

Indexing grey resources: considering the usual behaviour of library users and the use of Dublin Core metadata using a database of specialised vocabulary

Jacques Cuvillier

Université Paris 8

Le laboratoire Paragraphe, groupe Document Numérique et Usages, France

Abstract

Grey resources spread over a large set of matters. In each field of interest, people have their own concepts, each group will use their own vocabulary.

Most authors of Grey resources were not able to determine the Decimal or Dewey classification themselves, nor are they able to fix themselves the appropriate terms to embed in the DC elements.

It is then worth providing the webuser with an interface that makes him feel at home, by adapting automatically the terms designing the fields of the search form to those which are in use in his domain of activity.

The whole classification proposed cover about 110 different topics, each one having a set of terms chosen to entitle the fields of the interface of thematic search as well the submission forms.

Moreover, seven entry points are proposed in the first step of the connection, driving the user towards the appropriate context in accordance with the purpose of the document to find : teaching and research ; literature ; technological documentation and data sheets ; vulgarization and everyday life ; citizen and social activities ; arts , sports and spare time ; bank of images.

Indeed, the system must allow the flexibility that allow the replacement of any term by another if necessary. So the entire set of terms is put in a database. This system is also able to adapt the language chosen by the user.

The submission forms filled by authors are parsed according to the topic selected and the Dublin Core metadata file - as an XML file - is generated and stored in the indexing server. It is delivered on simple metadata request, and is reachable by an hyperlink included in any page of search results.

The experimental system has been built and evaluated over a period of five years, and was implemented in the site of the university of Poitiers since June 2005. The software is to be released under free software license (CECILL).

1 Introduction

Sharing Grey resources leads to conciliate the role of several partners : authors, library keepers and library users. In the following, the three will be considered as webusers.

Regarding authors and library keepers, the purpose of collecting a significant amount of Grey literature - even from authors among whom a part have various concerns for rigorous metadata - leads to conceive an adapted system that makes the indexation process as non-constraining as possible.

Moreover, we should keep in mind that classifying a document or embedding metadata within DC elements are not tasks that everyone could manage. Nevertheless, filling up the set of basic Dublin Core metadata is considered as a the lowest goal. The method used for collecting these metadata should naturally lead to convergent interpretations.

Regarding library users, the search task has different aspects depending on the widespread use of concerned documents. In a narrow scope, one or two fields of metadata should be sufficient, for example the author's name or the reference of a course. Also, the screens used for document search are up to the library keepers, and smart means of building up on-demand interfaces¹ have been implemented on the site <Document Libre>.

Nevertheless, thematic search based on the full set of available metadata is considered as the typical and more powerful solution, and will be considered in the following.

2 Grey resources and Grey behavior

Presenting to library users a convenient way to find the required resources is our main concern. A comparison can be done about the ways they got in use till now whether they are stepping in a real library or using an Internet browser.