



A Recommender System for the DSpace Open Repository Platform

Desmond Elliott, James Rutherford, John S. Erickson
Web Services and Systems Laboratory
HP Laboratories, Bristol
HPL-2008-21
April, 8 2008*

recommender
system, digital
repository,
user interface
design,
metadata,
collaboration,
bookmarking

We present Quambo, a recommender system add-on for the DSpace open source repository platform. We explain how Quambo generates content recommendations based upon a user selected set of examples, our approach to presenting content recommendations to the user, and our experiences applying the system to a repository of technical reports. We consider how Quambo could be combined with the peer-federated DSpace add-on to extend the item-space from which recommendations can be generated; a larger item-space could improve the diversity of the set from which to make recommendations. We also consider how Quambo could be extended to add collaboration opportunities to DSpace.

Internal Accession Date Only

Approved for External Publication

Submitted to Open Repositories 2008, Southampton, UK, April 1-4, 2008

© Copyright 2008 Hewlett-Packard Development Company, L.P.

A Recommender System for the DSpace Open Repository Platform

Desmond Elliott James Rutherford John S. Erickson
Hewlett-Packard Labs
Bristol, England; Vermont, U.S.A

April 7, 2008

Abstract

We present Quambo, a recommender system add-on for the DSpace open source repository platform. We explain how Quambo generates content recommendations based upon a user selected set of examples, our approach to presenting content recommendations to the user, and our experiences applying the system to a repository of technical reports. We consider how Quambo could be combined with the peer-federated DSpace add-on to extend the item-space from which recommendations can be generated; a larger item-space could improve the diversity of the set from which to make recommendations. We also consider how Quambo could be extended to add collaboration opportunities to DSpace.

1 Introduction

Recent studies [1] [2] of the use of institutional repositories in academic environments report that scholars see little added-value in incorporating their institution's repository as a regular part of their scholarly activities. We have developed an add-on to the DSpace open source repository which generates and presents content recommendations to make the institutional repository a more compelling resource for its users.

Quambo introduces changes to the DSpace user interface, extensions to the DSpace data model, and a recommender system. The DSpace user interface has been changed to show content recommendations on the item information

page and within a new usage paradigm we call *research contexts*. A research context is an extension to the DSpace data model which allows users to separate their research interests into distinct areas to help improve the quality of their personalised recommendations. A research context consists of three components: bookmarked items, recommended items, and keyword-value pairs representing the essence of research context. Quambo’s recommender system produces a set of recommended items using the Jaccard Similarity Index, modified to consider the weighting of metadata where appropriate.

Future research will examine the effectiveness of recommending items from within a federation of repositories and developing collaboration opportunities within DSpace. We are considering two approaches to recommending items within a federation: searching across a local cache of metadata harvested from “nominated” sources with a metadata harvesting interface, and federated search against keyword search equipped services. To facilitate collaboration, we will use Atom feed [3] representations of a research context to allow “mashups” with non-repository applications. We will also use Atom feeds to enable light-weight sharing of research contexts within DSpace.

The remainder of this report is structured as follows. Section 2 provides an overview of institutional repositories and recommender systems. Section 3 describes the add-on, Quambo, focusing on its additions to the DSpace data model, how Quambo generates content recommendations, and how users interact with Quambo in the context of the DSpace user interface. Section 4 describes our experiences with storing HP Labs Technical Reports on a local instance of DSpace with the Quambo add-on. Section 5 discusses further research projects that will focus on extending the reach of recommendations using federated caching or searching, and adding collaboration opportunities within and outside DSpace.

2 Related work

2.1 Institutional repositories

Jones et al. describe an institutional repository as a system to store, preserve, and provide access to the collection of articles, publications, reports, and grey literature output from an institution; an institution is typically a university, an industrial research laboratory, or an arts & culture organisation [4]. Davis and Connolly state that institutional repositories exist to offer management and access to the output of an institution [2].

OpenDOAR [5], an authoritative directory of academic open access repositories, reports there are over 1,000 open-access repositories across the world, at the time of writing. The most deployed platforms are DSpace [6] and EPrints [7] which each account for one quarter of the total number of deployed repositories. Repository66 [8] provides a Google Maps [9] visualisation of the OpenDOAR data and shows the repositories offering the most content do not use either DSpace or EPrints as a deployment platform, they use custom software and typically signify a *subject-based* collection. Examples of such collections include: ArXiv, primarily a repository of physics papers; PubMedCentral, a repository of open-access medical papers; and RePEc, a repository of economics papers. We know of no research which evaluates the reasons why these subject-based repositories have more content than institutional repositories; however, we believe the precedence amongst scholars in these areas leads them to deposit their papers in subject-based collections.

In a recent study, van Westrienen and Lynch report a positive trend in the deployment of institutional repositories [1]. The results of their survey show a 100% adoption rate in some countries; however, this trend was not universally observed and the adoption rate was found to be as low as 5% in some countries.

Institutional repositories are reported to suffer from a lack of use. Connolly and Davis observe the following: there is a personal investment in learning to use a new service which seems to provide no added value; scholars have concerns about violating publisher copyright over open-access; there is a perceived low quality of content provided by open-access repositories; and there is a fear of plagiarism as a result of “scooping” due to early release [2]. van Westrienen and Lynch observe the difficulties with convincing faculty about the value of institutional repositories and the difficulty of convincing faculty to contribute as an integral part of their scholarly process [1].

Despite these problems, Ferreria et al. report they attribute the success of their institutional repository to the combination of a promotional plan, an institution-wide mandated policy, and financial incentives [10]. They define success as the year-on-year growth of the number of items submitted to the repository by users and the year-on-year growth of the number of item downloads.

2.2 Recommender systems

Adomavicius and Tuzhilin define recommender systems as applications that identify sets of items of potential interest to a user based on the user's interaction with a system [11]. More formally, for the set of all users of a system and all items in a system, we can say a recommender system generates a set of recommended items for each user such that recommended items are maximally useful, for some definition of useful. They also define and describe three approaches to producing a set of recommended items: *content-based* recommendations; *collaborative filtering* recommendations; and a *hybrid* of content-based and collaborative filtering.

A content-based recommender system uses the properties of the distinct items a user has expressed an interest in to find similar items that may be useful to the user within a defined threshold of similarity [11]. Deshpande and Karypis' review of content-based recommender systems explains that these systems construct a user-item matrix and typically use a *cosine-based* similarity measure or a *conditional probability-based* similarity measure to generate recommendations for a user [12]. One problem with this type of recommender system is it is not able to scale effectively because the sparsity of the user-item matrix tends to increase as the number of users or items increases [12].

Adomavicius and Tuzhilin describe the collaborative filtering approach as a recommender system providing recommendations using the rating of items, as rated by other users who by some definition are similar, as the measure of the utility of an item to a user. The utility of an item to a user is determined by either a *heuristic-based* approach or a *model-based* approach. A heuristic-based approach uses the aggregation of ratings for the most similar users to determine the relevance of an item to a user. A model-based approach uses a trained model from the ratings using machine learning or statistical methods to determine the relevance of an item to a user [11].

One problem with both a collaborative filtering recommender system and a content-based recommender system is the *new user* problem. The new user problem occurs when a user who has no existing profile still wants to receive recommendations. Rashid et al. suggest an approach to solving the new user problem is to determine the most universally popular items, present these items to a new user to rate, and the results of the rating should be used to bootstrap the creation of a new user profile, which can then be used to provide recommendations [13].

3 Implementation

In this section we describe the data model and database schema of Quambo, how item recommendations are generated, and changes to the DSpace user interface which allow users to interact with Quambo.

3.1 Data model

Quambo adds research contexts, bookmarked items, and recommended items to the DSpace data model. A research context consists of a collection of bookmarked items, the metadata extracted from bookmarked items, and recommended items. Items are bookmarked through the user interface of the item display page and are said to be bookmarked into a research context. Recommended items are determined as explained in Section 3.3 and are stored as an isolated set to prevent a bookmarked item from also being a recommended item. A visual representation of how research contexts, bookmarks, and recommended items combine is presented in Figure 1.

The metadata extracted from bookmarked items forms the *essence* of a research context. The essence is a characterisation of a research context and is used to determine personalised recommendations based on the interactions of users. It contains a set of key-value pairs in which the key is a Dublin-Core *dc.subject* metadata field [14] and the value is the weight of the keyword in the essence. Bookmarking an item updates the essence of a research context with the keywords extracted from the newly-bookmarked item's metadata. This action either adds a new keyword-weight pair to the essence or increments the weight of an existing keyword-weight pair. In the case of an existing keyword, the weight of the keyword is incremented to signify the keyword is now more relevant in characterising the research context. We chose a value of 1 unit as the weight contributed by each occurrence of a keyword in a research context.

In the future Quambo may be extended to extract more metadata fields from items and research contexts to determine recommendations; these fields will be customisable by the scholar. This ability to customise will enable the scholar to express whether they want metadata of a particular type to be more or less significant in the calculation of recommendations, and hence the value of the weights will increase or decrease.

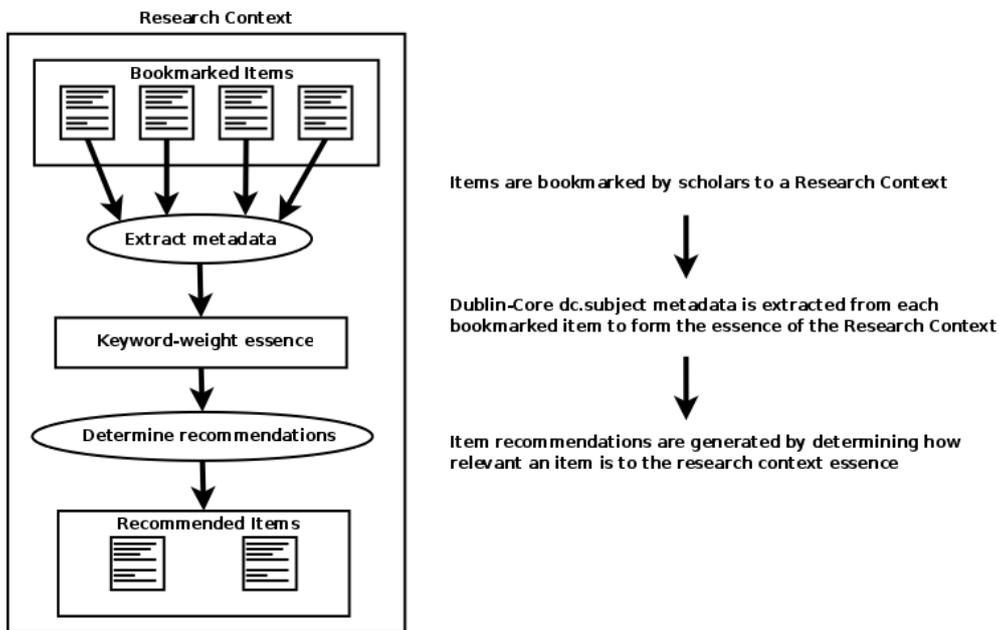


Figure 1: An abstract representation of the data model of a research context

3.2 Database schema

The data model of Quambo extends the DSpace database schema with five new tables. We use the “quambo_” prefix to prevent namespace clashing with existing or future additions to the DSpace database schema. Figure 2 visually describes the database schema additions described below.

The *quambo_research_context* table stores the name of the research context, the ID of the owner of the research context, the minimum similarity threshold of recommended items, whether the research context is publically accessible, and a UUID. The UUID will be used in the future to share a research context among users within DSpace, see Section 5.2 for more details.

The *quambo_research_context_keyword* table stores the ID of the research context this keyword belongs to, the text of the keyword, and the weight of the keyword. Multiple rows will exist in this table if a scholar has more than one research context which contains the same keyword.

The *quambo_bookmark* table stores the ID of the research context the item is bookmarked in, the ID of the bookmarked item, and the ID of the person who bookmarked the item.

The *quambo_item_recommendation* table stores the ID of the research context this item is recommended in, the ID of the item recommended, and the similarity rating calculated when determining the item recommendation. Re-calculation of item recommendations updates table rows as appropriate. Only the 6 most relevant item recommendations are stored per research context.

The *quambo_eperson_properties* table stores the ID of the user, the research context ID of the user's initial research context, and whether the user has ever logged in. We store whether the user has logged in because Quambo needs to create an initial research context for a user. Every user needs an initial research context so they can always bookmark an item with respect to some research context.

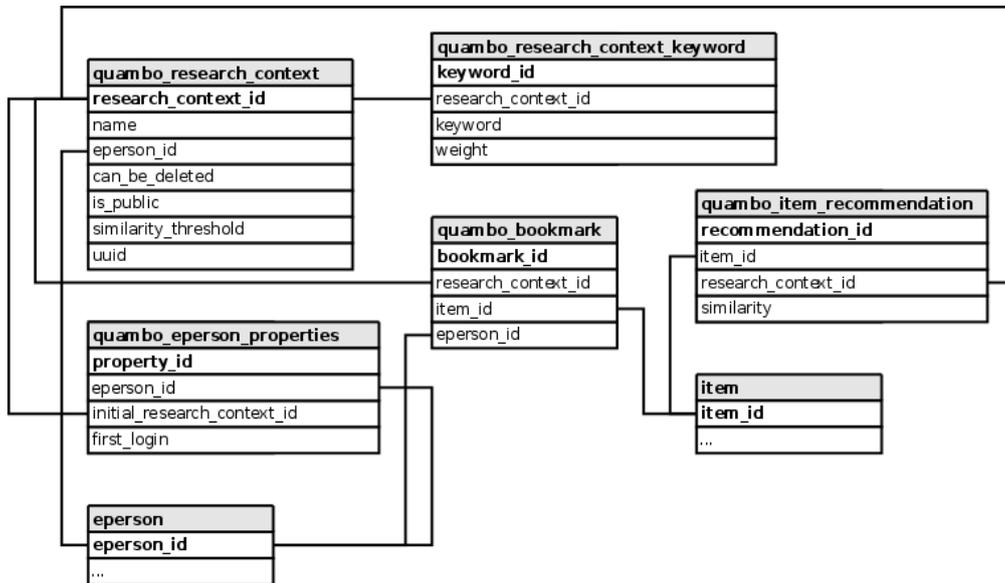


Figure 2: The database schema added by Quambo to the standard DSpace schema

3.3 Generating recommendations

In Quambo, content recommendations are either generated on an item-to-item basis for presentation on an item information page, or by determining how relevant an item is to a research context. Determining recommendations is a two step process: first, keyword metadata needs to be extracted from

each object being compared; second, an appropriate similarity measure needs to be applied to determine if a recommendation should be made.

The keyword metadata extracted from each object depends on the situation: in the case of an item-to-item comparison, we extract Dublin-Core *dc.subject* metadata fields and attribute a weight of 1 with them; when determining the relevance of an item to a research context, we extract Dublin-Core *dc.subject* fields from the item, with a weight of 1, and key-value pairs from the research context essence.

Item-to-item recommendations are generated using the Jaccard Similarity Index, which, for two sets is defined as the size of the intersection divided by the size of the union (Equation 1) [15]. When comparing two items, the size of the intersection and the size of the union are defined to be the sum of the weight of the keywords representing these sets.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

To determine the relevance of an item to a research context we cannot use the Jaccard Similarity Index as defined in Equation 1, because the weight of intersecting keywords are not necessarily equal. This is because the key-value pairs representing the weight of a keyword in a research context essence could be different to the weight of the keywords extracted from an item. To calculate the size of the intersection and the size of the union we need to consider how the weighting of these keywords affect the relevance of an item to a research context. We do this by modifying the Jaccard Index to take into account the weight of the keywords in the essence of the research context.

The Modified Jaccard Index presented as pseudocode in Algorithm 1. This algorithm is a modification of the Jaccard Index due to the process of calculating the size of the intersection and the size of the union. Lines 3 - 5 calculates the sum of the weight of the keywords in the intersection of the research context and the item. Lines 6 - 12 calculates the sum of the weights of the keywords in the union where lines 7 - 9 calculate the weight of the keywords that are present only in the item *dc.subject* fields, and lines 10 - 12 calculate the weight of the keywords that are present only in the research context essence.

We recognise the limited value that keywords have due to their variability and are considering which metadata we could extract from each object in the future to improve the quality of recommendations.

Input: A: an item
Input: B: a research context
Output: the relevance of item A to research context B:
similarity $\in \mathbb{R}$, $0 \leq \text{similarity} \leq 1$

```

1 begin
2    $\cap$ -weight  $\leftarrow 0$ 
3   foreach  $keyword \in A \cap B$  do
4      $\cap$ -weight  $\leftarrow \cap$ -weight +  $\text{weight}_B(\text{keyword})$ 
5   end
6    $\cup$ -weight  $\leftarrow \cap$ -weight
7   foreach  $keyword \in A - (A \cap B)$  do
8      $\cup$ -weight  $\leftarrow \cup$ -weight + 1
9   end
10  foreach  $keyword \in B - (A \cap B)$  do
11     $\cup$ -weight  $\leftarrow \cup$ -weight +  $\text{weight}_B(\text{keyword})$ 
12  end
13  return  $\frac{\cap\text{-weight}}{\cup\text{-weight}}$ 
14 end

```

Algorithm 1: Modified Jaccard Index

3.4 Interacting with Quambo

We have changed the standard DSpace user interface to enable users to bookmark items, interact with their research contexts, and receive content recommendations. The changes apply to the the *item display* page and the *MyDSpace* page.

Item bookmarking The item display page has been split to show item-to-item recommendations next to the standard item information display. If a registered user is logged in, they can bookmark or unbookmark items to any research context they can access. We experimented with highlighting the matching *dc.subject* elements of recommended items, but this was found to

clutter the presentation of the recommendations. A maximum of the six most similar items are recommended in-line based on the results of [16]. Figure 3 shows our modification to the display item page. The bookmarking tools and recommended items are presented in detail in Figure 4.

The screenshot shows a detailed view of an item page. On the left, there is a metadata section with the following information:

- Title:** Adaptive User Profiles for Enterprise Information Access
- Authors:** [Mühlmeier, Elke](#), [Cayzer, Steve](#), [Shahajee, Paul](#)
- Keywords:** user profiles, del.icio.us, tagging, folksonomy, web2.0, semantic web, information access, enterprise information management
- Issue Date:** 22-Nov-2007
- Abstract:** A major challenge for users of enterprise information is finding the information they want, presented in a way that makes sense to them. In this paper we tackle this problem by creating adaptive user profiles from implicit behaviour. In traditional approaches to information filtering, the user has to explicitly create his or her profile, and manually keep the profile up to date. Taking advantage of the popularity of collaborative tagging systems, we use the recorded tagging behaviour to construct an implicit, yet realistic and dynamic user profile. We present and evaluate algorithms for creating such profiles, characterizing their behaviour through statistical analysis. In addition, we present a visualisation tool which was used in a small scale user study to provide insight as to the effectiveness of our approach. Finally, we show how the profiles can be leveraged to enable personalised access to enterprise data sources.
- URI:** <http://hdl.handle.net/paperstore/81>
- Appears in:** MISG

Below the metadata is a table titled "Files in This Item":

File	Description	Size	Format
HPL-2007-72.pdf		1.53 MB	Adobe PDF View/Open

Below the table is a button: [Show full item record](#)

At the bottom, there is a small text: "Items in DSpace are protected by copyright, with all rights reserved, unless otherwise indicated."

On the right side of the page, there is an "Item Tools" section with a "Bookmark" button and a dropdown menu set to "Initial Context". Below it is an "Unbookmark" button and a dropdown menu set to "Quambo".

Below the "Item Tools" is a "Similar Items" section with a list of recommended items:

- [Clustering Tags in Web and Enterprise Folksonomies](#)
- [Digital Object Repositories Explained: An Intractable White Paper](#)
- [An Assessment of Tag Presentation Techniques](#)
- [Using Annotations in Enterprise Search](#)
- [Social Tagging and Folksonomy: steve.museum and Access to Art](#)

Figure 3: The modified Item Display page showing the bookmarking tools and the presentation of similar items.

This figure provides a detailed view of the "Item Tools" and "Similar Items" sections. The "Item Tools" section contains two buttons: "Bookmark" and "Unbookmark". The "Bookmark" button has a dropdown menu set to "Semantic Web". The "Unbookmark" button has a dropdown menu set to "Quambo".

The "Similar Items" section contains a list of recommended items:

- [Semantic Views for Controlled Access to the Semantic Web](#)
- [Perceived Values and Benefits of Institutional Repositories: A Perspective of Digital Curation](#)
- [the future of the institutional repository: making it personal](#)
- [The Future of DSpace: Making it Personal](#)
- [New Recommendation Techniques for Multicriteria Rating Systems](#)

Figure 4: Detail of the bookmarking and item recommendations

Research Contexts The MyDSpace interface provides a point of interaction for research contexts. Users can create and maintain research contexts, view content recommendations, and visualise the keyword-weight pairs that characterise the essence of their research context. In Figure 5, recommendations are shown in the *Recommended Items* section ordered from most to least relevant. Bookmarked items are shown in the *Bookmarks* section, and the metadata extracted from those items is shown in the *Essence* section.

We experimented with showing how relevant an item was to a research context using an interface similar to the Amazon.com star rating. Our test users did not understand what the star ratings meant and initially believed they were how other testers had rated a document. We abandoned the star-rating representation of the relevance of an item and are investigating other models of feedback.

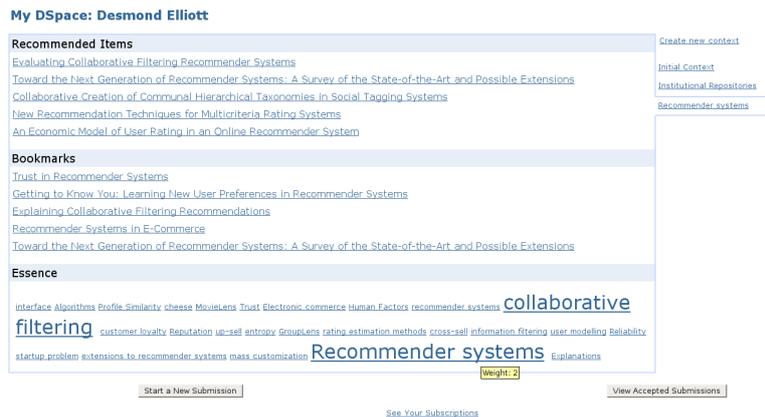


Figure 5: Changes to the MyDSpace interface, including research contexts available to this user, recommended items, bookmarked items, and the essence

Figure 6 shows how users can interact with more than one research context. The *Create new context* link allows a user to create a new research context, shown in Figure 8; there is a list of the research contexts available to this user and the selected research context is highlighted as being part of the main body of the page. Figure 7 shows the detail of the Essence, including a tooltip explaining the weight of the keyword “Recommender systems”. Figure 8 shows how users can create a research context. Users are prompted to attempt to bootstrap the recommendation process by including up to three

seeding phrases. This is our approach to dealing with the *new user problem* described in [11].

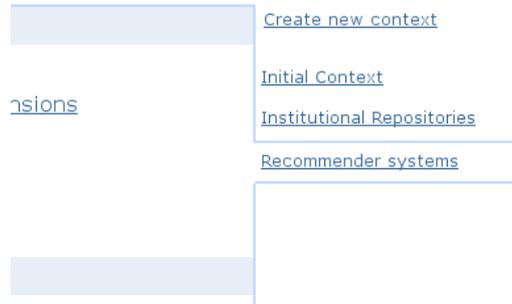


Figure 6: Detail of the user interface for multiple research contexts for a scholar



Figure 7: Detail of the visualisation of the essence of a research context showing a tooltip highlighting the weight of the Recommender Systems keyword

Create a new research context

Context Name

Seed Phrases

Figure 8: Detail of how research contexts can be created by users

4 Paperstore

Paperstore is an HP Labs internal DSpace service with the Quambo add-on. The purpose of Paperstore is to allow researchers at HP Labs to receive recommendations of related work produced by the organisation. It runs on an HP DL585 server, further details of the system can be found in the Appendix.

We hope Paperstore will eventually store all HP Labs Technical Reports, it currently holds 163 items in 9 collections. To achieve this we will ingest the Technical Reports into Paperstore using the DSpace *ItemImport* tool and an XSLT transformation of the XML representation of the Technical Report metadata provided by the HP Library.

The service has 13 registered users who have created 22 research contexts. 66 items have been bookmarked and 253 items have been recommended to users. The distribution of the similarity of the recommended items is shown in Figure 9. This figure shows many recommended items are not convincing recommendations, however, our dataset is small and the metadata we use to generate recommendations is limited.

The research context essences are characterised by 271 keyword-value pairs; the distribution of keyword-value pairs is shown in Figure 10. This figure shows that the weight of a keyword in a research context may not have a significant impact on calculating recommendations, however, we hope that a larger dataset will lead to richer research context essences and better recommendations.

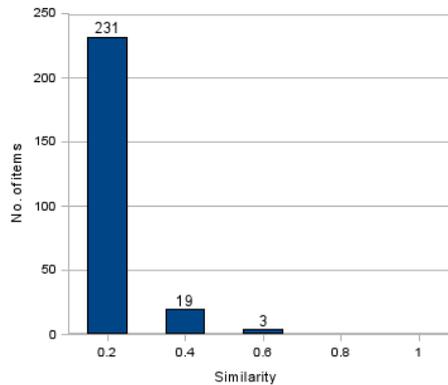


Figure 9: Distribution of the calculated similarities of recommended items

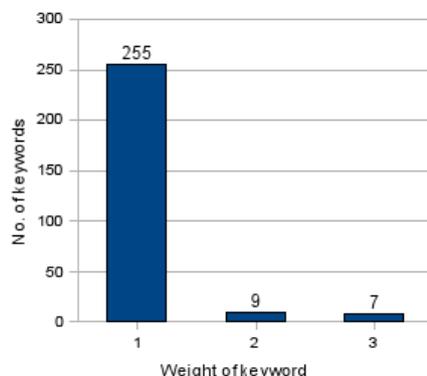


Figure 10: Distribution of the keyword-value pairs in research contexts

5 Further work

5.1 Federated recommendations

The value of recommendations is limited within an isolated repository because repositories rarely offer both a good breadth and good depth of content. We plan to extend Quambo to enable repositories to be combined to achieve a good breadth and depth of the domain upon which recommendations are based, improving the quality and quantity of recommendations. Our strategy is to increase the number of sources from which content metadata is harvested. One approach for achieving this is to locally cache metadata from repositories in a federation; another approach uses keyword-based searching of repositories in a federation, and other non-repository sources, such as blogs and wikis.

Local caching of item metadata harvested from repositories in a federation will allow Quambo to generate recommendations from a larger item-space. One problem with this approach is that local metadata storage for repositories in large federations may be excessive; however, the peer-federated DSpace add-on allows individual nodes to nominate which peers they want to harvest metadata from, thereby managing the scale of the harvest.

Keyword-based searching of repositories in a federation will allow Quambo to send targeted queries for content through a federation. A scholar's research context will have a keyword search query associated with it that would regularly search the federation for new potential content recommendations. The

result of these queries will be analysed by Quambo when calculating recommendations. The OAI-SQ standard [17] will be used to implement keyword-based searching because it is a simple extension to OAI-PMH [18], which is an existing protocol used in DSpace.

5.2 Socialising Quambo

We believe there is value in allowing research contexts to be shared between users within DSpace and with applications outside DSpace. Our strategy is to allow a research context to either be publically published in a repository, privately shared with other users, or exported from DSpace to be used in other applications. All three approaches will use Atom feeds for light-weight publishing of research contexts.

Publically publishing a research context is a read-only concept which will allow any repository visitor to view the contents of a research context similar to the way they view the list of items in a collection. The value in publishing a research context would, for example, allow artists at an arts & culture organisation to make public the sources used as inspiration for new works.

Privately sharing a research context among a group of registered users is a read-and-write concept which will add a noticeably different research context tab to a user's list of research contexts and allow a group to share resources for current projects.

Publishing research contexts from DSpace would allow users to use recommended items and bookmarked items outside DSpace in desktop or web applications, such as an RSS reader to receive regular updates on new additions to the research context without visiting DSpace.

6 Conclusion

We have developed an add-on to the DSpace open source repository platform which allows users to bookmark items in a repository and receive recommendations based on their bookmarked items to make the institutional repository a more compelling resource for scholars as part of their regular scholarly activities.

The open source nature of DSpace means we will share our work with the DSpace community so institutions can experiment with the changes introduced by the Quambo add-on. We hope to be able to apply the add-on to the HP Library DSpace server and hope that feedback received from system administrators will be valuable in helping us evaluate the performance of our recommender system over larger datasets and to receive more feedback from users to help us refine our user interface design.

7 Appendix

Paperstore runs on Red Hat Enterprise Linux 4, with the 2.6.9-42.0.10.ELsmp kernel, the PostgreSQL 8.2 database server, and the Tomcat 6.0.7 application container. No optimisations have been made to the PostgreSQL server or the Tomcat server. Descriptions of the CPUs and RAM appear in the following sections.

7.1 CPUINFO

For each processor, 0 to 7, the response of the following command is shown below:

```
cat /proc/cpuinfo
```

```
processor      : 0
vendor_id     : AuthenticAMD
cpu family    : 15
model         : 33
model name    : AMD Opteron (tm) Processor 880
stepping      : 2
cpu MHz       : 2199.964
cache size    : 1024 KB
physical id   : 0
siblings      : 2
core id       : 0
cpu cores     : 2
fpu           : yes
fpu_exception : yes
cpuid level   : 1
wp            : yes
flags         : fpu vme de pse tsc msr pae mce cx8 apic sep
```

```

                mtrr pge mca cmov pat pse36 clflush mmx fxsr
                sse sse2 ht syscall nx mmxext lm 3dnowext
                3dnow pn1
bogomips       : 3599.38
TLB size      : 1088 4K pages
clflush size   : 64
cache_alignment : 64
address sizes  : 40 bits physical, 48 bits virtual
power management: ts fid vid ttp

```

7.2 MEMINFO

The output of the following command is shown below:

```
cat /proc/meminfo
```

```

MemTotal:      10073700 kB
MemFree:       23092 kB
Buffers:       109536 kB
Cached:        2721100 kB
SwapCached:    0 kB
Active:        7174512 kB
Inactive:      2633628 kB
HighTotal:     0 kB
HighFree:      0 kB
LowTotal:      10073700 kB
LowFree:       23092 kB
SwapTotal:     2619248 kB
SwapFree:      2619020 kB
Dirty:         792 kB
Writeback:     0 kB
Mapped:        7034148 kB
Slab:          176692 kB
CommitLimit:   7656096 kB
Committed_AS: 15952392 kB
PageTables:    28252 kB
VmallocTotal: 536870911 kB
VmallocUsed:   3332 kB
VmallocChunk: 536866551 kB
HugePages_Total: 0
HugePages_Free: 0

```

Hugepagesize: 2048 kB

References

- [1] G. van Westrienen and C.A. Lynch. Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid-2005. *D-Lib Magazine*, 11, 2005.
- [2] P.M. Davis and M.J.L Connolly. Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*, 13, 2007.
- [3] J. Gregorio and B. de hOra. The Atom Publishing Protocol. <http://tools.ietf.org/html/rfc5023>.
- [4] R. Jones, T. Andrew, and J. MacColl. *The Institutional Repository*. Chandos, 1st edition, 2006.
- [5] OpenDOAR. Open Access Repositories. <http://www.opendoar.org/>.
- [6] DSpace. <http://www.dspace.org/>.
- [7] EPrints for Digital Repositories. <http://www.eprints.org/>.
- [8] Stuart Lewis. Repository 66. <http://www.repository66.org/>.
- [9] Google. Google maps. <http://maps.google.com/>.
- [10] M. Ferreira, A. A. Baptista, E. Rodrigues, and R. Saraiva. Carrots and Sticks. *D-Lib Magazine*, 14(1), January 2008.
- [11] G. Adomavicius and A. Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, June 2005.
- [12] M. Deshpande and G. Karypis. Item-based Top-N Recommendation Algorithms. *ACM Transactions on Information Systems*, 22(1):143–177, January 2004.
- [13] A. Rashid, I. Albert, D. Cosley, S. Lam, S. McNee, J. Konstan, and J. Riedl. Getting to know you: Learning new user preferences in recommender systems. pages 127–134, 2002.
- [14] Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, Version 1.1. <http://dublincore.org/documents/dces/>.

- [15] Paul Jaccard. The Distribution of the Flora of the Alpine Zone. *New Phytologist*, 11:37–50, 1912.
- [16] S. S. Iyengar and M. Lepper. When Choice is Demotivating: Can One Desire too Much of a Good Thing? *Journal of Personality and Social Psychology*, 79:995–1006, 2000.
- [17] Internet Scout Project. OAI-SQ. <http://scout.wisc.edu/Projects/OAISQ/>.
- [18] Open Archives Initiative. The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/pmh/>.