

1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010)

Linking OpenCourseWares and open education resources: creating an effective search and recommendation system

Brett E. Shelton*, Joel Duffin, Yuxuan Wang, Justin Ball

Center for Open Sustainable Learning, Utah State University, 2830 Old Main Hill, Logan, Utah, USA

Abstract

With a growing number of digital libraries and other open education repositories being made available, open education specific recommender and search tools play an important role in helping teachers and learners find relevant resources. This paper discusses the design and evaluation of *Folksemantic*, a system that integrates OpenCourseWare search, Open Educational Resource “more like this” recommendations, and personalized recommendation functionality into a single open source project. Ongoing research is described.

© 2010 Published by Elsevier B.V. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: digital libraries; open education; recommender system, opencourseware, oer

1. Introduction

This paper describes the implementation and evaluation of the *Folksemantic* project including the OCW Finder, OER Recommender, and *Folksemantic* websites. OCW Finder provides tag and keyword search interfaces for finding OpenCourseWares (OCWs). OER Recommender provides widgets and APIs that can be used to insert “more like this” recommendations into open education resource (OER) web pages. The *Folksemantic* website provides full text search of OERs, a way to add resources to the *Folksemantic* index and the ability for users to create profiles so that they can receive personalized recommendations. The goals of the *Folksemantic* project include:

- Provide widgets and programming interfaces that OCWs and digital libraries can easily use to add links from their resources to related resources in other OCWs and digital libraries.
- Study algorithms, architectures, and approaches to implementing a system that provides high quality recommendations for OERs.
- Research effective ways to present recommendations to teachers and learners who access OERs.
- Maximize adoption of the *Folksemantic* recommender system by OER websites and maximize its use by teachers and learners.

* Corresponding author. Tel.: 1-435-797-2393; fax: 1-435-797-2693.

E-mail address: brett.shelton@usu.edu

The *Folksemantic* project was motivated by a desire to increase cross-pollination between larger sized OCWs and smaller OERs. The inception and initial growth of the OCW movement, which makes university course materials freely available online, was made largely possible through significant funding by the Hewlett foundation [1]. Through programs such as the National Science Digital Library (NSDL), the NSF has invested in the development of a large number of high quality OERs for K-16 such as simulations, videos, animations, student activities, and teacher resources [2]. The premise of the NSDL *Folksemantic* project was that adding recommendations to OERs and OCWs could help connect learners and educators who are looking at online resources to additional resources to meet their needs. This work contributes to work by other researchers who have recognized that search and recommendation in the area of OERs has its own unique considerations [3].

2. Initial efforts at customized search tools

2.1. OCW Finder

OCW Finder was initially a static lightweight client-side OCW search tool, created early in the OCW movement. It attracted attention because of its simple multi-column tag interface for browsing and searching across multiple OCW collections [4]. *Folksemantic* extended the functionality of OCW Finder by adding the ability for people to register new collections and by putting a database and search engine behind the finder so that as new courses were added they would be immediately available to search. Fig. 1 shows the current OCW Finder interface.

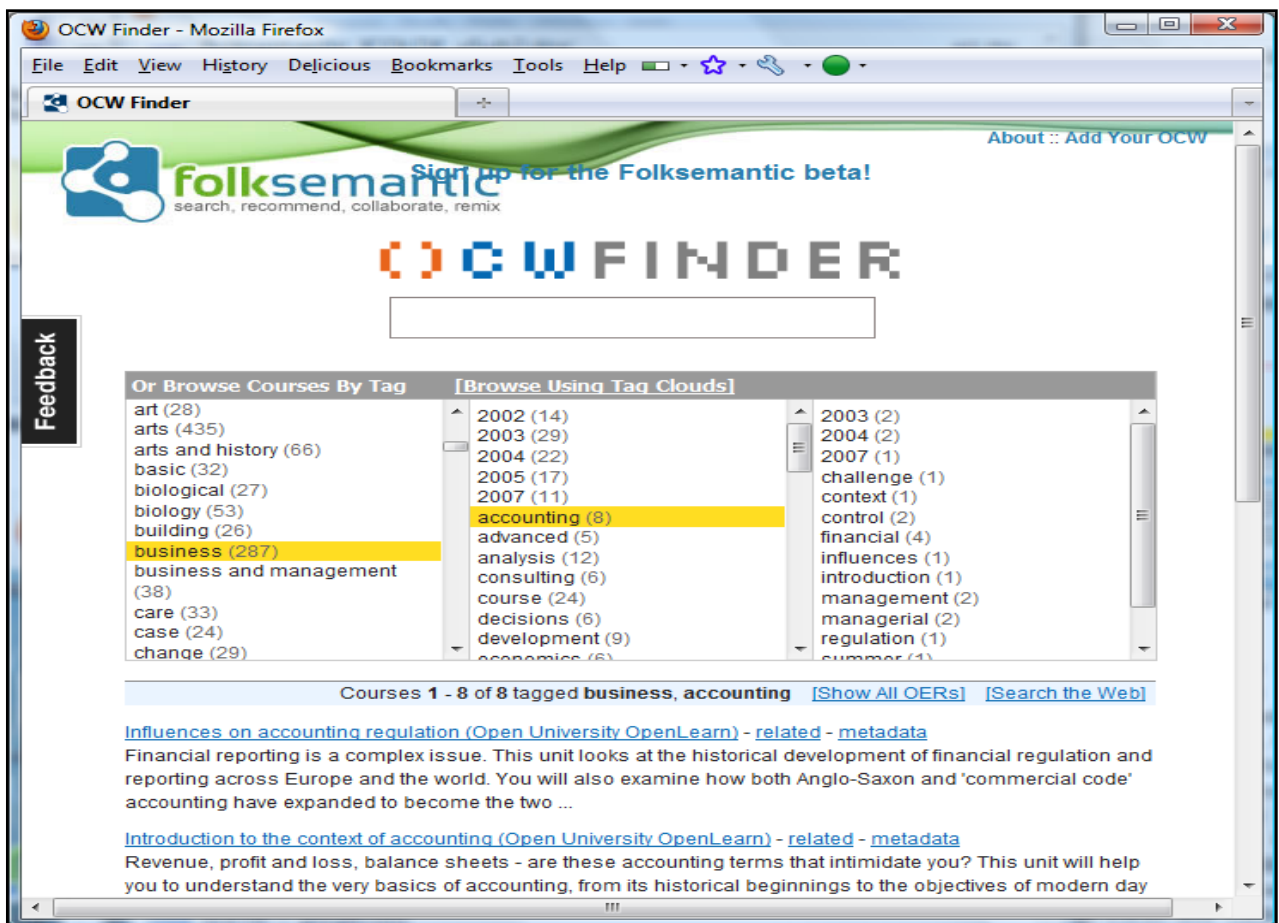


Fig. 1. OCW Finder interface

2.2. OER Recommender

OER Recommender is a content-based system that recommends related resources based on the semantic relatedness of their metadata. The *Folksemantic* aggregator harvests metadata from RSS feeds, OAI-PMH endpoints, and other types of data sources by utilizing and extending the ROME and other open source libraries for feed parsing. OAI-PMH is a protocol that the NSDL and many other digital libraries support to provide a way for their metadata to be harvested using HTTP requests [5]. *Folksemantic* researchers initially wrote their own indexing, search, and recommendation algorithms from the ground up and later switched to building on Lucene. Lucene is a popular open source library that supports full text indexing and searching [6]. The *Folksemantic* recommendation algorithm takes resource title, tags, and descriptions into account, weighting each differently. *Folksemantic* tracks user clicks and time on page data and uses it to adapt the ordering of recommendations based on this user data [7]. The simplest way to add *Folksemantic* OER recommendations to a web page is to place a small snippet of HTML in the web page. Additionally, recommendations can be retrieved in RSS, HTML, JSON, and XML formats. Fig. 2 shows an example of a web page that has included the OER Recommender HTML snippet.



Fig. 2. An example of integrating OER Recommender into a web page

To encourage adoption of the OER Recommender service by OER collections, a Greasemonkey [8] script, web browser extension was developed that can be used to easily demonstrate what it would be like to integrate the OER Recommender service into a website. The script includes the ability to retrieve recommendations calculated in real-time for resources not already in the *Folksemantic* index. OCW websites including ones at the Open University's OpenLearn and Utah State University integrated OER search and recommendations into their websites [9]. A popular platform for hosting OCWs called eduCommons added an option that allows a system administrator to easily turn on *Folksemantic* recommendations for an eduCommons site.

2.3. Learning from user feedback and dealing with ordering bias

OER Recommender tracks user clicks and time on page in order to use that implicit feedback to improve recommendations. Because users are naturally inclined to click on results that appear higher in a list of results, it can

create bias in systems that learn from click data. The OER Recommender system addresses this potential bias by grouping recommendations into categories and randomly ordering resources in those categories. The system begins by identifying the top 20 most semantically related resources. It then divides these into two groups called highly recommended and recommended. The highly recommended group contains resources with relevance scores at least one standard deviation higher than the average relevance score of the entire 20 resource. Before a resource has sufficient user feedback data gathered, OER Recommender displays resources in the highly recommended category first by randomly ordering them, followed by resources in the recommended category, which are also randomly ordered. The system uses user feedback data to develop a popular category. The popular category contains resources that have at least one standard deviation higher than the average number of clicks. Because the popular category represents user preference, the system displays resources in that category before resources in the highly recommended category and in strict order of user preference.

3. Folksemantic

In previous projects, *Folksemantic* researchers developed Web2.0 tools to increase the impact of open education by supporting human interaction around OERs [10]. In order to facilitate automated and human-to-human personalization of OERs, *Folksemantic* researchers decided to integrate OCW Finder and OER Recommender and social network components from previous projects into a single platform. Key aspects of the integration effort were to (1) merge the underlying index used by both OCW Finder and OER Recommender into a single index, (2) translate the OCW Finder interface into additional languages, (3) transition OCW Finder to use Lucene for its underlying indexing, searching, and recommendation functionality, (4) add a user account system to allow people to sign up and create profiles, and (5) make it possible for users to register feeds they produce such as blogs and bookmarks so they can be used to personalize recommendations.

To facilitate adding collections to the *Folksemantic* index, an interface was created that allowed registered users to submit collections. When system administrators and trusted users submit feeds for collections, the system immediately begins harvesting metadata from those feeds. When other users submit feeds, system administrators are notified and must approve the feed before the system begins harvesting metadata from it.

4. Evaluating usage of Folksemantic websites

In order to better understand how *Folksemantic* websites are being used and how well project goals are being met, researchers gathered and analysed usage data. To facilitate analysis, researchers implemented consistent logging and analysis methods across the *Folksemantic*, OER Recommender, and OCW Finder web sites. All three web sites share the same application database and Lucene indexes. In order to gather data from which to analyse the *Folksemantic* system, researchers integrated Google Analytics, implemented custom query logging, and archived standard web server log files. In addition, researchers used the web application database as a data source for analysis. Application database data analysed included click tracking and time on page data that the application uses to improve recommendations. The web application click tracking approach omits multiple clicks on the same recommendation in a given user session, an approach people might use to try to “game the system”.

Web server log files were analysed by writing custom scripts that matched and counted specified patterns. The scripts were then applied to archived log files to mine the desired data. Scripts identified and omitted requests from web crawlers such as the Googlebot crawler. The web application database was analysed by writing custom SQL queries and generating reports. The earliest month that statistics are given for is August 2009 because that is when the OER Recommender and OCW Finder websites were modified to use the same database and index and the *Folksemantic* website was launched.

The results shown in Table 1 provide insight into the number of indexed resources, website usage, OER provider adoptions, and recommendation usefulness. The number of registered collections and indexed resources has consistently grown, though it should be noted that previously the index contained more resources than it currently does (over 110,000), but some were lost when the system was transitioned from a custom indexing system to Lucene. During recent development efforts, the focus has not been on adding new resources. Researchers recognize that continually adding resources will be important to feeding interest and adoption of both “more like this” recommendations and personal recommendations. Usage of the *Folksemantic* websites as measured by unique

visitors has been consistent and grown incrementally, though not dramatically. A small number of OER Providers have integrated recommendations from OER Recommender into their websites. eduCommons has played an important role in a number of those adoptions, perhaps because administrators can easily enable recommendations from OER Recommender via a system option. A substantial number of recommendations from OER Recommender are being presented to users each month and a small but significant number (11%) are being clicked on. This percentage is encouraging. Additional research is needed to understand why people are exploring recommendations and how satisfied they are with recommendations once they have visited recommended resources.

Table 1. Folksemantic resources, adopters, and users.

Measure	Period	Total, Daily Average, Daily Std Dev
OpenCourseWare Collections †	Mar 2010	46
Total Collections †	Mar 2010	628
Indexed Courses †	Mar 2010	7160
Indexed Resources (All OERs) †	Mar 2010	87941
Searches £	Mar 2010	Aggregate: 10,377; 346, 100 Folksemantic: 4,258, 142, 83 Recommender: NA Finder: 6,119, 204, 75
Websites Using <i>Folksemantic</i> Widgets ‡	Mar 2010	18
Visitors (Unique IPs) *††	Aug 2009 Mar 2010	1893 8140
Requests (hits) ‡	Mar 2010	Aggregate: 874,613; 29,154; 9,097 Folksemantic: 316,329; 10,544; 6,226 Recommender: 501,270; 16,709; 4,269 Finder: 57,014; 1,900; 452
Registered Users †	Mar 2010	1763
Recommendation Lists Displayed ‡	Mar 2010	Aggregate: 437,890; 14,596; 3588 Folksemantic: 33,666; 1,122; 1,077 Recommender: 404,224; 13,474; 476 Finder: NA
Clicked Recommendations †	Mar 2010	47974 (11%), 3198, 838

For each measure, four sets of statistics are presented: (a summary of all 3 *Folksemantic* websites, followed by statistics for the *Folksemantic*, OER Recommender, and OCW Finder websites, listed in that order).

* Data was gathered from Google analytics. Note that this does not include visits to websites that display recommendations using *Folksemantic* widgets.

† Data was gathered from the application database.

‡ Data was gathered from the web server log files.

£ Data was gathered from custom query log files.

†† The maximum number of visitors in a single day was on July 8, 2009 when the sites had 600 unique visitors in response to being highlighted on a nationally syndicated technology radio program.

5. Ongoing research

Ongoing *Folksemantic* research focuses on system evaluation, increasing the number of indexed resources, expanding the number adoptions by OER providers, and extending recommendation algorithms to be able to provide personalized recommendations individual users. To evaluate the quality of the current recommendation algorithms, researchers have created user test protocols and questionnaires for use with teachers and OER providers. In order to facilitate aggregating additional resources, researchers are improving the user interface for registering feeds and adding the ability to index resources for which no metadata feed is available. In order to increase adoptions by OER providers, demonstration videos are being created and registered collections are being contacted. Ongoing work on personal recommendations is building upon the open source Mahout machine learning library. Personal

recommendations will be based on user interests identified from the feeds that the users register and their interactions with the *Folksemantic* system. The approach will model user interests using term vectors and use those term vectors to query the *Folksemantic* index for resources to recommend. When generating term vectors, each resource that a user has paid attention to will be weighted using attention type, attention recency, attention frequency, and resource recency values. Attention type refers to what type of attention a user gave to a resource (e.g. visited, or bookmarked). Attention recency refers to how long ago the attention occurred. Attention frequency refers to how often they have paid attention to the resource. Resource recency refers to how long ago a resource was added to the *Folksemantic* index. Users will be able to view and directly modify the term vectors representing their interests. These ongoing efforts will help the *Folksemantic* project increase the impact of the open education movement by providing open education specific search and recommendation tools.

Acknowledgements

The work described in this project was supported by National Science Foundation grant DUE-0532895.

References

1. T. Caswell, S. Henson, M. Jensen, and D. Wiley, Open Educational Resources: Enabling universal education., *The International Review of Research in Open and Distance Learning*, vol. 9, no. 1 (2008).
2. L.L Zia, The NSF National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) Program: A Progress Report. *D-Lib Magazine*, vol. 6, no. 10 (2000).
3. X. Ochoa and E. Duval, Relevance ranking metrics for learning objects. *IEEE Transactions on Learning Technologies*, vol. 1, no. 1, pp. 34–48, (2008).
4. D. Wiley, OCW Finder. iterating towards openness. <http://opencontent.org/blog/archives/175>, July 13, 2005.
5. C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, The Open Archives Initiative Protocol for Metadata Harvesting - Version 2.0. <http://www.openarchives.org/OAI/openarchivesprotocol.html>, 2002.
6. Apache Lucene, <http://lucene.apache.org/>.
7. J. Duffin, B. Muramatsu, and S.L. Henson-Johnson, OER Recommender: A recommendation system for open educational resources and the National Science Digital Library. White paper funded by the Andrew W. Mellon Foundation for the Folksemantic.org project, 2007.
8. A. Boodman, Greasemonkey, <https://addons.mozilla.org/en-US/firefox/addon/748>, 2007.
9. B. Muramatsu, J. Ball, J. Duffin, D. Wiley, J. Gray, A. Godwin, R. Reedstrom, C. Cooper, and C. Whipkey, Open Educational Resources Interoperability Sprint Outcomes: Linking OER Providers. *OpeniWorld: Europe*, Lyon, France, 2008.
10. S. Henson-Johnson, The folksemantic web: Tools for a human-centered approach to learning. *International Conference of the Learning Sciences*, Utrecht, Netherlands, 2008.