Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 387–396

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

# Automatic approach to enrich databases using ontology: Application in medical domain

Zina Nakhla[a*], Kaouther Nouira[a]

[a] *Université de Tunis , BESTMOD Laboratory Institut Supérieur de Gestion, 41 Avenue de la Liberté 2000 Bardo, Tunisia*

**Abstract**

The enrichment of databases is fundamental to maintain them, as well as the consistency and accuracy of the data. The database becomes useless if it is not up to date. Since there are a large number of databases, an automatic enrichment approach is required. However, until now no efficient approach has been provided in order to cope with this problem. In this paper, we propose a new approach to automate the enrichment of databases. It is based on an ontology, which model domains through sets of concepts and semantic relationships established between them. The proposed approach presents a set of rules to analyze ontologies and databases components and filter subsequently the necessary ones for the database enrichment of databases. We applied our approach in the medical domain that is a renewable domain. Also, it is characterized by a large number of databases and ontologies, and a large volume of data. For experimentations, a platform is developed to test rules using medical databases and medical ontologies. As a result we obtain enriched databases with new components that are either tables, attributes, or records.


Peer-review under responsibility of KES International

*Keywords:* Database; Ontology; Enrichment rules

## 1. Introduction

Database managment systems are used since 1970s to store various kinds of data for different purposes. They enabled information to be efficiently stored and queried. However, databases are not complete, the new elements of knowledge must be continually added[1]. On the other hand, ontologies have appeared as an alternative to databases in applications that require a more enriched meaning[2]. Several methods in the literature used ontology to model information systems in different domains[3], to develop decision support systems[4], to construct database based on ontologies[5], to improve the habitability of a natural language[6], to analyze data from the patient record[7]... Also, ontologies are widely used in different domains such as Semantic Web[8], Natural Language Processing[9], Medicine[10], Commerce[11],... An ontology can provide enough information about a domain, and even structure the appropriate terms of a domain. Besides the similarity between ontologies and databases, ontologies provide summarized contex-

* Zina Nakhla. Tel.: +2-161-105-3411 ; fax: +0-000-000-0000.
*E-mail address:* Zina.Nakhla@isg.rnu.tn

1877-0509 
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.221

tual information about the contents of databases. In this context, we propose to use ontologies for the enrichment of different components of databases. Researches in the literature have proposed several approaches to improve database structures and data using ontologies. An approach to build databases based on ontologies[12], an approach to create ontologies based on databases[13], an approach to use ontology to manage data and to make decisions[14]. In addition some researches proposed a semi-automatic approach based on ontologies to enrich databases[15]. Other works proposed approaches to enrich only a part of databases e.g tables, attributes, records[16]. The proposed works are limited, they only enrich one component of the database that is mostly attributes or records. But there is a lack of works concentrated on tables enrichment. Also, the approaches in the literature are semi-automatic which necessitate every time the intervention of an expert.

Our proposed approach presents a set of enrichment rules which automate the databases enrichment using ontologies. The proposed enrichment rules treated all the components of the database (tables, attributes, records). We focus in this paper on medical domain, which is characterized by an important volume of scalable information. Our goal is to have an updated database compared to the original one. The new database contains the missing information, and the novelties of the medical domain which are related to the database.

This paper is organized as follows: Section 2 presents works in the literature, which are related to using ontology to improve databases and methods to enrich databases. Section 3 explains the architecture of our proposed approach. Section 4 describes the enrichment rules. Section 5 presents the databases and ontologies used for the experiments and results. Finally, we give conclusion.

## 2. Related work

In the literature, ontologies are used to improve databases in different levels. It is used to remove ambiguity in the structured query language of the database[17]. Also, ontologies are used to build and to model databases which take ontology as input and generates a database schema based on it[18,12,19]. In addition, ontologies based databases are used to categorize web pages[20], and used to harmonize knowledge concepts in databases and models[21]...

For the database enrichment, researches used the records matching process. It consists on bringing together data from different databases about the same entity[22]. Matching records in databases is used in different domain, mainly medical domain because it has multiple legacy and information systems that support health care professionals[23]. Huang et al.[24] propose a Pathway And Gene Enrichment Database (PAGED) which is an online database that integrates gene-set-based prior knowledge as molecular patterns. The database resulted consists of disease-gene association data, curated and integrated from Online Mendelian Inheritance in Man database and the Genetic Association Database. Some approaches proposed to enrich databases using ontologies. Jesus et al.[16] proposed to extract semantic information from unstructured texts and transformed it in ontology. Then, ontology will be used to enrich the contents of textual attributes. Hamaz and Benchikha[15] proposed a reverse engineering process that aims to transform a relational database to an Ontology-Based Database. They attempt to add additional semantics behind applying the enrichment process. This enrichment process is semi-automated because that requires interactions of an expert. Yuan An et al.[25]developed the FormMapper system to enrich databases that accept user-created data entry forms, using ontologies and integrates them into existing databases in the same domain. The limitations of this approach are to include the missing correspondences and syntactic nature.

Most of aforementioned approaches have proposed a solution to enrich database. But these works are semi-automatic and enrich only a part of databases. These works need each time the intervention of experts, which is waste of time and effort. Also, some errors can be unnoticed by experts which engender errors in databases. Our work proposes an automatic and complete database enrichment approach.

## 3. Our proposed database enrichment approach

Usually we find different computer applications related to similar databases. Despite the different issues of applications, they use nearly similar databases which can contain the same tables, most of the same attributes and records. Each database is created separately, which is waste of time, effort and storage space. To resolve this issue, simply share the database to optimize the development of application. However, a richer description of the target database