

A Minimalist, Reusable Design for an online portal to manage a catalog of courses offered by an institution: a case study of COL Virtual Event Environment Portal

Ravi Mula
NPBridge Solutions
ravishankarmula@gmail.com

Ricky Zhiyong Cheng
Commonwealth of Learning
rcheng@col.org

Abstract

Online courses are increasingly becoming a way to deliver the content as they scale to a large number of users. Courses related to many subjects are being offered by universities as well as institutions. With the number of online courses increasing, there has been an increase in platforms, that allows delivering the content. While there are popular commercial online learning platforms like Coursera and Udacity, there are also open-source platforms like Moodle, Edx using which organizations can run a single course or a group of courses. It is likely that an organization chooses multiple platforms to offer content depending on the requirements. In such scenarios, it is important to maintain the metadata of the offering at an organizational level. Metadata helps in discovery and identification and as information to search and locate a resource, provide recommendations, analytics for decision making and many more. In this paper, we present different types of metadata associated with an online course, how the metadata can be put to use, architectural approach to building a metadata platform, and how to build such platform with minimum technical resources. We offer a case study using a course portal built for Commonwealth of Learning (COL), called COL Virtual Event Environment (COLVEE) portal. This portal has about 200 courses cataloged in it for easy access. A set of metadata fields were evolved in a consultative manner, with specialists in subject matters making a number of contributions. This facilitates ease of search as well as navigation. Where applicable, the competency classification is also provided, thus making this portal useful for courses with wide variation in credentials. Built on Drupal, a popular open-source content management system, it can be re-used by any institution that offers open courses. In other words, COL can offer this portal software as a service to partner institutions.

1. Introduction

Nowadays, users around the world are able to get quality content online due to online courses in other terms MOOCs (Massive Open Online Courses). Many institutes and organizations around the world offer online courses as they are accessible by a wide number of users. There are multiple open-source projects like Moodle and Open EdX, which an organization can use to offer the courses online and users can find and take a course from the institution. Many organizations offer courses through multiple platforms. Stanford, for example, offers courses via Coursera, Udacity and Stanford Lagunita (built on Open edX platform). In the scenario of multiple content delivery platforms by an institution, users need online search services like ClassCentral (“Class Central”, n.d.) and MOOCList (“MOOC List”, n.d.) to browse all the courses. Such services work well with large organizations - as courses they offered are popular thus the services will dedicate resources to monitor their course offerings.

In the scenario of multiple platforms by smaller institutions or institutions with limited resources such as open universities, there is a lack of a unified portal, similar to a library catalog system where all the books available in the institute can be found online. Some institutes may have developed some portals to address the issue. For example, Indira Gandhi National Open University (IGNOU) has built a wiki system for the course catalog. There are open-source ones like ClassCentral (“Source code for Class Central - GitHub”, n.d.) - but the open-source ones have an in-built schema, and changing the schema - like adding more information which is institute specific would need an investment in development.

2. Background

2.1. MOOCs

Massive Open Online Courses (MOOCs) currently offer about 11.4K courses running worldwide from various universities on the top 5 platforms which have a combined user reach of 101M users (“By The Numbers: MOOCs in 2018 — Class Central”, 2018). There are many other open platforms of choice. Choosing the platform depends on various factors like Technical Requirements, Funding Organisations, Collaborating

Organisations, Sponsors, etc. and it is a common practice by an organization to choose multiple platforms for different courses. Due to the number of increasing platforms, it is difficult for users to browse all courses by a particular organization. There are services like ClassCentral and MOOCList which list all courses from portals like Coursera and EdX, but none of them are tailored for small institutions.

2.2. Alternatives

Considering the requirement of an open-source portal - ClassCentral is one such alternative. The open-source version of class central solves the discoverability problem for the users. But it needs developer expertise for customizing the institutes' requirements such as adding new types of information that needs to be captured. The project is maintained by 2-3 users and is not under active development.

2.3. Need for an open-source portal

There is a need for a portal built on open source technologies for an organization of small size, which is customizable and maintainable with minimal resources. Institutes should have the option to customize various types of information that is captured. The information captured in the portal can also help in recommending courses to the users based on their behavior. COLVEE is one such portal aimed to maintain a catalog of all the courses by Commonwealth of Learning.

3. Proposed Solution

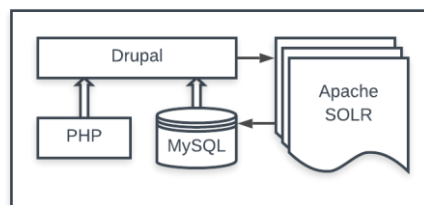
Commonwealth of Learning Virtual Event Environment (COLVEE) (“COLVEE, COL's Open Course Catalogue”, 2019) is one such portal designed for COL, which has the information of over 200 courses which were offered by COL. The portal is designed to be simple, modifiable and completely built on open-source platforms.

3.1. Features of COLVEE

- *Minimalistic*: The portal consists of only 2 platforms - Drupal & Apache SOLR; Drupal is a CMS to host the content and Solr to provide search interface on top of the data in the portal
- *OpenSource*: The portal is built on complete open-source technologies, allowing any level of customizations with no additional cost.
- *Search & Filter*: Users should be able to search for any keywords related to the course, either by filter or by free text search.
- *Easy to Manage*: Portal should be easy to manage - editors should be able to manage the metadata of the courses with a minimal learning curve.

3.2. Architecture and Implementation

The technology for the solutions is chosen based on five quality attributes: usability, reliability, security, maintainability, security, and modifiability. The portal is developed on top of Drupal (“Drupal”, n.d.), a popular open-source content management system written in PHP. MySQL is used as a database to store the content. Apache SOLR (“Apache Solr”, n.d.), an open-source search platform is used to provide search interface on the metadata that is captured in the portal. SOLR extension for Drupal allows a direct interface between content in Drupal and Solr. Below diagram represents the high-level architecture.



3.2.1 Search

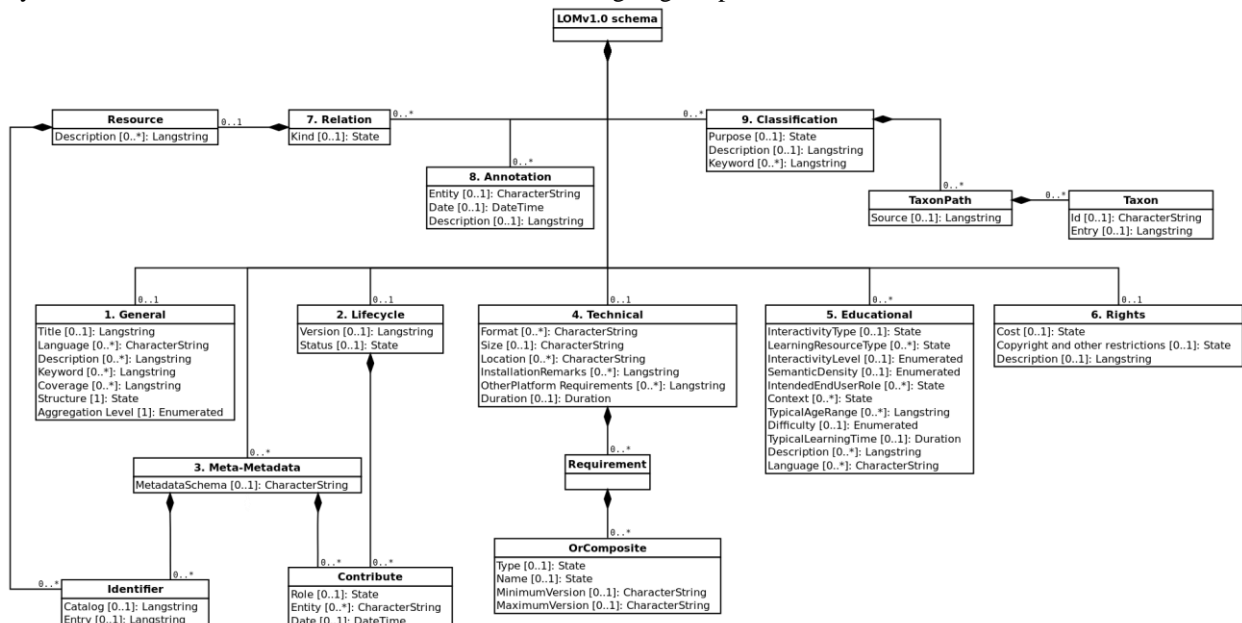
Drupal, by default, comes with inbuilt MySQL based full-text search. When the content grows it tends to become slow and doesn't work properly. It doesn't support for stemming, lemmatization and has limited support for facets. Scaling search independently is not possible as we have to scale the entire database.

Apache SOLR search application built on top of the Lucene indexer. Lucene treats each record as a document made up of any number of different fields and is capable of storing just about anything, as long as the resource can be broken up into fields and the textual data can be extracted from those fields. This makes it a good choice for indexing web-based content where you might be dealing with HTML, PDF, XML, Microsoft Word, and all kinds of other document formats. SOLR is an HTTP API for interacting with the Lucene application that makes it easier to create custom search applications. Compared to MySQL full-text search, Solr has best-in-class stemming and tokenization; scalable - both horizontally and vertically; built-in support for facets, geospatial searches, and other advanced query options.

3.2.2. Metadata

Metadata is the data about data (“Metadata - Wikipedia”, n.d.). Metadata summarizes basic information about data, which can make finding and work with particular instances of data easier. For example, author, date created, date modified and file size are examples of very basic document metadata. Having the ability to filter through that metadata makes it much easier for someone to locate a specific document.

Learning Object Metadata (LOM) (“1484.12.1-2002 - IEEE Standard for Learning Object Metadata - IEEE”, 2002) is the data model to describe learning objects and similar digital resources to support learning. The IEEE 1484.12.1-2002 Standard for Learning Object Metadata is an internationally recognized open standard for the description of "learning objects". Relevant attributes of learning objects to be described include the type of object; author; owner; terms of distribution; format; and pedagogical attributes, such as teaching or interaction style. Relevant attributes from the LOM are used for designing the portal.



A schematic representation of the hierarchy of elements in the LOM data model (“Learning object metadata - Wikipedia”, n.d.)

3.2.3. Types of Metadata Captured

LOM has classified metadata into 9 different categories. The attributes in each category are generalized for all learning resources. When we are referring to an Online Course in specific, the below are the possible attributes in each of the 9 categories

1. *General*: Course Name; Instructors
2. *Lifecycle*: Start Date; End Date
3. *Meta-metadata*: CourseID; URL; Session
4. *Technical*: Duration; Status; Access; Format
5. *Educational*: Credit Hours; Effort; TQF Level; Certificate; Language
6. *Rights*: License
7. *Relation*: Offering Organisation; Provider; Series
8. *Annotation*: Overview; Learning Outcomes; Syllabus; Competencies
9. *Classification*: Programme/Initiative; Subject; Category

3.2.4. Semi-automated ways to capture metadata

Many of the most recent platforms that are used to offer online courses follow a specific structure in representing the courses on that platform. We use the structure to generate a config per platform. Given the URL of a course, based on the config per platform we can fetch the general information such as Course Name, Description, Syllabus, Instructors, Partner Institution, Start Date, End Date, Duration, Level and Effort and feed it into the portal. Information like the Subject and Programme have to be manually mapped in the portal in some cases where the details not explicitly specified on the course page and human intelligence is required. Below is an example config for the platform MOOC4DEV (“MOOCs For Development”, n.d.) used by COL to offer courses.

```
{
  "course_name": "//div[@id=\"block-drupal8-zymphonies-theme-page-title\"]//h1//span",
  "syllabus": "//div[@id='block-course-content']//div//ul",
  "outcomes": "//div[@id=\"block-course-outcomes\"]//div//p",
  "start-date": "//div[@class=\"field--name-field-start-date\"]//time",
  "instructors": "//div[@class=\"field--name-field-team\"]//p"
}
```

The above config (“mvdobos/php-spider - GitHub”, n.d.) is read as follows: To get the Course Name, get the contents of the ‘<div>’ block whose id is ‘block-drupal8-zymphonies-theme-page-title’ whose id then gets the contents of ‘<h1>’ then get the text contents of ‘’

3.3. Discovery over metadata

Having the metadata (from 3.2.3) helps the users to discover their learning paths. Full-Text Search is enabled for the user to locate courses using either by looking for Course Name or by Programme or by Instructor or by any fields that belong to Annotations or Relation or General series. Attributes from the Classification and Educational Categories are used as facets to filter the courses from the search results or from the global level.

3.4. Recommender Systems

The attributes from the Classification category combined with the behavioral data that is captured when the user is interacting with the site - with the help of the advancement inventions in Artificial Intelligence (AI) and Machine Learning (ML) can be used to build course recommendation system. Data from the user registrations such as age, gender, country can be used to make the engine more personalized to the user (“Mittal, Jain & Majumdar”, 2014)(“Ruchika, Singh & Sharma”, 2017).

This platform acts as a base layer for analytics that can be built on top of the metadata that is captured from the courses.

4. Conclusions and Future Work

Having a metadata portal at an organizational level helps the users to discover the courses with ease. Aggregators that automatically collect information about various courses will be able to provide a list of courses of the organization from a single point of source, rather than multiple sources which may get missed or delayed.

COLVEE portal is one such a course aggregating service, which contains a list of all courses offered by Commonwealth of learning. This portal is built on open source technologies, enabling others to replicate the implementation to cater to their own needs. The technology to enable the platform allows to build and maintain such portals with minimum resources.

Here are some possible improvements in the future:

- Recommender system: The portal is designed to work with Anonymous users. Enabling the users to log in and tracking their behavior helps in recommending the courses that match their interests. Once registered, users will be able to get notified about upcoming courses they might be interested in.
- Learner analytics: Behavioral data that can be generated by this portal, can reveal insights about users interests across multiple countries.
- Automatic metadata gathering: In the current approach, the metadata is a pull-based mechanism. Extensions for the MOOC platforms can be developed so that the platform can send the metadata to the

portal whenever there is an update or addition of courses on the platform. This helps in keeping the metadata up-to-date.

- Open source: We plan to release the portal source code to github.com, which will allow interested institutions to reuse the portal and build their own course catalog system

References

- Class Central*. (n.d.). Retrieved from <https://www.classcentral.com/>
- MOOC List*. (n.d.). Retrieved from <https://www.mooc-list.com/>
- Source code for Class Central - GitHub*. (n.d.). Retrieved from <https://github.com/classcentral/class-central>
- By The Numbers: MOOCs in 2018 — Class Central*. (2018). Retrieved from <https://www.classcentral.com/report/mooc-stats-2018/>
- COLVEE, COL's Open Course Catalogue | COLVEE*. (2019). Retrieved from <http://www.colvee.org/>
- Drupal*. (n.d.). Retrieved from <https://www.drupal.org/>
- Apache Solr* -. (n.d.). Retrieved from <https://lucene.apache.org/solr/>
- Metadata - Wikipedia*. (n.d.). Retrieved from <https://en.wikipedia.org/wiki/Metadata>
- 1484.12.1-2002 - IEEE Standard for Learning Object Metadata (2002)*. Retrieved from <https://ieeexplore.ieee.org/document/1032843>
- Learning object metadata - Wikipedia*. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Learning_object_metadata
- MOOCs For Development*. (n.d.). Retrieved from <https://www.mooc4dev.org/>
- mvdbos/php-spider: A configurable and extensible PHP web - GitHub*. (n.d.). Retrieved from <https://github.com/mvdbos/php-spider>
- P. Mittal, A. Jain and A. Majumdar. (2014). *Metadata based recommender systems*. 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)
doi: 10.1109/ICACCI.2014.6968531
- Ruchika, A. V. Singh and M. Sharma. (2017). *Building an effective recommender system using machine learning based framework*. 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)
doi: 10.1109/ICTUS.2017.8286008