On the Identification of Several Key Issues on OER Discovery for Smart Learning Environments [†]

Pablo Molins-Ruano *, Francisco Jurado and Pilar Rodriguez

Dpto. Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, 28049, Spain; francisco.jurado@uam.es (F.J.); pilar.rodriguez@uam.es (P.R.)

* Correspondence: pablo.molins@uam.es; Tel.: +34-91-497-2292

† Presented at the 13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019, Toledo, Spain, 2–5 December 2019.

Abstract: The best Open Educational Resources (OER) for each final user can be hard to find. OER come from lots of sources, in many different formats, conforming to a diverse logical structure and each user may present different objectives depending on their role in the teaching/learning process and their context. Previous attempts have only focused on one kind of technologies. A new approach that embraces diversity, may gain from the potential synergies of sharing resources in the development of the final recommendation system and the exploitation of the data. In this work, we aim at identifying the main challenges facing the field of OER recommendation, for a potential architecture model.

Keywords: Open Educational Resources; recommendation systems

1. Introduction

Open Educational Resources (OER) are those *"teaching, learning and research materials in any medium, digital or otherwise, that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions"* [1]. Thanks to their open perspective, OER present an opportunity to make education more accessible, in many different ways. For the instructors, OER can be a great starting point, as the vast majority of teachers build on previous resources to prepare their own. For the students in an educational institution, having access to a more diverse and wide pool of resources improves the chance of finding the proper resource to gain new knowledge. In addition, for those not in an educational institution, but still interested in learning such as long-life learners or people without the material access to education, OER may be a great substitute.

To take advantage of all the potential of OER, some problems should be addressed. A revision of the literature concerning OER pointed to some open questions in the field [2]. One of them was referred to as the discovery problem, that is, what can we do to make the most relevant resources become easily accessible to the final user?

Finding the proper resource is challenging in multiple ways. First of all, as more and more OER become freely accessible, the discovery problem gets worse. For each user, the number of relevant resources is becoming more marginal compared to the total number of resources available. Comparing two resources to choose the most relevant is relatively easy. Comparing between millions is more complex. One of the more important advantages of OER, the vast number of them, is also one of the problems. Here is where recommendation technologies are a good starting point and have been widely tried in the OER field.

Two major characteristics of the OER recommendation field are especially relevant. The first of them is that in other fields the recommendations are provided to a set of final users that share some

goal, usually entertainment. In the OER field, different kind of final users coexist and their goals are more diverse. The second is that it is not easy to decide which the best resource for each circumstance is as the number of different variables that are important is much bigger than in other fields.

We will start with the former feature. A relevant part of any recommendation systems is knowing and meeting the expectations of the users interested in the recommender system. Fulfilling these expectations or even exceeding them, is the definition of success in this kind of system. To fulfil expectations, we need to acknowledge the objectives of the user and give answers to them. On the education context, we mainly find two different roles: students and teachers. When we have two roles we have two different sets of objectives, at least, and their needs may change over time. Students and teachers will expect different outputs from the recommendation system [3], such as annotation in context, find good items, find all good items, recommend a sequence, recommendations out of the box while the user is browsing, find credible recommender, find novel resources, find peers with similar interests, recommendation for alternative learning paths through learning resources.

We think that the quantity and type of variables that should be involved in an educational recommendation setting are also worth considering. It is harder to recommend an OER as the number of variables that are relevant to the recommendation process increases, and those variables are harder to quantify, and as so, they are harder to automatically being processed by a computer system.

In the education field, we need to take into account different variables, such as the student previous knowledge, social context (including, but not limited to, language), most effective ways of learning, the quality of the resource, personal taste, etc. Each of these variables is more abstract, and so, more difficult to consider in a computational model.

To address the discovery problem, many families of recommendation techniques exist. For example, [4] identify 11 different techniques. However, these previous attempts tend to have in common that they focus on a very specific task and in one family of the techniques. We think that this extensive set of independent solutions for different objectives is not a complete solution for final users, as they would need to jump from through multiple independent systems to find a proper resource.

In the next section, we will review some of the more relevant related works in the area of OER recommendation and personalization, for any of their possible contexts. In Section 3, we will introduce and discuss the main challenges associated with the OER discovery problem. Next, in Section 4, a model to face the identified challenges is presented with a brief discussion about the implementation possibilities and threats. Finally, we will present a discussion for the presented work in Section 4.

2. Related Works

Many different attempts to solve the discovery problem have been already explored. We may consider plain repositories of OER, such as MERLOT [5] or OER Commons [6], as the most basic and intuitive approach. These kind of repositories gather resources but they are quite limited on their recommendation or adaptation efforts as the user still needs much effort to find the right resources.

The discovery problem has been an active field during some time. We can find works trying solutions even before the UNESCO defined the term. Just to mention a few, we can highlight a system based on social navigation on repositories [7] or an approach based on social tagging and recommendation of similar resources to a given one [8]. More recently, once the term was coined by a UNESCO while also calling for solutions, even more works were developed.

Many of these works have focused on just one repository, repositories sharing a common taxonomy or repositories conforming to the principles of Linked Data. An example of the latter is [9], which presents a recommendation framework based on Linked Open Data to make to recommend new OER. Other related works focus mainly on a very specific task. For example, [10] present a system that uses machine learning technics to detect if, for the understanding of a given resource, some previous requisites should be meet and if another resources can be used to learn that previous requisites. By doing so, they hope to fill the knowledge gap a new resource may arise for a student.

Ref. [11] introduced CROERA, an aggregator of different OER repositories that was able to automatically aggregate the resources even when the repositories did not share a common taxonomy used to classify the resources. They did so while avoiding the traditional approach of a one-to-one matching situation, becoming easier to search and to aggregate new repositories in the future. To make the aggregation CROERA uses natural language processing (NLP) to extract features from the resources and categorize them by using support vector machines (SVM). This approach solves part of the discovery problem as makes it easier to search multiple repositories; but no recommendation, personalization or adaptation of the resources is presented, just an aggregation.

3. Challenges in OER Recommendations

As previously stated, the field of OER discovery presents some challenges. The resources are not as uniform as in other fields and there are, at least, two clear roles with different objectives using the same resources: teachers and students. In this section, we will discuss these main challenges: the source, format and logical structures heterogeneity and the multiple objectives.

Most of the previous research and developments have focused on some parts of the global problem. While a divide and conquer strategy may be the viable, multiple subsystems should coexist in any proposal to allow synergies to emerge between them, what would not arise if they were different systems. If the recommendation algorithms share information, efforts are saved and new opportunities are achieved.

3.1. Source Heterogeneity

Plenty of different OER repositories exist: those managed by educational institutions (e.g., MIT Open Courseware or MERLOT), by non-governmental organizations (e.g., [5]) or by public institutions. In some repositories, only certain people or groups can upload material, while in others anyone can do it. Some repositories are supervised by editors or librarians who review all the content and some repositories without any kind of prior supervision. There are thematic repositories and for many different topics. There are some repositories associated with OER producers and there are some that are collectors. There are so many, that repositories of repositories also exist.

The vast majority of them allow automatically access to resources, and some even offer public API to access them or their services, such as search engines. A few, ironically, although they collect open resources, but do not allow automatic access to third parties (an example of this would be the OER Commons Platform Terms of Use [12], so not all repositories serve as a data source.

Likewise, the way of presenting each OER in each source is very diverse, making difficult a direct and automatic comparison. Usually, there is some type of listing, either as a result of a search or by a content organization scheme based on categories from which hang specific pages for each resource.

These sites tend to label the resources with a set of common fields (title, date, license…), but many other fields are unique to each site or with substantial differences that do not allow for direct translation. For example, a field indicating the expected target audience is very common but it may indicate age, educational level (which change from one country to another) or by some type of category that attempts to represent difficulty.

What is common to all OER in all kind of repositories or sources, regardless of any other consideration, is that every OER has assigned a URL to identify the resource, because if the resource does not have a URL in the first place, the recommender system will not be able to access it, nor the final users. Although this designation is not unequivocal since a single resource can have several URLs different, more complex designations that solve this problem are not worth the effort.

3.2. Format Heterogeneity

One of the great opportunities offered by the OER is that for each topic it is possible to find very diverse resources trying to explain the same topic. Not only with different approaches but also in completely different formats. We can find videos, books, presentations, recorded talks, software simulations, etc. For example, in the MERLOT system, 22 different kinds of resources are

distinguished [13]. This diversity is very rare for a teacher to have material and temporary resources to create a course in which each resource is in various formats to please a different kind of students' preferences.

Nevertheless, this diversity of formats complicates the recommendation a lot since each format comes with its peculiarities. For example, a book-only recommender will create a model that includes information about the publisher, the date of edition or the number of pages, metrics that do not make any sense talking about videos, for example. In addition, it is not only difficult when it comes to building the models, but it is also complicated when working with the files of original data. The software needed to process or analyze video files is very different from necessary to do the same with a web page.

3.3. Diverse Logical Structures

It is not uncommon for OER to be aggregated in courses, books or compilations, and those are also OER equally valid. Some students will only need to read a chapter, but others the whole book. As such, it is necessary some kind of variable hierarchical structure to store these multilevel aggregation relationships.

The problem is that no standard regulates this in any way (and a new standard is not expected to be created and followed any time soon). In addition, by allowing the modification and redistribution of the modifications of OER, it is very likely to divide any resource into other new ones, and thus make new related OER by aggregation and division.

4. Description of a Proposed Model

After discussing the motivation and the challenges regarding the discovery problem in the OER field, now we will present a model. Our model is divided into 5 different stages. A diagram can be seen in Figure 1 to follow our explanation. It is remarkable that as new OER are created or updated constantly and the users may interact many times with the system, the model should be seen as a dynamic model, where the stages will be constantly producing results, it is not a static step-by-step process.
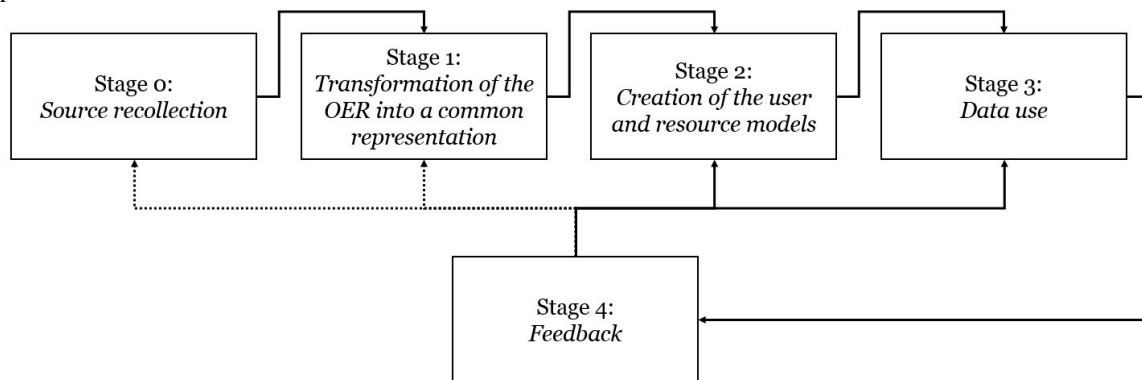

Figure 1. A diagram of the different stages of the model. The dynamic nature of the recommendation process generates a potentially never-ending loop.

4.1. Stage 0: Source Recollection

A recommendation system cannot recommend what it does not know. So in the initial stage, it is necessary to discover what can be recommended. When we talked about the challenge posed by the heterogeneity of sources, we already anticipated that there are so many models of OER repositories that it is impossible to look for a common model that abstracts them all. The only assumption we can make about what the repository will contain is that there will be resources and the resources will be identified by an URL.

The URL of the resource itself may not always be enough to process it. Many times, like in text files, this resource can be self-explanatory, but on other occasions, such as video or audio, these

resources can be accompanied by a page describing the resource without which the possibilities of understanding the resource (or know the license under which it is distributed) decrease. For these cases, we have contemplated the possibility that each URL of the resource is accompanied by the description URL. This URL can be an HTML focused on being seen by humans or can be linked data aimed at automatic processing. Be that as it may, it cannot be assumed that this field will contain information. Finally, the lists of ascending and descendant resources appear.

The founders of Google said in 1998 that "Crawling is the most fragile application since it involves interacting with hundreds of thousands of Web servers and various name servers which are all beyond the control of the system" [14] and this first stage is mainly about crawling. Any implementation should pay attention to crawlers' traps, the order of visit, any server limitations. redundant content found in multiple repositories and remember that the OER will change over time, being updated, so it will be necessary to consider strategies to control and maintain those changes.

4.2. Stage 1: Transformation of the OER into a Common Representation

The heterogeneity of formats discussed above, together with the diversity of objectives, forces us to seek a common representation of resources if we want to achieve a system with universal aspirations. The problem is that not all recommendation algorithms can work with the same starting representations. There are recommendation algorithms based on linked data, on the content of resources, on collaborative filtering, on models, etc. and as these representations are different, in the end, any recommendation system must choose what its approach will be from the beginning and is limited to recommendation algorithms based on it.

But, what if it is not necessary to choose? That is, we have options to create multiple representations of the OER. In a single system we may have OER represented in linked data models or in a textual representation based on its content, and possibly many others. If we accept that we can have multiple recommendation systems, we necessarily have to accept that all these representations can coexist, and thus we accept that there are several possible representations for each resource.

The question now is that if at a later stage we create as many models as necessary for each recommendation algorithm, what is the meaning of this previous common representation? Does not it pick up the diversity of models later and everything we need?

The answer is no. That is, the resource models that are created in the next stage are strongly linked to the needs of specific recommendation algorithms. The common representations that are created in this stage are not, and are thought independently and abstractly from a specific algorithm while being linked to a family of algorithms. The general rule to discern if something should be a model or a common representation is if it can be useful for more than one recommendation algorithm: if it can be, it should be a common representation and therefore shared, because resource models are unique to each recommendation algorithm. This distinction is what justifies this step.

4.3. Stage 2: Creation of the User and Resource Models

Although this and the next stage are different, they have a strong mutual dependence, since no resource and user models will be created if not to be used later by the recommendation algorithms of the next stage.

The idea is that the systems implementing the model at the end will start from a series of common representations of each resource (which are the result of the previous stage) and from each user (which is all the information collected by the system throughout their interaction) and instead of working on that primary model directly, each subsequent recommendation algorithm should create derived models or views of it to manipulate their information. These views allow models to store information without stepping on each other. It is important at this stage to pay special attention to avoid duplicated information. Firstly, because if something that a recommendation algorithm creates is useful for others it should be a common representation, not a model of its own. Secondly, because if there were duplicated information precedes inconsistent states that should be avoided.

Several recommendation algorithms sharing the same model of the original user can also be a great advantage in terms of the usability of the system compared to several that replicate the same

functionality. For example, if several recommendation algorithms share as input some data that they should ask the user, when sharing a single question is enough, which avoids saturating the user. Even if the system wants to avoid asking the user and tries to infer, the more data available previously, there will be more possibilities to find relationships that allow inference.

Depending on the exact set of algorithms chosen, the system will have different models. For example, if algorithms from the collaborative filtering family are used, the user model will gather information about the evaluation of each resource made by the user, but they will not be gather in the user models related in content based recommendations, as in those the user evaluation of the resource plays no part. For each task defined in the system, more than one of these models may be used, as we explain in the next stage.

4.4. Stage 3: Data Use

The ultimate goal of the model is to make recommendations and this is the stage where that goal is finally met. As we have already explained, in this stage multiple recommendation systems coexist. The input data can be shared (from the OER themselves, of from the user activities) and the output of some models may be the input of others, as many times as necessary, creating a living ecosystem that is growing towards an increasingly complete and useful system. For example, the system may use a contented based algorithm to give results to a user query in a search bar. That algorithm will give a sorted set of results that are used as an input to an algorithm that takes that sorted set and the user affinity to some type of resources (deduced by the kind of interaction with previous resources) and provides a refinement of the original sorted set of results. This refined sorted set is again used as input for an algorithm that also takes into consideration some rules established by the course teacher, and filters that refined sorted set, providing finally the results to the user. These algorithms will relay in different user and resource models, may share some common representations and don't need to be created at the same time, as each one can be reuse by features demands and programmers.

The number of recommendation algorithms that are here is immense. There are needs that will be covered by associating resources by thematic affinity, by the diversity of format or because they form a progression. Other algorithms could associate resources with users or even users with each other. The fundamental idea is that there are as many recommendation algorithms as can be useful for someone and that they carry as much information as possible.

4.5. Stage 4: Feedback

Once the recommendations are made, many models can take advantage of the feedback given by the users and it is important to consider this in order to allow a gradual improvement of the overall process. The idea is very simple: once a recommendation system has created a recommendation, some other model should enter to analyze the explicit or implicit feedback generated by the user. It can be as simple as offering "I like"/"I do not like" buttons or it can be more complex, such as analyzing user activity patterns in the system collected in registers implemented for it.

Regardless of the method of obtaining feedback, all the previous stages can benefit from the conclusions derived from this feedback, but there is a distinction between the first two and the last two. For stages 2 and 3, it makes sense to devote efforts to automate the incorporation of the information delivered by this feedback process but this is not the case for stages 0 and 1. The two first stages will be much more stable stages and that they need fewer changes, so it seems more reasonable that these changes will occur only after a human intervention, although always after analyzing the feedback.

5. Discussion

In this work, we have reviewed some of the challenges that make difficult the recommendation of OER. Also, we have presented a model for the automatic recommendation of OER by the integration of multiple components, whose purpose is to find the best-fit OER.

The model presented focuses on the integration of different recommendation algorithms in a single system. By doing so, we are able to better address the diverse objectives of the users; given that some techniques would yield better results for some objectives than others. Moreover, we are able to take advantage of potential synergies appearing from the sharing of resources for the development of a single system and the share of data. The model is divided into five stages: (1) the source recollection, (2) the transformation of the OER into a common representation, (3) the creation of the user and resource models, (4) the data use, and (5) feedback. This way, the model can integrate any existing or future recommendation technique.

The proposed model allows integrating multiple OER sources and to make recommendations in very diverse educational settings, making it possible to respond to users with different objectives. With the presented model it is possible to treat all the resource formats and all the sources, in the same way, thanks to the transformation of the resources into common representations. These common representations facilitate the introduction of new recommendation algorithms since they can take advantage of previous efforts with greater ease and reduce maintenance efforts. Based on these common representations the next stage is a separate stage for the creation of the final models used by the recommendation algorithms. We distinguish these two stages in order to emphasize the importance of the recommendation algorithms sharing as much as possible in order to eliminate duplicated efforts, guarantee that there are no incongruent states and to lay the foundations for synergies to emerge.

It is difficult to support the idea that creating an OER recommendation system that meets all the needs and objectives at first try is feasible. Anyway, with this highly modular architecture, we might build on previous models and efforts, and we hope to facilitate synergies. These synergies would be especially relevant to the possibilities that the feedback from the users offers. Every recommendation system is improved by this feedback and even it is not easy to gather nor to share this information between systems. In consequence, we believe that using our model to create a multipurpose system will create a system more interesting to the users, as they can achieve most of their objectives in a single system, and so it will be used in more situations creating even more feedback information that can be used to improve the system, attracting more users, etc. This should be, in the end, a positive feedback loop.

Some weaknesses of the proposed architecture should be discussed too. We would like to begin by turning our attention to a question that has nothing to do with the proposed model itself. It has to do with the fact of trying to solve a problem in an area where there are already exist many previous attempts with yet a new proposal. If one of the problems for the discovery of resources is that there are multiple sources, will the situation really improve by the creation of a new source? It is true that this proposal aims to group as many previous efforts as possible while creating new possibilities but, certainly, it is not the first attempt with this vision.

Looking to the proposed model itself, the idea of decoupling the stage of creation of common representations and the creation of models will make it easier for the recommendation algorithms to share data but, at the same time, there is a risk of the system becoming chaotic. Defining a clear set of rules, to decide what should be part of that common representation and what should not, will be a challenging task. Ideally, the common representations should be altered very little and be exclusively cumulative, that is, never eliminating content from the common representations, to avoid ending with orphaned recommendation algorithms. After all, it's always easier to ignore data than to have to guess it because it is missing. But storage and maintenance costs should also be considered.

As future work, plenty is ahead. First of all, the model should be empirically validated. The proposal should be effectively implemented and progressively evaluated to know the effectiveness of it in different contexts. Likely, the progressive integration of all these recommendation strategies will be hard. It will be important to also pay special attention to the creation of the interface and user experience, in order to make it suitable to all the different user profiles that may benefit from it, always maintaining a balance between functionality and acceptable complexity. Once the system begins to take sufficient form, the model should be evaluated to check its effectiveness and limitations in real environments and consider it the effort is worthy.

Given the highly integrative approach of the model, we expect that any future works should not be limited to the integration of previous methods of recommendation, but also focused on building new and improved methods, taking advantage of the synergies that we hope will arise.

Ideally, if tests with users work and can be extended to larger user communities, the data collected itself may be valuable and worthy of becoming an independent datasets divulged to the scientific community to encourage research in recommendation systems in educational contexts. The vast majority of datasets publically available, on which research is founded, are of films or songs, and introduction these are settings different from the educational. Testing recommendation algorithms with thematically limited sets may be hiding problems in the recommendation algorithms and these datasets can become a useful tool to discover and address them.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. UNESCO. *2012 Paris OER Declaration*; Paris, 2012. Available online: http://www.unesco.org/new/fileadmin/ MULTIMEDIA/HQ/CI/CI/pdf/Events/Paris OER Declaration_01.pdf (accessed on 20 October 2019).
2. Wiley, D.; Bliss, T.J.; McEwen, M. Open Educational Resources: A Review of the Literature. In *Handbook of Research on Educational Communications and Technology*; Springer: New York, NY, USA, 2014; pp. 781–789, doi:10.1007/978-1-4614-3185-5_63. ISBN 9781461431855.
3. Manouselis, N.; Drachsler, H.; Vuorikari, R.; Hummel, H.; Koper, R. Recommender Systems in Technology Enhanced Learning. In *Recommender Systems Handbook*; Springer US: Boston, MA, USA, 2011; pp. 387–415, doi:10.1007/978-0-387-85820-3_12.
4. Tarus, J.K.; Niu, Z.; Mustafa, G. Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning. *Artif. Intell. Rev.* **2018**, *50*, 21–48, doi:10.1007/s10462-017-9539-5.
5. Cafolla, R. Project MERLOT: Brining peer review to web-based educational resources. *J. Technol. Teach. Educ.* **2006**, *14*, 313–323.
6. OER Commons. Available online: https://www.oercommons.org/(accessed on 20 October 2019).
7. Brusilovsky, P.; Cassel, L.N.; Delcambre, L.M.L.; Fox, E.A.; Furuta, R.; Garcia, D.D.; Shipman, F.M.; Yudelson, M. Social navigation for educational digital libraries. *Procedia Comput. Sci.* **2010**, *1*, 2889–2897.
8. Shelton, B.E.; Duffin, J.; Wang, Y.; Ball, J. Linking open course wares and open education resources: creating an effective search and recommendation system. *Procedia Comput. Sci.* **2010**, *1*, 2865–2870.
9. Chicaiza, J.; Piedra, N.; Lopez-Vargas, J.; Tovar-Caro, E. Recommendation of open educational resources. An approach based on linked open data. In Proceedings of the 2017 IEEE Global Engineering Education Conference (EDUCON), Athens, Greece, 25–28 April 2017; pp. 1316–1321.
10. Gasparetti, F.; De Medio, C.; Limongelli, C.; Sciarrone, F.; Temperini, M. Prerequisites between learning objects: Automatic extraction based on a machine learning approach. *Telemat. Inform.* **2018**, *35*, 595–610.
11. Mouriño-García, M.; Pérez-Rodríguez, R.; Anido-Rifón, L.; Fernández-Iglesias, M.J.; Darriba-Bilbao, V.M. Cross-repository aggregation of educational resources. *Comput. Educ.* **2018**, *117*, 31–49.
12. OER Commons Platform Terms of Use. Available online: https://www.oercommons.org/terms (accessed on 20 October 2019).
13. MERLOT Help. Available online: http://info.merlot.org/merlothelp/topic.htm#t=MERLOT_Collection.htm (accessed on 20 October 2019).
14. Brin, S.; Page, L. The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw. ISDN Syst.* **1998**, *30*, 107–117.