Semantically enriched Massive Open Online Courses (MOOCs) platform

Leyla Zhuhadar^{a,*}, Sebastian Ryszard Kruk^b, Jerry Daday^c^a Dept. of Computer Information Systems, Western Kentucky University, Bowling Green, KY, USA^b Knowledge Hives sp. z o.o., Rokitnikowa 11, 81-589 Gdynia, Poland^c Dept. of Sociology, Western Kentucky University, Bowling Green, KY, USA

ARTICLE INFO

Article history:

Available online 16 April 2015

Keywords:

Massive Open Online Courses (MOOCs)
 Collaborative semantic filtering techniques
 Annotated information
 Computing for Co-learning
 Interpretability
 SemanticWeb
 Social networks

ABSTRACT

Massive Open Online Courses (MOOCs) are becoming an essential source of information for both students and teachers. Noticeably, MOOCs have to adapt to the fast development of new technologies; they also have to satisfy the current generation of online students. The current MOOCs' Management Systems, such as Coursera, Udacity, edX, etc., use content management platforms where content are organized in a hierarchical structure. We envision a new generation of MOOCs that support interpretability with formal semantics by using the SemanticWeb and the online social networks. Semantic technologies support more flexible information management than that offered by the current MOOCs' platforms. Annotated information about courses, video lectures, assignments, students, teachers, etc., can be composed from heterogeneous sources, including contributions from the communities in the forum space. These annotations, combined with legacy data, build foundations for more efficient information discovery in MOOCs' platforms. In this article we review various Collaborative Semantic Filtering technologies for building Semantic MOOCs' management system, then, we present a prototype of a semantic middle-sized platform implemented at Western Kentucky University that answers these aforementioned requirements.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Current MOOCs environments (i.e., Edx, Coursera, Udacity, Udemy, P2PU, etc.) utilize a hierarchical standalone structure. There is no semantic relationship among courses. The interface is more similar to the traditional digital library where searching for a specific resource would involve the process of finding an appropriate course among the listed categories of courses, then, searching for a specific learning object (LO), such as a video lecture, an article, an assignment, or a PowerPoint presentation. For instance, assume that a student started a course entitled "Data Mining", and he was struggling with an assignment for lack of knowledge in some statistical background. If MOOCs' learning objects were interconnected semantically and personalized recommendations were implemented, student would receive the right resource that he needed to revise to answer this assignment. Semantic technologies support more flexible information management than that offered by the current MOOCs' platforms. In this article we review various Collaborative Semantic Filtering technologies for building semantically Enriched MOOCs management system and then, we present

the prototype implementation of a semantic middle-sized platform at Western Kentucky University that answered these aforementioned requirements.

2. Background and related work to Collaborative Semantic Filtering techniques

In order to keep up with the growing amount of information published online, search engines cannot rely on the manual indexing process that was once used by online catalogs like Yahoo. However, completely automated indexing processes suffer from lower precision than manual ones. Collaborative filtering is the idea of automating the indexing process by using knowledge gathered in a social network. Since the time early implementations of collaborative filtering were developed, a number of methods have been proposed for collaborative filtering and social filtering. These methods are based on various statistical data like the precision and accuracy, while users' experience and knowledge is often omitted. In this article we present various aspects of social collaboration and describe an approach that improves on collaborative filtering techniques by constructing a network of collections maintained by the members of a social network, dubbed Social Semantic Collaborative Filtering (SSCF). Based on their level of expertise on

* Corresponding author.

a given topic, users collect small subsets of information and share these collections with other members of the social network.

2.1. Overview of collaborative filtering solutions

The term *collaborative filtering* was first introduced by (Goldberg, Nichols, Oki, & Terry, 1992) to denote a technique used to handle large amounts of emails using the Tapestry system developed at Xerox PARC. The authors contrast typical content-based filtering techniques with social filtering. They presented an algorithm for automating and scaling the process of information discovery through the social network of users with similar interests.

An early implementation of collaborative filtering was delivered by the GroupLens project. GroupLens included a Usenet news client that allowed users to annotate and rate each other's messages. It extended the ideas introduced by the Tapestry project and featured a scalable and open architecture. (Greeno et al., 1996) analyzed different item-based recommendation algorithms as a solution to generating recommendations over large data sets.

One of the first systems that successfully implemented collaborative filtering as an Internet service was Ringo, a music recommendation system developed by Upendra Shardanand. Ringo required an initial questionnaire in which users expressed their opinion on a set of music items registered in the Ringo database. The social recommendation-filtering algorithm implemented in Ringo exploited similarities between the tastes of different users to recommend music items. The algorithm was based on general trends and patterns within the taste of a person or group of people. The goal was to automate 'word of mouth' recommendations. The system was very popular in the 1990s and grew out of MIT into a commercial product called Firefly, which was later bought by Microsoft.

Some collaborative recommendation systems monitor the activities of their users and present recommendations based on the assumption that similar items interest those users who share similar interests. Systems by Amazon or Alexa Internet recommend items that were viewed by people who accessed a similar set of items. (Kamiya, Röscheisen, & Winograd, 1996) argue for a more holistic approach to capture user profiles in order to generate better collaborative filtering recommendations. They present the Grassroot system – a prototype implementation of their algorithms.

Netflix¹, a popular DVD and Blu-Ray renting service, uses a collaborative filtering technique for recommending interesting movies to its customers. Netflix uses the Cinematch recommendation algorithm, which analyzes accumulated movie ratings and, based on subscribers' interests, generates personalized predictions. The recommendations are provided based on the computed weekly list of similar movies combined with personalized real-time multivariate regressions.

In October 2006, Netflix began a competition, *The Netflix Prize*, for designing a recommendation algorithm that would outperform the precision of their own Cinematch solution. They released a training dataset with 100 million movie ratings, provided for nearly 18,000 movies by over 480,000 anonymized users; each rating was a quadruple of user, movie, date, and rating (between 1 and 5). This contest encouraged researchers to develop new, interesting recommendation algorithms, such as those presented by (Salakhutdinov, Mnih, & Hinton, 2007; Zhou, Wilkinson, Schreiber, & Pan, 2008). Finally, (Turnbull, 2003) presents a review of various information seeking models including the collaborative filtering techniques.

The collaborative filtering systems we have presented are based on automated approaches based on a statistical analysis of users' activities and interests. There is, however, another group of collaborative filtering algorithms; these approaches engage the users in the filtering and sharing process. They utilize the existing social connections provided explicitly by the user. (Maltz & Ehrlich, 1995) describe a system built on the common practice where people tell their friends or colleagues of interesting documents. The users of this system collect and share bookmarks on interesting World Wide Web pages that they have found. (Sugiyama, Hatano, & Yoshikawa, 2004) describe a social collaborative filtering system where users have direct impact on the filtering process. The changes in users' interests are exploited to provide thorough relevance feedback to the system. To format and distribute the collections of bookmarks a simple system has been developed. The *Simon* system allows users to create subject spaces; these spaces are lists of hypertext links to the WWW pages with annotations on them. Individual people can either use the bookmarks for keeping track of their own explorations or share their knowledge by sending it to the Simon server.

The Pointer system has been modeled after what people do informally when sharing information. Pointers can be distributed by: saving in a private database (bookmarking favorite documents), saving in public databases, emailing (to one user, a group of users, or distribution lists), or editing predestined documents called *Information Digests*. (Maltz & Ehrlich, 1995) discuss a solution for sharing information by finding a personal referral that can answer the given query. The quality and reliability of the answer depends on the distance, in the social network can be estimated based on the research on the small worlds phenomenon.

The network of relationships can also help in exploring the hidden web, the part of the Internet that is not indexed by search engines, as some information is deliberately not accessible outside the intranets. Studies have revealed that one of the most effectively used channels of dissemination of knowledge, especially in an organization, is an informal network of collaborators. In this approach, searching for information becomes a matter of searching the social network for an expert on the topic as well as providing a chain of personal referrals from a person searching for information to an expert.

(Basu, Hirsh, Cohen, & Nevill-Manning, 2002) introduce the hybrid filtering algorithm. It combines two techniques: item-based filtering and social filtering. The latter engages members of the social network in the process of filtering the information space. This technique, called *hybrid filtering*, is used to maximize the precision metric and ensure that the recall metric remains above the specified limit.

In the following sections, we present an algorithm for SSCF Model, then, we introduce the methodology we used to implement a semantically enriched MOOCs' platform as a prototype. Finally, we report on the results of the evaluation of SSCF.

3. Methodology for implementing semantically enriching MOOCs' platform

3.1. Social Semantic Collaborative Filtering (SSCF)

3.1.1. Scenario

In our example scenario (see Fig. 1) Alice writes a report on 'Mediation between Bibliographic Ontologies'. She registers with the semantically enriched digital library or a course in MOOC. She discovers that some of her friends are already registered as well. With features known from on-line communities, she connects her profile to her friends' profiles.

¹ Netflix: <http://www.netflix.com/>.

Fig. 1. The scenario of the basic SSCF model.

Later on, Alice starts to gather the information required for her report. She keeps links to resources she has found in collections managed by the online bookmarks system. Soon she discovers that resources that she has bookmarked do not cover the topic of the report at a satisfiable level. With the features provided by SSCF she tries to find other people within her neighborhood with higher expertise on the related topic.

3.1.2. SSCF Model

We introduce an approach to the problem based on the basic SSCF model. Each collection is annotated by the owner. The collaborative filtering feature in the digital library (MOOCs) lists all the collections, within the given range of social neighborhood, with topics related to the ones defined by Alice. The basic SSCF model is a tuple $M(P, C, G, F, T)$ where a set of users P , linked in a social network digraph $G_{peers}(P, F)$ through direct connections between peers F , maintains a set of collections P , each annotated with concepts from a graph T with various knowledge organization systems (taxonomies, thesauri, tags). We assume that each collection $c \in C$ has exactly one owner $P \in P$.

The basic model defines the following operations:

- **PeerCollection**: $P \rightarrow 2^C$ – returns all collections owned by the user.
- **OwnedBy**: $C \rightarrow P$ – returns the owner of the collection.
- **SubCollection**: $(C, C) \rightarrow \{0, 1\}$ – checks if one collection is a sub collection of the other.
- **Expertise**: $(P, C) \rightarrow [0, 1]$ – express the level of expertise the user P has in the topic represented by collection C ; in our model it denotes the quality of this collection.
- **Classification**: $C \rightarrow T$ – returns the list of topics describing the collection.
- **PeerDistance**: $(P, P) \rightarrow N$ – computes distance between two peers in the social network graph using Dijkstra algorithm.
- **Similarity**: $(T, T) \rightarrow [0, 1]$ – computes similarity level between two classification topics.
- **FinalRankingSM**: $(user, collection, PeerDistance, Similarity, Expertise) \rightarrow [0, 1]$ – computes ranking values for a collection $c \in C$ in the social network of a user $u \in P$ based on:
 - a distance to the owner = $OwnedBy(collection)$,
 - similarity level between classification topics owned by the user and the owner ($Similarity(T_{owner}, T_{user})$),
 - expertise measure of the owner ($Expertise(owner, collection)$).
- **knowsRange** – defines a maximal distance between two people when traversing the graph of social relations.

One of the possible ways to compute the expertise level of a collection is by analyzing the graph of collection inclusions. The more

people include the given collection in their collections the more important it is. The quality of the collection corresponds to the expertise level of the owner on a related topic. We are agnostic to which algorithm to use to compute the quality metric. For example it can be computed with the PageRank algorithm applied to the graphs of collection inclusions and the social network graph; (Marmolowski & Kielczyński, 2012) discuss other algorithms in more detail.

Algorithm 1: Find collections in the Basic SSCF Model

```

REQUIRE:  $p \in P$ 
ENSURE:  $C' \subset C$  owned by users in knowsRange degrees of
         separation from  $p'$ 
 $P_{used} = \emptyset$ 
FOR  $d = 1$  TO knowsRange DO
  FORALL  $p' \in P$  such that  $PeerDistance(p, p') = d$  DO
    IF  $p' \notin P_{used}$  THEN
       $P_{used} \leftarrow P_{used} \cup \{p'\}$ 
      FORALL  $c \in PeerCollection(p')$ 
        IF  $\exists c' \in C' \text{ SubCollection}(c', c)$  THEN
           $T_c = Classification(c)$ 
           $T_{c'} = Classification(c')$ 
          IF  $Similarity(T_c, T_{c'}) < \alpha_{threshold}$  THEN
             $C' \leftarrow C' \cup \{c\}$ 
          ENDF
        ENDF
      ENDF
    ENDF
  ENDF
   $C' \leftarrow \text{sort } C' \text{ according to } FinalRanking_{sm}$ 
RETURN  $C'$ 

```

In our scenario (see Fig. 1), Alice finds out that one of her friends, Caroline, gathers information on *digital libraries* and her expertise level on that topic is very high. Since Caroline included high-quality folders delivered by Damian and Eric (*libraries* and *Semantic Web* respectively) in her *digital library* folder, this information becomes automatically available for Alice as a recommendation from Caroline. Alice finds the *Semantic Web* collection from Eric very useful and she decides to link it directly under her *Bibliographic Ontologies Mediation*. She also links to the *Artificial Intelligence* folder from Bob. Alice can now take advantage of information gathered by her direct friends Caroline and Bob, as well as by other members of the social network (Damian and Eric) without bothering her direct friends.

3.2. A prototype for designing semantically enriched MOOCs' platform

In the previous section we have listed the current shortcomings of MOOCs' Platform. We have also presented an example of information seeking techniques: Social Semantic Collaborative Filtering. In this section we present JeromeDL, a prototype implementation of a semantic digital library that could be used to semantically enrich MOOCs.

JeromeDL is as prototype implementation of the concept of a semantic digital library; it is distributed on the BSD, open source license. It has been built as a joint initiative of the Gdansk University of Technology and DERI, National University of Ireland, Galway. The project started in 2004 and has been significantly redesigned compared to its predecessor: Elvis-DL². The

² Elvis-DL project: <http://elvis-dl.sf.net/>.

main goal of JeromeDL is to improve user experience in information seeking in digital libraries. JeromeDL has been designed and built with semantic web technologies in mind. It also supports communities of library users by implementing the social networking components. Compared to other solutions implementing the vision of semantic digital libraries, JeromeDL offers a complete, mature, and out-of-box solution. Compared to FEDORA, it does not require much time and effort to set up. Unlike (Aloia, Concordia, Meghini, & Barroso, 2007; Tummarello, Morbidoni, & Nucci, 2006; Caili, 2013), it has already been developed beyond the prototype stage.

3.2.1. Instantiation of Semantic Digital Library Architecture

JeromeDL is built with Java technology using a number of open source components, e.g., Lucene,³ Sesame⁴ and RDF2Go.⁵ AJAX communication is implemented using the prototype.js framework.⁶ The services offered by JeromeDL are implemented using Servlets/JSP technology. The business logic is implemented using Java beans and servlets; the results are delivered with JSP pages or serialization servlets. JeromeDL uses the content negotiation technique and a set of optional serialization parameters to deliver the results of an underlying business process execution in various RDF serializations (RDF/XML, N-Triples, N3, Turtle), subscription feeds (RSS, ATOM), and JSON and XML serializations. This clear cut between the view and the logic layers allows the current user interface provided by JeromeDL to be easily replaced with a custom user interface solution (UI agents). It has already been successfully exploited in the dContentWare project, which is built upon the JeromeDL infrastructure.

JeromeDL uses an RDF store for managing information about the library resources and users. JeromeDL (version 3.0) uses Sesame 2 with RDF2Go as an abstraction layer; it allows the RDF store to be changed effortlessly. The content of the binary resources (e.g., PDFs) and the RDF graph (store in the Sesame database) is fulltext indexed using the Lucene engine for faster and more reliable retrieval. We do not use a relational database due to a number of reasons:

- Using only an SQL database would limit expressiveness and flexibility of RDF and other semantic technologies.
- Based on the experience from the Elvis-DL project we learned that storing different pieces of information in different data models (RDMS, XML, RDF) poses integration problems.
- Industry-ready RDF storages like Oracle⁷ and Virtuoso⁸ provide RDMS functionality built into the store for more convenient access.

JeromeDL implements the semantic digital libraries architecture. It supports various types of digital library actors: from *communities of users* and *user interface agents* to *digital library designers* and *administrators* to *service developers* and *external services*. The data source components support different types of ontologies, including those identified in the article: structure, bibliographic, and community. The information is stored in the RDF repository and is indexed with the full-text engine. The high-level architecture of JeromeDL can be broken down into three layers of services and metadata management (see Fig. 5):

- **The bottom layer** handles typical tasks required from a digital objects repository, that is, keeps track of the physical representation of resources, their structure, and their provenance. The bottom layer utilizes structure ontology to provide a service for a flexible and extendable representation of library objects; this ontology is used to express the relations to other library resources.
- **The middle layer** raises the legacy bibliographic descriptions to the semantic level. It utilizes ontologies to represent the concepts defined in popular metadata formats, such as Dublin Core, MARC21, and BibTEX. The main advantage of the semantic layer is the services, which exploit machine-understandable, semantically rich relations between various kinds of resources; these services enhance the usability of information retrieval in the digital library, and provide interoperability with other digital libraries.
- **The top layer** utilizes the results of engaging the community of users to annotate and filter resources. On today's Internet the influence of user communities cannot be overestimated; collaborative efforts in information sharing and management proved to be the right way to go and have led to the success of many Web 2.0 sites.

In the following section we present the services implemented in JeromeDL on each of the above-mentioned layers; we will describe components used to deliver the semantic digital library services in JeromeDL.

3.2.2. Services implemented in JeromeDL

In this section we highlight the most interesting services offered by JeromeDL on each of three abstract layers (see Fig. 5).

3.2.3. Classic services

JeromeDL implements a number of services that are expected from a digital library management system. These services range from managing information objects to information retrieval to protecting content and services. JeromeDL uses the structure ontology for managing the information objects; it also uses the extensible access control ontology to define the access control policies for the library resources and services.

3.2.3.1. Managing information objects. JeromeDL allows the construction of complex library resources that can consist of articles, pages, multimedia parts, and attachments. JeromeDL utilizes the JOnto component⁹ to manage knowledge organization systems (KOS), such as thesauri, taxonomies, and authority files, used for annotating library resources (see Fig. 2). JOnto allows access to the index and the structure of KOS concepts represented in the SKOS ontology. JeromeDL provides special support for certain types of resources, such as antique books (using the Java Applet technology or the DjVu format¹⁰), Adobe Flash presentations, audio and video streams, and PDF files. These resources are full-text indexed and used as elements of articles or pages of other library resources. Other digital types can be attached to a library resource with the hasAttachment property. All binary resources are stored in the file system and their URLs in the RDF graph are relative to the storage directory.

JeromeDL uses the Lucene engine to maintain the full-text index and perform searches on the content of binary resources, and on the literal values and resources' names in the RDF graph with structure, bibliographic and community annotations. Additional libraries, e.g., PDFBox,¹¹ are used to index the aforementioned

³ Lucene Full-Text Indexing: <http://lucene.apache.org/>.

⁴ Sesame RDF Store: <http://www.openrdf.org/>.

⁵ RDF2Go: <http://ontoware.org/projects/rdf2go>.

⁶ prototype.js: <http://www.prototypejs.org/>.

⁷ Oracle RDF Store: <http://www.oracle.com/technology/tech/semantictechnologies/>.

⁸ Virtuoso: <http://www.openlinksw.com/virtuoso/>.

⁹ JOnto: <http://jonto.sf.net>.

¹⁰ DjVu: <http://djvu.org/>.

¹¹ PDFBox: <http://www.pdfbox.org/>.

Fig. 2. Annotating JeromeDL resources with the JOnto component.

special types of binary resources. The indexing service can also index web resources referenced by the library resources.

The backup service offered by JeromeDL allows the creation of backups of all resources, i.e., binary resources, library resources, dynamic collections, and user profiles, together with metadata. The backup is organized according to the main types of library resources; e.g., a book is backed up in a folder with other books. A single folder for each library resource contains all referenced binary resources and metadata in RDF. Additionally, the backup service can prepare an RDF document that contains only URIs with relative paths so that the prepared backup can be imported to another instance of JeromeDL. Finally, a library administrator can request the metadata to be exported in the DublinCore format compatible with DSpace.

3.2.3.2. Information retrieval. JeromeDL provides a simple full-text search interface using services offered by the Lucene engine. Additionally, users can choose to limit their results by searching only within given RDF properties; JeromeDL automatically translates properties defined in the DublinCore, BibTEX and MARC21 standards into the MarcOnt Ontology properties that are used during indexing.

JeromeDL implements complete support for the OAI-PMH standard. Digital library administrators can define which OAI-PMH providers should be harvested. By default the entire JeromeDL content can be harvested by other OAI-PMH services.

In addition to OAI-PMH, JeromeDL implements the HyperCuP P2P protocol. This protocol allows distributed querying in the network of digital libraries. JeromeDL instances can also be configured to allow browsing through the hierarchy of digital library systems. For example, a university instance can reference JeromeDL instances set up for the faculties and those in turn point to the department libraries. The users of the university library can then effortlessly browse through the collections delivered by the libraries set up in the departments. JeromeDL implements the OpenSearch protocol,¹² which allows, among other functions, a search through the library resources using the Firefox web browser on the A9 meta-search engine.¹³ Finally, JeromeDL allows browsing through the index of controlled vocabularies by simply clicking on a concept (such as keywords, topics, or authors) describing library resources (see Fig. 3).

3.2.3.3. Content and services protection. JeromeDL implements two types of content protection mechanisms. The first one allows

protection of the content of a certain type of resource against copying or printing; currently only resources with their source stored in the XSL:FO¹⁴ format can be protected from copying and printing, while PDF resources can only be protected from printing.

One of the optional modules for JeromeDL, called Extensible Access Control, allows definition of the access control policies that can be applied to both the resources and the REST services. This module uses concepts defined in the EAC ontology to define the access control policies. The EAC module provides a communication bus that allows the implementation of custom actions, such as applying the predefined licenses to the given types of resources or implementing the fair-use sharing policies.

3.2.4. Semantic services

JeromeDL extends the set of classic services with services that use semantic technologies. These services improve information seeking, support interoperability with other services, and allow the management of the identity of the library users. Although most of the information on which the semantic services operate is represented using concepts from the MarcOnt Ontology, these services are independent of the actual ontology used. Additionally, librarians are free to introduce their own concepts to the bibliographic description. These concepts are managed by the MarcOntX component and contribute to the MarcOnt Ontology development process in the MarcOnt Portal.

Certain information, such as the access control policies or the private data from users' profiles, have to be protected from the external services accessing the JeromeDL database through, e.g., the SPARQL endpoint. Therefore, most of the semantic services operate on the secure snapshot that contains only publicly available information.

3.2.4.1. Improving information seeking. JeromeDL delivers a number of search and browsing components, allowing users to choose the right tool for the current tasks:

The **Natural Language Query Templates** component allows digital library administrators to create templates (in multiple languages) for the most common questions users might ask. These templates are used to translate users' natural language questions into SPARQL queries. They allow complex questions, which could not be asked by a simple keyword-based search, to be answered; e.g., *what articles by students of Prof. Stefan Decker have recently been published?*

¹² OpenSearch protocol: <http://www.opensearch.org/>.

¹³ A9: <http://a9.com/>.

¹⁴ XSL Formatting Objects: <http://www.w3.org/TR/2001/REC-sx1-20011015/slice6.html#fo-section>.

Fig. 3. Navigation Filters in JeromeDL.

The **MultiBeeBrowse** component presented in article allows users to collaboratively browse the information space by filtering the current result set, browsing related resources, and finding similar resources according to given rules.

The **TagsTreeMaps** component (TTM) utilizes the relations between the controlled vocabulary concepts and renders using the treemaps algorithm. TTM allows the current view to be easily filtered by browsing through the taxonomy of concepts.

The **Dynamic collections** module allows the digital library administrators to create a hierarchy of views over the library database without being restricted to predefined taxonomies used when annotating the resources. The module takes into account annotations contributed by the community of library users.

The **Semantic Query Expansion** module refines a keyword-based query into a query with concepts from the knowledge organization systems (thesauri, taxonomies, authority files). It applies weights to the items in the query based on information from the user's profile, such as his/her bookmarks, recent searches, interests, etc.

The **Related resources** module displays library resources related to the one currently viewed. It uses semantic annotations on the library resources and can be expanded with plugins supporting new similarity metrics. Users can adapt the similarity function by adjusting weights used by each similarity plugin. Users can also define how many results each plugin should contribute to the overall recommendation.

Finally, JeromeDL can be extended with other information seeking solutions, for example, with the Exhibit faceted navigation component from the SIMILE project.

3.2.4.2. Interoperability with other services. One of the challenges for semantic digital libraries is to improve interoperability with other library systems and online services. JeromeDL realizes this goal by using the MarcOnt Mediation Services module, which allows translation between the semantic descriptions, used in JeromeDL and the legacy metadata formats, such as MARC21, BibTEX, and DublinCore. Additionally, JeromeDL exposes the semantics used to render each view (e.g., search, browsing, or filtering) using standard metadata publishing solutions, such as Microformats,¹⁵ eRDF,¹⁶ SIOC, RSS, or pointing to an RDF document using <link> element in the HTML page.

3.2.4.3. Managing identity of library users. Information about the creators and the contributors of library resources (the authority

files) is managed in JeromeDL using FOAF metadata. JeromeDL utilizes the FOAFRealm service for efficient management of this information. The same component is also used to manage the identities of the library users. FOAFRealm allows users to maintain their social networks; it also provides an indispensable service for the social services implemented in JeromeDL.

3.2.5. Social services

Social services implemented in JeromeDL further improve user experience in information seeking by capturing the community annotations and encouraging knowledge sharing. The ontologies used to express information on this layer are mainly the SSCF Ontology, the S3B Tagging Ontology, and the SIOC Ontology. The most valuable and the most appreciated service is the social semantic collaborative filtering component implemented in JeromeDL using the FOAFRealm service. It allows users to share their bookmarks with their friends. This component also provides valuable information to the Semantic Query Expansion service; the system can also compute an extrapolated user profile based on the profiles of their friends registered in the library. The SSCF component in JeromeDL has also been extended with the recommendation engine written in Prolog. This allows improvement of the knowledge flow in the social network by recommending the bookmark folders on topics interesting to the library user. This engine utilizes the semantic annotations on the users' profiles (e.g., foaf:workHomepage) and on the bookmark folders.

JeromeDL allows users to leave comments and rank library resources. It supports SIOC metadata to encode users' comments; it allows library resources to be represented as blog posts with comments from the library users. Users can also annotate excerpts of multimedia resources, i.e., a region of interest in a photo, a fragment of a video, or an audio stream.

4. Implementation of JeromeDL prototype at Western Kentucky University (WKU)

In 2007, Western Kentucky University started an initiative to provide online students with an open source repository of lecture. The platform is named the *HyperManyMedia*¹⁷ repository. This platform is running on a local server¹⁸ at WKU.¹⁹ Our designing approach is a *user-centered design* which is driven by users' needs; for more

¹⁵ Microformats: <http://microformats.org/>.

¹⁶ eRDF: <http://research.talis.com/2005/erdf/wiki/Main/RdfInHtml>.

¹⁷ We proposed this term to refer to any educational material on the Web (hyper) in a format that could be a multimedia format (image, audio, video, podcast, vodcast) or a text format (Webpage, PowerPoint).

¹⁸ <http://hmm.wku.edu>.

¹⁹ <http://www.wku.edu>.

Fig. 4. HyperManyMedia Open Source Courses.

details, refer to the following research (Zhuhadar & Nasraoui; Zhuhadar et al., 2009c; Zhuhadar, Nasraoui, & Wyatt, 2009a; Zhuhadar, Nasraoui, & Wyatt, 2009b). The objective of this design is to provide users with an easy access to online learning objects (LO) in a variety of formats (audio, video, podcast, vodcast, html, pdf, or PowerPoint). Over the last two years, the growth of content in this repository became an overload factor for users (learners) to find the needed information. We noticed that searching for learning objects in this repository via classical techniques of searching and browsing or through static taxonomies was insufficient. However, Web 2.0 and Web 3.0 introduced new paradigms of tools that provided interoperability between multiple platforms, integration of folksonomies represented as social tagging ("folksonomy is an Internet-based information retrieval methodology consisting of collaboratively generated, open-ended labels that categorize Web content"), social/collaborative (Kruk & McDaniel, 2008) bookmarking, dynamic taxonomies, semantic annotations, etc. Therefore, a new vision of our initiative was proposed. The main goal is to keep the interest of our online learning communities in our online learning materials without the need to duplicate the resources; in addition, we provide learners with advanced tools, such as social tagging, social bookmarking, question/answering querying based on natural language, etc.

We summarize the outcome of the new architectural design as an architecture that provides dynamic taxonomies and faceted search capabilities. We note that the previous architecture of the *HyperManyMedia* repository was designed as an application independent and reusable. The platform was designed in a way that we separated the resources (learning objects) from the design and implementation of the platform. This separation enables us to move from platform to platform without duplicating our resources. A reusability concept is considered as a building block in our ongoing architectural design.

HyperManyMedia repository consists of 11 field of disciplines (English, History, Mathematics, Chemistry, Management, Accounting, Engineering, SocialWork, Architecture and Manufacturing Sciences, Communication Disorders, Consumer and Family Sciences). Currently, we have 64 courses, 7424 learning objects (lectures), and each learning object represented in seven different formats (text, powerpoint, streamed audio, streamed video, podcast, vodcast, RSS). There is a total of ~51,968 individual learning objects. These materials were created by Western Kentucky University and located on the *HyperManyMedia* E-learning repository and augmented with external open source resources from MIT OpenCourseWare.²⁰

4.1. Redesigned architecture

A synergistic approach among the Semantic Digital Library *JeromeDL* and *HyperManyMedia* repository was proposed to provide users (online learners) with enhanced information discovery features for learning objects (lectures). The redesigned architecture serves as a platform to (a) author learning objects; (b) classify each learning object employing proper taxonomy using different libraries, such as DMOZ, ACM, UDC, LOC, or DDC; (c) bookmark sharing and collaborative filtering; and (d) provide natural language query templates.

The architecture of the Semantic Digital Library, *JeromeDL*, consists of three components: (1) system (semantic services and social services), (2) content (multimedia resources, dynamic objects, community annotations, and semantic annotations), and (3) users (community and aggregation) (Kruk, Gzella, Dobrzanski, McDaniel, & Woroniecki, 2007). This architecture assures a high level of usefulness, usability, and performance.

Our main objective for redesigning *HyperManyMedia* repository is to provide learners with a social semantic E-learning repository where each resource is described using three types of metadata: structure, learning objects-aware ontologies, and community-aware ontologies. *JeromeDL* delivers the three types of metadata in one platform; therefore, users are presented with ontological representation of each learning resource.

In addition, users are provided with social semantic collaborative filtering which enables learners to actively participate in the process of knowledge representation (Kruk et al., 2007). In the following section we provide the methodology we pursued to redesign *HyperManyMedia* repository.

In addition, users are provided with social semantic collaborative filtering which enables learners to actively participate in the process of knowledge representation (Kruk et al., 2007). In the following section we provide the methodology we pursued to redesign *HyperManyMedia* repository.

JeromeDL, the Semantic Digital Library, consists of a three-layered architecture of metadata management. We used the bottom layer (Digital Library Services) to represent the *HyperManyMedia* resources (learning objects); it uses ontology to define each learning object. This method provides flexibility for other services to interact with those resources and to provide links to other resources. The middle layer, the main objective of this layer is to provide bibliography for existing resources in the digital library, such as a book, article, etc. We modified the usage of this layer to provide the semantic description of the learning materials, such as audio, video, podcast, vodcast, or a text document.

²⁰ <http://ocw.mit.edu/OcwWeb/web/home/home/index.htm>.

Therefore, the main purpose of this layer is to (a) store the resource; (b) deliver metadata about the resource using popular existing format (Dublin Core,²¹ MARC21,²² or BibTEX²³); (c) manage the resource; (d) information retrieval services, such as semantic search, natural language search, etc.; and (e) provide social networking using FOAF ontology (Kruk et al., 2006). The upper layer provides community-oriented services, such as tagging, blogging, and collaborative filtering for the online students. Fig. 4 represents the HyperManyMedia platform. The platform is located on a local server²⁴ at Western Kentucky University. The methodology used to construct the learning objects is the following:

- **Cover:** represents a thumbnail picture of the lecture.
- **Abstract:** represents a snippet from the lecture.
- **Author information:** (author, and/or editor, and/or publisher).
- **Domains:** suggests the taxonomies to which the lecture belongs.
- **Keywords:** provides an easy way to search for the learning object.
- **RDF:** presents each learning object with its own RDF.
- **Bookmarks:** provides methods for social bookmarking.
- **Resource:** links each resource URI (Our main goal was not to duplicate the already existing resources, so we provided a direct link to our learning objects in the HyperManyMedia repository).
- (for example, Fig. 5 represents an instance of adding a resource to the College of History).

Algorithm 2: Augmented Metadata Search Engine

```
<html xmlns="http://www.w3.org/1999/xhtml">
<!-- InstanceBegin template="/Templates/1.252J.dwt"
codeOutsideHTMIsLocked="false" -->
<head>
<meta name="college" content="ingenieria">
<meta name="course"
content="planificaci3ndeltransporteurbano">
<meta name="instructor"
content="proffredericksalvuccimikelmurga">
<meta name="language" content="spanish">
<!-- InstanceBeginEditable name="metalect" -->
<meta name="lecture"
content="sistemasdetransporteinteligentes"/>
<!-- InstanceEndEditable --> <meta name="format"
content="text">
<meta http-equiv="Content-Type" content="text/html;
charset=ISO-8859-1">
<!-- InstanceBeginEditable name="doctitle">
<title> Curso: Planificaci3n del transporte urbano. Oto3 2002;
Prof.
Frederick Salvucci, Mikel Murga; Conferencia: Sistemas de
transporte
inteligentes ; Formato: texto</title>
```

4.2. Adding resources to the semantic HyperManyMedia

Fig. 6 describes the process of uploading a resource. This process divided into two sections:

- **Providing general information:** In this section, we define

- The digital type of the learning object: This can be one of the following (URI, xslfo, xml, pdf, rtf, or swf). In Fig. 7 we define the type as URI and we provide the actual location of the learning object.
- The type of the resource: We predefined 11 types of resources; in this special case, the resource belongs to the History resources and it is a document (lecture). Therefore, we select “History Lectures”. Belong to Collection: In this section, we emphasize to which collection from the College of History this resource belongs, more specifically, to which course in History. In this special case, we define it under the course “Western Civilization since 1648”.
- Title: We provide the lecture’s title.
- Abstract: A small meaningful snippet from the lecture.
- Cover: Thumbnail picture from the real lecture.
- Creator, publisher, and editor.
- **Providing additional information:** This section is the most important element in defining the resource. Fig. 8 illustrates this section:
 - Folksonomies: It links the resource to specific taxonomies to which this resource belongs. JeromeDL provides five different folksonomies libraries from which that we can choose (ACM, UDC, DMOZ, LOC, DDC). This definition is very essential since it is linked to the way the user links the resource to other services in the library. It provides the ontology of this resource.
 - Keywords: Users can add keywords that present this resource, such as (most frequent keywords used in the lecture, the name of the professor, the name of the course, etc.).

4.3. Methodology used for designing the ontology

We used Protégé,²⁵ an open source ontology editor and knowledge-based framework that supports two ways of modeling ontologies – (1) Protégé-Frames and (2) Protégé-OWL editors – to design and build the structure of HyperManyMedia ontology. Our current ontology consisting of ~32,000 lines of code.²⁶ The main question is how to design an ontology that can summarize the whole domain? We used two concepts: *Formal Context Representation (FCR)* and *Semantic Factoring (SF)*, refer to Phase VII: *Visual Ontology-based Search Engine*, for more details on the definitions and the visual implementations.

Constructing the HyperManyMedia Ontology: Fig. 8 depicts the upper-level of HyperManyMedia Ontology in Protégé. This figure describes the classification of the ontology. The highest level is “Thing” and underneath it is the definition of the five major entities (College, Course, Language, Lecture, and Professor). However, since we extended the domain ontology into the multilingual domain (English and Spanish), we need to define the same entities in Spanish. Protégé provides the user with the capability to create any type of relationship that fits any structure needed. In our case, we defined the following entities: *has_College*, *has_Course*, *has_Language*, *has_Lecture*, *has_Professor*, *sub_Class_Of*. In addition, each entity has different characteristics (*Functional*, *Description*).

In Fig. 9, for example, we illustrate the characteristics of *Entity = has_College*: (*Description*: College, *Equivalent classes*: Colegio, *Superclasses*: Thing, *Members*: Accounting, Architecture_ and Manufacturing, Biology, etc., *Disjoint*: Sub classes).

Two most important definitions used in our ontology design are the following:

²¹ Dublin Core: <http://dublincore.org/documents/dcq-html/>.

²² MARC21: <http://www.loc.gov/marc/bibliographic/>.

²³ BibTEX: <http://www.bibtex.org/>.

²⁴ <http://161.6.105.165:8080/jeromed1/>.

²⁵ <http://protege.stanford.edu>.

²⁶ Double Blind Peer Review policy-should not be disclosed to the reviewer.

Fig. 5. HyperManyMedia Open Source repository based on a Semantic Digital Library (JeromeDL).

Uploading resource

General information

digital type of resource:

uri:

type of resource:

belongs to collection: ☐ english, ☐ Working With Young Children & Families, ☐ Mgmt Of Nonprofit Org, ☐ Management, ☒ Western Civilization since 1648

title:

abstract:

cover:

protection type of resource:

preprinted? ☐

Creators: [Search]

Available persons:

..Creators..
WKU Distance Learning[x]

..Publishers..
foefadmir[x]

..Editors..
Leyla Zhuhadar[x]

Fig. 6. Uploading a resource to HyperManyMedia (Part-I).

Fig. 7. Uploading a resource to HyperManyMedia (Part-II).

- 1) *Equivalent classes*: Equivalent classes: equal to $\equiv$ relation, to mention some of these entities (*College* $\equiv$ *Colegio*, *Engineering* $\equiv$ *Ingenieria*, *English* $\equiv$ *Ingles*, ..., *Social Work* $\equiv$ *Trabajo Social*, *Chemistry* $\equiv$ *Quimica*, etc.)
- 2) *Sub_Class_Of*: Related to the hierarchy design of our domain: <Cluster_Of Course>, <Course is – a sub_Class_Of College>, etc.

Examples of Cluster Features: Fig. 10 depicts the location of the descriptive features that have been obtained from the Cluster Analysis phase. These features have been added to each lecture belongs to the course “Game Theory for Managers” under the Accounting domain.

5. Evaluations

5.1. Evaluation of Social Semantic Collaborative Filtering

The SSCF approach differs from classic collaborative filtering because it utilizes the social relations given explicitly by the users

instead of computing recommendations based on a social network artificially created by the collaborative filtering algorithm. In this section we report on a model evaluation of the previously presented SSCF algorithms.

5.1.1. Hypothesis

Our claim is that the overall social network becomes better informed when using the social semantic collaborative filtering technique to disseminate information. In other words, the hypothesis claims that the members of a social network gain access to higher-quality information contributed by the domain experts.

5.1.2. Simulation model

The simulation model was based on the similar ideas defined in the Referral Web project. The main difference between SSCF and the Referral Web is that in the Referral Web the user performs project the process of finding an expert on a certain topic manually. In SSCF, semantic annotations on the knowledge provided in the social network are used to automate the process of finding

Fig. 8. Augmented Ontology: level-1.

the highest quality of information. The simulation model itself might be similar to the one presented in, so we just need to prove that it is possible to find an expert within the given maximal degree of separation. Our model is based on two networks: a social network and a network of collections with information gathered by the users. Fig. 11 depicts an example of a social network (seen from the perspective of a single person) overlaid with the network of collections. Additionally, each collection is annotated using Dewey Decimal Classification. This simplifies the model in the sense of computing similarities between topics. Although in the real world implementation categories are described with other knowledge organization systems, DDC seems to be enough for the modeling purposes (see Fig. 11).

The network of collections is created based on the taxonomical relation between DDC categories, i.e., similarities between topics. Each collection is owned by one (and only one) member of the social network. An expertise level of the given user on the topic associated with his/her collection is represented by a real number in the range [0, 1].

5.1.3. Questions for evaluation

We have identified two questions for this evaluation. Answers to these questions will test the given hypothesis:

Can a user access, within six degrees of separation, information gathered by the domain experts traversing his/her social relations and the network of collections?

Is average expertise, to which members of a social network gain access, higher than the average expertise of single (not connected) members?

This first question derives from the Milgram experiment: we want to determine if the expertise passed on through the network of collections can be reached by all members of the social network within six degrees of separation. With the second question, we want to evaluate if the average expertise of each member of the social network is improved by passing information through the network of collections.

5.1.4. Assumptions for the evaluation model

In our model of SSCF, each user manages collections with information on selected topics. The different users represent different expertise on the given topic. We assume that:

- The quality of the information provided by a user on a certain collection is proportional to the expertise level of the user on the topic of the collection.
- It is possible to find a user with high expertise on the given topic within the social network.

According to the simple social collaborative filtering model, the simulation environment includes a set of users and a set of collections managed by those users. The quality of the collection is based on the user's expertise on a related topic. Each user knows a

Fig. 9. Augmented Ontology: level-2.

number of other users; however, the social relation is not implicitly considered as symmetric.

Although according to the Small World Phenomena the distribution of the degree of the social connections is power-law based (Zipf's distribution), we have decided to perform a second set of experiments where the degree of social connections is a bell-curve shaped (normal random variable); this distribution was suggested by some researchers to apply to specific types of social networks, e.g., a network of academics.

We use Lotka's Law to define the distribution of expertise on a certain topic within the social network. The number of authors making n contributions is n^α times less than those making only 1 contribution; where α is often nearly 2. We assume that the expertise on the given topic is proportional to the number of high-quality publications. Hence, the distribution of expertise in the social network can be modeled using Lotka's Law.

5.1.5. Definition of the experiment

During the experiment each user ($p \in P$, $\text{sizeOf}(P) = N_p$) tries to find, in the social network within a given range R , a collection that provides information on the topic $t \in T_p$. The topic is randomly selected from the list of topics associated to collections owned by the user. The average value of the highest expertise $\bar{E}_{\max}(R)$ level found within the given range (R) is computed using Algorithm 3.

Algorithm 3: AverageMaximalExpertise (R) – calculating average value of the highest expertise level found within a given degree of separation

```

REQUIRE:  $R, N_p$ 
ENSURE:  $\bar{E}_{\max}$  (that is the average value of the highest
expertise level in a given range  $R$ )
FORALL  $p' \in P$  WITH SELECT  $t \in T_p$  FIND  $c$  THAT
     $t = \text{Classification}(c)$ 
     $\text{PeerDistance}(p, \text{Owner}(c)) < R$ 
     $e = \text{expertise}(\text{Owner}(c), c)$  IS MAXIMAL
     $\bar{E}_{\max} += \frac{e}{N_p}$ 
ENDFOR
RETURN  $\bar{E}_{\max}$ 

```

In our experiment the social network consisted of $N_p = 1000$ users. Each user in our social collaborative filtering environment had only one associated collection. This simplification was correct since during the experiment we were only looking for collections with exactly the same topic as the one selected; therefore, collections associated with each topic create a subgraph that is independent of the actual number of collections owned by each user. The expertise level for each collection has been randomly selected according to the power law distribution. In the first experiment

Fig. 10. Descriptive features added to *Game Theory for Managers*.

Fig. 11. Evaluation model from the perspective of a single person.

the degree of social connections was randomly selected according to normal distribution ($m = 25$, $s = 12.5$) in the second experiment the power law distribution ($\theta = 1.9$) was applied. During each experiment average maximal expertise values $\bar{E}_{max}(R)$ were calculated for the given degree of separation $R \in [1, 6]$.

5.1.6. Results of the experiment

Based on the data gathered during the experiment we have calculated an average maximal expertise within the given scope (R) of the social network (see Table 1 and Fig. 12). These data will help us to answer the questions for the evaluation:

- 1) **Can a user access, within six degrees of separation, information gathered by the domain experts traversing his/her social relations and the network of collections?** In the social networks of Zipf's distribution of social relations the maximal expertise within six degrees of separation is 91%, which can be interpreted such that users can access the expertise of a domain expert through the network of collections. In the case of the special types of social networks with

Fig. 12. An average maximum expertise level in the social neighborhood of the given size.

Table 1

Results of the experiment – average maximal expertise $\bar{E}_{\max}(R)$.

R	Zipf ($\theta = 1.9$)	Bell ($\sigma = 12.5$)
0	0.00749	0.00749
1	0.01694	0.06750
2	0.09800	0.69145
3	0.31440	0.96291
4	0.62273	0.96789
5	0.82938	0.96789
6	0.91352	0.96789

bell-curved distribution of social relations, a member of the social network can gain access to even higher expertise within even three degrees of separation.

- 2) **Is the average expertise to which members of a social network gain access higher than the average expertise of single (not connected) members?** We have computed the average expertise of members of the social network ($R = 0$). In case of both types of distributions of social relations the

average expertise of a single member of a social network is much lower than in the network within even one degree of separation.

Taking into account the positive answers to the above-mentioned questions, we can conclude that our hypothesis holds, i.e., the overall social network becomes better informed when using the SSCF approach for disseminating information.

Following the experiments by (Kautz, Selman, & Shah, 1997), we have constructed a similar social collaborative filtering model. The results revealed that each user is able to find (on average) the best quality of information provided by other users within the sub-graph of a social network limited to six degrees of separation. These experimental results proved that the constructed social network model corresponds to the small world phenomena. Hence, the assumptions underlying the SSCF approach have been fulfilled: the overall social network is better informed and it is possible to find an expert (with an average expertise level above 90%) within the social network neighborhood.

5.2. Evaluation of the HyperManyMedia platform

JeromeDL platform provides three types of search facets:

- 1) Simple Search
- 2) Advanced Search, and
- 3) Semantic Search.

We examined each one of them as follows: We tested the simple search engine with keywords extracted from the contents of the learning objects. The search engine was able to retrieve documents with high accuracy. In advanced search, users can search by titles, authors, editors, publishers, dates, etc. Fig. 13 illustrates the results retrieved for keywords search “Russian Civil War”, with additional properties (search by title).

The retrieved results present the following information: title, authors, abstract, and the URI for the actual document that contains the query term as title. In general, the Semantic search engine provides the user with two types of search:

Results of search for resources (advanced)

Search for in content

Search for in

☐ All query parameters must be matched in the results page \n

☐ Perform distributed search

Search results

Results 1 - 1 of 1

Search results

Title	Russian Civil War
authors	WKU Distance Learning
abstract	The Whites were led by an admiral by the name of Alexander Kolchak and he was also supported by Japanese forces again the whole premise was to get the Russians so involved that they in some parts might be eventually persuaded and concerned to enter back into World War I. The Whites came from Siberia and were almost able to carry their war to the Volga River in the center of Russia by mid 1919 Whites under General Anton Denikin and he would be supported by the French carry their forces all the way through to the Ukraine and almost to Moscow although three of the White Bolshevik armies aided by the Brits seemed to be ready to claim victory the Red Bolsheviks pushed them back and by 1920 the Ukraine was taken followed by Georgia, Russian Armenia, and Azerbaijan in 192.
URIs	
accuracy: 100.0% publication type: History Lectures presentation form: Link to external resource (URI) Keywords: Czar, Western Civilization since 1648, Hardin, Red Bolsheviks Domains: Social sciences, Social sciences, World history and history of Europe, Asia, Africa, Australia, New Zealand, etc., Geography & history, Recreation, World	

Fig. 13. Advanced search engine.

Results of search for the resources with RDF and NL interface

Natural Language Query
RDF Query

Select template: Show me all publications in the library. and tweak it ...

Show me all: Show me all publications written by ...

Show me all publications written by friends of ...

Show me all publications written by students of ...

Find articles related to the subject of ... (in the context of ... domain)

Search results

- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*Hitler and Nazis Germany* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*Rhyme Schemes Sonnets* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*Foundations Of World Dominance* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*Constitutional Monarchy in England* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*Poetry Lines* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*S-O* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*World War II* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*Poetry Lines Person Irony* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*Beginnings of the Cold War* [Edit] [Remove]
- ▶ WKU Distance Learning , Leyla Zhuhadar (Ed.) :*Accounting Third* [Edit] [Remove]

Fig. 14. Semantic search engine.

- 1) Based on Natural Language Query (in this case, additional help is provided to tweak the query, as shown in Fig. 14, the user can search the entire courses or by resources written (uploaded by) specific user, or written by friends, and by
- 2) RDF Query (in this case the user can choose one of the following query languages, such as SERQL, SRQL, or RDQL).

6. Conclusion

In this article we reviewed various Collaborative Semantic Filtering technologies for building Semantic MOOCs management system, then, we present prototype implementation of a semantic middle-sized platform at Western Kentucky University that answers these aforementioned requirements. We showed how we envisioned the new generation of MOOCs' platform. We illustrated how semantic technologies support more flexible information management than that offered by the classical MOOCs' platforms. We also presented how annotated information about learning resources can be composed from heterogeneous sources, including contributions from the communities in the forum space. These annotations, combined with legacy data, build foundations for more efficient information discovery in MOOCs' platforms. In our future work, we will incorporate newer infrastructures, such as, Alice (Zhang, Liu, Ordóñez de Pablos, & She, 2014).

References

- Aloia, N., Concordia, C., Meghini, C., & Barroso, J. (2007). Querying a bricks digital library.
- Basu, C., Hirsh, H., Cohen, W. W., & Nevill-Manning, C. (2002). Technical paper recommendation: A study in combining multiple information sources. *Journal of Artificial Intelligence Research*, 14, 241–262.
- Caili, X. (2013). A research on CallimachusDL model based on Semantic Web. *Research on Library Science*, 3, 008.
- Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70.
- Greeno, J. G., Collins, A. M., Resnick, L. B., et al. (1996). Cognition and learning. *Handbook of Educational Psychology*, 15–46.
- Kamiya, K., Röscheisen, M., & Winograd, T. (1996). Grassroots: A system providing a uniform framework for communicating, structuring, sharing information, and organizing people. *Computer Networks and ISDN Systems*, 28(7), 1157–1174.
- Kautz, H., Selman, B., & Shah, M. (1997). The hidden web. *AI Magazine*, 18(2), 27.
- Kruk, S. R., Samp, K., O Nuallain, C., Davis, B., McDaniel, B., & Grzonkowski, S. (2006). Search interface based on natural language query templates.
- Kruk, S. R., Gzella, A., Dobrzanski, J., McDaniel, B., & Woroniecki, T. (2007). E-learning on the social semantic information sources. *Lecture Notes in Computer Science*, 4753, 172.
- Kruk, S. R., & McDaniel, B. (2008). *Semantic digital libraries*. Springer Verlag.
- Maltz, D., & Ehrlich, K. (1995). Pointing the way: Active collaborative filtering. *Paper presented at the proceedings of the SIGCHI conference on Human factors in computing systems*.
- Marmolowski, A., & Kiełczyński, W. (2012). Modelling of the Heat Flux Density Distribution for Laser Beam Welding. *Paper presented at the Solid State Phenomena*.
- Salakhutdinov, R., Mnih, A., & Hinton, G. (2007). Restricted Boltzmann machines for collaborative filtering. *Paper presented at the Proceedings of the 24th international conference on Machine learning*.
- Sugiyama, K., Hatano, K., & Yoshikawa, M. (2004). Adaptive web search based on user profile constructed without any effort from users. *Paper presented at the proceedings of the 13th international conference on World Wide Web*.
- Tummarello, G., Morbidoni, C., & Nucci, M. (2006). Enabling Semantic Web communities with DBin: An overview. *The Semantic Web-ISWC 2006* (pp. 943–950): Springer.
- Turnbull, D. (2003). Augmenting information seeking on the world wide web using collaborative filtering techniques. <http://downturn>.

- Zhang, J. X., Liu, L., Ordonez de Pablos, P., & She, J. (2014). The auxiliary role of information technology in teaching: Enhancing programming course using Alice. *International Journal of Engineering Education*, 30(3), 560–565.
- Zhou, Y., Wilkinson, D., Schreiber, R., & Pan, R. (2008). *Large-scale parallel collaborative filtering for the netflix prize*. *Algorithmic Aspects in Information and Management*. Springer (pp. 337–348).
- Zhuhadar, L., & Nasraoui, O. Personalized Cluster-based Semantically Enriched Web Search for E-learning. *Paper presented at the CIKM 08: ONISW the 2nd international workshop on ontologies and information systems for the Semantic Web*. <<http://dl.acm.org/citation.cfm?id=1458498>>
- Zhuhadar, L., Nasraoui, O., & Wyatt, R. (2009b). *Visual Ontology-Based Information Retrieval System* (pp. 419–426). <http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5190855&abstractAccess=no&userType=inst>
- Zhuhadar, L., Nasraoui, O., & Wyatt, R. (2009). Dual representation of the Semantic user profile for personalized web search in an evolving domain. *Paper presented at the proceedings of the AAAI 2009 spring symposium on social Semantic Web, where Web 2.0 meets Web 3.0*. <<http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB4QFjAA&url=http%3A%2F%2Fwww.aaai.org%2FPapers%2FSymposia%2FSpring%2F2009%2FSS-09-08%2FSS09-08-021.pdf&ei=Cl0cVef6LJ2FNsyzgagC&usg=AFQjCNFktxXufR1duUXZA8QZljRJW7nKkw&sig2=RCt3QgZnF13KQxQICEFdDA&bvm=bv.89744112,d.eXY>>
- Zhuhadar, L., Nasraoui, O., & Wyatt, R. (2009a). Metadata as seeds for building an ontology driven information retrieval system. *International Journal of Hybrid Intelligent Systems*, 6(3), 169–186. <http://dl.acm.org/citation.cfm?id=1735969>.