

Designing a collection layout for DG-CO collections

Radoslav Pantic

Master student in Electrical Engineering
Ecole Polytechnique Federale de Lausanne
radoslav.pantic@epfl.ch

CERN, summer 2014

1. Introduction

Over many years of CERN existence an enormous amount of resources is collected at the CERN Document Server (CDS). Many of the CERN related documents and data are saved there. Thus the problem of successful browsing and finding desired content is becoming more and more difficult.

Two main limitations arise here. First, the large amount of available resources make the search difficult to focus. Thus there should exist preselection of documents, such that the user should already be in the narrowed field related to his interests before starting the search. The other problem is that the current search engine is not friendly to unexperienced external visitors since it offers only keyword searches.

The solution for these problems can be divided into 2 parts (Figure 1). First one is to create specific collections for each department/service (for example DG-CO, the communication group) and preselect the desired documents. In this way every department/service will be able to make its own collection and include the documents by desire. Along with adding the document, person in charge should be able to add new metadata to each document which will be used later in browsing through this collection. This is called *Back office*.

The other part of solution is called *Front office* and we focus on it in this report (creating one for DG-CO). Front office is the webpage, interface for user, which allows him to browse through the specific collection created and filled with content in the *Back office*. We will present the webpage layout and the search engine functionality.

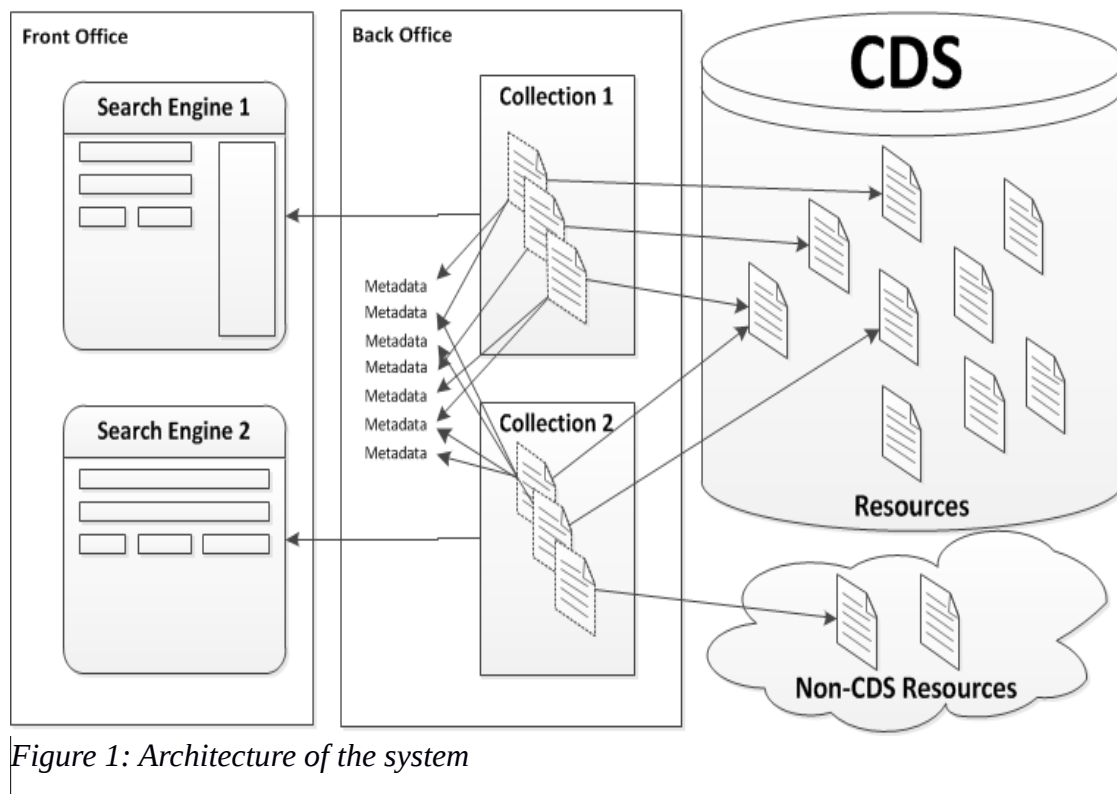


Figure 1: Architecture of the system

2. Design of the layout

The idea is to integrate the webpage to the existing website, thus keeping the general appearance of the original. Changes were made on the records list presentation and on the search engine which can be seen in Figure 2. At first we wanted to remove the existing search tool and to create new one to avoid confusion of the user. Eventually, this is solved by keeping the existing search bar, while the upgrade was built and integrated with it.

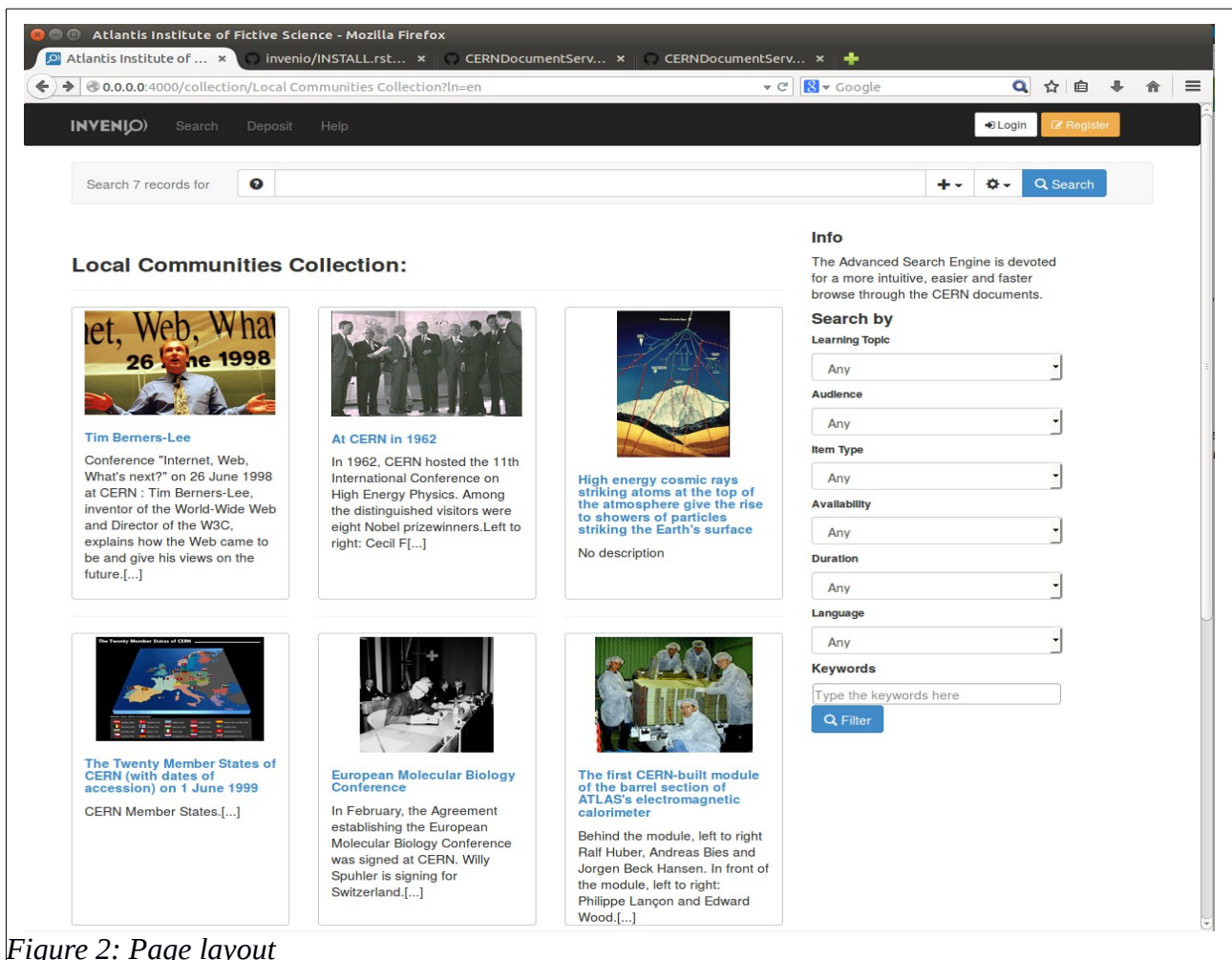


Figure 2: Page layout

2.1. Records list

As can be seen in Figure 2 the classical appearance of the list, with all the details about the records, is replaced with thumbnail list of records with only basic information. This should ease and speed up the search of the user group that is interested in finding the content faster rather than getting detailed information for all the records. In the case of non-picture record we can choose either to have predefined pictures for each data type, either we display only the rest of information (this can be modified additionally).

2.2. Search engine upgrade

The search engine upgrade consists of selection fields on the right side of the window. In figures 3

and 4 the appearance of the search upgrade is shown. User can filter the collection using different

Filter	Label
Learning Topic	<i>topic</i>
Audience	<i>audience</i>
Item Type	<i>type</i>
Availability	<i>availability</i>
Duration	<i>duration</i>
Language	<i>041__a</i>
Keywords	<i>title</i>

Table 1: Filters and their labels

categories, like topic, language of the content, audience to which the content is proposed etc. In addition one can also enter the keywords to the search filter. Once the user press “Filter” button, the filter information of all selection boxes is gathered and passed to the existing search engine, where the search is triggered. Basically, if the user chooses an option (for example *Learning Topic->Cosmology*), it will be passed to the search engine together with the label of the selection (in this case *topic:”cosmology”* is passed to the search line). An example of filter query can be seen in Figure 5. Currently only labels for language and title exist (*041__a* and *title*, respectively). All the other labels are not usable by the current search engine since there is still not any of this metadata on the content. In order to make this filter work properly these labels have to be defined in the *Back office*. Corresponding labels and filters are listed in Table 1.

Info

The Advanced Search Engine is devoted for a more intuitive, easier and faster browse through the CERN documents.

Search by

Learning Topic

Any

Audience

Any

Item Type

Any

Availability

Any

Duration

Any

Language

Any

Keywords

Type the keywords here

Filter

Figure 3: Search upgrade layout

Info

The Advanced Search Engine is devoted for a more intuitive, easier and faster browse through the CERN documents.

Search by

Learning Topic

Any

Audience

Any
Families
Journalists
Visitors
Children (6 to 12 years)
Teenagers (12 to 15 years)
Young adults (16 to 25 years)
Physics Teachers
Teachers
Animators

Language

Any

Keywords

Type the keywords here

Filter

Figure 4: Filter options

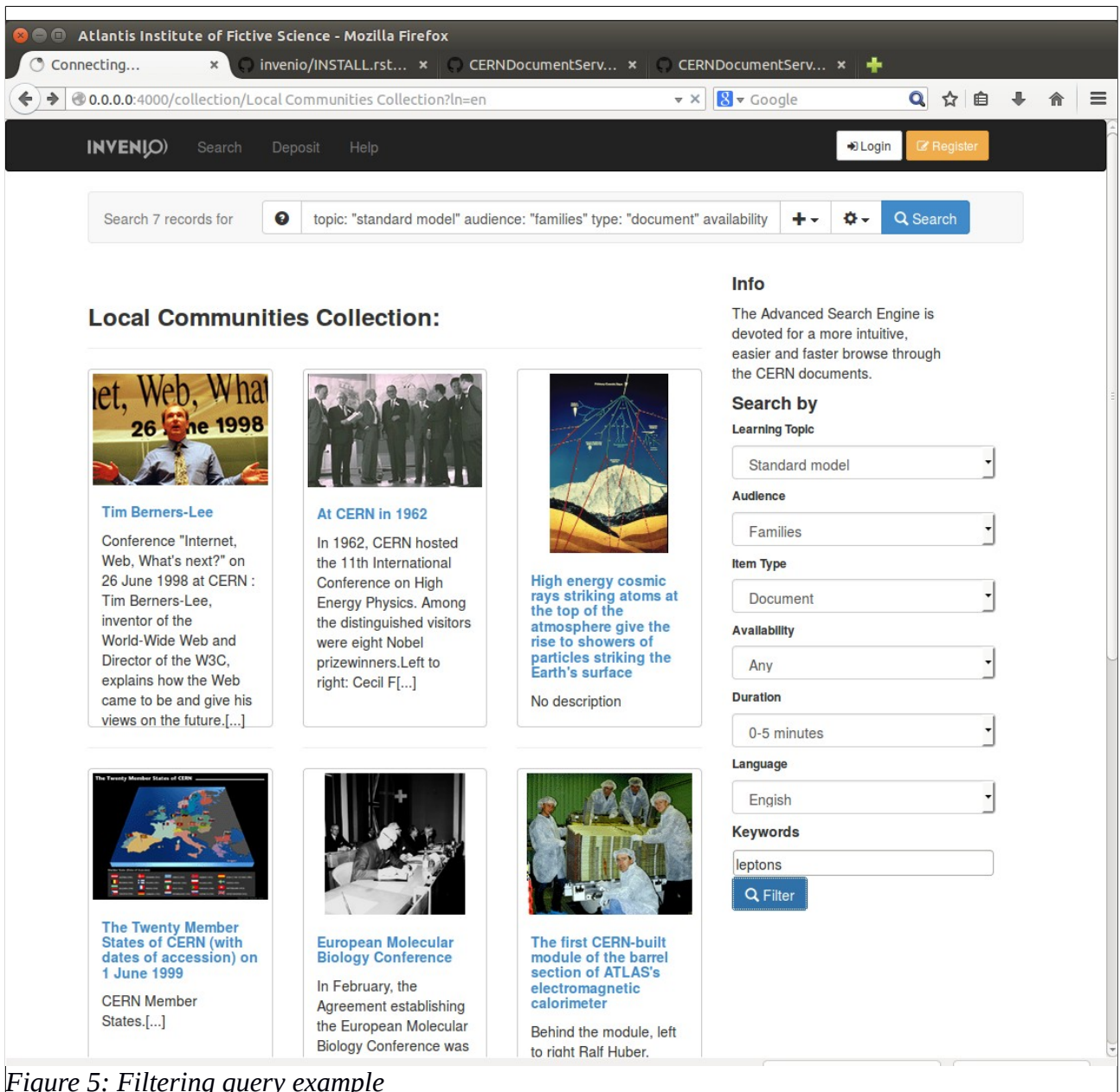


Figure 5: Filtering query example

3. Future work

In this report the layout design of the *Front office* is shown. Next thing that is to be developed is the development of the *Back office*. This would mean creating an interactive form where a person in charge will be able to add documents into the dedicated collection. For every document to be added, the values (metadata) for the labels in Table 1 (except for language and title) have to be added. Once a document like this is inside the collection, it can be searched for using the provided filtering engine.

Regarding the search results, next to each record there should be a possibility to display metadata or the sorting tree can be implemented where the end user will be able to display all the content with certain metadata value by clicking on the option.