A Portable Natural Language Interface
for Diverse Databases Using Ontologies*

J. Antonio Zárate M.[1], Rodolfo A. Pazos R.[1],
Alexander Gelbukh[2], J. Isabel Padrón C.[1]

[1] Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET)
Interior Internado Palmira S/N,
Ap. Postal 5-164, 62490, Cuernavaca, Mor., México
{jazarate,pazos,jipc}@sd-cenidet.com.mx,

[2] Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico
gelbukh@cic.ipn.mx, www.gelbukh.com

Abstract. The growth of user needs for accessing information resources, the technological advance in this field, and the limitations of graphical and form-based interfaces, motivate the proposal of new solutions and the revision of several others in order to solve one of the main problems in computer applications: human-machine interface. Natural language processing has experienced a new impulse in recent years, and it is proposed as the best solution for the aforementioned problem. The first results of a project for developing a natural language interface to databases are presented, which is an extension of a larger project aimed at developing user interfaces for facilitating access to databases via Internet. In this project the use of ontologies is proposed as a means for making the interface portable to different databases, contributing in this manner to facilitate the configuration task for this type of interfaces, which is one of the main factors that have limited their application. In this paper the conceptual architecture of a natural language interface to databases on the Internet is described as well as the development attained.

1 Introduction

The fast growth of the Internet is creating a society where the demand for storage services, organization, access and analysis of information is constantly increasing. The advent of the Internet has completely changed the research directions in all areas of computer science, especially those related to databases as can be seen in the Asilomar report [5].

The growing need by users without wide knowledge of computers to access data over the Internet has resulted in the development of many types of interfaces, such as QBE (query by example) [42], form-based interfaces, restricted natural languages, etc. These tools, despite all the facility they provide to users, always imply some degree of difficulty when translating what the user would normally express to another person, into a structured form appropriate for the query engine.

A solution for the problem of enabling any user to express a data query easily is natural language interfaces to databases (NLIDB's). This topic has attracted interest since the 70's, and has motivated a large series of projects. However, these have not been widely used because of their being considered "exotic" systems, and the complexity and tediousness of their start-up configuration, in order to make the interface work for the first time with a specific database or for using the interface with another database, whose semantics is different from the original.

2 Previous Work of the Authors

The first project developed at the beginning of the last decade by the Distributed Systems Group of CENIDET, Mexico, was the Distributed Database Management System (SiMBaDD) [39]. In recent years, the group has focused on the problems of data access via Internet, with particular interest in interfaces to databases that are friendly enough for the great number of new Internet users, who are usually inexperienced in handling databases. Some examples of projects developed for this purpose are the following:

- A query by example (QBE) tool for databases on the Internet. Its main objective was to develop a tool that enabled inexperienced and casual users to access databases via Internet, in a platform-independent way (which was achieved through its implementation in Java) [32].
- A query by example (QBE) tool for multidatabases on the Internet. This project improved some features of the interface, such as the possibility of processing a query that involves tables in different databases, subquery processing, help windows, and a new three-tier architecture, which is the basis of the current project [21].
- An EzQ query tool for multidatabases on the Internet. The purpose of this project was to improve the human-machine interface, mainly concerning the ease with which inexperienced users can formulate queries that involve *joins*, without the user having to master the complex join concept [6].

These developments have led us to conclude that the next step is the integration of a NLIDB, a point of view shared by several investigators [1], since we consider that we have exhausted the possibilities of other types of database interfaces, either by using formal query languages, like in the project SiMBaDD [39], or using graphic tools for inexperienced users [32, 21, 6]. The current architecture of the QBE tool is shown in Figure 1.

It is well-known that NLIDB's are not the panacea for solving all the problems of human-machine interaction, as shown in a study [37]. However, in the same study it is demonstrated that in the cases when several tables are involved or when the solution is not similar to the examples previously known by the user, NLIDB's prove to be simpler than graphical interfaces or formal query languages.

An experiment carried out using Intellect, concluded that natural language is an effective method for the interaction of casual users with a good knowledge of the database in a restricted environment. The evaluation criteria of such type of interfaces are defined in [1].
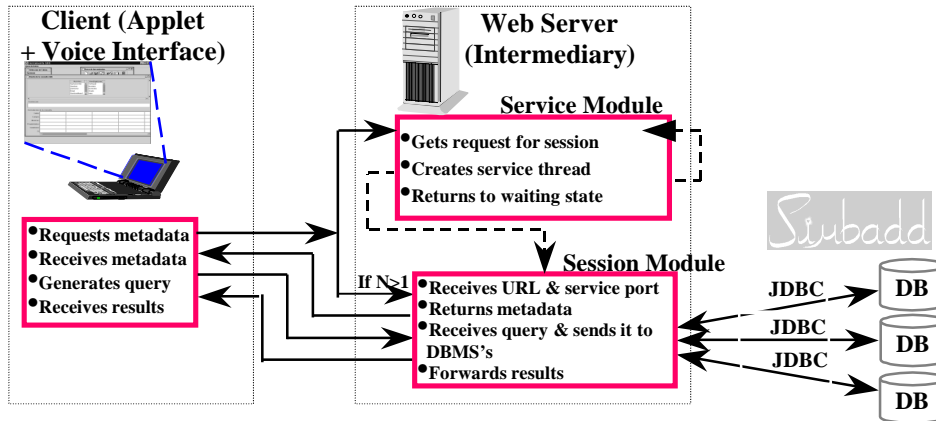
Fig. 1. Architecture of the Query by Example tool

3 Previous Work on NLIDB's and Ontologies

3.1 NLIDB's Projects

There exists in the literature a series of analysis about NLIDB's [1, 36, 41] which describe their evolution over the last four decades. Since our NLIDB is being developed for the Spanish language, we will limit our description to a few of the most important projects related to NLIDB's for Spanish:

- NATLIN (Universidad de las Américas-Puebla, Mexico). This system is a NLI for accessing databases expressed in logic using Sun Prolog (BIMprolog). NATLIN accepts questions for a geographical database domain. A module was added recently for translating into SQL the queries generated by NATLIN [35].
- INTERNAT (Ministry of Industry and Energy, Spain). INTERNAT is based on a translation approach from natural language to a formal language like SQL. It was implemented in C and can interact with dictionaries and other applications based on NLP and menu-based database access systems. INTERNAT was installed at AENA where its operation was validated [27].
- Silvia-NLQ (Universidad Autónoma de Madrid, Spain). The project was aimed at developing a database query system using natural language. The system had a multilingual character (English, German, French, Spanish and Italian) and its final goal was to develop a commercial product (LanguageAccess), which was introduced on the market. The participation of the Universidad Autónoma de Madrid consisted of developing the Spanish grammars for analysis and generation, as well as a dictionary and conceptual models for interpreting certain linguistic subdomains [20].
- GILENA (Universidad de Concepción, Chile). Taking as input several parameters (data dictionary, keywords and grammatical rules), it automatically generates all the source programs for a NLI. This tool was used to implement several interfaces:

expert system for failure diagnosis of electric equipment, teaching system of Indian cultures for grammar schools, NLIDB's for products and/or services, and command shell for Unix in Spanish. The parser was implemented using a nondeterministic augmented transition network [3].

It is worth mentioning that none of the aforementioned NLIDB's was designed for easy porting to databases different from the one for which it was originally developed nor the dictionary and the NLIDB knowledge base were designed for reuse or sharing, which are objectives that are pursued in our project.

3.2 Ontology Projects

Some of the most relevant projects aimed at using ontologies for achieving interoperability among applications are the following:

- Process Interchange Format (PIF). Its purpose is to exchange business process models using different representations. It uses an interlingua with local translators between PIF and local process representations [30].
- Knowledge Sharing Effort (KSE). DARPA (Defense Advanced Research Projects Agency) project aimed at the search of solutions for sharing knowledge among heterogeneous systems [24].
- Knowledge Representation Specification Language (KRSL). Language developed for representing plans and planning information. Its purpose is to provide a common vocabulary for concepts, relationships and common conditions for planning activities. KRSL considers two main aspects: an abstract ontology with the main categories (time, space, agents, actions, and plans) and a set of specialized modular ontologies with alternative concepts and theories common to planning systems (e.g. specific ontologies for time instants, time relationships, etc.) [2].
- Knowledge Interchange Format (KIF). It is a language that intends to represent through ontologies most of the current concepts and distinctions of the most recent languages for knowledge representation. It mainly intends to serve as a bridge between ontologies using proprietary language translators to/from KIF. It is a language based on predicate logic extended for definition of terms, metaknowledge, sets, nonmonotonic reasoning, etc. [19].
- Common Object Request Broker Architecture. Permits to retrieve and invoke operations on objects through a network. It provides a mechanism where objects can issue requests and receive responses transparently. CORBA defines an Interface Definition Language (IDL) that specifies objects and operations for remote/ distributed applications and incorporates informal notions of ontologies [16].
- CYC (project of Microelectronics and Computer Technology Corporation). Its ontology is organized as a microtheories network, where each microtheory captures the knowledge and reasoning needed for a specific domain, such as space, time, causality, or agents. A microtheory can show particular views related to a specific domain, therefore in one domain may coexist several microtheories [16].
- Toronto Virtual Enterprise (TOVE). Intends to develop ontologies for businesses using first order logic and permits to infer answers to common sense questions [15].

- Simple HTML for Ontology Extensions (SHOE). This was the first ontology and Web page tagging developed at the University of Maryland in 1996. SHOE can define ontologies and tags (which are meaning bearing XML tags), and a knowledge representation language based on HTML. One of its limitations is that it does not permit to define class negation and disjunction [38].
- Ontology Interchange Language (OIL). This language intends to combine Internet models with logic representations and the descriptive structures of ontologic approaches. OIL makes possible to infer conclusions about contents represented in this language [25].
- Resource Description Frame (RDF). It is a model for defining semantic relationships among different URI's. RDF is based on the XML syntax and permits to describe semantically a URI associating to it a set of properties and values. RDF models are constructed as directed graphs specifying triplets (URI, property, value). The metadata specified with RDF are understood by computers, and therefore, can be processed automatically [33].
- DARPA Agent Markup Language (DAML). Consists of a formalism that permits software agents to interact with each other. The DAML language is also an extension of XML and RDF. It provides a large number of constructions for ontology definition and increases semantic information to make it legible and understandable for computers. Ontology definition with DAML provides a new way to face the challenge of large scale integration of services. The proposed extensible interoperability network provides the necessary mediation level to solve the semantic differences among all the value chain participants. It has currently been used for annotating ontologies on the semantic Web [8].
- EuroWordNet. The objective of EuroWordNet is the multilingual extension of the English Word Net for the different languages involved (Italian, Spanish and Dutch). EuroWordNet has been proposed as a standard for the semantic codification of texts, and it is intended to be used as interlingua in multilingual systems for information retrieval and automatic translation [23].

4 Natural Language Query Processing System

The system will be used for the Spanish language spoken in Mexico, and will have additional elements with respect to other similar systems [3, 7, 10, 35]: a better language coverage, much better portability of DBMS and operating system, and transparent access through Internet.

The architecture used previously (Fig. 1) was substantially modified. The three-level client-intermediate-server structure is preserved, but the functionality of each level has been changed. The client functions will be much simpler, which will partially solve the problems of the current QBE interface, at the expense of a more active role of the intermediary level. The new architecture of the natural language query processing system for Web databases is shown in Fig. 2.

At the onset of a session with the interface, the client will present to the user an ontology (that will be stored in a repository), which represents the knowledge stored in the database dictionary. This differs from the QBE interface that shows the database designer's abstractions through tables, which most of the times are difficult

to understand by the inexperienced users and also lack a lot of very important semantic information. The presentation of this ontology permits the user to better understand the contents of the database, which facilitates the user to formulate his query.
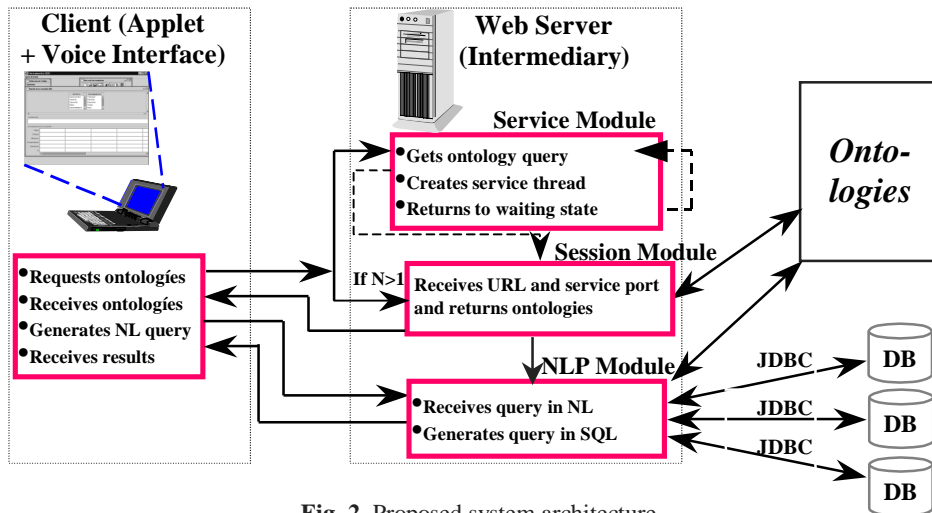

Fig. 2. Proposed system architecture

After receiving the ontology, the user issues a query using a voice interface. The output of this voice interface is received by the client and passed on to the service module of the intermediary, which in turn passes it to the natural language processing module (NLP). The architecture of the NLP module is quite standard, except that the lexicon consists of two parts: a general linguistic ontology based on the approach of the WordNet project [23], and a domain ontology that describes the semantics of the database (Fig. 3).

Upon reception at the NLP module, the natural language query is syntactically and semantically parsed in order to transform it into an internal representation, which in turn is translated into structured query language (SQL). The SQL query is sent back to the session module, which forwards it for evaluation against a database management system (DBMS). The session module forwards the result generated by the DBMS to the final user through the client interface.

Even though, there exist other proposals for separating linguistic and domain knowledge [12, 22], none of those possess a knowledge representation that is reusable, shareable and implemented according to a standard. In contrast, this project proposes the implementation of a lexicon following the outline of the WordNet project [23], which is becoming a de-facto standard, and the development of the domain ontology based on the DARPA Agent Markup Language (DAML) [6], which is backed by DARPA and is being used for tools development and different applications.

An important advantage that we find in DAML is that it can be used to implement both the ontologies and the inference mechanisms that utilize the ontology information. One innovative aspect of our proposal is that the semantic parser will be

implemented taking maximum advantage of the possibilities offered by DAML, such as described in [29].
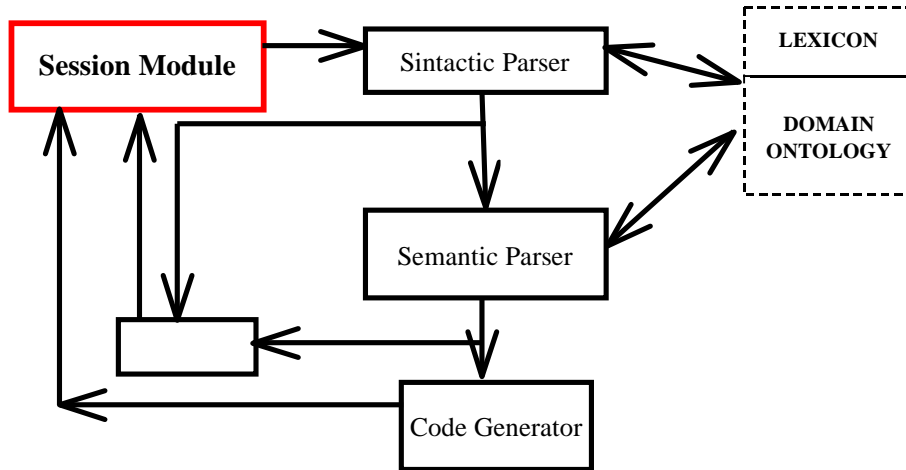

Fig. 3. Natural language processing module

5 Current Progress

Currently eighteen noun categories and seven verb categories have been input to the lexicon, which were coded following the outline of WordNet. Additionally, the lexical parser has been implemented [28], while the syntactic parser is being coded and will use the Spanish grammar developed in a previous work [13]. Table 1 shows an example of the grammar categories that are recognized by the lexical parser.

**Table 1.** Example of the lexical analysis of a sentence

WORD	CLASSIFICATION
la 'the'	Article, feminine, singular.
peregrina 'pilgrim'	Common noun, feminine, singular.
llevaba 'wore'	Verb llevar, 3rd person, preterit singular, imperfect indicative, first conjugation.
sombrero 'hat'	Common noun, masculine, singular.
negro 'black'	Color adjective, masculine, singular.

The general operation of the lexical analyzer is shown in Fig. 4. The user inputs his question by dictating it to the Dragon Naturally Speaking interface. This software translates the question into text and stores it in a file. The question may contain words that are not useful for the lexical analyzer, and therefore they are eliminated from the question in the text file. The list of irrelevant words was determined through a survey and can be modified as necessary. Examples of this type of words are the following: *quiero* (I want), *dame* (give me), *lístame* (list for me), *muéstrame* (show me), etc.

After the previous filtering, the sentence is passed on to the lexical analyzer, which classifies its words and tags them with syntactical information. Each tag includes all the syntactical categories to which the word belongs.
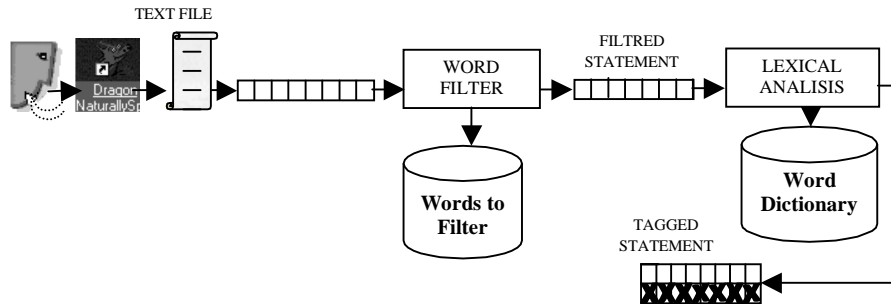
Fig. 4. General operation of the lexical analyzer
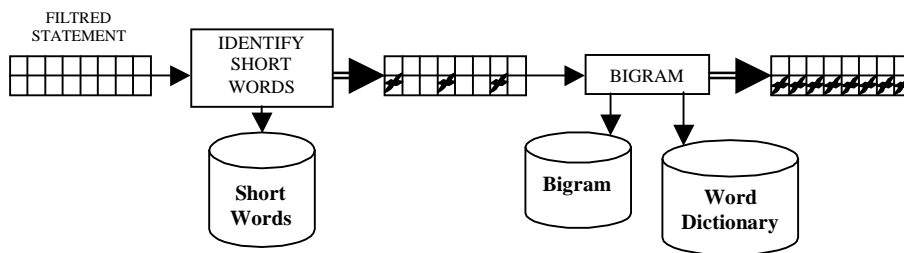
Fig. 5. Internal process of the lexical analyzer

The algorithm that performs the tagging (Fig. 5) consists of two steps. In the first step short words (articles, conjunctions, unions, etc.) are tagged, since they have a small probability of being ambiguous. In the second step, each word following a tagged word is assigned a tag using a Markov grammar [13], which permits to predict with some certainty the syntactic category of a word based on the syntactic category of the preceding word (these two words are also known as bigram). Some examples of the bigrams used are shown in Table 2. This last process is repeated for each remaining untagged word.

Table 2. Fragment of the bigram table

Utterance	Word	Word Tag	Next Word Tag
3	El	A&MS	N&13MS
2	El	A&MS	R&13MS
1	El	A&MS	V&13MS

It is quite possible that some lexical ambiguities arise when using bigrams, since there might exist several alternative categories possible for a word. In order to improve the disambiguation capabilities of the tagging algorithm, it is being expanded

to work with trigrams and higher order N-grams. It is important to point out that a working assumption is that the database semantics defined by the ontology will substantially limit the ambiguity possibilities in the query analysis.

The algorithm stops when the entire sentence is completely tagged, then it can be checked by the syntactic parser, so it can be restructured and a syntactic tree can be generated.

The lexicon was implemented using different tables for different word types: nouns, verbs (including the derivations corresponding to all tenses and persons), short words (articles, pronouns, conjunctions, and unions, which are shorter than six characters), adjectives, and adverbs. We chose this implementation instead of lemmatization at execution time, because search time for words in these tables is very fast, considering that today's main memory sizes permits to keep entire tables in main memory.

numero	Verbo	presente indicativo primera persona	segunda	tercera	cuarta	quinta	sexta
5	partir	parto	partes	parte	partimos	partís	parten

copretérito indicativo primera persona		segunda	tercera	cuarta	quinta	sexta
partía		partías	partía	partíamos	partíais	partían

pretérito indicativo primera persona		segunda	tercera	cuarta	quinta	sexta
partí		partiese	partió	partimos	partisteis	partieron

futuro indicativo primera persona	segunda	tercera	cuarta	quinta	sexta
partiré	partirás	partirá	partiremos	partiréis	partirán

pospretérito indicativo primera persona	segunda	tercera	cuarta	quinta	sexta
partiría	partirías	partiría	partiríamos	partiríais	partirían

presente subjuntivo primera persona	segunda	tercera	cuarta	quinta	sexta
parta	partas	parta	partamos	partáis	partan

pretérito subjuntivo primera persona		segunda	tercera	cuarta	quinta	sexta
partiera o partiese		partieras	partiera	partiéramos	partierais	partieran

futuro subjuntivo primera persona	segunda	tercera	cuarta	quinta	sexta
partiere	partieres	partiere	partiéremos	partiereis	partieren

presente imperativo segunda persona		tercera	cuarta	quinta	sexta
parte		parta	partamos	partid	partan

infinitivo	gerundio	participio
partir	partiendo	partido

Fig. 6. Example of the verb conjugation dictionary

Verb stems were input manually into the verb table (Fig. 6), while the morphological variants of each stem were generated and inserted automatically into the table. Short words and their corresponding tags and bigrams were input manually into the corresponding table. A fragment of this table is shown in Table 3.

Table 3. Fragment of the short words table

Num.	Word	Type	Word Tag	Next Word Tag
1	las 'the'	Article	A&FP	N&&FP
2	la 'the'	Article	A&FS	N&&FS
3	el 'the'	Article	A&MS	N&&MS
4	unas (some)	Indefinite	E&FP	N&&FP

In order to obtain the noun variants corresponding to number and gender, it was necessary a syllable divider. There exist 15 rules in the Spanish grammar for syllable division [9]. Table 4 shows some of the results obtained.

Table 4. Results of the syllable divider

Word	Syllable Division	Rules
Adscripción	Ads-crip-ción	VCC-CCVC-CDC
Inconsciente	In-cons-cien-te	VC-CVCC-CDC-CV
Costumbre	Cos-tum-bre	CVC-CVC-CCV

Where:
V for vocal,
C for consonant,
D for diphthong.

6 Final Remarks

The importance of developing natural language interfaces is explained by the need to make available computational resources to any user. This means that the language for accessing computers has to be the same as human language, either in written or spoken form.

A study conducted by a group of information system administrators on the usefulness of different applications of natural language interfaces concluded that those used for obtaining information from databases was preferred by users over those for information retrieval and text preparation [37]. This type of interfaces left very far behind other applications such as language translation.

Two aspects of this work are worth mentioning: the use of ontologies which is scarce for Spanish NLIDB's [22] and the portability of NLIDB's over different domains that can be achieved. The first aspect is very important because ontologies are being used for a wide variety of research topics (knowledge management, NLP, etc.). Equally important, the lack of portability together with other NLP problems has resulted in little use and popularization of NLIDB's.

In order to provide portability to our interface an ontology editor is being implemented [10], which will help define ontologies for the database domains. This editor will incorporate some features of other projects such as Protegé-2000 [31] and OntoWeb [26]. Additionally, the ontology design method Methontology [4] will be used for the domain ontology design and construction.

The work on natural language interfaces is necessary because there are more and more people that need access to computer resources, but do not have experience in this nor usually time to acquire it. Also, being Spanish the third language in the world by the number of native speakers (around 390 million), it is very important to develop appropriate tools for this huge market.

References

1. Androutsopoulos I., Ritchie G., Thanisch P.: Natural Language Interfaces to Databases, An Introduction, Department of Artificial Inteligence, University of Edinburg; http://citeseer.nj.nec.com/androutsopoulos95natural.html, 1992.
2. Allen J., Lehrer N.: DRAFT of the DARPA/Rome Laboratory Planning and Scheduling Initiative Knowledge Representation Specification Language (KRSL), Version 2.0.1 Reference Manual. ISX Corporation, June 26, 1992. http://www-ksl.stanford.edu/knowledge-sharing/papers/krsl.tar.Z.
3. Atkinson J.: GILENA (Generador de Interfaces en Lenguaje Natural), M.S. dissertation, Universidad de Concepción, Chile http://www.dai.ed.ac.uk/homes/atkinson/private/gilena, 1991.
4. Blázquez M., Fernández M., García-Pinar J. M., Gómez-Pérez A.: Building Ontologies at the Knowledge Level Using the Ontology Design Environment, Laboratorio de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, mayo de 1997, http://delicias.dia.fi.upm.es/articulos/kaw98/FernandezEtAl97.
5. Bernstein P., Brodie M., Ceri S., DeWitt D., García-Molina H., Gray J., Held J., Ellerstein J., Jagadish H.V., Lesk M., Maier J., Naughton J., Pirahesh H., Stonebraker M., Ullman J.: The Asilomar Report on Database Research, Sept., 1998.
6. Carreón G.: Herramienta para Consultas EzQ para Multibases de Datos en Internet, M.S. dissertation (to be published), Departamento. de Ciencias Computacionales, Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Mor.
7. Chay E.: Una Interfaz en Lenguaje Natural en Español para Consultas a Bases de Datos. M.S. dissertation, Monterrey Institute of Technology, Cuernavaca Campus, 1990.
8. DAML, Darpa Agent Markup Language, http://www.daml.org/.
9. Diccionario Básico de Gramática Práctica, ed. Oceano, Barcelona, España 1999.
10. Domínguez A. P.: Implementación de un Analizador Gramatical en Lenguaje Español. M.S. dissertation (to be published), Instituto Tecnológico de Cd. Madero. Jun. 2001.
11. EuroWordNet, supported by the European Union, http://www.hum.uva.nl/~ewn/.
12. Fox, M. S.: "The Tove Project: Toward a Common-Sense Model of the Enterprise." Lecture Notes in Computer Science 604 (1992) 25-43.
13. Galicia S. N.: Análisis Sintáctico Conducido por un Diccionario de Patrones de Manejo Sintáctico para Lenguaje Español, PhD. dissertation, Centro de Investigación en Computación del IPN, Aug. 2001.
14. García M.: Herramienta para la Generación de Ontologías a Partir de un Diccionario de Datos, M.S. dissertation in progress, Departamento de Ciencias Computacionales, Centro Nacional de Investigación y Desarrollo Tecnológico, Sept. 2002.
15. Gruber A.: A Translation Approach to Portable Ontology Specifications, Technical Report KSL 92-71, Knowledge Systems Laboratory, Stanford University, Sept. 1992.
16. Huerta V. O.: Un Método para el Reconocimiento a Bases de Datos en Interrogaciones en Lenguaje Natural, M.S. dissertation, Monterrey Institute of Technology, Cuernavaca Campus., 1989.
17. InBase, Russian Research Institute of Artificial Intelligence, Moscow, http://www.inbase.artint.ru/article/voicnl-eng.asp.
18. Jurafsky D., Martin J.H.: Speech and Language Processing, ISBN 0-13-095069-6, Prentice Hall, 2000.
19. KIF/ontolingua, "Knowledge Interchange Format", http://elies.rediris.es/ elies9/ 5-4.htm.
20. Marcos F., Olmeda C., Martínez J., Guilarte S.: Proyecto Sylvia-NLQ, Laboratorio de Lingüística informática, Universidad Autónoma de Madrid.

21. May A.: Herramienta para Consultas Basadas en Ejemplos (QBE) para Multibases de Datos en Internet, M.S. dissertation, Departamento de Ciencias Computacionales, Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Mor., Apr. 2000.

22. Morales E.: Curso de Representación de Conocimiento, Master in Science Faculty, Monterrey Institute of Technology Campus Cuernavaca, 1999, http://w3.mor.itesm.mx/ ~rdec/node1.html.

23. Miller G.: Wordnet. Wordnet, a Lexical Database, Cognitive Science Laboratory, Princeton University. http://www.cogsci.princeton.edu/~wn/.

24. Neches R.: The Knowledge Sharing Effort, an Overview, Knowledge Systems Laboratory, Stanford University, http://www-ksl.stanford.edu/knowledge-sharing/papers/kse-overview.html.

25. OIL , Ontology Interchange Language, http://www.ontoknowledge.org/oil/.

26. OntoWeb. http://delicias.dia.fi.upm.es/ontoweb/sig-tools/.

27. Palomar M., Moreno L., Molina A.: SISCO, Sistema de Interrogación en Lenguaje Natural a una Base de Datos Geográfica, GPLSI Grupo de Investigación en Procesamiento del Lenguaje y Sistemas de Información, Universidad de Alicante, http://gplsi.dlsi.ua.es/gplsi/ menupublicf.htm.

28. Padrón J. I.: Analizador Morfológico-Sintáctico de Oraciones en Lenguaje Español. M.S. thesis in progress, Departamento de Ciencias Computacionales, Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Mor., Oct. 2002.

29. Pease A.: Why Use DAML?, Teknowledge on 10 April, 2002, http://www.daml.org/ 2002/04/why.html.

30. Process Interchange Format, supported by: ARPA, NSF, Corporate Sponsors of the MIT Center for Coordination Science. http://ccs.mit.edu/pif/.

31. Protégé-2000. http://protege.stanford.edu/index.html.

32. Rasgado F.: Herramienta para Consultas Basadas en Ejemplos (QBE) para una Base de Datos en Internet, M.S. dissertation, Departamento de Ciencias Computacionales, Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Mor., Dec. 1999.

33. Resource Description Framework (RDF), World Wide Web Consortium (W3C), http://www.w3c.org/RDF/.

34. Reis P., Matias Nuno J.: Edite - A Natural Language Interface to Databases, a New Dimension for an Old Approach. INESC Av. Duque dÁvila 23, 1000 Lisboa, Portugal.

35. Rocher G. : Traducción de Queries en Prolog a SQL, B.S. dissertation. Computer Systems Enginnering Dept, Engeneering Faculty, Universidad de las Américas-Puebla. Sep. 1999.

36. Rojas J., Torres J.: A Survey in Natural Language Databases Interfaces, Proc. of the Eigth. International Congress on Computer Science Research. Instituto Tecnológico de Colima, Colima, México, Nov. 2001, pp. 63-70.

37. Sethi V.: Natural Language Interfaces to Databases: MIS Impact, and a Survey of Their Use and Importance, Graduate School of Business, Univ. of Pittsburgh. Pittsburgh, PA 15260.

38. SHOE, Simple HTML Ontology Extensions, http:// www.cs.umd.edu/ projects/plus/SHOE.

39. SiMBaDD, Sistema Manejador de Bases de Datos Distribuidas, Departamento de Ciencias Computacionales, Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Mor., http://www.sd-cenidet.com.mx/simbadd.

40. Validation and Business Applications Group: PASO-PC315 PROJECT, Generator of Natural Language Databases Interfaces, http://www.vai.dia.fi.upm.es/ing/projects/paso.htm.

41. Zárate J.A, Pazos R., Gelbukh A.: Natural Language Interface for Web-based Databases, In Proc: 2nd WSEAS Int. Conf. on Robotics, Distance Learning and Intelligent Communication Systems (ICRODIC 2002), http://www.wseas.com/conferences/2002/ skiathos/icrodic/.

42. Zloof M.: Query By Example: a Database Language. IBM Sys. Journal 16. 4 (1977) 137 - 152.