

CERN SUMMER STUDENT 2013 REPORT

Book shelf virtual browse interface and searching for records with physical copies on CDS.

Christopher Amanor (Supervisor: Jens Vigen)

(17th June to 9th August)

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to everyone, whose support and guidance enabled me to attend the 2013 edition of the CERN Summer Students Programme.

First of all, I would like to sincerely thank Jens Vigen, my supervisor and Head of GS-SIS, for providing this great opportunity and also for ensuring a pleasant time for me, here at CERN. I would also like to thank Prof. S. K. Danuor, Head of Physics Department at KNUST, for his assistance, which made participation in this programme possible.

Many thanks to Anca Burghart, Tullio Basaglia, Raja Sripada and Laura Garcia for their continuous guidance and support during my time at CERN. I greatly appreciate the efforts of all lecturers, professors and everyone who took us through lectures, workshops and visits to the experiments at CERN. Thanks also to all my friends and colleagues in the GS-SIS department for the time i got to spend with them and all that I have learnt from them during the summer programme.

TABLE OF CONTENTS

Abstract	1
1. Introduction	2
1.1 Invenio	2
1.2 MARC standards	2
1.3 Git	2
2. Tasks	3
2.1 Tagging Records of documents on CDS	3
2.2 Research web technologies for browsing	3
3. Results and Conclusion	4

ABSTRACT

This project provided the opportunity to learn how the CERN Document Server functions by creating a local invenio instance. With the python programming language, a feature was implemented on CDS in conjunction with other developers. This feature makes it possible for CDS users to search for records using their shelf number and for records which have a physical copy attached to them. This was done by adding the fields 852__2, 852__h and 340__p, obtained from a bibirculation module's database table to the marc data of the records. Research was also done into ways of creating a virtual browsing system for CDS users to quickly browse through shelves of books while online.

1. INTRODUCTION

The CERN Document Server (CDS) consists of books, reports, articles, periodicals, archives, multimedia etc. These are mostly about high energy physics. However, there are documents covering a variety of subjects as well. CDS is an instance of the digital library software invenio and has over 430, 000 bibliographic records and 170, 000 full-text documents.

1.1 Invenio

Invenio is essentially a python app that offers all the services of a digital library management system. These services include the collection, preservation and dissemination of bibliographic data. It is open source and was developed by the CERN Document Server team. Invenio currently runs on python 2.7.
[2]

1.2 MARC standards

MARC standards are simply a way of representing and communicating bibliographic information in machine-readable form. They hold information specific to each item in a repository. MARC stands for Machine-Readable Cataloging and it is the metadata format used by Invenio. MARC standards define three aspects of a MARC record: the field designations within each record, the structure of the record and the actual content of the record itself. [2]

1.3 Git

Due to the large number of people developing with invenio, a source code management system was required. The reason is to prevent repetition, loss of code and other problems associated with development in large teams. This calls for the use of a source code management system such as Git. With Git multiple independent local branches of code can be created, merged and deleted within seconds. [3]

2. TASKS

2.1 Tagging records of documents on CDS

Currently, users of CDS cannot search for an item (such as a book, periodical etc.) in the CERN library by using its shelf number. Also, CDS users should be able to identify whether an item is physically present in a library or only an e-book is available.

In order to make this possible, the marc of all records on CDS which have a physical copy should be appended with a field of 852 \$\$2 udc \$\$h <shelf number>. This will indicate whether the record has a physical copy or not. The tag 340__p: physical is also added to the marc data when there is a physical copy.

2.2 Research web technologies for browsing items

Currently, CDS users can not browse through books on CERN library shelves digitally. To make this possible, a python sorting algorithm had to be created for the Universal Decimal Classification (UDC) system that the CERN library uses. The UDC employs the use of symbols and punctuations for facilitating the ordering of books and other library items. The solution would enable CDS to create a virtual browsing system to help CDS users browse book covers online in the same order as they are in the actual library.

3. RESULTS AND CONCLUSION

To enable searches by shelf number on CDS and to determine whether a document has a physical copy or not, the bibcirculation daemon task was updated using python. Essentially, this task fetches recently modified records from the database, composes a marcxml for each individual record and then schedules a bibupload task. The daemon task runs daily so that any newly modified or created item in the repository is retrieved and its corresponding records' marc is appended with a couple of fields.

Two technologies that were found to help in displaying of book covers in the digital library are Amazon Carousel Widget and the Roundabout jQuery plugin.

The Amazon widget makes it possible to customize the display layout and size for easy integration into any website. However, this carousel has the limitation of only displaying items that are on the Amazon website by using the items' ISBN (International Standard Book Number).

Roundabout jQuery plugin seems a more feasible and a more likely choice because it can be easily integrated into any system. It converts lists and other nested HTML structures into a region or space where content can revolve according to the way a user interacts with it. It provides a variety of movements besides the turntable movement and also easily integrates with other plugins.

REFERENCES

1. Invenio: <http://invenio-software.org/wiki/Development/GettingStarted>
2. MARC standards: <http://www.loc.gov/marc/>
3. Git <http://git-scm.com/>
4. Amazon Carousel Widget: <https://widgets.amazon.com/Amazon-Carousel-Widget/>
5. Roundabout jQuery plugin: <http://fredhq.com/projects/roundabout/#/download>