Library & Information Science Research 38 (2016) 39–51

Contents lists available at ScienceDirect

Library & Information Science Research

SciELO suggester: An intelligent support tool for cataloging library resources

Natalia L. Mitzig [a,b], Mónica S. Mitzig [b], Fernando A. Martínez [a], Ricardo A. Piriz [a], Víctor M. Ferracutti [a,b], María Paula González [b,c], Ana G. Maguitman [b,c,*]

[a] Biblioteca Central de la Universidad Nacional del Sur, Avda Alem 1253, 8000 Bahía Blanca, Argentina
[b] Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Avda Alem 1253, 8000, Bahía Blanca, Argentina
[c] Consejo Nacional de Investigaciones Científicas y Técnicass, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 28 October 2013
Received in revised form 14 July 2015
Accepted 24 January 2016
Available online 21 February 2016

ABSTRACT

Existing cataloging interfaces are designed to reduce the bottleneck of creating, editing, and refining bibliographic records by offering a convenient framework for data entry. However, the cataloger still has to deal with the difficult task of deciding what information to include. The SciELO Suggester system is an innovative tool developed to overcome certain general limitations encountered in current mechanisms for entering descriptions of library records. The proposed tool provides useful suggestions about what information to include in newly created records. Thus, it assists catalogers with their task, as they are typically unfamiliar with the heterogeneous nature of the incoming material. The suggester tool applies case-based reasoning to generate suggestions taken from material previously cataloged in the SciELO scientific electronic library. The system is implemented as a web service and it can be easily used by installing an add-on for the Mozilla Firefox browser. The tool has been evaluated through a human-subject study with catalogers and through an automatic test using a collection consisting of 5742 training examples and 120 test cases from 12 different subject areas. In both experiments the system has shown very good performance. These evaluations indicate that the use of case-based reasoning provides a powerful alternative to traditional ways of identifying subject areas and keywords in library resources. In addition, a heuristic evaluation of the tool was carried out by taking as a starting point the Sirius heuristic-based framework, resulting in a very good score. Finally, a specially designed cognitive walk was completed with catalogers, providing additional insights into the strengths and weaknesses of the tool.

1. Introduction

Although many standardized resources and well-established practices are commonly used to generate library records, the process of cataloging remains a bottleneck in library management. Organizing resources associated with diverse topics is a difficult and costly task for the cataloger, who is typically unfamiliar with incoming resources due to their heterogeneous nature. A variety of solutions based on information technologies have been proposed to assist in the cataloging process (Buckland, 1992; Levy & Marshall, 1995; Park & Lu, 2009; Sølvberg, 2001).

The SciELO Suggester system is an innovative tool developed to facilitate the process of cataloging resources arriving at a library. The task of cataloging involves associating a set of metadata with incoming resources. For example, a thesis is associated with an author, an advisor, a title, an abstract, one or more subject areas, a small set of keywords (words or short phrases that are used to describe the topic of a resource), and a date of publication. Some of these data, such as the title, author, advisor, abstract, and date are explicitly given in the digital resource itself, while other data, for instance, subject areas and keywords, typically need to be inferred by the cataloger.

The proposed tool applies ideas from case-based reasoning (CBR) to assist catalogers, supplementing traditional cataloging tools by identifying appropriate subject areas and keywords for incoming material. The suggester tool operates as an experience-based system by presenting suggestions taken from material previously cataloged in the SciELO scientific electronic library (http://www.scielo.org).

2. Problem statement

Existing cataloging interfaces are designed to reduce the bottleneck of creating, editing, and refining bibliographic records (Gómez, 2015; Reese, 2015). These interfaces provide a convenient framework, but ease of data entry provides only a partial solution to the problems of cataloging—the cataloger still has the harder task of deciding what information to include. The intellectual effort which is expended at this

* Corresponding author at: Departamento de Ciencias e Ingeniería de la Computación Universidad Nacional del Sur, Avda Alem 1253, 8000 Bahía Blanca, Argentina.
E-mail address: agm@cs.uns.edu.edu (A.G. Maguitman).

http://dx.doi.org/10.1016/j.lisr.2016.01.001
0740-8188/

stage is time-consuming, costly, and leads to bottlenecks in resource processing.

Informal discussions with catalogers indicate that when cataloging, they often pause for significant amounts of time wondering what information to include. They usually look at existing catalogs (e.g., the Library of Congress Catalog, https://catalog.loc.gov/) and other information on the Web for metadata to associate with incoming resources. Through bibliographic database services such as those provided by OCLC, Inc. (2015a, 2015b), catalogers have easy access to up-to-date records for many kinds of resources. However, these tools require the user to explicitly request a bibliographic record, a request that can only be fulfilled if the resource is already cataloged in the databases of bibliographic records that are accessible to the cataloger.

Some of the resources that arrive at a library, such as doctoral and master's theses or other rare material, are not cataloged in these databases. In spite of that, bibliographic records of material that is topically similar to the to-be-cataloged resource can be helpful at the moment of generating metadata such as subject areas and descriptive keywords. Thus, intelligent tools that identify similar material and generate suggestions could provide substantial benefits for cataloging. This approach motivated previous work done to expand cataloging tools with intelligent aides (Delgado, Maguitman, Ferracutti, & Herrera, 2011; Dini, Varela, Antúnez, Maguitman, & Herrera, 2010).

Such tools are necessary to optimize cataloger productivity and can save libraries the burden of investing in a task that can be replaced to a large extent by automatic processing. The functionality of these tools not only increases the efficiency of the cataloging process, and thereby saves money and reduces bottlenecks, but it can also improve the quality of the catalog itself and enable more complete cataloging.

The work described here is part of an effort carried out as a collaboration between the main library of the Universidad Nacional del Sur (Bahía Blanca, Argentina) and members of the Knowledge Management and Information Retrieval Research Group. The SciELO Suggester tool is one in a series of prototypes developed with the purpose of leveraging existing bibliographic records to assist the cataloger.

3. Literature review

Unlike manual catalog creation, semi-automatic generation of catalog entries relies on support tools that assist the cataloger to effectively identify the most appropriate metadata for the resource under analysis. For several years, in domains other than library science, intelligent support tools have served the purpose of expanding the user's natural capabilities, for example by acting as intelligence or memory augmentation mechanisms (Engelbart, 1962; Licklider, 1960). Many of these systems are highly autonomous and are based on the intelligent agent metaphor (Bradshaw, 1997; Laurel, 1997; Maes, 1994; Negroponte, 1997) while others adopt a user-driven approach and need to be initiated by commands or direct manipulation interfaces (Shneiderman, 1992; Sutherland, 1963; Ziegler & Fahnrich, 1988). An intermediate group of support tools reconciles both approaches, giving rise to mixed-initiative user interfaces (Horvitz, 1999). In general, these tools complement the

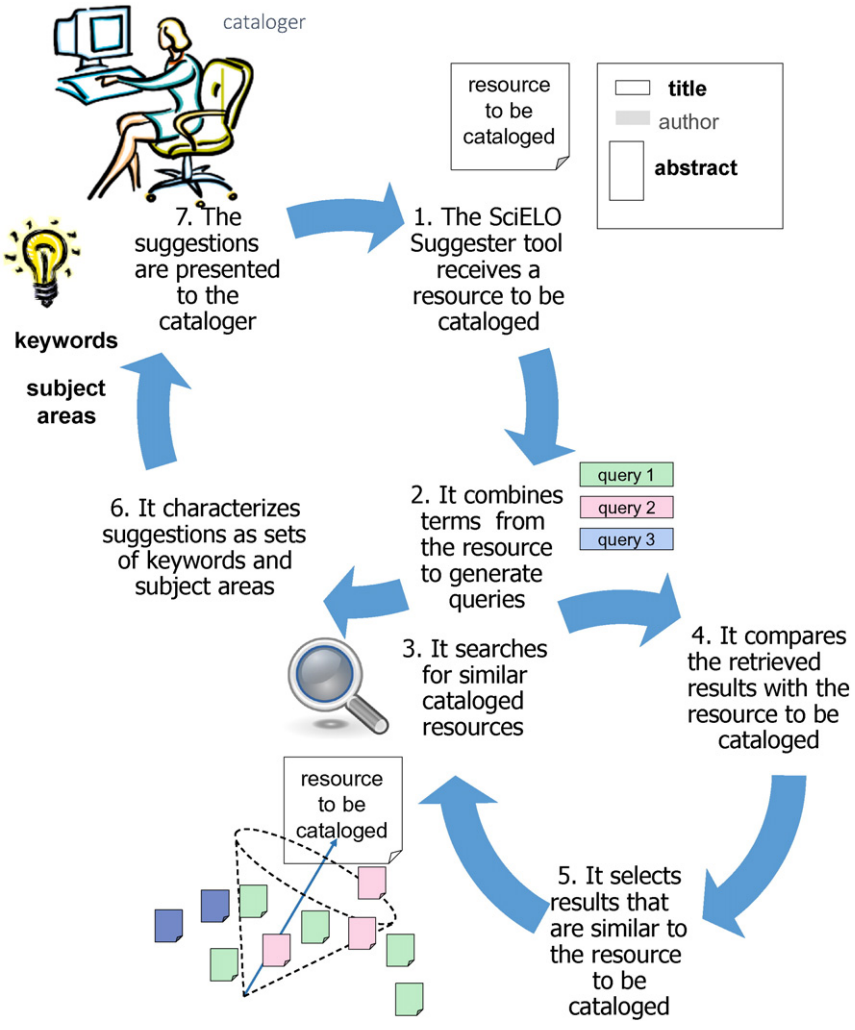

Fig. 1. The SciELO Suggester's cycle.