Enterprise Knowledge model for Legacy database applications using protégé
Case study of Libyan dates (Tomour) industry

Ahmed A. Arara and Amel Hijazi

Abstract—Ontologies to describe the contents of information sources in enterprise has become the key issue to intelligent retrieval and access of information. Ontologies, as an explicit specification, are widely used in many disciplines to support the semantics and to provide a richer model of reality. Domain ontology as an emerging new discipline has a great potential in providing such explicit specification to develop knowledge models of legacy databases. Legacy databases represent crucial information sources to enterprises, and they are in need for enrichment to permit information access and retrieval from the web. A case study is used to demonstrate a proposed knowledge model. In this paper, a Libyan date (Tomour) domain is explored as an example to develop knowledge model based on existing legacy database.

Index Terms— Legacy database, domain ontology, dates, database mapping.

I. INTRODUCTION

DATABASES are widely used by enterprises, and are the basic constituent of any information systems. In recent years, there has been a wide use of computer networks and Internet technology. The inception of web-based information systems must provide answers to web-based query from existing information sources mostly databases. Information sources were often developed for obsolete environment, and they are considered as legacy information sources. End-users who have an access to the networking computing and Internet need not (and should not) be knowledgeable of all technical details attached to information sources. An explicit specification [4] of such legacy databases can be collected and managed as a source of knowledge to shield off end-users from all lexical and syntactical details of information source, and to remove many ambiguities of terms and domain concepts. Moreover, a knowledge base will provide shared common understanding of a specific domain.

Protégé2 , as a knowledge editing tool, is used to construct domain models and knowledge based applications with ontologies[1]. Protégé can be adapted to provide domain friendly support for creating knowledge and entering data based on existing legacy databases. Hence, exploiting protégé by discovering and building a KB for a new domain namely dates (Tomour) domain will be of great challenge and great benefits added to web-based Tomour industry site in Libya. The paper outlines the various problems encountered, and practical experiences gained when developing a knowledge base from existing legacy information source. Domain ontology for Libyan dates (Tomour) industry was captured using protégé tool. The knowledge base, as a domain ontology, is accessible via a web site where users are able to navigate and traverse the KB taxonomy. This paper is organized as it follows:- In section (2), we pinpoint the problem and present our motivations of building a knowledge model for existing information sources for specific domain. Section 3 focuses on KB building based on existing ER diagram and data integrity constraints. Section 4 discusses the proposed system architecture and its main components. In section 5, a formal model of Tomour Ontology based on description logics (DLs) is presented. Section 6 skims over some related works, and section 7 will conclude the paper.

II. PROBLEM DEFINITION

Semantic web services and semantically described web pages [12] have now demonstrated their benefits and maturity in a number of applications. Ontologies are important in supporting to organize large information repositories and to access such repositories in an efficient manner [10][11]. Ontologies can play a major role in the key aspects namely: searching, accessing web-based information, interpretation and reasoning about information [5]. In this research, we investigate the problem of transforming monolithic legacy databases into semantic web application. In specific, we focus on domain ontology creation from existing legacy database schema for Tomour (dates) industry in Libya. Our approach shall focus on transforming legacy RDB schema into knowledge base.

III. MAPPING DATABASES ONTO KNOWLEDGE BASES

A huge number of legacy databases, existing within various organizations and enterprises, have become the subject of re-use and modification to cope with the advancement of computer networks and distributed environments [3]. Hence, it is almost impossible to ignore or isolate such legacy systems from the vast expansion of computer networks and distributed computing environment sources. Therefore, a concrete approach of preserving and supporting legacy databases is significantly important to allow for better access, exchange and share-ability of this great wealth of accumulated information. Every database system is a representative of a particular view of the real world, since it describes implicitly the meanings and relations among terms originated from that particular instance of the real world. Domain ontology, however, will be constructed by capturing concepts from databases that are “just needed” to interlink the various applications of an enterprise. Ontologies can be built
and modeled with several types of database models, such as: Entity/Relational model (ER) diagram, object oriented database models or deductive database models. In general, all database models can be explored, and used to build domain ontologies [6]. Relational data bases (RDB’s) mappings onto ontologies is established by considering the database schema (represented as ER-diagram in our case) as an input to the ontological tool extractor (protégé as the kernel of the proposed framework). RDB schema is usually available for any applications. System tables are a good source in any RDBMS. Protégée plug-ins need to be developed to capture all ontological concepts from RDB. For instance, entities become concepts, attributes becomes properties, and relationships are kept nearly the same in both models. The mapping procedure from RDB schema onto KB taxonomy is shown in figure 1.

![ER Diagram DB model](image)
![RDB model tables](image)
![Knowledge base model](image)

Figure 1 The mapping process DB onto KB

A mapping of such schema onto domain ontology may lead to a rich knowledge model. Moreover, intelligent queries based on the developed ontology shall be formed.

IV. KNOWLEDGE BASE MODEL ARCHITECTURE

The framework in figure 2 describes how to build domain ontology from existing legacy relational database. Building Ontology through using a suite of integrated tools may assist in making web-based information systems more intelligible[5][12].

![Figure 2 Knowledge base model architecture](image)

As shown in figure 2 above, domain experts provide their contributions in terms of ontological concepts, relationships, attributes, and constraints. It should be noted out that Knowledge Base Management system (KBMS) and Database Management system (DBMS) are part of the framework, and they can be accessed by the ontology integrated tool namely protégé.

V. TOMOUR DOMAIN ONTOLOGY FORMALIZATION

Description logics (DLs) are commonly used to design KB of applications. A DL KB consists of two components TBOX and ABOX. First, the component TBOX contains intentional knowledge in the form of terminology. Second, the ABOX component or the assertion BOX which contains existential knowledge which is directed to individuals or instances of the KB. A sample of Tomour ontology as a TBOX and ABOX is shown in table 1.

<table>
<thead>
<tr>
<th>TBOX</th>
<th>ABOX</th>
</tr>
</thead>
<tbody>
<tr>
<td>TamrAndRutabAndBallah = Ballah ( \cap ) Rutab ( \cap ) Tamr</td>
<td>hasStage (Hura, Rutab). VITAMIN (A1, A2, A7, B).</td>
</tr>
<tr>
<td>BallahAndRutab = Ballah ( \cap ) Rutab ( \cap ) Tamr.</td>
<td>TYPE (Hallaw y, Helloo, Lawzy, Sugary).</td>
</tr>
<tr>
<td>Product = Food (Date) ( \cup ) PalmTree.</td>
<td>STAGE (Gmace, Helalia Tamr, Rutab, Ballah).</td>
</tr>
<tr>
<td>Tommor = PalmTree</td>
<td>REGION (Centare, Coast, Janub)</td>
</tr>
<tr>
<td>Date = Tommor = Product</td>
<td>COUNTRY (Tunisia, Libya, Algeria)</td>
</tr>
<tr>
<td>Nakhla = PalmTree.</td>
<td>Nakhla (tabuni,)</td>
</tr>
<tr>
<td>Janub = TommornRegion</td>
<td>JJanUB (Kufra)</td>
</tr>
</tbody>
</table>

Table 1 Tommour ontology

VI. RELATED WORK

ONTO [2] is a tool developed as a protégé plug-in. It makes use of the Protégé OWL plug-in API for handling ontologies in conjunction with a JDBC-based module for handling relational databases. Christopher [6] developed an approach for extracting knowledge from textual DB in medical domains; she constructed a set of graphical patterns that indicate the presence of casual relation in sentences. These patterns are matched with a syntactic parse tree of sentences. However, these patterns lacked to semantic web in sentences. Christopher concentrates on syntactic part and ignored the semantic one in presenting sentences in opposite manner. Kevin Knight [7] focused on building ontology by merging various online dictionaries database, semantic networks, and bilingual resources, through semi-automatic methods. First method provides a common inventory of semantic tokens, used in both analysis and generation of “lexicon”. Second method describes semantic tokens that are naturally related to each other, depending on “grammar” of the Interlingua tool. Peter Haase [8] worked at generating domain ontologies from textual resources by applying natural language processing and machine learning techniques. However, his work ignored domain description of Textual database resources.

[6] and [8] focused on the textual part of the DB, [2] payed attention to the implementation issue. [7], however, concentrated on the merging of ontologies. Our work is similar to [2],[6],[7] and [8] but we consider DB schemas and their descriptions as a good target to build reasonable ontology.

VII. CONCLUSION AND FUTURE WORK

In this paper, we aimed at investigating the issue of ontology development from an existing information sources, namely database repositions. The concepts of knowledge bases and
semantics become the savior of many database shortcomings such as inferencing and data elements explications. In a network, distributed environment, a Protégé platform (ontology editor), was used to extract concepts (element of knowledge) in a Libyan date (Tomour) domain which is explored as an example of building knowledge model based on legacy database. A mapping process from database to knowledge base was presented as semi-automatic approach. Full automation of the process of mapping is still a research problem but with existing software tools and plug-ins mapping can become more feasible. As for future work, we plan to upload the derived knowledge in Tomour website. Such knowledge can assist end-users to have an answer to several knowledge-based queries instead of the limited SQL database queries. We, also feel that task ontology in addition to domain ontology can enrich the knowledge base concepts. This becomes very important when we consider the e-Commerce aspects of Tomour domain.

REFERENCES


[6] Christopher S.G. Khoo, Syin Chan and Yun Niu "Extracting Causal Knowledge from a Medical Database Using Graphical Patterns", Centre for Advanced Information Systems, School of Computer Engineering, Blk N4, Rm2A-32, Nanyang Avenue Nanyang Technological University


