Extending MOOC ecosystems using web services and software architectures

Abstract

This paper presents a research project that aims to extend MOOC ecosystems by integrating external tools like social networks. This integration is developed through a software architecture that mediates between different systems and platforms, facilitating communication workflows and analyzing the information retrieved. This kind of system is applied in a real case, allowing teachers and managers of the MOOC platform to obtain enhanced information and insights about users' interaction with contents and MOOC tools, as well as metrics impossible to retrieve or calculate manually in eLearning platforms with a high number of users.

Categories and Subject Descriptors


General Terms

Measurement, Human Factors.

Keywords

iMOOC, MOOC, Moodle, Web Services, Software Architectures, HCI, eLearning.

1. INTRODUCTION

The analysis of interaction among users and systems provides great insights about how users use, understand, and take advantage of tools and platforms they utilize to perform any kind of task.

The fact of analyzing interaction and trying to extract valuable knowledge from it, have real application in many areas of knowledge and business, such as in digital marketing, in education (Learning Analytics), etc. Although in some fields as education, this type of behavior analysis, and interaction analysis is increasingly common, the approaches and tools developed should be updated and adapted to new systems, platforms, and paradigms in eLearning. In these new types of learning platforms and paradigms can be highlighted the MOOCs, because they expand the traditional limits in students' interaction with teachers, contents, and online learning platforms. Furthermore, MOOCs leverage other platforms (even those that are not purely intended to be applied in education) like the social networks and other online tools, applying by this way multi-platform and multi-context approaches that can improve and upgrade the learning experience.

It is because of this use of multiple tools and multiple context that it is necessary to design and implement new ways of interaction analysis and platforms that allow to perform it. These new ways and platforms will manage the acquisition of knowledge regarding the learning and interaction with platforms, establishing convergence of knowledge between different learning vectors and context, to finally allow teachers and managers to learn, explore, and implement possible improvements that help in the learning process, the design of content, and the motivation of students.

The goal of this paper is to explain a modular software architecture implemented to allow teachers and managers of a MOOC to retrieve knowledge about how users enrolled in a MOOC course utilize some tools external to the MOOC platform, getting by this way insights about what did users on these external tools, what kind of interaction they perform inside them, and thus, discover possible improvements and solutions for eLearning processes to be applied later in the MOOC platform and its courses.

To explain these contents, the paper is divided into the following sections: section 2 (Aims and goals) presents the main aims and goals of the research presented. Section 3 (iMooc Platform) describes real MOOC platform where is being tested the software architecture proposed and the integration with other online tools. Section 4 (Software Architecture Proposal) explains the software...
architecture designed to tackle the integration with other tools and
gain knowledge about the MOOC users and their interaction with
contents inside or outside the MOOC platform. Section 5
(Services and Crawlers) explains how is being developed the
software architecture integration with other tools and platforms
(eLearning platform and social networks). Section 6 (Results)
shows partial results of the application of this software
architecture retrieving data from the integrated environments.
Finally, the Section 7 (Conclusions) presents several conclusions
about the research work and potential work for the future

2. Aims and goals
The MOOC courses are characterized by a high drop out rates
primarily due to the heterogeneity of participants is why further
information on their activity on the platform will assist meet the
shortcomings of course adapted to the student needs.

It aims to create a model to obtain the maximum information from
the MOOC platform combining it with the results obtained in the
external elements that support the learning of the course, such as
social networks, to give a feedback to the educational process,
analyze performing users to adapt the content to their interaction,
study the behavior to provide the extra functionality MOOC
integrating user feedback analysis.

Thus the analysis of interaction in media is emphasized to obtain a
deeper comprehension of users or the use of content analysis
possible interest. Detecting those topics that interest students
while those resources contributed by students for future editions
are reused.

It is important to locate the types of learning according to the
proposed activities. It is in the non-formal and informal learning
where these foster interactions between users

Because of the massiveness of these many interactions resources
and resources making it necessary to use tools to filter and collect
all the information generated to recover this. This paper is based
on the use of hashtags associated with publications to quickly
retrieve the generated content. Hashtag is a character string with a
topic associated preceded by a # sign and is used to mark tweets
[14], part of the text of these. Although their use is widespread in
twitter have extended to other social networks such as Google+.
A study of hashtags is posed for the types of learning identified in
the elements and course activities.

3. iMOOC Platform
3.1 The iMOOC approach
MOOC courses offer new opportunities for learning, features like
massiveness of participants, peer-to-peer interactions, free-of-
charge, openness or scalability [28]. There are two main types of
courses MOOC, the xMOOC with a behavioral approach (occurs
in traditional online courses) and eMOOC with a connectivism
and networking based approach [17]. Taking advantage of
features of both types of MOOC can apply the cooperative model
of Fidalgo et. al. [18] which can be defined by three layers: The
first one is the “technological layer” linked to the platform where
will find the course content and the social network that will
support the learning community. The second layer named
“training strategy” refers to the instructional design of the course.
This layer is divided, according to the two types of MOOC
exposed, into a “behavioral strategy” focused on the acquisition of
basic common knowledge and into a “connectivist strategy”
dedicated to the generation of educational resources by
participants. At this point, the resources generated from both
strategies need organized. Finally, the “cooperative layer” shows
the outcomes and the content generated with the cooperation of
instructors and participants of the course [18]. For more
connectivist orientation course based on the interaction between
participants will require the use of social tools, a good selection of
these is a good way to engage students and promote their
participation in the course [11] using these tools we can create
virtual learning communities (VLC), it is defined as communities
of people who share common values and interests, and that are
connect via different communication tools that such networks
offer, whether synchronous or asynchronous [5].

The incorporation of VLC may provide greater interaction between
participants, support and guidance to people with difficulties and
may increase collaborative processes between participants.

Over the life of people not only learn in schools or Universities,
so do in other situations in life, this idea is called lifelong learning
MOOC one of the developments associated with this item. Based
on the concept of lifelong learning are identified three types of
learning associated with this concept [16]. Their differences [15]
are based on the following items: where does the learning take
place, in the case of online learning, the platform becomes as the
institutional place for learning; if the learning is structured or
unstructured; whether the learning is guided or not by faculty; and
if there is any certification [2].

Taking these differences the Formal learning is that occurs in an
organized and structured environment and leads to validation and
certification, it is the most institutionalized [2]. The following
types of learning take place in less organized spaces. Then, with
opposite sense to the formal is informal learning, which is
voluntary and unstructured. It is based on the intrinsic motivation
[11] of the student. In this type of learning, the student chooses
the way to acquire the knowledge, the learning is given
everywhere and applies to any situations in common life [15] and
is not evaluated. Finally there is the non-formal learning, which is
structured and guided by the faculty, but is generally voluntary
and is usually not evaluated. MOOC elements can be associated
with the three types of learning, finding the more formal part in
the course platform (xMOOC) and the informal part and non-
formal community-associated learning (cMOOC).

3.2 Description and main features of the
platform
In 2013, because of the agreement between the Technical
University of Madrid, the University of Zaragoza and the
University of Salamanca, the platform iMOOC or intelligent
MOOC [32] emerges. Based on Moodle 2.6.5 platform [27] was
chosen for its versatility. The main MOOC platforms such as edX
[13], Coursera [6] or MiriadaX [24] focused on Spanish speaking
require a single path for the course. Given the heterogeneity of the
participants this can be in many cases the abandonment of them to
take the part that interests them, one of the main distinguishing
features of the iMOOC platform in front those with a more
traditional approach is adaptivity of courses for students. This
adaptivity is based on three variables: Depending on the user
profile, according to an itinerary chosen by the student or for
student progress within the course and the knowledge that he is
 gained. To achieve this functionality is necessary to use Moodle
features such as conditionals and groups, as well as external
plugins to create groupings. On the other hand iMOOC provides
an appropriate environment to an Informal and cooperative learning.

Will make use of the tools offered by the platform such as profiles, forums, workshops and secondly external social networks for informal.

3.3 MOOC course “Social Networking and Learning”

To take advantages of iMOOC platform in early 2015 a course based on a test offered by the faculty version along two editions in the Iberoamerican MOOCs platform Miriada X [4] was launched. The course of one-month duration began January 12 and ended on February 8, leaving an extra week to allow students to finish it. A Total of 793 students were enrolled for the course, more than 400 started it, and 183 students finally accomplish the goal.

The VLC was created using the tool communities of Google +, space where students can interact with each other and with content in the network sharing it. Here they can also discuss, submit questions and publish the result of voluntary exercises throughout the course. To do this has been created 9 default categories (presentations, announcements, discussions, questions, resources, Activities and exercises, application examples, contests and more) with which to classify each publication, besides the possibility to use hashtags.

This course is based on the cooperative model [18] at which has been added a fourth layer called "Gamification layer" that interacts with the other three layers, which aims to improve motivation [3]. Depending on types of learning can be divided the course study into three parts, the first section concerns the iMOOC platform itself, which takes place the more formal or academic study, the second section is focused on the social networks and finally reviewed with less detail other valid spaces. As is the case of Google +, which allows for recognition of credits, can not be considered as formal courses by not offering an official certifications, for example to obtain a certificate of participation offered for free.

Table 1. Relationship of course activities on Google +

<table>
<thead>
<tr>
<th>Activity</th>
<th>Category</th>
<th>Hashtag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search examples of social networks</td>
<td>Activities and exercises</td>
<td>#RSEEjemplossRS</td>
</tr>
<tr>
<td>Bad practices in the use of social networks</td>
<td>Activities and exercises</td>
<td>#RSEMalaPractica</td>
</tr>
<tr>
<td>Measuring influence in social networks using Klout</td>
<td>Activities and exercises</td>
<td>#RSEMiKlout</td>
</tr>
<tr>
<td>Using Twitter in education</td>
<td>Activities and exercises</td>
<td>#UsosTwitterEnseñanza</td>
</tr>
<tr>
<td>Discussion about the possibility of replacing a learning management system (LMS) through a social network</td>
<td>Discussion</td>
<td></td>
</tr>
<tr>
<td>Discussion on digital identity</td>
<td>Discussion</td>
<td></td>
</tr>
<tr>
<td>Hangout</td>
<td></td>
<td>#RSEHangout</td>
</tr>
</tbody>
</table>

These activities can also be performed on Twitter although the official platform is Google +. Course evaluation is based on four questionnaires one per module that participants must overcome to obtain a certificate of participation offered for free.

So this part can be associated to a non-formal learning, since such courses by not offering an official certifications, for example to allow for recognition of credits, can not be considered as formal learning.

About informal learning this occurs in the community of Google + and is associated with the interactions among participants not addressed by the teaching staff, similar to learning which could result in informal settings outside the classroom such as between colleagues in an office. Three types of interactions:

- Creating new hashtags and subsequent interaction
- Discussion proposed by participants
- Resources contributed

The last two actions are included in the categories created for this publication ("debates" and "resources").

4. Architectural proposal

Following the previous experience of the authors in similar cases, where they apply software architectures to extend the functionality of eLearning ecosystems [7-10; 21], authors decided to use the core of a software architecture they built in 2014.
Several layers compose this core, one to retrieve data from each external platform or tool, other that wipes and stores the information retrieved, another to push analyzed information to other platforms, and others that enable searches and interaction between information and users.

The core of the architecture is a system that acts as a mediator between the different social networks and learning platforms that will be interconnected (Figure 1). This mediator system communicates with each external tool through using web services (REST APIs commonly) and crawlers; retrieving data and information from them and analyzing the information in order to convert the raw data in valuable information for teachers and managers of the iMOOC platform (based on Moodle).

In order to implement this software architecture and its layers, has been used for development several technological components and technologies that are listed below:

- **Django Framework [12]**: This web framework is used to build the software layers and to coordinate the information workflows between the components and systems of the architecture.
- **MongoDB [25]**: This NoSQL database is used to store the data without the traditional restrictions of the SQL databases, and allows to adapt the database storage schemas to each kind of content retrieved from external tools and platforms [7].
- **REST APIs**: these web services are used to serve as communication channels between components and systems involved in the MOOC ecosystem. In case of those tools and systems that do not provide this kind of facilities, will be used crawlers to retrieve information (this will be explained in the following section).

In order to implement the information workflows shown in the Figure 1 between the software architecture and the different systems and social networks, researchers need to establish the proper communication channels for the information. These communication channels are based, in this case, in services and crawlers:

- The services are facilities provided by third-party software to facilitate the communication and interconnection with other systems, applications or clients. In this case, researchers have used services for communicating with Moodle and Twitter.
- The crawlers are software applications that find automatically information in third-party systems when they do not provide services for pull and push information between systems. In this case we are working on crawlers for getting information from Google+ Communities (Google+ does not provide API or other services to get and post activities and other information within the communities).

In the subsections below are explained how have been used these services and crawlers, and how they are implemented within the software architecture.

### 5.1 Moodle

Moodle provides several API services and API architectures; allowing users and third party applications and systems interact with courses, administration settings, users and configuration information. The API used in this work is based on Representational State Transfer (REST) architecture [19], and it allows several actions in both directions of communication (GET and POST actions, as well as DELETEs, etc.); the full documentation and functionalities of this API can be found at [26].

For example, these API endpoints and functionality allow managers and teachers of the (i)MOOC course to make automatic checks about the tasks completed by users, automatized (and simply) assessment about their participation in the MOOC, etc. In a regular course on Moodle, this usage of the API is not a key aspect, most of these checks and assessment is performed manually by the teachers, but in a MOOC course with more than 700 hundreds (in this case, several thousands in bigger MOOC courses) turns out this resources as a key factor to evaluate the users’ interaction with the MOOC and for assessing their learning within the course.

Bellow is presented an example code that allow teachers to retrieve the full list of users enrolled in the course; this result list of users enrolled, for example in the case of the iMOOC course was used to check what users filled their profile with the links to their personal social networks profiles, which was proposed as an activity of the Twitter in education lesson. As previously explained, authors implemented the software architecture using Python language, so the code is formatted in the `phytonic` way and includes the main software library used, Requests [30] that allows to implement easily the API consumption.

```python
import requests, json
```
Extending MOOC ecosystems using web services and software architectures
J. Cruz-Benito | O. Borrás-Gené | F. J. García-Peñalvo | A. Fidalgo Blanco | R. Therón

parameters = {'wsfunction': 'core_enrol_get_enrolled_users', 'courseid': 'id', 'moodlewsrestformat': 'json', 'wstoken': 'xxxxx'}
url = "http://gridlab.upm.es/iMOOC/"
response = requests.get(url, params=parameters)
if response.status_code == 200:
    results = response.json()
    for result in results:
        print result
else:
    print "Error code %s" % response.status_code

5.2 Twitter

Regarding the Twitter data retrieval implementation, the authors have implemented a collector that gets tweets on live based on their hashtags. This implementation is possible thanks to the Twitter REST APIs [31] and Tweepy library for Python [29]. Using both facilities (specially the Twitter Streaming API) authors built a software that is able to retrieve in real time tweets tagged [23] with the any of hashtags proposed in the MOOC course and storing the tweets in the software architecture database (enabling by this way MOOC user matching, etc). As example of how is done this data retrieval, below is attached a simplified version of the code:

```python
from __future__ import absolute_import, print_function
from tweepy.streaming import StreamListener

auth = OAuthHandler(consumer_key, consumer_secret)
l = StdOutListener()

if __name__ == '__main__':
    auth.set_access_token(access_token, access_token_secret)
    stream = Stream(auth, l)
    stream.filter(track=['#RSEEjemplosRRSS', '#UsosTwitterEnseñanza'])
    time.sleep(60)

    results = response.json()
    if response.status_code == 200:
        return True
    else:
        print "Error code %s" % response.status_code
```

About Twitter integration in the system, should be highlighted that the MOOC managers and teachers must get permission of the users about storage their tweets, or simply anonymize the personal data present in each tweet (name and username, location, etc.), because the social network specify in their policy rules this restriction.

5.3 Google+

About Google+, the situation is totally different. This social network provides APIs and methods to retrieve information about users, posts, comments, etc. [22], but it does not allow to retrieve information from the users communities within the social network. This disables the possibility of use the same way to get information about conversation and interactions in the communities, regarding this, teachers and managers from the MOOC course were searching other tools that let them to retrieve the desired information; for example, they use currently the tool AllMyPlus (http://www.allmyplus.com/) that allows them to retrieve information of the learners community related to MOOC. This is not the best solution, because it convert the ideal automatic process indeed in a manual process, so the authors are trying to develop a crawler that enables them to retrieve information directly from Google+ website or AllMyPlus website.

6. RESULTS

By using the software architecture and the other tools (AllMyPlus), was possible to retrieve information about users’ posts on social networks, information about their profiles on the iMOOC platform, etc. As example below of the utility of this kind of software architecture supporting and expanding MOOC functionalities, are shown several metrics retrieved from the complete learning ecosystem:

- During the course were recorded in the Google+ community 302 publications belonging to 140 users, 57 of whom have used a hashtag. Table 2 shows the times that have been used hashtags and how many of them are depending on the type of learning.
- During the course were retrieved more than 200 interactions of MOOC users with the content and hashtags in the social network Twitter.
- Also, to evaluate the completion of the MOOC activities, for example as mentioned in previous section, related to the number of users that fill their MOOC profiles with the links to their social networks profiles. In this case, the number of users that accomplish this activity was 275, a 33'86% of the total users.
- Hashtags usage and temporal evolution (Figure 2) of proposed hashtags by the teachers (non-formal) and those used by students initiative (informal).

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Diffrent</th>
<th># misspelled</th>
<th>Users using #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-formal</td>
<td>128</td>
<td>8</td>
<td>8</td>
<td>57</td>
</tr>
<tr>
<td>Informal</td>
<td>144</td>
<td>82</td>
<td>-</td>
<td>43</td>
</tr>
<tr>
<td>Total</td>
<td>272</td>
<td>90</td>
<td>-</td>
<td>23 (both types)</td>
</tr>
</tbody>
</table>

Table 3. Interactions with MOOC contents and proposed hashtags in Twitter

<table>
<thead>
<tr>
<th>Total interactions</th>
<th>Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publications</td>
<td>108</td>
</tr>
<tr>
<td>Replies</td>
<td>17</td>
</tr>
<tr>
<td>Retweets</td>
<td>42</td>
</tr>
<tr>
<td>Favorites</td>
<td>45</td>
</tr>
</tbody>
</table>
Extending MOOC ecosystems using web services and software architectures

J. Cruz-Benito | O. Borrás-Gené | F. J. García-Peñalvo | A. Fidalgo Blanco | R. Therón

7. CONCLUSIONS
This paper explains a software architecture designed and developed to extend the MOOC ecosystems functionalities and utilities by integrating external tools like social networks. This integration is built by using a software architecture that mediate between the different systems and platforms establishing communication workflows and analyzing the information retrieved. This kind of system is applied in a real case, and it allows teachers and managers of the MOOC platform to get enhanced information and insights about users interaction with contents and MOOC tools, as well as some metrics impossible to retrieve or calculate manually in this kind of eLearning platforms with high amounts of users.

In order to demonstrate the utility of this kind of software architectures, showing also the possibilities and new metrics that could be gathered using it, authors show some data gathered from the iMOOC platform and social networks, showing in these data how the application of this software architecture can help to measure elements difficult to estimate because the vast amount of users enrolled in MOOC courses or because they implicate the usage of external tools like the social networks that teachers can not track without this kind of tools and systems.

8. ACKNOWLEDGMENTS
Authors would like to thank to the GRIAL and LITI research groups by their support in this research, also they would like to thank to the Technical University of Madrid by its financial support throughout the “Diseño y desarrollo de MOOC universitarios” (Ref. PT1415-05000) Project and to the Vice Chancellor for Academic Policy at the University of Salamanca for financing the project under the Teaching Innovation program ID2014/0281 that allowed the authors to develop the software architecture described in this research work.

9. REFERENCES

Figure 2. Evolution of informal and non-formal hashtags related to MOOC contents usage
Extending MOOC ecosystems using web services and software architectures

J. Cruz-Benito | O. Borrás-Gené | F. J. García-Peñalvo | A. Fidalgo Blanco | R. Therón


