OERScout: Widening Access to OER through Faceted Search

Ishan Sudeera Abeywardena, Wawasan Open University, ishansa@wou.edu.my
Chee Seng Chan, University of Malaya, cs.chan@um.edu.my
V. Balaji, Commonwealth of Learning, vbalaji@col.org

Sub theme: Promoting Open Educational Resources (OER)

ABSTRACT

In recent years, the Open Educational Resources (OER) movement has achieved considerable success within the academic community with respect to advocacy of the concept. As a result, many organisations such as the Commonwealth of Learning (COL), UNESCO and the International Development Research Centre (IDRC), in partnership with academic institutions, have produced large volumes of OER. However, due to the disconnected nature and the constant expansion of volume, many repositories hosting these resources are less frequented or completely ignored by OER users. i.e. only the more popular OER repositories such as Connexions and WikiEducator are frequent stops in the search for academically useful resources. This limitation, in turn, reduces the access to high quality resources hidden away in isolated repositories hosted by lesser known sources. Furthermore, the time and labour required to trawl these repositories with a view of identifying the most suitable OER is tantamount to creating ones' own material from scratch. As a solution to these issues, this paper discusses how the OERScout technology framework uses a "faceted search" approach to locate the most desirable OER from sources spread throughout the globe. It also highlights how focused searching can greatly improve access to OER readily useable in teaching and learning.

Keywords: OERScout, OER Search, OER Curation, Faceted Search, Access to OER

INTRODUCTION

OER are fast gaining traction amongst the academic community as a viable means of increasing access and equity in education. The concept of OER is of especial significance to the marginalised communities in the Global South where distance education is prominent due to the inability of conventional brick and mortar institutions to cope with the growing demand (Lane, 2009). However, the wider adoption of OER by academics in the Global South has been inhibited due to various socio, economic and technological reasons. One of the major technological inhibitors is the current inability to search for OER which are academically useful and are of an acceptable academic standard.

In his study on "*which inhibiting factors for reuse do content developers in developing countries experience with open content?*" Hatakka (2009) points out that the most inhibiting factor is the inability to locate 'relevant' material for a particular teaching or learning need. The subjects of Hatakka's study attribute the inability to locate relevant material to (i) the inability to locate resources which fit the scope of the course in terms of context and difficulty; (ii) the lack of awareness with respect to how 'best' to search for material on the Internet; and (iii) the inability to choose the most appropriate resources from the large number of resources returned by search engines such as Google. Affirming these statements, Shelton et al. (2010, p. 316) argue

"Well-studied and commercialized search engines like Google will often help users to find what they are seeking. However, if those searching do not know exactly what they are looking for, or they do not know the 'proper' words to describe what it is that they want, the searching results returned are often unsatisfactory".

In an attempt to identify how effective mainstream search engines such as Google are with respect to locating relevant OER, Dichev et al. (2011) of the Winston-Salem State University conducted an

experiment by putting Google head to head against native search mechanisms of OER repositories. To make the Google search narrower to OER, the advanced search feature 'free to use, share or modify, even commercially' was used. Alongside Google, native search mechanisms of 12 OER repositories were used to search for material in the computer science domain. The repositories were namely: Connexions, MIT OpenCourseWare, CITIDEL, The Open University, OpenLearn, OpenCourseWare Consortium, OER Commons, Merlot, NSDL, Wikibooks, SOFIA, Textbook Revolution and Bookboon. From their comparison between Google and native OER search mechanisms with respect to locating relevant material, it is apparent that native search mechanisms fair better than Google in terms of locating relevant material.

Commenting further on the inability of mainstream search engines such as Google to effectively locate OER, Pirkkalainen & Pawlowski (2010, p. 24) state that *"… searching this way might be a long and painful process as most of the results are not usable for educational purposes"*. Furthermore, they argue that search mechanisms native to OER repositories are capable of locating resources with an increased relevance. However, the problem is which repositories to choose within the large global pool. Levey (2012, p. 134) relates to this from working in the African 'AgShare' project. She states her experience as

"Despite numerous gateways, it is not always easy to identify appropriate resources. How a resource is tagged or labelled is one problem. Poor information retrieval skills is another. Furthermore, academics are busy".

This inadequacy with respect to searching for OER from a diversity of sources gives rise to the need for new alternative methodologies which can assist in locating relevant resources. Ideally these search tools should return materials which are relevant, usable and from a diversity of sources (Yergler, 2010). Yergler further suggests that the reliance on a full text index and link analysis of mainstream search engines impede the process of discovery by including resources not necessarily educational. As such, *"increasing the relevance of the resources returned by a search engine can minimize the time educators need to spend exploring irrelevant resources"* (Yergler, 2010, p. 2).

The UNESCO Paris OER Declaration (2012) , which is a global non-binding declaration signed by many governments, declares the need for more research into OER search. The recommendation reads

"i. Facilitate finding, retrieving and sharing of OER: Encourage the development of user-friendly tools to locate and retrieve OER that are specific and relevant to particular needs. Adopt appropriate open standards to ensure interoperability and to facilitate the use of OER in diverse media".

This declaration is the culmination of a global effort towards establishing a roadmap for the future development of the OER movement. The above recommendation made with respect to OER search reaffirms the need for new and more effective OER search methodologies within the context of locating relevant material for particular teaching and learning needs.

THE FACETED SEARCH APPROACH

Search engines have undergone rapid evolution in the past decade due to global technological giants such as Google. In his book 'Faceted Search'*,* Daniel Tunkelang (2009) of Google explains how previous search technologies morphed into the faceted search approach. According to Tunkelang, the earliest search engines used the Boolean retrieval model which limited the flexibility and increased the complexity of the search query. Abandoning this method, information retrieval (IR) researches adopted a free-text query approach which provided increased flexibility in creating search queries. This method cast a wide net to return results based on rank. Although not as accurate as Boolean retrieval, many search engines still follow the free-text query approach incorporating the ranked retrieval framework. Another approach used in searching for information, especially on the World Wide Web (WWW), is the directory approach. The advantage of this approach is the organisation of content based on set taxonomies. This allowed users to navigate categories and sub categories to ultimately arrive at the information they are after. However, Tunkelang highlights that the creators of the taxonomies themselves and the users frequently

disagree on the categorisation of the content. i.e. users will have to learn to think like the creators to find the relevant information.

Faceted search is a hybrid search approach which combines parametric search and faceted navigation (Tunkelang, 2009). According to Dash et al. (2008, p. 3) *"First, it smoothly integrates free text search with structured querying. Second, the counts on selected facets serve as context for further navigation".* Marti Hearst of UC Berkley, who was the lead researcher in the popular *Flexible information Access using Metadata in Novel COmbonations* (Flamenco) faceted search project, argues *"a key component to successful faceted search interfaces (which unfortunately is rarely implemented properly) is the implementation of keyword search"* (Hearst, 2006, p. 4)*.* In simpler terms, modern faceted search combines free-text querying to generate a list of results based on keywords which can then be refined further using a Boolean, structured or directory approach. To achieve this functionality, faceted metadata need to be extracted from documents using text mining techniques. A few general strategies are (i) exploit latent metadata such as document source, type, length; (ii) use rule based or statistical techniques to categorise documents into predetermined categories; and (iii) use an unsupervised approach such as terminology extraction to obtain a list of terms from the document (Tunkelang, 2009).

Typical interaction between a faceted search interface and the user is explained by Ben-Yitzhak et al, (2008) as (i) type or refine a search query; or (ii) navigate through multiple, independent facet hierarchies that describe the data by drill-down (refinement) or roll-up (generalization) operations. Koren et al. (2008, p. 477) further explains this interaction as

"The interfaces present a number of facets along with a selection of their associated values, any previous search results, and the current query. By choosing from suggested values of these facets, a user can interactively refine the query."

Ultimately, faceted search allows users to quickly drill down into a more focused set of search results using the initial results set.

SEARCHING FOR OER WITH OERSCOUT

The 'OERScout' technology framework (Abeywardena, Chan, & Tham, 2013) is a comprehensive solution to the current OER search dilemma (Abeywardena & Chan, 2013). It uses text mining techniques to autonomously mine specific keywords which accurately describe the academic domains of a particular OER. In essence, OERScout (i) "reads" textual educational resources; (ii) "understands" the content; and (iii) "recommends" the most useful resources for a particular teaching or learning need. The usefulness of a particular OER is parametrically measured using the 'Desirability' framework (Abeywardena, Raviraja, & Tham, 2012) which takes into consideration the (i) openness; (ii) accessibility; and (iii) relevance attributes of a resource. The system then creates a searchable dataset called the Keyword-Document Matrix (KDM) which is used by the OERScout client interface to effectively search for resources.

In addition to the text mining techniques employed at the server end to create the KDM, the user interface of OERScout equally contributes to the novelty of this solution. The faceted search approach available to the users via the client interface is a far cry from the conventional free text search method where users are presented with a static list of search results spread across hundreds of pages. It is also superior to the directory search method where users are forced to manually drill down multiple layers before arriving at the resources they are after.

The searching for desirable OER using the OERScout interface is threefold. Firstly, the user inputs a search query into the free text search box. Unlike in existing search methodologies where the accuracy of the search query governs the relevance of the search results, OERScout extracts the key terms from the search query by removing stop words to form multiple focused search queries. These queries are then executed on the KDM to generate a list of 'Suggested Terms'. The suggested terms act as the first facet which allows the user to select from a broad list of domains autonomously mined by OERScout. Secondly, the user selects a particular area of interest from the list of suggested terms. This action

creates the second facet which lists the 'Related Terms' to the selected suggested term. Thirdly, the user hones in on the exact subject domain he/she is after in the related terms facet to generate a ranked list of desirable resources.
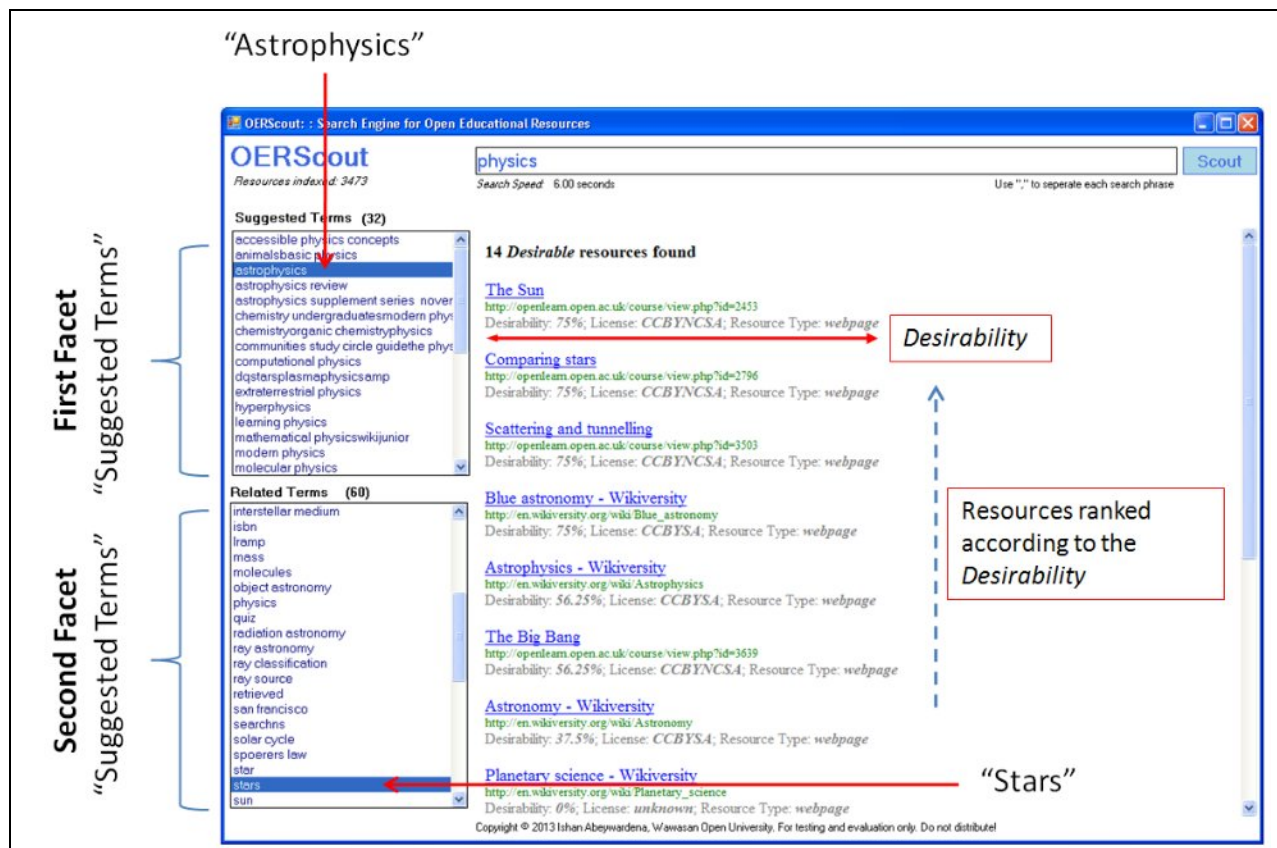

Figure 1 OERScout faceted search user interface. The figure shows a search conducted for *Physics: Astrophysics: Stars*.

Figure 1 shows an example of a faceted search conducted on OERScout to locate resources in "Physics". The suggested terms facet has listed 32 different topic areas identified by the system in the domain of "Physics". According to the selection in the first facet which is "Astrophysics", 60 related topics have been listed in the second facet. Based on the selection in the second facet, a list of desirable resources have been presented to the user which covers the topic "stars". The resources are arranged in descending order of the Desirability. The Desirability, license type and resource types are also indicated to the user to facilitate faster selection.

Referring to Figure 1, the top three resources returned are from the OpenLearn repository of The Open University which is highly reputed for the quality of its academic content. From this example, it is apparent that the OERScout faceted search interface allows users to quickly and effectively hone in on desirable OER required for their teaching and learning needs. It also spares users from reading a large number of resources returned by a search engine to ascertain their usefulness for a particular academic purpose.

BENEFITS OF OERSCOUT TO THE COMMUNITY

A number of technological factors have contributed to the current OER search dilemma. The first among these is the inability of mainstream search engines such as Google to effectively locate useful resources for academic purposes. Furthermore, the dependence of these search engines on human annotated metadata and commercial page ranking algorithms force them to give prominence to the widely popular

OER repositories such as Wikipedia, WikiEducator and Connexions. However, these repositories might not contain material which are the most desirable for a particular academic need. As such, the use of these search engines limits the access to the wider resource pool available globally. OERScout addresses this issue by autonomously mining metadata for search purposes. As a result, the annotation of resources becomes consistent and uniform. Furthermore, the learning aspect of the algorithm constantly strives to identify the most accurate metadata for a particular OER. When combined with the Desirability framework, OERScout objectively determines the usefulness of a resource based only on the needs of the user. In turn, resources from less popular repositories will be given the same search visibility as the resources from the more popular. This aspect of the system significantly increases the access to the global pool of quality OER.

The repository independence of OERScout is another key benefit to the OER community. Traditionally, content creators would have to archive their material on a repository to enhance searchability. Furthermore, they would have to provide the necessary metadata and comply with the repository's technological requirements. Due to the heterogeneity of these repositories, this task becomes a time consuming and slightly complicated one. Additionally, the heterogeneity of these repositories contributes to the inconsistencies in the search process. These issues once more result in limiting access to OER. In contrast, OERScout is not affected by the heterogeneity of the repositories. It further promotes the decentralisation of resources. As such, content creators can opt to make available their resources via a personal blog, personal website, institutional website or even a cloud space. OERScout will allow users to easily locate these resources through its faceted search interface whereby the visibility of these resources is increased.

In sum, the OERScout technology framework provides a viable solution to the current OER search dilemma. Through the use of the Desirability framework and the faceted search approach, it allows users to locate OER which were previously invisible in the searchsacpe. We see it as a game changer in terms of widening access to desirable OER for academic purposes. The current version of the system is only available as a prototype. We intend to provide a publically accessible faceted search interface in the near future.

REFERENCES

Abeywardena, I.S., Raviraja, R., & Tham, C.Y. (2012). Conceptual Framework for Parametrically Measuring the Desirability of Open Educational Resources using D-index. *International Review of Research in Open and Distance Learning , 13* (2), 104-121.

Abeywardena, I.S., & Chan, C.S. (2013). Review of the Current OER Search Dilemma. *Proceedings of the 57th World Assembly of International Council on Education for Teaching (ICET 2013).* Nonthaburi, Thailand: ICET.

Abeywardena, I.S., Chan, C.S., & Tham, C.Y. (2013). OERScout Technology Framework: A Novel Approach to Open Educational Resources Search. *International Review of Research in Open and Distance Learning , 14* (4), 214-237.

Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., et al. (2008). Beyond basic faceted search. *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 33-44). Palo Alto: ACM.

Dash, D., Rao, J., Megiddo, N., Ailamaki, A., & Lohman, G. (2008). Dynamic faceted search for discovery-driven analysis. *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 3-12). Napa Valley: ACM.

Dichev, C., Bhattarai, B., Clonch, C., & Dicheva, D. (2011). Towards Better Discoverability and Use of Open Content. *Proceedings of the Third International Conference on Software, Services and Semantic Technologies S3T* (pp. 195-203). Berlin Heidelberg: Springer.

Hatakka, M. (2009). Build it and they will come?–Inhibiting factors for reuse of open content in developing countries. *The Electronic Journal of Information Systems in Developing Countries , 37* (5), 1-16.

Hearst, M. (2006). Design recommendations for hierarchical faceted search interfaces. *In ACM SIGIR workshop on faceted search*, (pp. 1-5).

Koren, J., Zhang, Y., & Liu, X. (2008). Personalized interactive faceted search. *Proceedings of the 17th international conference on World Wide Web* (pp. 477-486). Beijing: ACM.

Lane, A. (2009). The impact of openness on bridging educational digital divides. *The International Review of Research in Open and Distance Learning , 10* (5).

Levey, L. (2012). Finding Relevant OER in Higher Education: A Personal Account. In J. Glennie, K. Harley, N. Butcher, & T. van Wyk (Eds.), *Open Educational Resources and Change in Higher Education: Reflections from Practice* (pp. 125-138). Vancouver: Commonwealth of Learning.

Pirkkalainen, H., & Pawlowski, J. (2010). Open Educational Resources and Social Software in Global E-Learning Settings. In P. Yliluoma (Ed.), *Sosiaalinen Verkko-oppiminen* (pp. 23-40). Naantali: IMDL.

Shelton, B. E., Duffin, J., Wang, Y., & Ball, J. (2010). Linking OpenCourseWares and Open Education Resources: Creating an Effective Search and Recommendation System. *Procedia Computer Science , 1* (2), 2865-2870.

Tunkelang, D. (2009). Faceted search. In G. Marchionini (Ed.), *SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES* (Vol. 5, pp. 1-80). Morgan & Claypool.

UNESCO. (2012). *Paris OER Declaration.* Paris.

Yergler, N. R. (2010). Search and Discovery: OER's Open Loop. *Proceedings of Open Ed 2010.* Barcelona.