Ontologies. She has many publications to her credit**.**