Multilingual Question/Answering: the DIOGENE System

Bernardo Magnini, Matteo Negri, Roberto Prevete and **Hristo Tanev**
ITC-Irst, Centro per la Ricerca Scientifica e Tecnologica
Via Sommarive, 38050 Povo (TN), Italy
{magnini,negri,prevete,tanev}@itc.it

Abstract

This paper presents the DIOGENE question/answering system developed at ITC-Irst. The system is based on a rather standard architecture which includes three components for question processing, search and answer extraction. Linguistic processing strongly relies on MULTIWORDNET, an extended version of the English WORDNET. The system has been designed to address two promising directions: multilingual question/answering and question/answering on the Web. The results obtained in the TREC-10 main task will be presented and discussed.

1 Introduction

In the last few years, under the promotion of the TREC-8 (Voorhees and Tice 1999) and TREC-9 (Voorhees 2000) competitions, there has been a large interest in developing QA systems. Being our first participation at TREC-QA we adopted, for the DIOGENE system, a rather common architecture with three basic components: a *question processing* component, based on several linguistic processors and resources including WORDNET; a *search component*, based on information retrieval techniques; and an *answer processing* component, which exploits the similarity between the question and the documents to identify the correct answer.
In order to identify which modules are appropriated both previous experiences in QA and modules already available at ITC-Irst have been considered. Among the proposals that we have found of interest for the linguistic analysis, we have considered the taxonomy of question types implemented in LASSO (Moldovan et al. 1999) for identifying both the question type and the answer type; the term extraction capabilities developed for the QALC system (Ferret et al. 2000); and the use of lexical knowledge contained in WORDNET for the retrieval of semantic relations, as proposed in (Moldovan et al. 2000).
In addition, during the design phase of the system, two directions for future developments have been taken into consideration: *multilingual QA* and *QA on the Web*. Multilinguality is a crucial aspect when the language of the search question and the language of the text collection are different. We experimented with an Italian/English scenario, where the question can be posed either in English or Italian, the search is performed either in English or Italian, and the answer is given in the language of the question. In the paper we will discuss the solutions adopted for multilinguality, even if they are currently out of the scope of the TREC competition.
As for QA on the Web, this perspective raises a number of specific issues, the most evident being that the implication of an answer with respect to its question for a Web user is generally weaker than for controlled text collections, where human judges apply rigid tests. A reason for this is that the relevance of a Web document relies on several factors. For instance, a retrieved document could not include the answer in itself, but could nevertheless provide links to other documents useful to find the answer. Elements that provide implicit knowledge are the document structure, hypertextual links, multimedia co-reference, and generally, a strong use of contextual information.

This paper is structured as follows. Section 2 introduces the overall architecture of the system. Section 3 addresses the linguistic analysis of the question, including word sense disambiguation, answer type identification and query expansion. Section 4 describes the searching modalities. Section 5 presents the answer extraction

component, including paragraph filtering, named entities recognition and answer identification. Section 6 illustrates the TREC results, with a discussion on strengths and weaknesses of DIOGENE.

2 Architecture

DIOGENE relies on a rather standard architecture based on three basic components (see Figure 1): a *question processing* component, a *search component* and an *answer processing* component.

The ***question processing*** component is in charge of the linguistic analysis of input questions that can be formulated either in English or in Italian. At this step of the process, we confront the multilinguality problem using language specific resources. The analysis is performed sequentially by the following modules.

- ***Tokenization and pos tagging***. First the question is tokenized and words are disambiguated with their lexical category by means of statistical part of speech taggers. The Treetagger developed at the University of Stuttgart (Schmid, 1994) is in charge of the disambiguation of English words. Italian questions are processed by a part of speech tagger developed at ITC-Irst.
- ***Multiwords recognition***. About five thousand multiwords (i.e. collocations, compounds and complex terms) have been automatically extracted from different resources and are recognized by pattern matching rules. English multiwords have been extracted from WORDNET (Fellbaum, 1998); Italian multiwords have been extracted from a monolingual Italian dictionary (Disc, 1997)
- ***Word sense disambiguation***. This module, described in Section 3.1, disambiguates words in the query with respect to their senses. MULTIWORDNET (Pianta et al., 2002) was adopted as a sense repository. Word sense disambiguation is crucial for providing reasonable keyword expansions and correct translations between the two languages.
- ***Answer type identification***. The answer type for a question represents the entity to be searched as answer. This module relies on a taxonomy of answer types and a pattern matching rule system, both described in Section 3.2.

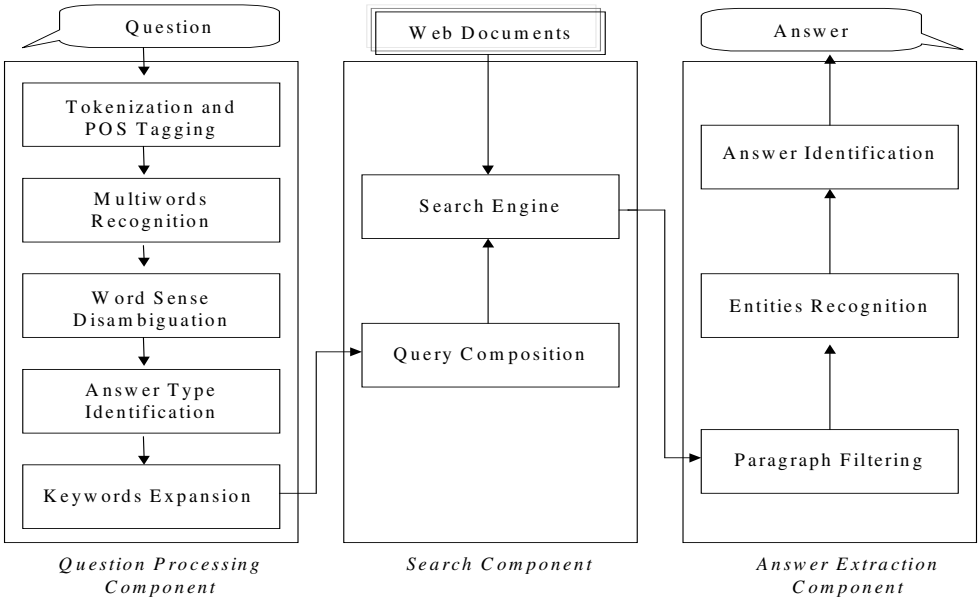

Figure 1. System Architecture.

- ***Keywords expansion***. Two kinds of expansions, described in Section 3.3, are carried out: word synonyms and morphological derivations. Also in this case, multilinguality is guaranteed by the use of MULTIWORDNET.

The ***search*** component first composes the question keywords and their lexical expansions and then performs the document retrieval. For the participation at TREC-10 the ZPrise search engine, described in section 4, has been used.

The ***answer extraction*** component implements a paragraph filtering module that extracts text paragraphs from the top scored retrieved documents. This is done by maximizing the number of keywords and expansions produced in the question processing phase within a window of a fixed length of words. The output is composed by the text paragraphs that should contain the answer to the question. Then, a *named entities recognition* module identifies the entities in the text paragraphs corresponding to the answer type category. We are using an adaptation of *Learning*-PINOCCHIO (Ciravegna, 2000), which makes use of learning algorithms to recognize named entities, such as persons, organizations, locations, measures and dates. Finally, the *answer identification* module highlights the portion of text containing the answer to the question which is then presented to the user.

3 Linguistic Expansions

3.1 Semantic Disambiguation

The identification of the correct sense of a word in a question is necessary to add either synonyms or translations for that word without the risk of introducing disturbing elements in the search query. There are two crucial questions to address: first, a repository of word senses has to be identified; second, it is important to develop a disambiguation technique able to cope with the specificity of questions, particularly with the availability of a limited context (i.e. few words).

As for sense repository we have adopted MULTIWORDNET (Pianta et *al*. 2002), a multilingual lexical database including information about English and Italian words. MULTIWORDNET is an extension of English WORDNET (Fellbaum, 1998), a semantic network of English words grouped into synonym sets called synsets. Words and synsets are linked by means of various relations, distinguished into semantic relations, which link concepts, and lexical relations, which link individual words. The main lexical relations represented in WORDNET are synonymy and antonymy, while hyponymy, hypernymy, meronymy, entailment and conceptual opposition are the main semantic relations that link the synsets. MULTIWORDNET has been developed keeping as much as possible of the semantic relations available in the English WORDNET: Italian synsets have been created in correspondence with English synsets, importing lexical and semantic relations from the corresponding English synsets. The Italian part of MULTIWORDNET currently covers about 40,000 lemmas, completely aligned with the English WORDNET 1.6 (i.e. with correspondences to English senses).

As far as word disambiguation is concerned we have applied Word Domain Disambiguation (WDD), a technique already experimented for the disambiguation of short news (Magnini, Strapparava 2000), and further extended by adding domain frequency information. Word Domain Disambiguation is a variant of Word Sense Disambiguation (WSD) where for each word in a text a *domain* label, (among those allowed by the word) has to be chosen instead of a *sense* label. Domain labels, such as MEDICINE and ARCHITECTURE, provide a natural way to establish semantic relations among word senses, grouping them into homogeneous clusters. In MULTIWORDNET the synsets have been annotated with one or more domain labels selected from a set of about two hundred labels hierarchically organized (see (Magnini, Cavaglià, 2000) for the annotation methodology and for the evaluation of the resource).

The WDD algorithm works in two steps. First, for each content word in the query and for each sense of the word, the corresponding domain labels in MULTIWORDNET are collected with a score determined by the frequency of the label

among the senses of the word. Let us consider as example an ambiguous query from the TREC corpus. In "*What is the brightest star visible from Earth*?" the situation after the first step of the WDD algorithm is represented in Figure 2.

star	star#1: celestial body	ASTRONOMY
	star#2: an actor who play	ART
bright	bright #1: brilliant shining	PHYSICS
	bright #2: popular glorious	FACTOTUM
	bright #3: promising auspicious	FACTOTUM
visible	visible #1: conspicuous obvious	PHYSICS
	visible #2: visible seeable	ASTRONOMY
earth	earth #1: Earth world globe	ASTRONOMY
	earth #2: estate landed_estate …	ECONOMY
	earth #3: clay	GEOLOGY
	earth #4: soil dirt	GEOLOGY
	earth #5: ground_earth	ELECTRICITY
	earth #6: dry_land solid_ground …	GEOGRAPHY
	earth #7: land ground soil	GEOGRAPHY
	earth #8: earth ground	GEOLOGY

Figure 2. Word Domain Disambiguation.

At the second step, all the possible tuples of domain labels for each word are scored by means of a similarity function, and the best tuple is selected. The similarity between two domains is computed according to the probability of the two domains to co-occur within a text. This information has been computed over several balanced corpora, both for English (i.e. the Brown corpus, the LOB corpus and the Reuters news corpus) and for Italian (i.e. the Elsnet corpus and a large collection of newspaper news). In our example, the algorithm selects ASTRONOMY for "star", PHYSICS for "bright", ASTRONOMY for "visible" and ASTRONOMY for "earth", which correspond to our intuition about the involved word senses.

Results obtained over the 200 TREC questions, previously manually annotated with the correct domain label for each keyword, are very encouraging, showing a limited loss in accuracy with respect to WDD over longer texts, where larger pieces of context are available for disambiguation.

3.2 Answer Type Identification

The answer type for a question represents the entity to be searched as answer. In order to extract this information, a taxonomy of *answer types* was manually defined starting from the 200 TREC-8 questions. The taxonomy includes categories such as "LOCATION", "PERSON", "TIME-PERIOD", "MEASURE" and "GENERIC". Then, each category is associated with a set of rules that check different features of the question; in particular a rule may detect the presence of a particular word occurrence, of words of a given part of speech, and of words belonging to a given semantic category. For instance, the rule described in (1) matches any question starting with "quale" ("*what*"), whose first noun, if any, is a person.

(1) RULENAME: WHICH-WHO
TEST: ["which" [¬NOUN]*
[NOUN:person-p]ɹ +]
OUTPUT: ["PERSON" J]

Rule (1) matches questions like "*What famous communist leader died in Mexico City?"* because the first noun encountered after "what" (i.e. "leader") satisfies the person-p constraint. The same rule does not match the question "*What large U.S. city had the highest murder rate for 1988?*", because "city" does not satisfy the person-p predicate. Semantic predicates (e.g. location-p, person-p, time-p, etc.) are defined on the MULTIWORDNET taxonomy, already described in Section 3.1. Each predicate checks if the sense of a word referred in a rule is subsumed by at least one high level synset manually defined for that predicate. As an example, for person-p, we identified synsets like person#1 ("human being") and group#1 ("any number of entities considered as a unit"). Then the predicate is satisfied if leader#1 is subsumed by at least one of these synsets.

While rules are mostly language dependent, semantic predicates defined on MULTIWORDNET are reusable for English. For instance, rule (2) is the corresponding rule for Italian:

(2) RULENAME: QUALE-CHI
TEST: [“quale” [¬NOUN]*
[NOUN:person-p]ɹ +]
OUTPUT: [“PERSON” J]

The output of a rule gives two pieces of information: the category of the answer type and the focus of the question (Moldovan et *al*. 1999), i.e. the word that expresses the answer type in the question. In the examples above, the answer type is the category PERSON, while the focus is the word “leader”. This information will be used to retrieve the correct answer to the question in the documents retrieved by the search engine. In particular, knowing the category of the entity we are looking for (i.e. a PERSON) the focus will be used to determine if any “candidate answer” we find in a document (i.e. person names) is an appropriate instantiation of that category. This can be done by accessing the MULTIWORDNET taxonomy and checking if the candidate answer is a synonym of the focus or is subsumed by the focus.
Currently we use about 90 answer type rules for Italian and about 70 for English. They have been checked on the TREC corpus resulting respectively in a 93% accuracy and 91%. Failures are due mainly to pos-tagging and disambiguation errors.

3.3 Keyword Expansion

At this step of the linguistic processing of the question, a stop words filter is applied that cuts off both non content words and non relevant content words. The remaining words (we call them “basic keywords”) are then passed to an expansion phase which considers both morphological derivations and synonyms.

Morphological derivation. The approach adopted is answer oriented, in that it considers the expansions with the higher probability to appear in the answer. For instance, given the question “*Who invented the electric light?*”, five expansions are automatically generated for the basic keyword “invent”: the past participle masculine “inventato”, because this is the actual form of the lemma; the past participle female “inventata”, because the direct object of the verb is female; the past indicative “inventò”, because in Italian it can substitute the past participle; the noun “inventore”, because it is the nominalization of the subject of the verb; finally, the noun “invenzione”, because it is the nominalization of the object of the verb. Derivations have been automatically extracted from an Italian monolingual dictionary (Disc, 1997).

Synonyms. The approach adopted for disambiguation, i.e. word domain disambiguation, is in line with this assumption: domains allow the clustering of related WORDNET senses. Once a domain label for a word is selected by the disambiguation algorithm, synonyms are collected from all the MULTIWORDNET synsets for that word belonging to the selected domain. For instance, given the morphological expansions described above for the verb “invent”, a number of synonyms extracted from MULTIWORDNET are added, including *discover* and *discoverer*.

3.4 Definition questions

In the TREC10 main-task there are several definition questions, whose answer requires a definition of the focus. For these questions a number of specific patterns have been defined, which consider the following features:

- definition questions typically begin with “Who” or “What”;
- they confine to the pattern “Who/What be sNP?”, where “be” stands for the different possible forms of the verb “to be” and sNP is a simple noun phrase, which consists of a noun or adjective + noun (eventually preceded by article). For example: “Who was Galileo?” or “What is an atom?”.
- Anther specific pattern for definition question is “What does ABBREV stand for?” (e.g. “What does NASA stand for?”)

The peculiarity of these questions is that the keywords themselves do not provide enough information for extracting the proper answer. For the above mentioned questions the keywords are just “Galileo”,“atom” and “NASA”.
By restricting ourselves to simple NPs consisting of noun or adjective + noun, we avoid the

improper coverage of non-definition questions. For example the *what-where* questions described in (Moldovan et *al.* 1999) are not considered definitions because they usually use complex NPs (e.g. "What is the capital of Uruguay?").
In the case of a definition question, DIOGENE makes use of the WORDNET glosses to expand the focus of the question. The intuition behind the WORDNET gloss expansion is that because the expected answer for a definition question will be a definition of the question focus, it is reasonable to expect that it contains words that also appear in the WORDNET gloss of the focus (i.e. its definition). For instance, in the case of "Who is Galileo?" the related keywords and multiwords extracted from the gloss of "Galileo" are "astronomer", "Italian", "mathematician", and "refracting telescope". The gloss keywords are used to extend the query and they are also considered during answer extraction.
The gloss keywords extraction takes into account the gloss nouns, because we consider nouns more informative than verbs, adjectives and adverbs. All the other parts of speech are considered only if they begin with a capital letter (in the Galileo gloss – "Italian"), because usually they denote names.

4. Search Component

At this moment DIOGENE makes use of the ZPrise (Dimmick, 1998, Downey, 1999) search engine developed by the Retrieval Group at NIST. In particular, we used PRISE 2.0, which is part of the Z39.50/PRISE2.0 package. ZPrise is based on vector techniques and supports term feedback; on the other hand it does not support Boolean search and has restricted phrase search capabilities.
Query composition is performed after keywords are extracted from the question and every word is supplied with an expansion list containing its word form and its expansions. Let's see how the query composition looks in the case of the question "Who is the inventor of the electric light?" After part-of-speech tagging and multiword extraction, the following keywords (key-phrases) are taken: "electric light" and "inventor". The lexical expansion produces the following expansion list for every key-phrase: "electric light", "electric light bulb", "incandescent lamp", and also morphological derivatives for these nouns: for "inventor" the expansion is "inventor" "artificer" "discoverer" "inventors" "artificers" "discoverers". For definition questions, gloss keywords are also extracted.
Previous experiments on search modalities (Magnini and Prevete, 2000) have proven the advantage of a Cartesian product search modality, in which a Boolean expression based on the Cartesian product of the expansion lists is generated. Because of the lack of Boolean expression support in ZPrise we realised the AND expressions as a single ZPrise query and the OR expressions were transformed into multi-line queries. Given the weak support of phrases in ZPrise we had to decompose multiword units into separate words, which diminishes the precision of the search. In addition, in the process of query composition no more than ten keywords for a question are taken into account because ZPrise slows down significantly when a big number of keywords are presented in one query line. Keywords are scored taking into account their part of speech (e.g. nouns score better then verbs), their polysemy (i.e. polysemous words score worse that less polysemous) and the first character of the word.
The speed of the search engine turned out to be a set-back during the time of the TREC-10 questions processing. To speed up the performance we divided the document collection into four sub-collections and created a separate index for every sub-collection. This made possible a parallel search with four instances of ZPrise running on four different computers simultaneously, while every instance processes its own sub-index. After that the results are unified using UNIX shell scripts specifically developed for this purpose. This strategy significantly improved the time for the search. The method of parallel work was also used in the final stage of the question processing.

5 Answer Extraction Component

5.1 Paragraph Filtering

A paragraph filtering module extracts text paragraphs from the top scored retrieved documents. This is done maximizing the number of keywords and expansions produced in the question processing phase, within a window of a fixed length of words. The output is composed by text paragraphs that should contain the answer to the question. Paragraph filtering can be tuned with three parameters: (i) the length of the paragraphs to be extracted, which was set at 200 words; (ii) the percentage of keywords and expansions that need to be present in the paragraph; (iii) a list, eventually empty, of obligatory keywords, whose presence is necessary.

5.2 Named Entities Recognition

Once the relevant paragraphs have been selected, the named entities recognition module identifies, among possible candidate answers, the entities that match the answer type category (i.e. PERSON, ORGANIZATION, LOCATION, MEASURE and DATE). The task is performed by *Learning*-PINOCCHIO (Ciravegna, 2000), a system for adaptive information extraction developed at ITC-Irst. *Learning*-PINOCCHIO learns template filling rules that insert SGML tags into texts without any other user intervention than corpus tagging.

During the training phase the system induces a large set of rules by bottom-up generalization of instances in a tagged training corpus where texts have been manually marked with SGML tags locating the information to be extracted (in our case the tags <*person*>, <*person*>, <*organization*> <*organization*>, <*location*> <*location*>, <*measure*>, <*measure*>, <*date*> and <*date*> have been used). Rule induction is also supported by dictionaries, such as repositories of person's first and last names, geographical and organization names extracted from gazetteers, online resources and WORDNET.

The training corpus used (~400Kb) was randomly extracted from the TIPSTER collection, part of the TREC material at our disposal. As expected, system performances are highly affected by the dimension, format and domain of the training corpus: documents contained in the TIPSTER collection proved to be rather heterogeneous, hampering the construction of a representative training corpus. Results obtained on the training corpus vary among categories, ranging from 74.5% precision and 82% recall for the category DATE, to 60% precision and 57% recall for the ORGANIZATION category.

5.3 Answer Identification

Answer identification is performed after named entities are recognised by *Learning*-PINOCCHIO. In case the answer type of the question is consistent with the type of an entity *c* in a given paragraph *p*, the entity is accepted as a candidate answer. Then, the candidate entity is scored considering the presence of relevant keywords in a fifty byte interval around it. More formally, given:

- $t_1, t_2, \ldots, t_m$ is the sequence of tokens composing the paragraph *p*, with *i* denoting the position of the i-th token t_i.
- *k*, *s* and *e* are functions such that *c* is composed of *k(c)* tokens, *s(c)* is the position of the first token and *e(c)* is the position of the last token of *c* in *p*.
- *kl(p)* is the list of keywords belonging to *p*.
- *len* is the *kl* length
- member(i,p) is a function which gives as output 1 if the token t_i is a *kl(p)* member, otherwise 0.

We define the *left key density* and the *right key density* of the answer candidate *c* in paragraph *p*, indicated respectively LKD(c,p) and RKD(c,p), using the following pair of functions:

$$\text{LKD(c,p)} = \frac{1}{len} \sum_{i=1}^{s(c)-1} \frac{\sqrt{member(i,p)*\alpha}}{s(c)-i}$$

$$\text{RKD(c,p)} = \frac{1}{len} \sum_{i=e(c)+1}^{m} \frac{\sqrt{member(i,p)*\alpha}}{i-e(c)}$$

Where α is a tuned parameter.

In order to build a 50 byte answer for a candidate entity *c*, we initially consider a default string A composed of the words in *c* (an entity can be composed of more than one word, for instance "Bill Clinton" is an entity whose type is PERSON). If A is composed of less than 50 characters other characters are appended to A both on its left and on its right. The number of characters added to the left is defined by LKD(c,p), while the number of characters added to the right is defined by RKD(c,p).

6 Results and Discussion

Being our first experience with a QA system and given the high number of different modules assembled in DIOGENE, we did not expect high performance. The system correctly answered just 10% of the questions of the TREC main task. A manual analysis over a small set of questions was carried out to evaluate problems related to single modules. Four modules were considered: answer type identification, search engine, paragraph filtering and named-entities recognition. Each module was evaluated in term of its error rate, given a small amount of correct inputs. The estimated error rate for the answer type module was 11%, which is comparable with the performances calculated at training time (see Section 3.2). ZPrise produced a very high error rate (53%), which means that for less than half of the questions the search engine retrieved a document containing the answer. The main reason is that filters (i.e. no more than 10 documents for a question and no document with length exceeding 2000 words) applied to ZPrise output are actually too restrictive. They were implemented to take the processing time under control, but their effects were underestimated. As for paragraph filtering, its estimated error rate is around 40%, which also indicates that the techniques for paragraph extraction can be significantly improved. Finally, the estimated error rate for *Learning*-Pinocchio was around 60%, which was mainly due to the low homogeneity between training and test corpus.

7. Conclusion and Future Work

The system described in this paper participated to the main task of the TREC-10 competition. Even though it relies on a rather standard architecture, DIOGENE deals with two important issues related to QA: multilingual QA and QA on the Web. For these purposes, some modules already available in ITC-Irst have been used. The results obtained by the whole system and the performance of each module have been described in the paper.

Another crucial issue for the future is the automatic evaluation of a QA system. The basic idea is to develop an evaluation methodology that does not rely on the human judgment of thousands of answers. Although there is some recent work in this direction (Breck et al. 2000), the approach we are testing considers the Web as the main information source to evaluate the relevance of an answer. Our work is based on the assumption that if a certain answer is relevant with respect to a given question, then there should be many documents containing keywords extracted both from the question and from the answer. Moreover, these documents should be semantically similar documents, i.e. they should maximize the overlapping of semantic features, such as the set of domains labels (see Section 3.1) extracted from each text.

References

Breck, E.J., Burger, J.D., Ferro, L., Hirschman, L., House, D., Light, M., Mani, I.: How to Evaluate Your Question Answering System Every Day …and Still Get Real Work Done. Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation (2000) 1495-1500.

Ciravegna, F.: Learning to Tag for Information Extraction from Text. In Ciravegna, F., Basili, R., Gaizauskas, R. (Eds.) ECAI Workshop on Machine Learning for Information Extraction, Berlin (2000).

Dimmick D., O'Brien G., Over P. and Rogers W.: Guide to Z39.50/PRISE2.0: Its Installation, Use and Modification.

Gaithersburg, USA, unpublished resource, (1998).
Downey L., Tice D.: A Usability Case Study Using TREC and ZPrise , Information Processing and Management, 35 (5), 589-603 , (1999).
Disc. Dizionario Italiano Sabatini Coletti. Giunti, Firenze (1997)
Fellbaum, C.: WORDNET, An Electronic Lexical Database. The MIT Press (1998).
Ferret, O., Grau, B., Hurault-Plantet, M., Illouz, G., Jacquemin, C., Masson, N., Lecuyer, P.: QALC – The Question Answering System of LIMSI-CNR. Proceedings of the Ninth Text Retrieval Conference (TREC-9), Gaithersburg, MD. (2000).
Harabagiu, S., Pasca, M., Maiorano, S.: Experiments with Open-Domain Textual Question Answering. Proceedings of Coling-2000 (2000a).
Magnini, B., Cavaglià, G.: Integrating Subject Field Codes into WORDNET. Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation (2000) 1413-1418.
Magnini, B., Prevete, R.: Exploiting Lexical Expansions and Boolean Compositions for Web Querying. Proceedings of the ACL workshop on Recent Advances in Natural Language Processing and Information Retrieval, Hong Kong (2000) 13-21.
Magnini, B., Strapparava, C.: Experiments in Word Domain Disambiguation for Parallel Texts. Proceedings of the ACL workshop on Word Senses and Multilinguality, Hong Kong (2000) 27-33.
Moldovan, D., Harabagiu, S., Pasca, M., Mihalcea, R., Goodrum, R., Girju, R., Rus, V.: LASSO: A Tool for Surfing Answer Net. Proceedings of the Eight Text Retrieval Conference (TREC-8) (1999) 65-74.
Moldovan D., Harabagiu S., Pasca M., Girju R., Goodrum R., Rus V.: The Structure and Performance of an Open-Domain Question Answering System. Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL), Hong Kong, (2000).
Pianta E., Bentivogli L., Girardi C.: MultiWordNet: Developing an Aligned Multilingual Database. Proceedings of the 1st International Global WordNet Conference, Mysore, India (2002)
Schmid, H.: Probabilistic Part-Of-Speech Tagging Using Decision Trees. Proceedings of the International Conference on New Methods in Language Processing (1994).
Voorhees, E.M., Tice, D.M.: The TREC-8 Question Answering Track Evaluation. Proceedings of the Eight Text Retrieval Conference (TREC-8) (1999).
Voorhees, E.M.: Overview of the TREC9 Question Answering Track. Proceedings of the Ninth Text Retrieval Conference (TREC-9) (2000).