

# AMSABEL: A tool for Multilingual Information Access\*

*AMSABEL: Una herramienta para el acceso a información multilingüe*

Alejandro Revuelta-Martínez, Luis Rodríguez and Ismael García-Varea

Departamento de Sistemas Informáticos

Universidad de Castilla-La Mancha

02071 Albacete, Spain. Tel +34 967 599200

{Alejandro.Revuelta,Luis.RRuiz,Ismael.Garcia}@uclm.es

**Resumen:** El acceso a la información es una actividad muy común actualmente. Las nuevas tecnologías, y sobre todo Internet, permiten que se tenga disponible una cantidad de información considerable. Sin embargo, la mayoría de los usuarios no tienen las habilidades necesarias para acceder a toda esta información. Una solución es el uso de interfaces basadas en lenguaje natural que podrían utilizarse inmediatamente, y además, añadiendo un sistema de diálogo se podría reducir considerablemente el tiempo de aprendizaje inicial.

**Palabras clave:** Acceso a información multilingüe, procesamiento de lenguaje natural, SQL

**Abstract:** Accessing to information is very usual nowadays. New technologies, especially the Internet, allow for the availability of huge amounts of data. However, most people lack the skills needed to properly access to this information. Natural language interfaces could be immediately employed in this case and, additionally, the use of a dialogue system could significantly reduce the training time required by new users.

**Keywords:** Multilingual information access, natural language processing, SQL

## 1 Introduction

Nowadays there are a large number of databases publicly available but their actual usefulness is somehow limited by the implemented user interfaces. It is not realistic to request the users to have any prior knowledge on the database technical features. Ideally, they could only be expected to have an initial idea on the contents of the database.

In the case of the widely employed relational databases, Structured Query Language (SQL) is used in order to provide a formal mechanism to retrieve data. SQL requires the user to know the logical structure of the data as well as the language syntax. Nevertheless, both requirements are usually not fulfilled by ordinary users.

Recently, in (Revuelta-Martínez, Rodríguez, and García-Varea, 2010) a

multilingual information access system is proposed. This system is able to translate queries from natural language into SQL by means of state-of-the-art phrase-based statistical machine translation techniques (Koehn, Och, and Marcu, 2003). Here, we present an user interface prototype implementation based on that architecture, which allows users to query a relational database using both English or Spanish. Besides, two different options are provided to use the prototype: a web page or a standalone graphical application.

## 2 Prototype description

The prototype has been designed following a modular and distributed client-server architecture. Communication among components is based on the TCP protocol allowing, this way, to deploy all the modules in different machines without additional changes. This implementation comprises the following sub-systems:

\* Work co-funded by the European Social Fund and the Spanish Consejería de Educación y Ciencia de la Junta de Comunidades de Castilla-La Mancha under grant PBI08-0210-7127

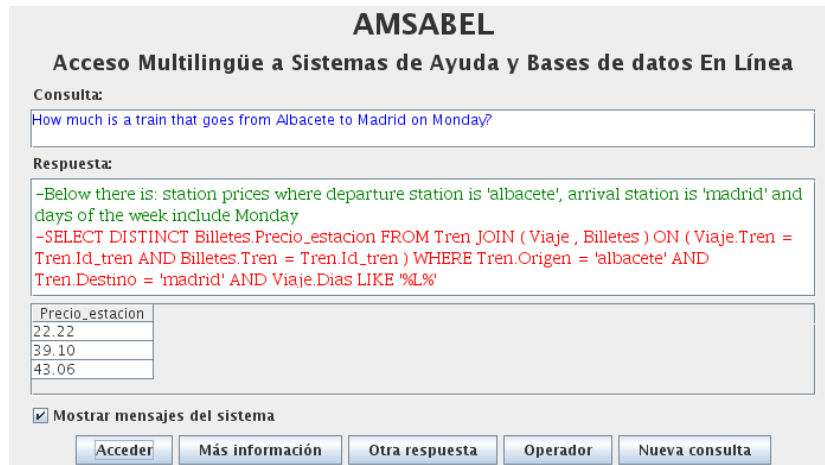


Figure 1: Screenshot of the prototype after answering a query

- The *user client* (figure 1) provides the user interface and the underlying communication with the database and translation servers.
- The *translation server* translates from natural language into SQL. It implements, as well, the user-system dialogue component.
- The *operator server and clients* allow human operators to assist the users in case they request help (see section 2.1).

In addition to these modules, one more server is used to store a relational database with the actual information. This version has been developed using *mySQL*<sup>1</sup> although the prototype could be easily adapted to any database management system, as long as it supports accessing through SQL and provides a TCP interface.

Two kinds of user clients have been implemented: the first one is a regular standalone application and the second one is an applet embedded into a web page. Despite that difference, both clients provide the same functionality.

## 2.1 Dialogue system

The prototype incorporates a dialogue component aimed at aiding the user within the query process. Specifically, a user is able to:

- Query the database in English or Spanish.
- Display additional information by adding related data.

- Request a new translation of the natural language query.
- Request assistance from a human operator who will receive the natural language query and will manually translate it into SQL.

Furthermore, the system tries to keep the user informed on the query process by returning supplementary information (e.g., error messages or natural language descriptions of the SQL query).

## 3 Demonstration

The presentation of the system will include an explanation of its main features and the performance of some sample queries, with an estimated duration of 15 minutes.

## References

- Koehn, P., F.J. Och, and D. Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 48–54, May.
- Revuelta-Martínez, A., L. Rodríguez, and I. García-Varea. 2010. Multilingual access to online help systems and databases. *Procesamiento del Lenguaje Natural*, 44:115–122, March.

<sup>1</sup><http://www.mysql.com/>