



# Detecting Similar Areas of Knowledge Using Semantic and Data Mining Technologies

Xavier Sumba<sup>a,1</sup>, Freddy Sumba<sup>a,2</sup>, Andres Tello<sup>a,3</sup>,  
Fernando Baculima<sup>a,4</sup>, Mauricio Espinoza<sup>a,5</sup> and  
V́ctor Saquicela<sup>a,6</sup>

<sup>a</sup> *Department of Computer Science, University of Cuenca, Cuenca, Ecuador*

---

## Abstract

Searching for scientific publications online is an essential task for researchers working on a certain topic. However, the extremely large amount of scientific publications found in the web turns the process of finding a publication into a very difficult task whereas, locating peers interested in collaborating on a specific topic or reviewing literature is even more challenging. In this paper, we propose a novel architecture to join multiple bibliographic sources, with the aim of identifying common research areas and potential collaboration networks, through a combination of ontologies, vocabularies, and Linked Data technologies for enriching a base data model. Furthermore, we implement a prototype to provide a centralized repository with bibliographic sources and to find similar knowledge areas using data mining techniques in the domain of Ecuadorian researchers community.

*Keywords:* Data Mining, Semantic Web, Linked Data, Data Integration, Query Languages.

---

## 1 Introduction

The number of publications is rapidly increasing through online resources such as search engines and digital libraries, making more challenging for researchers to pursue a topic, review literature, track research history because the amount of information obtained is too extensive. Moreover, most of the academic literature is noisy and disorganized. Currently, certain information about researchers and their bibliographic resources are scattered among various digital repositories, text files or bibliographic databases.

---

<sup>1</sup> Email:[xavier.sumba93@ucuenca.ec](mailto:xavier.sumba93@ucuenca.ec)

<sup>2</sup> Email:[freddy.sumbao@ucuenca.ec](mailto:freddy.sumbao@ucuenca.ec)

<sup>3</sup> Email:[andres.tello@ucuenca.edu.ec](mailto:andres.tello@ucuenca.edu.ec)

<sup>4</sup> Email:[fernando.baculima@ucuenca.edu.ec](mailto:fernando.baculima@ucuenca.edu.ec)

<sup>5</sup> Email:[mauricio.espinoza@ucuenca.edu.ec](mailto:mauricio.espinoza@ucuenca.edu.ec)

<sup>6</sup> Email:[victor.saquicela@ucuenca.edu.ec](mailto:victor.saquicela@ucuenca.edu.ec)

When you need to propose projects with several researchers in a specific area belonging to different Higher Education Institutions (HEI), different questions arise. For instance, who works in similar areas of research? or, how can I create a network of researchers in a common knowledge area? Then, detecting similar areas based on the keywords it could help governments and HEI to detect researchers with interests in common, opening an opportunity to generate new research projects and allocate efforts and resources to them. In that case, we could detect potential collaboration networks.

The expansion of this knowledge base will allow our academic community to have a centralized digital repository which has information of Ecuadorian researchers based in bibliographic resources. The collaborators are identified through a semantic enrichment of scientific articles produced by researchers who publish with Ecuadorian affiliations. This work aims to encourage institutions to collaborate and obtain a semantic repository to identify researchers working in similar areas and, provide updated information accessible and reusable. Enhancing the generation of research networks with academic peers in the region could provide a greater opportunity for collaboration between the participating institutions.

Obviously, there are many tools and services currently available in the web which already provide a wide variety of functionalities to support the exploration of academic data. Each tool or service operates in different ways, that in some cases complicate the literature review or utilization data. These tools or services allow search publications using keywords, author names, conferences, authors affiliations through *Applications Programming Interface (APIs)*. They have started using semantic technologies that helps to describe their resources, but each source is different. Our approach use these characteristics, to retrieve and enrich bibliographic data from several bibliographic sources to detect similar areas.

The rest of this paper is organized in the following way: section 2 presents the related work. We outline the architecture in section 3, detecting similar areas in the domain of Ecuadorian Researchers and detecting potential networks of collaboration, using semantic technologies to enrich data extracted from different bibliographic sources in a common model. Conclusions and future work are presented in section 4.

## 2 Related Work

This section introduces tools and services used for searching publications, unification of publications, authors disambiguation, and approaches related to the identification of similar research areas.

Some bibliographical sources have tools that allow access to data, but others sources do not have. For example, *Google Scholar* does not have an API that allows an automatic retrieval of publications. *Microsoft Academics Search* provides an API to search for publications, and they also provides a variety of tools for visualizations such as co-authorship graphs, trending publications, and co-authorship paths between authors. However, they have data from 2013, which actually is out-

Download English Version:

<https://daneshyari.com/en/article/4950046>

Download Persian Version:

<https://daneshyari.com/article/4950046>

[Daneshyari.com](https://daneshyari.com)