Ecology of social search for learning resources

Citation for published version (APA):

DOI:
10.1108/10650740910984619

Document status and date:
Published: 28/08/2009

Document Version:
Peer reviewed version

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
https://www.ou.nl/taverne-agreement

Take down policy
If you believe that this document breaches copyright please contact us at:
pure-support@ou.nl
providing details and we will investigate your claim.
Ecology of social search for learning resources

Abstract

Purpose: This paper deals with user-generated Interest indicators (e.g., ratings, bookmarks, tags). We answer two research questions: can search strategies based on Social Information Retrieval (SIR) make the discovery of learning resources more efficient for users, and can Community browsing help users discover more cross-boundary resources. By cross-boundary we mean that the user and resource come from different countries and that the language of the resource is different from that of the user’s mother tongue.

Design: We focus on a portal that accesses a federation of multilingual learning resource repositories. A measure for user’s efficiency in finding relevant resources was defined. We then collected users’ attention metadata and use this empirical data to answer two hypotheses.

Findings: We show that users are more efficient with Social Information Retrieval strategies, however, Community browsing alone does not help users discover a wider variety of cross-boundary resources.

Practical implications: By social tagging and bookmarking resources from a variety of repositories, users create underlying connections between resources that otherwise do not cross-reference. This is important for bringing them under the umbrella of SIR methods. Future studies should include testing SIR methods to leverage these user-made connections between resources that originate from a number of countries and are in a variety of languages.

Originality: The use of attention metadata to model the ecology of social search adds value to the actors of learning object economy, e.g. educational institutions, digital libraries and their managers, content providers, policy makers, educators and learners.

Keywords: learning resources, social information retrieval, social tagging, metrics, efficiency, Research paper

1. Introduction

Learning resource repositories and libraries make educational material and/or its metadata available in digital format, the sharing of which is their core raison d’etre. Their reuse has been touted for enabling cost savings because the creation of high quality material is costly, hence the focus on standards that enable interoperability (Cambell, 2003) even across repositories (Ternier, et al., 2008). Traditionally, metadata and/or web directories are used for searching and exploring educational content. Currently, novel exploratory search systems are developed for learning resources to assist users in obtaining content that meets their information needs. Such systems include social navigation and collaborative recommender systems, both of which belong to the family of techniques called Social
Information Retrieval (Goh et al., 2007).

Social navigation involves using the behaviour of other people to help navigate online. It is driven by the tendency of people to follow other people’s footprints when they feel lost (Dieberger et al., 2000). Such footprints in an online environment are what Claypool et al., (2001) define as Interest indicators which can be acquired either directly from the user (e.g. rating) or indirectly (e.g. time spent on an object). Collaborative recommender systems, on the other hand, use explicit ratings to find like-minded users (Adomavicius and Tuzhilin, 2005). Evaluation of recommender systems traditionally focuses on the algorithms and their performance (Herlocker et al., 2004), similar to exploratory search systems (White et al., 2008). Evaluating recommenders from the user perspective has received less attention (McNee, 2006).

Within the field of Technology Enhanced Learning (TEL) such systems exist. Rafaeli et al., (2005) introduced a system to harness the social perspectives in learning where the learner could choose from whom to take recommendations (friend or algorithm). Koper (2005) used indirect social interaction in choosing a path that allows successful competition of a learning task. Drachsler et al., (2008) took this research further showing that users employing a recommender system, that offers navigation support in self-organised Learning Networks, were more efficient time-wise in completing an equal number of Learning Activities. Farzan and Brusilovsky (2005) studied social navigation and found that adding the time spent reading each page provides more precise insight into the intention of the group of users and more accurate information about pages selected from search results. Jung et al., 2007 studied implicit click data to increase both precision and recall of the feedback data on a university search portal. Tang and McCalla (2009) studied the pedagogical value while using collaborative filtering to recommend papers for learners, and Manouselis et al., (2007) used multi-criteria ratings to recommend resources to teachers.

Both the field of recommender systems and social navigation, however, suffer from the same problems: how can Interest indicators be gathered without being too intrusive, and yet, at the same time, remain accurate enough in guiding users in their choice of product or navigational path. The sparse data and new items often are problematic too (Herlocker et al., 2004, Adomavicius, Tuzhilin, 2005, Rafaeli et al., 2005). Social bookmarking and tagging can offer new prospective thanks to tags, in which users describe their interest. This creates a triple (user, content, annotations) which indicates user’s relationship between resources, users, and tags (Golder and Huberman, 2006, Marlow et al., 2006, Sen et al., 2006). Such underlying structure allows flexible social navigation (e.g. tag-item, tag-user, user-item), but could also be a source for collaborative recommender systems by linking like-minded users not only through resources, but also through tag-based interest sharing (Santos-Neto et al., 2009). The idea of social tagging and bookmarking has been implemented in the TEL context (Maier, Thalmann, 2008, Vuorikari, Poldoja, 2008), in digital libraries (Puspitasari et al., 2007) and for scientific papers (Farooq et al., 2007, Gatutto, et al., 2008).

Millen et al., (2007) studied the use of social bookmarking at the enterprise level and suggest that integrated with traditional search engines, it has the potential to solve commonly known enterprise search problems, e.g. content from heterogeneous repositories that do not cross-reference via hyperlinks (Mukherjee, 2004). We study a similar implementation for multilingual learning resources within a federation that has social
bookmarking and tagging features (Vuorikari & Van Assche, 2007). Our aim is to study such a hybrid system to understand how it is used, how different variables are interconnected, and finally, how the behaviour of previous users could be leveraged to support and enhance the discovery process of educational resources for all users of the system.

From previous studies (e.g. McGormick et al., 2004, Vuorikari and Koper, submitted) we have evidence that users of educational content use cross-boundary resources to a certain extent, but their reuse remains rather low.

In this study our specific questions focus on the efficiency of resource discovery and how we can support the users in discovering cross-boundary resources. We have defined our two hypotheses as following:

\[ H1: \text{The search methods that take advantage of Social Information Retrieval yield more relevant resources with less effort from the users than the methods based on conventional text based search.} \]

\[ H2: \text{The users who take advantage of Community browsing discover more cross-boundary learning resources than those who use conventional text base search.} \]

![Learning resource lifecycle (Van Assche, Vuorikari, 2006)](image)

We focus on a specific moment in the lifecycle of a learning resource, namely when the user discovers the learning resource and evaluates whether it matches with the information seeking need at hand (Figure 1).

In the following section we introduce the study methodology and the data set. The next section focuses on the results followed by the discussion. The paper is concluded with an
outlook for future work.

2. Context of the study and its method

The portal that we study makes open educational resources available from 19 content providers from Europe and elsewhere. These resources exist in different languages and conform to different national and local curricula. The portal, developed in the MELT project (Figure 2), offers three different categories of searches (Millen et al., 2006).

Figure 2. Melt portal.

Explicit search: comprises the traditional search box with text and filtering options based on multilingual metadata. "Find by subject" offers browsing through pre-defined categories. The results are shown on the Search Result List (SRL) with metadata and annotations, if available.

Community browsing: these are social navigation features such as accessing resources through tagclouds and specific lists of most bookmarked resources, but also "pivot browsing" which means using tags or usernames as a means to reorient browsing. Figure 2 shows examples such as the multilingual tagcloud and different tabs (e.g. most bookmarked).

Personal search: Looking for bookmarks from one's own personal collection of bookmarks (MyFavourites).

In this study, Social Information Retrieval methods mean all the Community browsing features, and also comprise of retrieved resources that contain user-generated Interest indicators. These are Interest indicators like a rating on a scale 1 to 5 (1="of no use" to 5="very useful") or a bookmark with tags (called Favourites). These public annotations are also called contributing actions. By clicking on the link the user plays a resource and
generates click-through which is an implicit Interest indicator. We use the term “cross-boundary discovery” when the user bookmarks or rates resources that come from different country than she does and that are in a language other than that of the user’s mother tongue.

We use the following metric for relevance: when a resource is added in the Favourites, or the resource is rated with the value of 3 or greater. Such relevance represents the relationship between the object and the information need, as perceived by the user. We do not focus on the other types of relevance, such as the query and object match or topical relevance (Borlund, 2003). With "less effort from the user" we mean that users will play and annotate relevant resources with less search effort (i.e. the number of executed searches).

2.1 Description of the data

We have defined our own logging scheme for users’ attention metadata, exportable in a Contextual Attention Metadata compliant format (Najjar et al., 2007). It was first used on the Calibrate pilot and now we redefined the scheme to cover a range of contributing actions and social search. Table 1 describes the scheme that includes 3 main units: the resources, the user and user’s action. The latter includes three main categories: search, click and contributing actions. Each time a user arrives on the portal, a new session is started, regardless if the user logs in. Table 1, for example, shows that metadata about “languages” is gathered for the resource, the user (mother tongue and spoken languages) and for the tags.

Table 1. Logging scheme on the portal to capture users' attention metadata. LO stands for Learning object (resource).

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Unique ID</th>
<th>Languages</th>
<th>Location</th>
<th>Time stamp</th>
<th>Other metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Resources</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>LRE Application profile</td>
</tr>
<tr>
<td>2. Users</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Country, school, topics of interest.</td>
</tr>
<tr>
<td>3. Actions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1 Session</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>ID, if logged in.</td>
</tr>
<tr>
<td>3.2 Search</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Type of search: advanced, browse topics, tag search</td>
</tr>
<tr>
<td>3.3 Click</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>The LO, tag, rating, other user clicked on</td>
</tr>
<tr>
<td>3.4 Bookmark</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>User ID, LO ID</td>
</tr>
<tr>
<td>3.5 Tag</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>User ID, LO ID</td>
</tr>
<tr>
<td>3.6 Rate</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>User ID, LO ID, value, comment</td>
</tr>
</tbody>
</table>

The registered users, mostly primary and secondary teachers, were invited to use the portal after an initial pilot. They came from 11 different European countries (Austria, Belgium, Estonia, Finland, France, Germany, Greece, Hungary, Slovenia, Sweden and United Kingdom).

The attention metadata logs were collected from the period of Oct 1 to Dec 18 2008. After excluding the project staff and users who did not search and click resources, we were left...
with 82 users who had clicked at least on one resource. Table 2 presents the data used for this study which include registered users (82), but also sessions from unregistered users.

**Table 2. Data description of all recorded attention and actions included in the study.**

<table>
<thead>
<tr>
<th></th>
<th>Sessions</th>
<th>Executed searches</th>
<th>All plays</th>
<th>Distinct resources played</th>
<th>Distinct resources bookmarked and rated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unregistered users</td>
<td>2036</td>
<td>7846</td>
<td>1854</td>
<td>1547</td>
<td>Not available to users</td>
</tr>
<tr>
<td>Registered users (82)</td>
<td>310</td>
<td>1863</td>
<td>974</td>
<td>687</td>
<td>394</td>
</tr>
</tbody>
</table>

During the pilot, a selection of more than 30 000 distinct resources was made available. 565 resources (less than 2% of all resources) had been annotated prior to the period of study by the project staff. Figure 3 shows that the growth of new users and tags follow one another. Farooq *et al.*, (2007) explain that linear tag growth suggests that the tag vocabulary is still maturing and has not yet reached its relatively stable stage, where the tag growth diminishes over time (Marlow *et al.*, 2006), a view that we also adopt for our system.

During the pilot, a selection of more than 30 000 distinct resources was made available. 565 resources (less than 2% of all resources) had been annotated prior to the period of study by the project staff. Figure 3 shows that the growth of new users and tags follow one another. Farooq *et al.*, (2007) explain that linear tag growth suggests that the tag vocabulary is still maturing and has not yet reached its relatively stable stage, where the tag growth diminishes over time (Marlow *et al.*, 2006), a view that we also adopt for our system.

![Growth in MELT](image)

Figure 3. Tag and user growth in MELT. The portal was not in use of the summer months.

**2.2 Data processing**

We tested the normality of our data using the Kolmogorov-Smirnov test, which indicates a highly significant deviation from normality. As the assumption of normality was not tenable, we used methods for non-parametric data (e.g. Kendall tau). For testing the association between the type of search and the number of actions that followed, we used the Pearson Chi-Square test (p<0.001).

**3. Results**

We first give descriptive results on our three main actions: how do users search, play and contribute, and then look how new resources generated clicks and annotations. Following,
we present the model of the process that produced this data and then focus on our two hypotheses.

### 3.1 Descriptive results

#### Search

We found three main groups of search behaviour among the registered users: the ones who only used Explicit (47%) or only Community browsing features (6%) and about half of the users who used mixed methods (47%). Of all searches executed on the portal, 82% were Explicit methods (53% advanced searches; 29% browsing by discipline) and 18% Community browsing.

#### Click-through

Users generate click-through when they “play” a resource. Counting all click-through from all users, 1547 distinct resources were played 2828 times. Different user behaviour was observed among registered users: 52% played resources only on the Search result list, 10% only in Community browsing areas, and the rest (38%) in different parts of the portal.

Table 3 shows that 69% of the click-through was generated on Search result list on resources with no Interest indicators, whereas 31% of the click-through came from resources that had Interest indicators. As our Explicit search function does not take advantage of Interest indicators, we consider the latter to be an outcome of Social Information Retrieval.

#### Table 3. Amount and percentages by areas of the portal where click-through was generated.

<table>
<thead>
<tr>
<th>Click-through</th>
<th>Registered users</th>
<th>Other</th>
<th>All</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Result List</td>
<td>728</td>
<td>1235</td>
<td>1963</td>
<td>69.4%</td>
</tr>
<tr>
<td>Search Result List with Interest indicators</td>
<td>40</td>
<td>85</td>
<td>125</td>
<td>4.4%</td>
</tr>
<tr>
<td>Tagcloud</td>
<td>103</td>
<td>124</td>
<td>227</td>
<td>8.0%</td>
</tr>
<tr>
<td>“Travel well” list</td>
<td>68</td>
<td>300</td>
<td>368</td>
<td>13.0%</td>
</tr>
<tr>
<td>Favourites</td>
<td>20</td>
<td>34</td>
<td>54</td>
<td>1.9%</td>
</tr>
<tr>
<td>“Most bookmarked” list</td>
<td>15</td>
<td>76</td>
<td>91</td>
<td>3.2%</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>974</strong></td>
<td><strong>1854</strong></td>
<td><strong>2828</strong></td>
<td><strong>99.9%</strong></td>
</tr>
</tbody>
</table>

#### Annotations

77% of registered users who played resources also annotated them: 44% both rated and tagged resources, whereas 33% only rated or tagged resources. Table 4 shows that users rated and bookmarked very similarly. Two users out-performed others with 120 and 108 annotations. In general, the ratings are positive; 84% voted 3 or greater (≥ 3) and only 16% of ratings indicate a resource “of no use” (=1 or 2).
Table 4. Descriptive statistics and distribution of bookmarks and tags.

<table>
<thead>
<tr>
<th>Annotations:</th>
<th>Bookmarks and tags</th>
<th>Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>48</td>
<td>46</td>
</tr>
<tr>
<td>Number of actions</td>
<td>350</td>
<td>384</td>
</tr>
<tr>
<td>Number of tags</td>
<td>1507</td>
<td>Rating ≥3 (84%)</td>
</tr>
<tr>
<td>Minimum of actions</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Maximum of actions</td>
<td>65</td>
<td>56</td>
</tr>
<tr>
<td>Mean of bookmarks</td>
<td>7.29</td>
<td>8.34</td>
</tr>
<tr>
<td>Median</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Mode</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Standard Deviation of annotations</td>
<td>12.19</td>
<td>12.53</td>
</tr>
</tbody>
</table>

Annotations took place in different parts of the portal: 40% of users only annotated resources in the Search result list, 24% only in Community browsing parts of the portal and 36% of users annotated in both places, which resulted in most of annotations (70%). Most resources that were annotated (67%) had no previous annotations and were found through Explicit search (Table 5).

Table 5. Place of the portal where annotations took place.

<table>
<thead>
<tr>
<th>Bookmarks obtained after..</th>
<th>Ratings done in..</th>
</tr>
</thead>
<tbody>
<tr>
<td>..Explicit search no annotations</td>
<td>236 (67%)</td>
</tr>
<tr>
<td>..Explicit search with annotations</td>
<td>49 (14%)</td>
</tr>
<tr>
<td>..Community browsing</td>
<td>65 (19%)</td>
</tr>
<tr>
<td>..Personal Search</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Interest indicators, click-through and contributions

We found a tendency that resources with Interest indicators generated more actions: 14% of all played resources that had Interest indicators generated 29% of all plays. Table 6 shows that we found a small correlation coefficient between resources that had existing Interest indicators and bookmarking action (0.329, p<0.01), similarly between actions play and bookmark (0.327, p<0.01) and some between bookmark-rate (0.300, p<0.01). These correlations between actions can be further studied for generating implicit Interest indicators that are less intrusive for users.

Table 6. Correlation coefficient between different actions.

<table>
<thead>
<tr>
<th>Correlation coefficient between</th>
<th>Kendall Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource w/Interest indicators – bookmark action</td>
<td>(0.329, p&lt;0.01)</td>
</tr>
<tr>
<td>Click-through -bookmark action</td>
<td>(0.327, p&lt;0.01)</td>
</tr>
<tr>
<td>Bookmark action –rate action</td>
<td>(0.300, p&lt;0.01)</td>
</tr>
<tr>
<td>Resource w/Interest indicators and click-through</td>
<td>(0.253, p&lt;0.01)</td>
</tr>
<tr>
<td>Click-through –rate action</td>
<td>(0.233, p&lt;0.01)</td>
</tr>
<tr>
<td>Resource w/Interest indicators and rate action</td>
<td>(0.196, p&lt;0.01)</td>
</tr>
</tbody>
</table>
3.2 Modelling users’ actions: ecology of social search

A model was created of the process that produced the above described data to study how processes are interlinked (i.e. ecology); for example, how newly created annotations become part of the search process where users interact and eventually annotate them.

Figure 4 shows authenticated and unauthenticated users on the portal and the percentage of their actions by each category (search-play-contribute). In principle, these two groups search rather similarly: Explicit search is preferred by both groups and Community browsing methods account for 21% of all search actions.

When looking into details, the main difference is that the unauthenticated users tend to explore resources through browsing (51%): 30% browse by discipline and 22% Community browsing, whereas the authenticated users browse less (38%), but additionally use Personal search (9%). The Pearson Chi-Square test was significant ($p<0.001$) for these differences.

Users play resources differently, unauthenticated users play 71% of resources in the Search Result list (SRL) after an Explicit search and 29% in Community browsing areas. Authenticated users, on the other hand, play 79% of resources in the SRL and only 21% in Community browsing areas. The Pearson Chi-Square test was significant ($p<0.001$) for these differences.

Only users who are authenticated to the system can contribute. Out of all the actions recorded in the system, contributing actions amount to 16%. Most contributing actions are annotations (67%) on newly discovered resources, whereas 33% is generated in the Community browsing areas on resources that have previous annotations. The annotations are generated following a rather regular pattern (grey boxes in Figure 4), however, for ratings, a small amount is created in the user’s Favourites.

The model shows that the annotation (i.e. Interest indicators) play an integral part in creating a social search ecology and offer more diverse ways to discover resources. 16% of all the search-play-annotation actions contribute to this ecology, which are used to create novel ways to discover resources. 21% of all search actions took advantage of these
features, and they supported 31% of plays. Our finding points to the same direction as Glahn et al., (2008) who found that accessing of tagged resources is independent from the contribution level of a participant. However, we did not find that all participants use tags similarly while searching or accessing tagged resources: we observed different search preferences by users.

3.3 Measure of efficiency: Search-play-contribute ratio (H1)

Hypothesis 1 (H1) states that the search methods based on Social Information Retrieval yield more relevant resources with less effort from users. We use the above model to create a measure to test our hypothesis by studying the ratio between search, play and contributing actions from users who were authenticated on the portal. We are inspired by the Click-through rate that measures the success of an online advertising campaign. The rate can be obtained by dividing the number of users who clicked on an advertisement by the number of times it was delivered (e.g. Ward and Kirithi, 2007).

Our application of the efficiency measure shows how many search actions it takes to play and/or annotate a resource. For both search methods we created the following efficiency ratios: search-play, search-rate (3 or higher), search-bookmark, play-rate (3 or higher), and play-bookmark. The lower the figure the better, as it indicates the number of search actions that it takes the user to achieve the goal.

Table 7 presents the two baselines against which we measure. Explicit search indicates the results from a text-based search excluding the plays on resources that contained Interest indicators (i.e. ratings, tags). Community browsing indicates the results from tagcloud, lists and pivot browsing. SIR methods, against which we measure the efficiency, include results from Explicit search with Interest indicators (i.e. ratings, tags) and from Community browsing. On the top row, different ratios for actions are indicated, e.g. how many searches does it take to play a resource (2.1 Explicit searches to one play). An Efficiency rate for each search method is displayed, which is an average of all ratios. This allows for quick comparison. Lastly, Explicit search (comparison) reports the same ratios from another portal where no Interest indicators were made public (see Vuorikari, Ochoa, 2009).

Comparing the two top rows we compare the efficiency of Explicit search with Community browsing. We find that the search-play, search-bookmark and play-bookmark ratio is almost identical using both methods, whereas Community browsing is more efficient for rating.

<table>
<thead>
<tr>
<th>Actions for authenticated users in MELT</th>
<th>Search: play ratio</th>
<th>Search: rate (≥3) ratio</th>
<th>Search: bookmark ratio</th>
<th>Play: rate (≥3) ratio</th>
<th>Play: bookmark ratio</th>
<th>Efficiency rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit search (baseline)</td>
<td>2.1:1</td>
<td>7.0:1</td>
<td>6.4:1</td>
<td>3.3:1</td>
<td>3.1:1</td>
<td>4.4</td>
</tr>
<tr>
<td>Community browsing (baseline)</td>
<td>2.0:1</td>
<td>5.3:1</td>
<td>6.5:1</td>
<td>2.6:1</td>
<td>3.2:1</td>
<td>3.9</td>
</tr>
<tr>
<td>SIR methods</td>
<td>1.7:1</td>
<td>4.0:1</td>
<td>3.7:1</td>
<td>2.3:1</td>
<td>2.2:1</td>
<td>2.8</td>
</tr>
<tr>
<td>Explicit search (comparison)</td>
<td>0.5:1</td>
<td>21.6:1</td>
<td>11.3:1</td>
<td>42.6:1</td>
<td>22.2:1</td>
<td>19.6</td>
</tr>
</tbody>
</table>
Comparing the Explicit search baseline to SIR methods shows an efficiency gain, the Efficiency rate drops from 4.4 to 2.8. Search-rate ratio comes down from approximately 7 searches to 4 searches, whereas play-rate from 3 searches to 2 searches.

We conclude that on the MELT portal, Explicit search and Community browsing perform very similarly. However, Explicit search compared with SIR methods shows an efficiency gain of 1.6 units. For both the Explicit search data and SIR data, Pearson Chi-Square test was highly significant (p<0.001). We thus conclude that the findings support H1 and we can accept H1.

3.4 Discovery of cross-boundary resources (H2)

Hypothesis 2 (H2) states that the users who use Community browsing discover more cross-boundary learning resources. In addition to using the above measures, we add a component indicating the cross-boundary nature of a bookmark. To do this, we compared the country of the resource to that of the user and the language of the resource to that of the user’s mother tongue as in Vuorikari and Koper (submitted). Additionally, we created a “search method” profile and “cross-boundary resource profile based on bookmarks” for each user. These profiles are comprised of percentages of different search methods by the user, and the percentage of resources that cross national and linguistic boundaries.

57% of all bookmarked resources by users, total of 198, are cross-boundary discoveries (Table 8). Out of 48 users who bookmarked these resources, 40 had cross-boundary resources (i.e. 83% with cross-boundary profile and 17% with only national profile). This split reflects the goal of the portal which promotes the discovery of educational material from different countries. Such cross-boundary profiles are usually also “cross-repository” profiles, information of which can be used to make link structures between content in different repositories that otherwise is not cross-referenced via hyperlinks nor metadata.

<table>
<thead>
<tr>
<th>Cross-boundary discovery</th>
<th>Distinct LOs</th>
<th>Distinct LOs w/SI</th>
<th>“New” discoveries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>198 (57% of all)</td>
<td>73 (37%)</td>
<td>125 (63%)</td>
</tr>
</tbody>
</table>

72% of these bookmarked cross-boundary resources were discovered in the Search result list as a result of Explicit search, whereas 28% were a result of Community browsing. Within Community browsing, 23% were discovered in the tagcloud and 5% chosen from the “Travel well” list. These findings show that most often users discover cross-boundary learning resources as a result of Explicit search, and when the resource is deemed relevant, they bookmark it in the Search result list. This finding from our current data points to the direction that H2 cannot be supported, and prompts us to reject H2.

4. Conclusion

In this paper we used empirical data in the form of attention metadata to track users’ actions. We then use this data to model the ecology of social search of a learning resource
portal integrated with a social bookmarking and tagging system. We conclude that Interest indicators have an important role as a part of the social search ecology, and we expect that studies into inter-relations of these variables will offer interesting further insights. By studying the cross-boundary discoveries, we find that users create underlying connections between resources that come from a number of countries and are in a variety of languages, which is important for bringing them under the umbrella of SIR methods.

Hypothesis 1 was accepted showing that the search taking advantage of Social Information Retrieval (SIR) methods yield more relevant resources with less effort from the user. Despite this edge, users have a strong search preference for Explicit search methods (2/3 of all executed searches). These conventional search methods strongly proved their role in discovering new resources and thus also alleviating the “cold start problem”. This also led us to reject our second hypothesis (H2): most cross-boundary resources are discovered using Explicit search. Encouraged by the H1, though, we believe that leveraging both implicit and explicit Interest indicators to support cross-boundary discovery (e.g. indicating the cross-boundary nature of resource discoveries, tag clouds filtered by language and by the country of users), and collaborative filtering methods for like-minded users based both on item and tag-based interests, are worth studying further.

A limitation in this research was that we studied a system that was still evolving. This has the positive effect of allowing us to revisit our SIR strategies for both cross-boundary and within-boundary discoveries to better support users. A second limitation is that as the study is based on server-side log-files, we have left out subjective measures such as user satisfaction or cognitive load while searching for cross-language content, which would also add important information in studying such system. Future studies should include testing wider range of SIR methods to leverage the underlying connections that users have created through social tagging and bookmarking between resources that originate from a number of repositories and are in a variety of languages.

References


Greece. CEUR, 307.


MELT, http://info.melt-project.eu, the public portal at http://lreforschools.eun.org


