The DIOGENE Question Answering System at CLEF-2004

Hristo Tanev, Matteo Negri, Bernardo Magnini, Milen Kouylekov

{tanev|negri|magnini|kouylekov@itc.it}
ITC-irst, Centro per la Ricerca Scientifica e Tecnologica
Via Sommarive, 38050 Povo (TN), Italy

Abstract. This paper presents the ITC-irst Multilingual Question Answering system DIOGENE. The system was used successfully on the CLEF-2003, TREC-2003, TREC-2002 and TREC-2001 QA tracks. DIOGENE relies on a classical three-layer architecture: question processing, document retrieval, answer extraction and validation. DIOGENE uses MultiWordNet [Pianta et.al. 2002] (http://multiwordnet.itc.it) which facilitates the transfer of knowledge between languages. For answer validation we used the Web. This year we used also a set of linguistic templates for answering specific questions like definition questions, location questions, and a subset of who-is and what-is questions. DIOGENE participated in both monolingual Italian-Italian task and in the cross-language Italian-English task. A co-operative participation together with the Bulgarian Academy of Sciences was carried out in the cross-language Bulgarian-English QA task.

1   Introduction

Research in Question Answering (QA) has received a strong boost in recent years by the QA track organized within the TREC conferences [Voorhees 2004], which aims at assessing the capability of systems to return exact answers to open-domain English questions. However, the TREC conferences are concentrated exclusively on the English language. In contrast, the CLEF conferences provide a multilingual forum for evaluation of NLP systems in languages other than English. Multilinguality has been recognized as an important issue for the future of QA [Burger et.al.2001]. In CLEF-2003 a multilingual QA task was introduced for the first time. Our system showed promising results in CLEF-2003 in the monolingual Italian and cross-language Italian-English tasks. This encouraged us to participate also in the Bulgarian-English cross-language task, promoted this year.

The multilingual version of DIOGENE was built upon the same well tested three-layers architecture of the English version [Magnini et.al.2002a]. Figure 1 shows the main constituents of this common backbone: these are the *question processing* component, the *document retrieval* component, and the *answer extraction and validation* component. In all the monolingual and cross-language modalities DIOGENE relies on the knowledge in multilingual ontology MultiWordNet [Pianta et.al.2002], manually created rules for named entity recognition and question type identification, a set of handcrafted answer extraction templates and statistical information collected from the Web and off-line multilingual corpora. DIOGENE CLEF-2004 architecture was similar to the CLEF-2003 version [Negri et.al. 2003]. A novel feature for the CLEF-2004 version is the answer extraction and validation via linguistic templates. Linguistic templates were particularly important for the definition questions.

The following sections will provide a general overview of our participation to the *monolingual Italian* (M-I), *bilingual Italian/English* (B-I/E) and *bilingual Bulgarian/English* (B-B/E) tasks of the multiple-language QA track at CLEF-2004. In all the three tasks 200 questions were posed to the QA systems. For the M-I task questions were posed in Italian and the answer had to be sought in an Italian text collection (the 193Mb corpus of the 1994 year of *La Stampa* newspaper and the 172Mb corpus of the 1994 and 1995 *SDA* press agency),. For the bilingual B-I/E and B-B/E tasks questions were posed in Italian, respectively Bulgarian, and the answer had to be recovered from English corpus (the 425Mb corpus of the 1994 year of *Los Angeles Times* and 157Mb corpus of the 1995 year of *Glasgow Herald*). The rest of the paper is structured as follows: Section 2 provides

high-level description of the basic components of the DIOGENE architecture. Section 3 elaborates on the use of linguistic templates for answer extraction. Section 4 describes the Web validation approach which we used this year. Section 5 discusses the results in all the tasks where we participated. Finally, we describe the future directions for system development.
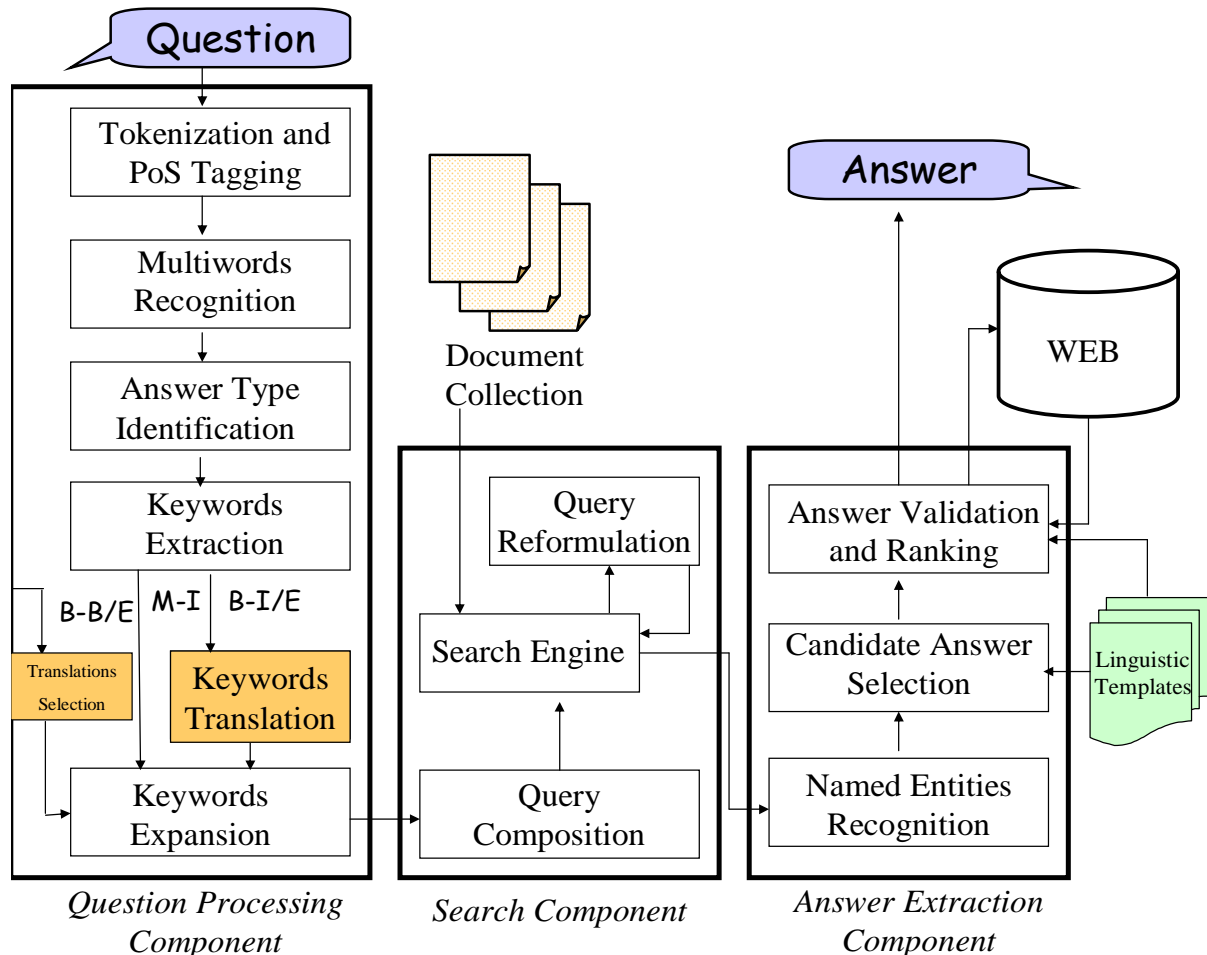

Figure1: The architecture of the DIOGENE system

2   System Architecture Overview

The overall system architecture remained the same as the architecture with which we participated in CLEF-2003 [Negri et.al.2003], apart of the linguistic templates which we plugged in the system and the new Web validation procedure which we used. DIOGENE is capable of processing questions in Italian and English and to search the answers in text collections in Italian and English. The system has multilingual architecture – the same modules work both for English and Italian, using language-specific rules and resources when necessary.   In the bililngual Bulgarian-English task (B-B/E) the questions were processed in the Bulgarian Academy of Sciences. For each question in Bulgarian, DIOGENE obtained a *question type* (what the question is about – person, location, etc.) list of keywords with all the possible translations in English. Next, DIOGENE chose the right keyword translations using a statistical approach described earlier in [Negri et.al.2003].   In both cross-language tasks after the question processing, all the other modules work in the same manner as the monolingual English version of DIOGENE (see [Magnini et.al.2002a] and [Negri et.al.2003] for more details).

2.1 Question processing

Question processing has five basic stages: question pre-processing (part-of-speech tagging, multiword recognition); *answer type identification* which defines what is the question about: person, location, date, definition, etc.; *keyword extraction* from the question; *keyword translation* for cross-language tasks; *keyword expansion* with synonyms and morphological derivations. All these components were described earlier in our CLEF-2003 system report (see [Negri et al. 2003]). Here we will just sketch the most interesting stages of the question processing.

Answer type identification. For Italian we used hand-crafted, language specific rules for answer type identification similar to the English ones. These rules operate on part-of-speech tagged questions. Semantic predicates defined over the synsets of the MultiWordNet hierarchy are used in the answer type identification. For instance, the rule described in (1) matches any question starting with "quale" ("*what*"), whose first noun, if any, is a person.

(1) **RULENAME: QUALE-CHI**
   TEST: ["quale"   [¬NOUN]* [NOUN:person-p]$_J$ +]
   OUTPUT: ["PERSON" J]

For instance, the rule matches the question *"Quale presidente americano è stato renitente alla leva?"* ("*Which American President failed to report for military service?*"*),* since the predicate *person-p* returns *true* for "*presidente*". However, the rule will not be activated on the question. *"Qual è il partito di Charles Millon?"* (*"What is the party of Charles Milton?"*), since *"partito"* (*"party"*) in MultiWordNet is not a hyponym of the concept *"person"* . Rule (1) gives on its output the type of entity which the question asks for ("*person"* in this case).

Since in MultiWordNet both English and Italian synsets are aligned, the predicates in the answer-identification rules like *person-p* can be used for both the languages.

Keyword translation. Both B-I/E and B-B/E require translation of the question in the target language (English). Since state-of-the-art translation systems are not optimized for translation of short pieces of text as the questions are, we have developed a methodology specifically designed for keyword translation. First, for each keyword from the question, all the possible English translations are found using bilingual dictionaries and MultiWordNet (for Bulgarian this has been done at the Bulgarian Academy of Sciences). Next, the most plausible combination of keyword translations is chosen. We chose the combination of keyword translations (k_1, k_2,...,k_n) with the highest frequency of co-occurrence in an English corpus (we used the AQUAINT and TIPSTER collections). The main assumption is that the more frequently a keyword translation combination appears with the translations close to each other (in one and the same paragraph) , the more plausible this combination is.

2.2. Search Component

DIOGENE at the QA tasks within CLEF-2004 relied on the same search component developed for the English version of DIOGENE, as it is described in [Magnini et al. 2002a]. Such component first combines the question keywords and their lexical expansions in a Boolean query; then performs document retrieval accessing the target document collections.

The search is performed by Managing Gigabytes (MG) [Witten et.al.1999], an open-source indexing and retrieval system for text, images, and textual images covered by a GNU public license and available from *http://www.cs.mu.oz.au/mg/*. MG allows for a fast and customizable indexing. We opted to index the document collection at the paragraph level, using the paragraph markers provided in the SGML format of the document collection. This way, although no proximity operator is implemented in MG, the paragraph index makes the "AND" Boolean operator perform proximity search. In order to divide very long paragraphs into short passages, we set 20 text lines as the limit for paragraphs' length.

The document retrieval module uses the Boolean query mode of MG. At the first step of the search phase all the basic keywords are connected in a complex "AND" clause, where the term variants (morphological derivations and synonyms) are combined in an "OR" clause. As an example, given the question *"Quando morì Lenin?"* (*"When did Lenin die?"*), the basic keywords resulting from the translation process (i.e. *"die"* and *"Lenin"*) are expanded and combined into:

[**Lenin** AND (**die** OR **dies** OR **died** OR **dying** OR **death** OR **deaths**)]

However, Boolean queries often tend to return too many or too few documents. To cope with this problem, we implemented a feedback loop which starts with a query containing all the relevant keywords and gradually simplifies it by ignoring some of them. Several heuristics are used by the algorithm. For example, a word is removed if the resulting query does not produce more than a fixed number of hits. Other heuristics consider the capitalization of the query terms, their part of speech, their position in the question, WORDNET class, etc. [Magnini et al. 2002a].

A post-processing procedure finally orders the paragraphs on the basis of the number and proximity of the keywords and their synonyms which are present.

2.3 Answer Extraction Component

Two types of questions were present in the questions set this year: *factoid questions*, which usually ask for a named entity: person, location, organization, etc., and *definition questions* (e.g. *"Who is Valentina Tereshkova?"*) which ask for person or concept definitions. For the factoid questions, the answer extraction component first performs a selection of the answer candidates through named entities recognition and linguistic templates; then, a Web-based procedure for answer validation is applied over the selected named entities to choose the best one. As for definition questions, DIOGENE extracts the answers using linguistic templates and then chooses the most plausible definition using semantic and syntactic clues.

Named Entities Recognition (NER). The named entities recognition module is in charge of identifying, within the relevant passages returned by the search engine, all the entities that match the answer type category (*e.g.* person, organization, location, measure, etc.). The Italian version of the NER module tested on a 77Kb text corpus[1] revealed a performance comparable to the English NER [Magnini et al 2002c], with an overall F-Measure score of 83%.

Linguistic templates. In order to increase the system precision we have applied linguistic templates to several question types: definition questions (both for English and Italian), dove-è (where-is) questions (only for Italian) and quale/chi-è (what/who-is) questions (only for Italian). Regarding the definition questions, the templates were the only source of information for answer extraction. The other two types of templates were applied for the respective question types, and if an answer was extracted, it was validated on the Web through answer validation templates; if no candidate was captured by the patterns, the classical DIOGENE answer extraction and validation was applied.

Answer Validation. In CLEF-2003 we used AltaVista for answer validation. Since this year AltaVista has changed its access interface and has not provided any further support of the proximity search on which we base our statistical approach, we opted for an alternative method based on the analysis of the snippets returned by AllTheWeb. Each named entity returned as a candidate answer to a factoid question was tested for close co-occurrence with the question keywords. This provides DIOGENE with clues for the plausibility of the candidate answer.

3   Linguistic Templates for Answer Extraction and Validation

Linguistic templates has proved to be an appropriate technique for certain question types such as definition and location questions. For our CLEF-2004 participation we plugged in DIOGENE linguistic templates which perform answer extraction and validation for: (i) definition questions, (ii) location questions of the type "Where is <LOCATION>?", and (iii) what-is or who-is questions of the type "(What | Who) is <NOUN PHRASE>?" (e.g. *"What is the Iraq currency?"*). For definition questions we created bi-lingual templates which work both for English and Italian. Regarding the other two classes of questions we created manually templates for Italian.

---

[1] Reference transcripts of two broadcast news shows, including a total of about 7,000 words and 322 tagged named entities, were manually produced for evaluation purposes and have been kindly provided by Marcello Federico and Vanessa Sandrini.

3.1. Definition Questions

In the CLEF-2004 QA track 10% of the questions (20) were definition questions (e.g. *"What is UNICEF?"* *"What is yakuza?"* *"Who is Jorge Amado?"*). While definition questions are among the most natural and frequent kinds of queries posed by humans, they raise specific issues in the development of the QA systems. First, the answer of a definition question is not a named entity. Next, while for most of the questions we have many content words whose co-occurrence may indicate the position of the answer, for the definition questions we have only one content word or multiword (i.e.the *focus* of the question, that is the entity for which the question asks for a definition).

CLEF-2004 QA track organizers stated in the guidelines that questions will be about persons and organizations. This makes the definition extraction more feasible, since capturing information about organization and persons is usually easier than finding definition of random concepts.

We adopted the approach described in [Tanev et al. 2004]. Our approach relies on linguistic regular expressions, much more expressive than the string templates introduced by [Ravichandran&Hovy 2002]. Since the syntactic structure of the definitions in Italian and English is not much different, we aligned the templates for English and Italian, obtaining multilingual templates. For example the following bilingual template (2) was used for capturing canonical definitions in English and Italian:

(2) **[~ Prep] <FOCUS> [~ Noun](1) [eng: lemma:be | ita: lemma:essere] [~ Prep Verb Conj](3) Noun**

This pattern captures the following sequence of words: a word which is not a preposition ([~ Prep]); followed by the focus (<FOCUS>) of the question; possibly followed by one word which is not a noun ([~ Noun](1)); followed by the auxiliary verb "be" appearing in one of the considered languages; followed by at most 3 words ([~ Prep Verb Conj](3)), none of which is a preposition, a finite verb form or a conjunction; followed by a noun (Noun). This patter captures a broad range of canonical definitions of the type *"yakuza is the Japanese mafia"* (English*), or "yakuza è la mafia giapponese"* (Italian). Correctly, it will not capture *"The members of yakudza are…"*, since no preposition is allowed before the focus.

Different templates have different levels of reliability, therefore each extracted definition obtains a syntactic score depending on the template with which it was extracted. Currently, we have manually defined the reliability score for the definition extraction templates; however, we consider learning it automatically.

Our experiments revealed that the syntactic score does not provide reliable ranking. Therefore, we applied complementary scoring strategy based on MultiWordNet (this multilingual ontology gave us possibility to work both for Italian and English):

1. If the question is about a person we search in the extracted definition a concept which is a hyponym of the concept "person" in MultiWordNet.
2. If the question is not about a person, it will be about organization, so we search for hyponyms of the concept "organization".
3. If the focus of the definition question is present in MultiWordNet, an additional score is given if terms from the gloss of the focus or its hypernyms appear in the candidate definition.

In case the focus of the definition question was not present in MultiWordNet, we searched in the Wickipedia database and assigned additional score if a term with hyperlink from the Wickipedia article (usually these hyperlinks mean that the term is important) appeared in the candidate definition.

On the definition questions we gained 40% accuracy in the monolingual Italian task and 25% on the cross-language Italian-English task. Although there is a lot of space for improvement, the definition question accuracy was higher than the accuracy of the answers of the factoid questions, especially in the monolingual Italian task.

3.2. Location questions

We have developed templates for location questions only for Italian. We noted that in many cases the answer of questions such as: *"Dove si trova la Valle dei Re?"* (*"Where is the Valley of the Kings?"*) or *"Dove si tiene il Motorshow?"* (*"Where does Motorshow take place?"*) is expressed through phrases like *"La Valle dei Re in Egitto"* (*"Valley of the Kings in Egypt"*) or *"Motorshow a Bologna"* (*"Motorshow in Bologna"*). Such answers are captured easily through superficial patterns like:

（3） **<FOCUS> (in|nel|nella) <LOCATION>**

<FOCUS> a <LOCATION>
<FOCUS> si trova in <LOCATION>
<FOCUS> ed in tutt(a|o|i) (il| l'| la | gli | i)? <LOCATION>

The extraction of the focus was carried out through specific question processing patterns.   For   each answer we count also how many times it appears in a location template; if it appears too little or many candidates are extracted, we further validate the answers querying the Web:

"<FOCUS> in <LOCATION>" OR   "<FOCUS> nel <LOCATION>" OR "<FOCUS> nella <LOCATION>"

For example: "Motorshow in Bologna" OR "Motorshow nel Bologna" OR "Motorshow nella Bologna" OR "Motorshow a Bologna"

The number of documents returned by the search engine as a response to this query together with repetition of the answer within templates in the local corpus are clues to the reliability of the answer.

3.3 Questions of the type "(what|who) is <NOUN PHRASE>"

We implemented templates for these questions only for Italian where these questions begin with "*Chi è*" or "*Qual è*" . Questions of this class do not contain any other verb except the auxiliary one. Examples of such questions from CLEF-2004 monolingual Italian test set are:

Qual è la unita di frequenza? ("What is the frequency unit?")
Qual è la capitale della Russia?("What is the capital of Russia?")
Chi è il ministro delle finanze russo?("Who is the Russian minister of finances?")

These questions are somehow opposite to the definition questions. In effect, such questions represent short definitions and the answers are entities which can be described via these definitions. Such kind of questions we call *inverted definition questions,* since if you ask a definition question about their answer, the focus of the inverted definition question will represent correct definition. For example

Question: *Chi è il ministro delle finanze russo?("Who is the Russian minister of finances?")*
Answer: Boris Fiodorov

Definition question: *Chi è Boris Fiodorov? ("Who is Boris Fiodorov?")*
Answer: *il ministro delle finanze russo ("the Russian minister of finances")*

Taking this into account, we applied patterns similar to the definition question patterns described in Section 3.1. For this question type we used also Web validation via patterns when necessary.

4   Using the Web to Validate Answers

When the named entity recognizer returns a list of candidate answers, those which are closer to the question keywords (and therefore considered more reliable) are passed to the answer validation algorithm which chooses the best candidate (if such exists).

The basic idea behind our approach to answer validation is to identify semantic relations between the question and each candidate answer by searching for their co-occurrences in a large document collection. In this framework, we consider the Web as the largest open domain text corpus containing information about almost all the different areas of the human knowledge.

In our previous participation at CLEF-2003 we used a Web validation method based on a co-occurrence statistical formula (see [Magnini et al. 2002d] and [Negri et al. 2003] for details). The frequency information used in this formula was taken from AltaVista. We used AltaVista's proximity operator "NEAR" which allowed

for identifying the number of pages in which certain words co-occur close to each other. However, AltaVista changed its interface, providing no further support for proximity searches, nor we were able to find a public available search engine which offers the same feature. Therefore, we opted for *content based answer validation*, whose main idea we described earlier in [Magnini et al. 2002b] and we used the AllTheWeb search engine (www.alltheweb.com). Our Web validation algorithm performs the following basic steps:

1. It queries the Web with the question keywords *QK* and the answer *a*. For example, for the question *"Quanti anni di prigionia ha subito Nelson Mandela?"* (*"How many years did Nelson Mandel spend in the prison?"*) and the (correct) candidate answer *27*, we have:

$QK=\{anni, prigionia, subito, Nelson, Mandela\}; a=27$

2. The top 100 hits returned by AllTheWeb are explored and for each text fragment where the answer *a* co-occurs with some of the *QK* words we calculate a score on the basis of the distance between *a* and the number of keywords present in *QK* which also appear in the snippet, according with the following formula:

$$score(snippet) = \prod_{k \in snippet \cap QK} 2^{1+|a\,k|},$$

where |a k| is the distance in tokens between the candidate answer *a* and a question keyword *k* which appears in the snippet.

For example for the text fragment:

"Nelson Mandela viene liberato dopo 27 anni di dura prigionia e di torture"

this formula assigns a score of 55.08, while for the snippet

"Nelson Mandela, che a Robben Island, l'isola-prigione a largo di Citta' del Capo dove ha trascorso 18 dei suoi 27 anni di prigionia"

we assign 43.14, since the distance between *27* and both keywords *Nelson* and *Mandela* is greater.

3. The score gained from different fragments are summed up for each candidate answer.
4. The candidate answer which gains the highest score is chosen.

DIOGENE returns as answer the candidate for which the answer validation returns the highest score. If the answer validation module returns zero for all the candidate answers, DIOGENE returns NIL as answer. After some normalization the answer validation score is returned as a confidence score as required this year by the CLEF QA track guidelines.

5 Results and Discussion

DIOGENE was evaluated in five runs: two in the monolingual Italian QA task, two in the Italian-English task, and one in the Bulgarian-English task. In each task the system had to answer to 200 questions, returning one answer per question. For some questions NIL was allowed, which is interpreted as "no answer exists to the question". The following table shows the results in all these tasks.

Run	R	W	U	X	Factoid	DEF	NIL	Overall Accuracy	Conf. Score
irst041itit	56	131	2	11	48/180(26.67%)	8/20 (40%)	6/22 (27%)	28.00%	0.156
irst042itit	**44**	**147**	**0**	**9**	**36/180(20%)**	**8/20 (40%)**	**4/6 (66.7%)**	**22.00%**	**0.107**
irst041iten	45	146	3	6	40/180(22.22%)	5/20(25%)	6/25(24%)	22.50%	0.122
irst042iten	**35**	**158**	**2**	**5**	**30/180(16.67%)**	**5/20(25%)**	**6/25(24%)**	**17.50%**	**0.075**
bgas041bgen	26	168	1	5	21/180(11.67%)	5/20(25%)	8/59(13.6%)	13.00%	0.056

Table 1: Results of DIOGENE at the CLEF-2004 QA track.

The first two rows of Table 1 show our results in the monolingual Italian task, the second two show the results from the cross-language Italian-English task, and the last row shows the results from our joint participation at the Bulgarian-English task. The first runs in both tasks (irst041itit and irst041iten) use the Web validation to calculate the final score of the answer candidates. The second runs (irst042itit and irst042iten) combine the results from the Web validation with the keyword density in the paragraph of the local text collection where the answer was found. The first column in table 1 is the run tag, the second column (R) shows the number of correct answers, the third column (W) contains the number of the wrong answers, the fourth column (U) indicates the number of unsupported answers (these are correct answers but extracted from document which do not support them), the fifth column (X) contains the number of inexact answers, the sixth column (Factoid) contains the number of the correctly answered factoid questions against all the factoid questions and the precision, the seventh column (DEF) contains the number and the precision of the correctly answered definition questions, the eighth column (NIL) contains the number and precision of the correctly returned NIL, the ninth column (Overall Accuracy) shows the main evaluation criteria - the percent of the correctly answered questions and the last column (Conf. Score) shows the value of the confidence weighted measure calculated by the CLEF judges.

The table shows that our results for this year QA tracks are lower than our CLEF-2003 results. We can explain that with the increased difficulty of this year questions. The main obstacle in front of DIOGENE were factoid questions which do not require a named entity as an answer (e.g. "*Quale animale tuba?*" ("*What animal coos?*")) .

The difference of the accuracy for monolingual Italian and Italian-English tasks is not big, which means that our translation mechanism works satisfactory.

Results show that pure Web based validation works 5-6% better than the combination with the keyword density.

Overall accuracy in the Bulgarian-English task was 10% under our best run on the Italian-English test set. However, accuracy on definition questions is equal for both cross-language tasks. Taking into account that this is one of the first QA experiments for Bulgarian (only one similar work has been currently reported - [Tanev 2003]) and the difficulty of this year questions, results are encouraging.

Definition question accuracy is better than the accuracy on the factoid questions. In particular, the results on the definition questions in the monolingual Italian task (40%) are much higher than the corresponding factoid accuracy (26.67% in the best run). There is a significant difference in the precision on the definition questions between monolingual and cross-language runs (monolingual run was 15% better). We did not study thoroughly the reasons for this difference.

6   Conclusion and Future Work

We described the architecture of our multilingual system DIOGENE, stressing on the improvements for CLEF-2004. The system combines the multilingual knowledge in MultiWordNet, a set of linguistic templates and information mined on the fly from the Web.

Our best run was on the monolingual Italian task (28% overall accuracy); on the definition questions we have 40% accuracy in the same task. While these figures leave a lot of space for improvement, they demonstrate the feasibility of our multilingual approach. In particular, our results on the definition questions show that putting together the knowledge encoded in an ontology and a set of linguistic templates can be an efficient strategy which should be studied further.

In the future development of our QA system we intend to strengthen the linguistic infrastructure for the Italian language by trying to integrate parsing and by using more extensively MultiWordNet. We would like DIOGENE to answer also to questions which do not have named entity as answers. We intend to create more templates which cover broader range of questions and to evaluate their performance.

Finally, we would like to extend the applicability of our QA system into the multilingual dimension by considering other languages.

References

Burger, J., Cardie, C., Chaudhri, V., Gaizauskas, R., Harabagiu, S., Israel, D., Jacquemin, C., Lin, C.-Y., Maiorano, S., Miller, G., Moldovan, D., Ogden, B., Prager, J., Riloff, E., Singhal, A., Shrihari, R., Strzalkowski, T., Voorhees, E., Weishedel, R.: Issues Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A), 2001.
URL: *http://www-nlpir.nist.gov/projects/duc/papers/qa.Roadmap-paper_v2.doc.*

Magnini, B., Negri, M., Prevete, R., Tanev, H.: Mining Knowledge from repeated Co-occurrences: DIOGENE at TREC-2002 Proceedings of the Eleventh Text Retrieval Conference (TREC-2002), Gaithersburg, MD., 2002a.

Magnini, B., Negri, M., Prevete, R., Tanev, H.: Comparing Statistical and Content-Based Techniques for Answer Validation on the Web. Proceedings of the VIII Convegno AI*IA, Siena, Italy, 2002b.

Magnini, B., Negri, M., Prevete, R., Tanev, H.: A WORDNET-Based Approach to Named Entities Recognition. Proceedings of SemaNet02, COLING Workshop on Building and Using Semantic Networks, Taipei, Taiwan, 2002c.

Magnini, B., Negri, M., Prevete, R., Tanev, H.: Is It the Right Answer? Exploiting Web Redundancy for Answer Validation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002), Philadelphia, PA., 2002d.

Manning, C., Shutze, H.: Foundations of Statistical Natural Language Processing. MIT Press, 1999.
Negri M., Tanev H., and Magnini B.: Bridging Languages for Question Answering: DIOGENE at CLEF-2003 Proceedings of CLEF-2003, Trondheim, Norway, August 2003.

Pianta, E., Bentivogli, L., Girardi, C.: MULTIWORDNET: Developing an Aligned Multilingual Database. Proceedings of the 1st International Global WordNet Conference, Mysore, India, 2002.

Ravichandran D., Hovy, E.: Learning Surface Text Patterns for a Question Answering System. In Proceedings of the 40th ACL Conference. University of Pennsylvania, Philadelphia, 2002

Tanev H.: Socrates - a Question Answering Prototype for Bulgarian, Proceedings of RANLP-2003, Borovets, Bulgaria, September, 2003.

Tanev H., Kouylekov M., Negri M., Coppola B., Magnini B.:Multilingual Pattern Libraries for Question Answering: a Case Study for Definition Questions, Fourth International Conference on Language Resources and Evaluation (LREC-2004) Proceedings, Lisbon, Portugal, May 26-28, 2004.

Voorhees, E.: Overview of the TREC 2003 Question Answering Track Proceedings of the Sixth Retrieval Conference (TREC-2003) , Gaithersburg, MD., 2004.

Witten, I. H., Moffat, A., Bell T.: Managing Gigabytes: Compressing and Indexing Documents and Images (second ed.), Morgan Kaufmann Publishers, New York, 1999.