Question-answer matching: two complementary methods

K. Lavenus, J. Grivolla, L. Gillard, P. Bellot

Laboratoire d'Informatique d'Avignon (LIA)
339 ch. des Meinajaries, BP 1228
F-84911 Avignon Cedex 9 (France)
{karine.lavenus, jens.grivolla, laurent.gillard, patrice.bellot }@lia.univ-avignon.fr

Abstract

This paper presents different ways at different steps of question answering process to improve question answer match. First we discuss about the role and the importance of question categorization to guide the pairing. In order to process linguistic criteria, we describe a question pattern based categorization. Then we propose a statistical method and a linguistic method to enhance the pairing probability. The statistical method aims to modify weights of keywords and expansions within the classical Information Retrieval (IR) vector space model whereas the linguistic method is based on answer pattern matching.

Keywords

Question answering systems, categorization, pairing, pattern-matching.

1 Question categorization in TREC Q&A systems

1.1 The Question Answering tracks

The Natural Language Processing community began to evaluate Question Answering (Q&A) systems during the TREC-8 campaign (Voorhees: 2000) that started in 1999. The main purpose was to move from document retrieval to information retrieval. The challenge was to obtain 250-byte document chunks containing answers to some given questions from a given document collection. The questions were generally fact-based. In TREC-9, the required chunk size was reduced to 50 bytes (Voorhees: 2001) and, in TREC-11, systems had to provide the exact answer (Voorhees: 2003). The TREC-10 campaign introduced questions whose answers were scattered across multiple documents and questions without answer in the document collection. For the more recent campaigns, questions were selected from MSN and AskJeeves search logs without looking at any documents. The document set contained articles from the Wall Street Journal, the San Jose Mercury News, the Financial Times and the Los Angeles Times and newswires from Associated Press and the Foreign Broadcast Information Service. This set contains more than 900,000 articles in 3 Go of text and covers a wide spectrum of topics (Voorhees: 2002).

1.2 Question categorizers

A classical Q&A system is composed of several components: a question analyzer and a question categorizer, a document retrieval software that retrieves candidate documents (or passages)

according to a query (the query is automatically derived from the question), a fine-grained document analyzer (parsers, named-entity extractors, …) that produces candidate answers and a decision process that selects and ranks these candidate answers.
Most of TREC Q&A question categorizers take natural questions as input to produce answer categories used by an entity extraction component. However, the expected answer may not be a named entity but a specific pattern. This kind of answer must be taken into account by the categorizer: a particular question category is frequently defined. Consequently, question categories strongly depend on the named-entity set of the extraction component employed to tag the documents of the collection. Depending on the system, several entity sets were employed. IBM's 2002 Q&A system (Ittycheriah & Roukos: 2003) subdivides entity tags along five main classes: Name Expressions (person, organization, location, country…), Time Expressions (date, time…), Number Expressions (percent, money, ordinal, age, duration…), Earth Entities (weather, plants, animals, …) and Human Entities (events, diseases, company-roles, …). Some other participants defined a larger set: 50 semantic classes for Univ. of Illinois (Roth *et al.*: 2003), 54 for Univ. of Colorado and Columbia Univ. (Pradhan *et al.*: 2003). G. Attardi *et al.* employed 7 general categories (person, organization, location, time-date, quantity, quoted, language) and some specific ones gathered from WordNet's taxonomy (Attardi *et al.*: 2003). Clarke *et al.* matched questions to 48 categories, many standards in Q&A systems (date, city, temperature…), a few inspired by TREC questions (airport, season…), and two (conversion and quantity) parameterized by required units (Clarke *et al.*: 2003). Li and Roth proposed a semantic classification of questions in 6 coarse classes and 50 fine classes and show the distribution of these classes in the 500 questions of TREC-10 (Li & Roth: 2002).

In order to categorize questions, most of participants developed question patterns based on the TREC collection of questions and employed a tokenizer, a part-of-speech tagger and a noun-phrase chunker. In our case (Bellot *et al.*: 2003), we decided to define a hierarchical set of tags according to a manual analysis of the previous TREC questions. The hierarchy was composed of 31 main categories (acronym, address, phone, url, profession, time, animal, color, proper noun, location, organization…), 58 sub-categories and 24 sub-sub-categories. For example, "Proper Noun" has been subdivided in 10 sub-categories (actor/actress, chairman, musician, politician…) and politician in some sub-sub-categories (president, prime minister…). For categorizing new questions, we developed a rule-based tagger and employed a probabilistic tagger based on supervised decision trees for the question patterns that did not correspond to any rule. The main input of the rule-based tagger was a set of 156 manually built regular expressions that did not pretend to be exhaustive since they were based on previous TREC questions only. Among the 500 TREC-11 questions, 277 questions were tagged upon theses rules. The probabilistic tagger we employed was based on the proper names extractor presented during ACL-2000 (Béchet *et al.*: 2000). This module used a supervised learning method to automatically select the most distinctive features (sequence of words, POS tags…) of question phrases embedding named entities of several semantic classes. The result of the learning process is a semantic classification tree (Kuhn & De Mori: 1996) that is employed to tag a new question. By using a subset of 259 manually tagged TREC-10 questions only as learning set, we obtained a 68.5% precision level for the missing 150 TREC-10 questions. This experiment allows to confirm that the combination of a small set of manually and quickly built patterns and a probabilistic tagger gives very good categorization results (80% precision with several dozens of categories) even if an extensive rule-based categorizer may perform even better (Yang & Chua: 2003). Sutcliffe writes that a simple ad-hoc keyword-based heuristics allowed his system to correctly classify 425 of the 500 TREC-

11 questions among 20 classes (Sutcliffe: 2003). The Q&A system QUANTUM (Plamondon *et al.*: 2003) employed 40 patterns to correctly classify 88% of the 492 TREC-10 questions among 11 function classes (a *function* allows to determine what criteria a group of words should satisfy to constitute a candidate valid answer). They added 20 patterns for the TREC-11 evaluation. Last but not least, the MITRE corporation's system Qanda annotates question with part-of-speech and named entities before mapping question words in an ontology of several thousands words and phrases (Burger *et al.*: 2003).

1.3 Several categories for several strategies

Some question categorizers aim to find both the expected answer type and the strategy to follow for answering the question. The question categorizer employed in the JAVELIN Q&A system (Nyberg *et al.*: 2003) produces a question type and an answer type based on (Lehnert: 1978) and (Graesser *et al*: 1992). The question type is used to select the answering strategy and the answer type specifies the semantic category of the expected answer. For example, the question type of the questions "*Who invented the paper clip*" and "*What did Vasco da Gama discover*" is "event-completion" whereas the answer types are "proper-name" for the first question and "object" for the second one. The LIMSI's Q&A system QALC determines whether the answer type corresponds to one or several named entities and the question category helps to find an answer in a candidate phrase: the question category is the "form" of the question (Ferret *et al.*, 2002). For the question "When was Rosa Park born", the question category is "WhenBePNBorn".
Finally the type of the question may be very helpful for generating the query and retrieving candidate documents (Pradhan *et al.*: 2003). For example, if the answer type of a question is "length", the query generated from the question may contain the words "*miles, kilometers*". A set of words may be associated to each answer type and be candidates for query expansion.

1.4 Wendy Lehnert's categorization

Wendy Lehnert's question categorization (Lehnert, 1978) groups together questions under 13 conceptual categories. This categorization inspired the TREC organizers to create their own set of test questions (Burger *et al.*: 2000, p. 34).
However, this categorization reflects only partially the type of questions asked within the Q&A framework. Indeed, some types of question found in TREC are not included in Wendy Lehnert's categorization: "why famous person" questions and questions asked in order to find out about an appellation or a definition[1], or the functionality of an object.
Besides, Lehnert's categories could have been defined differently. Thus, the "concept completion" category (What did John eat?; Who gave Mary the book?; When did John leave Paris?) may be divided into different categories according to the interrogative pronoun and the target[2] of the question: food, person name, date. Actually this categorization corresponds to the application it has been made for. Within the framework of artificial intelligence research, Lehnert proposed a Q&A system called QUALM in 1978, in order to test story comprehension. This context explains the existence of the "disjunctive" category (Is John coming or going?) and the importance given to questions about cause or goal. Besides, examples about cause or goal (4 categories: "causal antecedent", "goal orientation", "causal consequent", "expectational") sometimes seem irrelevant, because the difference between cause and goal, cause and manner, cause and consequence may be slight. The application context does not justify the existence of

[1] This type of question is nevertheless present in Graesser's categorization (Burger, 2000: 35), which can be considered as an enriched categorization with 18 categories.
[2] We define the target as the clue that indicates the kind of answer expected.

the “request” category in any case, as the performative aspect can not be realized. In the TREC competition, questions about causes are factual in order to be easily assessed. It is also why the “judgmental” category (What should John do now?) has disappeared.
Finally, Lehnert’s yes/no question categories have been deleted from TREC: “verification” (Did John leave?) and “request” (Would you pass the salt?) which implies an action as well.
We already have an idea of the importance of the role played by categorization in the Q&A frame. Let’s see precisely in section 2 why categorization is crucial to retrieve a good answer, and how we can refine it. Then in section 3, we will describe how question answer matching can be improved thanks to statistical and linguistic methods.

2 Our categorization

2.1 Role and importance of categorization

Question answering (Q&A) systems are based on Information Retrieval (IR) techniques. This means that the *question* asked by the user is transformed into a *query* from the very beginning of the process. Thus, the finest nuances are ignored by the search engine which usually :

1) transforms the question into a « bag of words » and therefore loses meaningful syntactical and hierarchical information;
2) lemmatizes the words of the query, which deletes information about time and mode, gender (in French) and number (singular *vs* plural);
3) eliminates “stop words” although they may be significant.

However, if the user has got the opportunity to ask a question thanks to a Q&A system, it is not only to obtain a concise answer but also to express a complete and precise question. But when the question is transformed into a bag of words, a lot of information is lost. For instance, the question How much folic acid should an expectant mother get daily ? [3](203), becomes: folic + acid + expectant + mother + get + daily when transformed into a query. Even if there are six terms, it is not enough to know what the user is seeking exactly. Thus, the Google search engine retrieves documents about the concerned topic, in the top results, without giving any information about the daily quantity to absorb. The answer *400 micrograms,* introduced by « get », is found in the fifth document of the first results page. To obtain this snippet from the very beginning of the process, it is necessary to indicate to the system that we are looking for a quantity. It is precisely what categorization can do.
As stop words do appear on many occasions, they are considered less significant than other words and are not taken into account by search engines. However, stop words play an important role in Q&A. First, their meaning can be useful during the categorization phase. Secondly, they can help locate the answer during the extraction phase. In this case, stop words must be kept in the query. For example, the question How far away is the moon ? (206) could become a one-keyword query: moon. It is difficult from this simple query, without any other information, to find an answer to question 206 in a document collection. In order to find the right answer, we need to add information about the answer type. For question 206, we could mention that we are looking for a distance: the distance which exists between the Earth (implicit data which needs to be made explicit!) and the moon. Six of the eight different answers given by TREC-9 competitors contain the stop word “away” . One contains the stop word “farther”, a derivative of

[3] From the 3rd section, questions quoted in this paper are from the TREC-9 test questions collection.

"far". In five answers out of eight, the stop word "away" is located just after the closing tag which encloses the exact answer (</AN>)[4]. Therefore, we can consider that it is possible to retrieve relevant passages and to locate the exact answer thanks to the stop word "away" .
Subtleties that can not be processed by a search engine when the question is transformed into a query must be taken into account during the categorization of the question. Based on the content of the question, this step allows to group information about the answer type and characteristics, before the pruning involved by the transformation of the question into a query.

To categorize questions, we have grouped together questions with common characteristics which concern - in the Q&A frame - the type or nature of the sought answer. The question type can be inferred in many cases. For instance, we assume that for questions called "why famous Person" like : Who is Desmond Tutu ? (287), we are looking for the job, function, actions or events in relation with the person mentioned.
Questions are mainly categorized according to the semantic type of the answer, which does not depend exclusively on the interrogative pronoun or on the question's syntax. Questions that begin with the same interrogative pronoun can belong to different categories such as questions beginning with "who". Sometimes we want to know why somebody is famous: Who is Desmond Tutu ? (287), which is equivalent to Why is Desmond Tutu famous? And sometimes we want to know the name of someone specific (which is, in a way, the opposite of the previous category): Who is the richest person in the world? (294), which is equivalent to What is the name of the richest person in the world?
As we can see, the single interrogative pronoun does not allow to detect the question type. Thus, the automatic learning of lexical syntactic patterns associated with question categories could be efficient (see section 3.2.4).

2.2 Linguistic criteria for categorization

2.2.1 Target and question categorization

As mentioned before, our categorization is mainly semantic and based on answer type. Thus, in order to know the answer type and to categorize a question, we need to detect the *target*, which is an interrogative pronoun or/and a word which represents the answer (i.e. is a kind of substitute). The target is printed in bold in the following examples :
1) **Name** a Salt Lake city **newspaper**. (745)
2) **Where** is Trinidad? (368)
« Name » indicates that we are looking for a name, and *serves as a variable* for the newspaper's name it stands for. In the same way, "Where" indicates that we are looking for a location and *serves as a variable* for this location.
Based on the target detection of a sample of the 693 TREC-9 questions, we have found six different categories, which are more or less important: named entities (459 questions); entities (105); definitions (63); explanations (61); actions (3); others (2). By "entities" we mean answers that can be extracted like "named entities". But as they do not correspond to proper names, they do not belong to this category. However, entities can be sub-categorized and grouped under general concepts (like animals, vegetables, weapons, etc.). Sekine [2002] includes them in his hierarchical representation of possible answer types.

[4] Answers given by the TREC-9 competitors can reach 250 bytes. In these chunks, we used regular expressions - provided by the organizers- to tag the exact answers.

2.2.2 Target and clues for answer retrieval

Here are several questions from the “entities” category. All these questions can be represented by the same pattern. The target of the question (in bold) matches with the direct object (NP2) introduced by the interrogative pronoun “what”.

Sem. Type	target	**Q-A** Link	**Pattern of the questions**	**Questions**
Entity	Sport **NP2**	hypo	What **NP2** aux *NP1* V?	What **sport** do the *Cleveland Cavaliers* play?
Entity	Animal **NP2**	hypo	What **NP2** aux *NP1* V?	What **animal** do *buffalo wings* come from?
Entity	Instrument **NP2**	hypo	What **NP2** aux *NP1* V?	What **instrument** does *Ray Charles* play?

Table 1: Question categories, question patterns and Q&A link

In the “Q-A link” column, we can see that the answer is the hyponym of a target. For example, in the case of the first question, if the system finds a hyponym for “sport” near the focus “Cleveland cavaliers” in a document, this hyponym may constitute the answer.

For many of the questions seeking a location, it is possible to find or to check the answer using a Named Entity tagger and WordNet. Depending on the pattern of the question and the syntactic role of the selected terms (target or focus), the answer will be a holonym or a meronym. For example, “What province is Edmonton located in?”: first the answer can be a holonym for “Edmonton”, and secondly a meronym for “province”.

Most of the links useful to answer this kind of questions are available in WordNet. Here are some examples of these links extracted from the TREC-9 corpus of questions and exact answers:

- Synonymy: Aspartame is known by what other name? (707): < AN>*NutraSweet*</AN>. Sometimes the user seeks a synonym which belongs to another language level: What's the formal name for Lou Gehrig's disease? (414): <AN>*amyotrophic lateral sclerosis*</AN>.
- Hyponymy: Which type of soda has the greatest amount of caffeine? (756): <AN>*Jolt*</AN>: Jolt can be considered as a “soda” hyponym.
- Hyperonymy: A corgi is a kind of what? (371): <AN>*Dogs*</AN>.
- Holonymy: Where is Ocho Rios? (698): <AN>*Jamaica*</AN>
- Meronymy: What ocean did the Titanic sink in? (375): <AN>*Atlantic*</AN>
- Antonymy: Name the Islamic counterpart to the Red Cross. (832): <AN>*Red Crescent*</AN>
- Acronym, abbreviation: What is the abbreviation for Original Equipment Manufacturer? (446): <AN>*OEM*</AN>.Conversely, it is also possible to obtain the spread form of an

acronym: What do the initials CPR stand for? (782): <AN>*cardiopulmonary resuscitation*</AN> : both are available with Wordnet in most of the cases.

Some other links are not directly available in Wordnet but may be found in the gloss part:
- Nickname: What is the state nickname of Mississippi? (404): <AN>*Magnolia*</AN>
- Definition: What is ouzo? (644): <AN>*Greek liqueur*</AN>
- Translation: What is the English meaning of caliente? (864): <AN>*Hot*</AN>

Finally, information can be added to our semantic question categorization. Depending on the question's semantic type and pattern, we can orient the search for the answer using semantic links relating a keyword to a potential answer. In order to locate and delimit the answer more precisely, we can use other information elements: some "details" generally ignored by search engines when they automatically transform the question into a query. These shades of meaning concern the number of answers (requested number; possible number); ordinal and superlative adjectives and modals.

2.2.3 Taking shades of meaning into account

Sometimes the user seeks a lot of information in one question. For example, the answer to the question What were the names of the three ships used by Columbus? (388) must include three different names of ships.

Many different but valid answers can also be given to questions using an indefinite determiner: Name a female figure skater. (567). When the confidence weighted score is calculated, this fact can be taken into account, as answers looking very different can yet be validated.

Some questions restrict the potential answers to a small sample: Name one of the major gods of Hinduism? (237). The answer must be composed of the name of one of the *major* gods: Brahma; Vishnu; Shiva. Therefore, many answers can be accepted as long as they respect the restriction printed in italic.

In the same way, ordinal and superlative adjectives used in the question show that the user is seeking a precise answer: Who was the *first* woman in space? (605). The name of a woman sent in space will not satisfy the user as he needs the name of the *first* woman in space. It is the same for the question What state has the most Indians ? (208): the user expects a precise answer, the name of the (American) state which comprises the highest number of Indians.

Lastly, modals have to be taken into account. In the case of How large is Missouri's population? (277), the user needs an up-to-date number. This can seem trivial, but numbers concerning the beginning of the XXth century will not interest him. In the example: Where do lobster like to live ? (258), the user wants to know where lobster *like* to live, which does not mean that they actually live there. In order to answer correctly, a Q&A system must detect these shades of meaning and manage them.

2.2.4 Creation of question patterns

If we want to place a question in the appropriate category and possibly disambiguate it, we need to create patterns which also represent shades of meaning. First we tried to factorize (i.e. we have not developed elements like noun phrases, which can be separately rewritten). But we have realized that it is necessary to keep some relevant and discriminating features , if we want to put the question in the right category. For example, the pattern "What be PN" is not subtle enough: it matches Definition question: What is a nematode? (354), Entity question: What is

California's state bird? (254); Named Entity question: What is California's capital? (324) and Entity question containing nuances: What is the *longest* English word? (810).
Moreover, in order to distinguish between similar structured questions which belong to different categories, we need to include lemma or words in the pattern of the question. These words are interchangeable insofar as they belong to the same paradigm, which limits the number of patterns. For example, the pattern: What be [another name| a synonym| the (adj) term | noun] for GN ? can match with these questions: What is the collective noun for geese?; What is the collective term for geese?; What is a synonym for aspartame?; What is another name for nearsightedness?; What's another name for aspartame?; What is the term for a group of geese?

Thus, a balance must be found between a global, abstract and a sharp representation of the question, which would be too precise to be reused in order to automatically categorize new questions.

	A	B	C	D	E	F
1	IDQ	QUESTION	Question Pattern	Cat.	Sub-cat.	Link Q-A
2	201	What was the name of the first Russian astronaut to do a spacewalk	What be the name of S NPp1 Vinf NP2?	SNE	person	
3	202	Where is Belize located?	Where aux NPLocation pp ?	NE	location	holonym
4	203	How much folic acid should an expectant mother get daily?	How much NP2 aux NPp1 V adv ?	NE	quantity	
5	204	What type of bridge is the Golden Gate Bridge?	What kind|type of NP be NP ?	DEF		hyponym
6	205	What is the population of the Bahamas?	What be the population of NPlocation?	NE	population	
7	206	How far away is the moon?	How far away be NP1?	NE	length	
8	207	What is Francis Scott Key best known for?	What aux NPP known for?	EXP	WFP	
9	208	What state has the most Indians?	What state have S NPp2 ?	SNE	state	
10	209	Who invented the paper clip?	Who V NP2 ?	NE	person	
11	210	How many dogs pull a sled in the Iditarod?	How many NP1 V NP2 ?	NE	number	
12	211	Where did bocci originate?	Where aux NP1 V ?	NE	location	
13	212	Who invented the electric guitar?	Who V NP2 ?	NE	person	
14	213	Name a flying mammal.	Name a NP	E	animal	hyponym
15	214	How many hexagons are on a soccer ball?	How many NP be (CCL|there)?	NE	number	
16	215	Who is the leader of India?	Who be NPp NPprep?	NE	person	
17	216	What is the primary language of the Philippines?	What be S NP ?	SE	language	hyponym
18	217	What is the habitat of the chickadee?	What be the N NPprepOrg ?	E	habitat	hyponym
19	218	Who was Whitcomb Judson?	Who be NPP?	EXP	WFP	
20	219	What is the population of Japan?	What be the population of NPlocation?	NE	population	
21	220	Who is the prime minister of Australia?	Who be NNp NPprep?	NE	person	
22	221	Who killed Martin Luther King?	Who V NP2 ?	NE	person	
23	222	Who is Anubis?	Who be NPP?	EXP	WFP	
24	223	Where's Montenegro?	Where be NPlocation ?	NE	location	holonym
25	224	What does laser stand for?	What aux (the initials| acronym| abbreviation|	DEF		spread form

CatégorisationETpatrons

Table 2: question patterns and categorization (sample)

The tag NP1 represents a Noun Phrase Subject, NP2 a Noun Phrase Object, NPprep a Noun Phrase introduced by a preposition, NPP a Noun Phrase which represents a Person name.
We can see that some terms are not tagged: What be the population of … ? In fact, as "population" represents the target and associates the question to Named Entity Number answer, we need to keep this word in order to categorize the question efficiently.
In the same way, specific features like superlatives are mentioned by the letter "S": What state have **S** NPp2 ? for What state has **the most** Indians ? (208). In order to locate these specific terms, we can tag lexical clues like "most" or spot "er" or "est" suffixes added to an adjective, or create exceptions lists.

Noun Phrases (NP) representing people are mentioned by NPp, which often corresponds to a function, a nationality or a profession: What state have S **NPp**2 ? for What state has the most **Indians** ? (208). This tag is useful to know that we are looking for a Person Named Entity. For example, if we know that "astronaut" refers to a person in What was the name of the first Russian *astronaut* ?, we can infer that we are looking for a person's name (vs What was the name of the first *car* ?).
Locating Named Entities *in the question* can be useful, in particular when the question is about the location of a place (see section 3.2.2). Depending on the syntax of the question and on the NP considered, we can find or check the answer searching for a meronym or a holonym in Wordnet.
Answers to questions containing the pattern « what kind | type | sort » can also be hyponyms of the term introduced by this pattern.

3 Pairing: statistical and linguistic criteria

3.1 Keywords and expansions to select

As information retrieval models have been created in order to find documents about a topic – which is very different from finding a concise answer to a precise question – we thought it would be interesting to modify the classical IR vector space model, in order to adapt it to Q&A systems. Taking into account the syntactic role of question words, the kind of keyword expansion and the question type, we could attribute different weights to the words of the question.

3.1.1 Keywords

To carry out this study, we have first automatically transformed each POS tagged TREC-9 question into a query: we have kept only nouns, proper nouns, adjectives, verbs, and adverbs.
Then, we have automatically sought the keywords and their expansions (given by WordNet 2.0) in the TREC-9 250 bytes *valid* answers corpus. First this has allowed us to know which keyword is near an answer in the strict sense (between tags <AN>), and how often. A complementary study will indicate if the number of occurrences can be related with the syntactic role of the keyword in the question and with the type of the question.

We can see in table 3 that we obtained 2425 keywords for the 693 TREC-9 questions (3,49 keywords per question). As we have considered the verbs « to be » and « to have » as stop words, only 307 verbs remain for 693 questions (13,48 % of the keywords). Question keywords are mainly composed of nouns (39,83%), proper nouns (33,65%) and adjectives (9,65%) (in bold), which is not surprising. But if we have a look at the keyword distribution within the answers, we can see that the number of proper nouns improves (58,32 %) as the number of nouns (30,41 %), adjectives (6,09%) and verbs (4,45 %) and other categories sinks. It confirms that proper nouns are good criteria to find the exact answer. So questions containing this kind of terms may be easier to process.

Keyword distribution within questions			Keyword distribution within answers		
tag	number		tag	number	percentage
CD	20		CD	79	0,44%
JJ	**208**		**JJ**	**934**	**5,22%**
JJS	**26**		**JJS**	**156**	**0,87%**
NN	**844**		**NN**	**4685**	**26,19%**
NNS	**122**		**NNS**	**755**	**4,22%**
NP	**808**		**NP**	**10395**	**58,11%**
NPS	**8**		**NPS**	**37**	**0,21%**
RB	36		RB	130	0,73%
RBS	2		RBS	1	0,01%
VV	72		VV	154	0,86%
VVD	96		VVD	187	1,05%
VVG	16		VVG	28	0,16%
VVN	94		VVN	277	1,55%
VVP	45		VVP	91	0,51%
VVZ	28	1,15%	VVZ	57	0,32%
	2425	100,00%		**17887**	100,00%

Table 3: Keyword tag distribution within questions and answers

KW distr. before exact answer			KW distr. within exact answer			KW distr. after exact answer		
tag	number	percentage	tag	number	percentage	tag	number	percentage
CD	30	0,37%	CD	1	0,11%	CD	48	0,54%
JJ	**391**	**4,84%**	**JJ**	**54**	**5,82%**	**JJ**	**489**	**5,45%**
JJS	54	0,67%	JJS	6	0,65%	JJS	96	1,07%
NN	**2017**	**24,98%**	**NN**	**175**	**18,86%**	**NN**	**2493**	**27,81%**
NNS	245	3,03%	NNS	63	6,79%	NNS	447	4,99%
NP	**4942**	**61,22%**	**NP**	**602**	**64,87%**	**NP**	**4851**	**54,11%**
NPS	17	0,21%	NPS	0	0,00%	NPS	20	0,22%
RB	53	0,66%	RB	1	0,11%	RB	76	0,85%
RBS	0	0,00%	RBS	0	0,00%	RBS	1	0,01%
VV	69	0,85%	VV	16	1,72%	VV	69	0,77%
VVD	59	0,73%	VVD	0	0,00%	VVD	128	1,43%
VVG	9	0,11%	VVG	2	0,22%	VVG	17	0,19%
VVN	131	1,62%	VVN	5	0,54%	VVN	141	1,57%
VVP	33	0,41%	VVP	3	0,32%	VVP	55	0,61%
VVZ	23	0,28%	VVZ	0	0,00%	VVZ	34	0,38%
	8073	100,00%		**928**	100,00%		**8965**	100,00%

Table 4: Keyword tag distribution before, within and after the <AN> tag which indicates the exact answer

First, we can see in table 4 that most of the keywords stand mainly before (44,93%) or after (49,89%) the exact answer which contains only 5,16% of question keywords.
Whereas the percentage of adjectives found in the different parts of the answer is stable, there are more nouns before and mostly after the answer than within. At the opposite, proper nouns are more numerous within and before the answer than after.

tag	-1 ; 1 distribution	percentage	**0 distribution**	percentage of 0
NP	120	36,59%	73	15,53%
NN	117	35,67%	41	8,72%
NNS	22	6,71%	12	2,55%
JJ	21	6,40%	5	1,06%
VVD	13	3,96%	0	0,00%
VVN	10	3,05%	2	0,43%
VV	6	1,83%	1	0,21%
VVP	5	1,52%	2	0,43%
RB	5	1,52%	1	0,21%
VVG	4	1,22%	2	0,43%
VVZ	3	0,91%	0	0,00%
JJS	1	0,30%	2	0,43%
CD	1	0,30%	1	0,21%
	328	100,00%	142	30,21%

Table 5: tag distribution around (-1; 1) and within (0) exact answers

Proper nouns seem to introduce or to follow the exact answer (36,59%) but in fact nouns are more numerous to stand just before or just after the exact answer (NN + NNS = 42,38 %). And verbs, with a sum of 12,49 %, are much more present before and after than in the whole answers (4,45 %). 30,21 % of the keywords that are very near from the answer are contained *in* the <AN> tags and are mainly composed of proper nouns (15,53%) and nouns (11,27%).

tag	before	within	after	mean of the number of occurrences
CD	40,74%	1,67%	57,59%	6,58
JJ	47,26%	1,99%	50,75%	7,13
JJS	40,10%	4,97%	54,93%	8,21
NN	45,25%	4,22%	50,53%	8,25
NNS	32,41%	12,63%	54,96%	8,88
NP	50,27%	4,72%	45,01%	14,46
NPS	36,59%	0,00%	63,41%	9,25
RB	45,73%	1,04%	53,23%	5,42
RBS	0,00%	0,00%	100,00%	1,00
VV	50,17%	3,38%	46,46%	5,13
VVD	36,53%	0,00%	63,47%	3,82
VVG	27,99%	11,11%	60,90%	3,11
VVN	49,27%	3,13%	47,60%	5,54
VVP	44,59%	2,22%	53,18%	3,37
VVZ	35,93%	0,00%	64,07%	6,33

Table 6: Tag distribution in answers

For each keyword tag, we have estimated the number of times it appears before, within and after the exact answer. The length of the exact answer is less important than the chunks that stand before or after the exact answer. That is why numbers in the "within" column are weak. But some lack of balance occurs sometimes between "before" and "after" distributions: keywords appear mainly after the exact answer (see CD, NNS, NPS, VVD, VVG, VVZ). This table also shows the mean of keyword tag occurrences in the answer set for a given question.

3.1.2 Keywords expansions

Secondly, we have counted the expansions occurrences in the answer snippets. The WordNet relations tags were simplified in order to see if we can find synonyms, hyperonyms, hyponyms, holonyms and meronyms NEAR and WITHIN the answer in the strict sense (as far as we assume that a holonym can answer a Location Type question, see section 3.2.2). Further results will present the most important links that should be taken into account according to the question type (making the difference between expansions that introduce the answer and expansions that constitute the answer).

Expansion tag	number	percentage
Also see	19	0,16%
attribute	23	0,19%
cause	82	0,69%
entailment	204	1,71%
holonym	1240	10,41%
hyperonym	2524	21,19%
hyponym	3426	28,76%
meronym	1426	11,97%
Similar to	86	0,72%
synonym	2136	17,93%
Verb group	745	6,25%
SUM	**11911**	**100,00%**

Table 7: Expansion distribution

Hyponyms are the most frequent kind of expansions, followed by hyperonyms, synonyms, meronyms and holonyms. But it is more interesting to have a look at these expansion distributions according to their position in the answer:

Expansion tag	number BEFORE	percentage	tag	number WITHIN	Percent.	tag	number AFTER	percentage
Also see	9	0,18%	Also see	3	0,19%	Also see	7	0,13%
attribute	10	0,20%	attribute	2	0,12%	attribute	11	0,21%
cause	29	0,59%	cause	2	0,12%	cause	51	0,95%
entailment	106	2,14%	entailment	1	0,06%	entailment	97	1,81%
holonym	415	8,39%	holonym	356	**22,10%**	holonym	469	8,76%
hyperonym	1020	20,63%	hyperonym	297	18,44%	hyperonym	1207	22,54%
hyponym	1391	28,13%	hyponym	488	30,29%	hyponym	1547	28,89%
meronym	557	11,26%	meronym	223	13,84%	meronym	646	12,06%
Similar to	39	0,79%	Similar to	1	0,06%	Similar to	46	0,86%
synonym	1020	**20,63%**	synonym	217	13,47%	synonym	899	16,79%
Verb group	349	7,06%	Verb group	21	**1,30%**	Verb group	375	7,00%
	4945	**100,00%**		**1611**	**100,00%**		**5355**	**100,00%**

Table 8: Expansion distribution before, within and after the answer

41,51% of the expansions stand before the exact answer, 13,52% within and 44.95% after. Holonyms are mainly within the answer; synonyms are mainly before the answer, verb groups are mainly before or after the answer. Other data do not revel significant differences according to the position.

Thirdly, we plan to measure the distance between keywords, expansions and the real answer (each word between the keyword or the expansion keyword and the answer counting for 1). This third step will allow us to do a crossed analysis (between the most numerous keywords or expansions and the nearest ones).
In order to check our hypothesis (to improve the weight of some keywords or expansions), we will launch the same process on the *invalid* TREC-9 answers.
It would have been useful to obtain the morphological derivation of keywords but at the moment, these links are not available in WordNet.

3.2 Answer patterns

Another method – a linguistic method- consists in associating a question type with answer patterns. Thus, it is often possible to associate interrogative noun phrases with answer patterns. In the following example, "What task" seems to be related to the answer pattern "used … for".

Q: What task does the Bouvier breed of dog perform? (672)

A: a **bouvier** des Flandres , a **breed** long **used** in Belgium **for** *herding* cattle

A: a rare **bouvier** des Flandres **breed** , distinguished by a rough , wiry coat and long **used** in Belgium **for** *herding* cattle

The association between the interrogative noun phrase and the pattern introducing the answer (in italics) could be reused for similar questions.
From many answer samples corresponding to questions seeking periods, we have noticed that numerous patterns can introduce the answer: x (days | months | years…) ago; x year old; in the x century; during (Christmas | the summer…); the x's[5]; date-date; from date to date; between date and date; starting in x and ending in y; etc,. Likewise, in order to find other temporal data and/or duration, we can use patterns like: since date; for duration; from date to date; etc.
Most of the answers giving a definition can be located thanks to extraction patterns, mostly when the term is defined by an apposition: MC, a definition; MC – definition –; MC (definition); MC, also (called | known as) synonym; etc.
Why Famous Person questions can also be answered with patterns. For instance, we took a set of question variants about the reason why Jane Goodall is famous:
Who was Jane Goodall? (419)
Why is Jane Goodall famous? (748)
What made Jane Goodall famous? (749)
What is Jane Goodall famous for? (746)
What is Jane Goodall known for? (747)

***NPP*[6]*,* a NP** (JG, a leading chimpanzee specialist): 419
NP *NPP* (naturalist JG): 419
***NPP* as the NP** (JG as the most recognizable scientist): 746
***NPP*'s NP** (JG's study of wild chimpanzees): 747

[5] For instance, "the 70's".
[6] NPP: noun phrase made of a proper name, like Jane Goodall (JG).

***NPP* be the first to VG** (JG was the first to study apes): 746, 748, 749
***NPP* who VG** (JG, who pioneered the study of primates): 419

4 Conclusion

Question answering systems use IR and IE (information extraction) methods to retrieve documents containing a valid answer. As IR methods were made to retrieve documents - and not *information*- we need to use other methods. Thus, we have demonstrated the importance of the role played by categorization in the Q&A frame. Our categorization, based on question patterns, allows us to delimit the target of the question in order to classify it semantically. In some cases, patterns contain clues for answer retrieval (links available in WordNet). Moreover, our categorization takes shades of meaning into account. Furthermore, we have carried out a study to modify the classical IR vector space model. If we attribute weights to question keywords and keyword expansions related to question category, we may adapt this model to Q&A systems. We have not yet finished this study but we have already get precise and interesting data about keywords and expansion distribution within valid answers. This statistical method can be completed with a linguistic method, which give goods results for some question categories.

5 Appendices

Tags	Meaning
CD	Cardinal number
JJ	Adjective
JJS	Adjective, superlative
NN	Noun
NNS	Noun, plural
NP	Proper noun, singular
NPS	Proper noun, plural
RB	Adverb
RBS	Adverb, superlative
VV	Verb base form
VVD	Verb, past tense
VVG	Verb, gerund or present participle
VVN	Verb, past participle
VVP	Verb, non-3rd person singular present
VVZ	Verb, 3rd person singular present

Table 9: Meaning of the tags

References

Attardi G., Cisternino A., Formica F., Simi M., Tommasi A. (2003), "PIQASso 2002", in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Béchet F., Nasr A., Genet F. (2000), "*Tagging Unknown Proper Names Using Decision Trees"*, in Proc. of ACL'2000, Hong-Kong, China, p.77-84.

Bellot P., Crestan E., El-Bèze M., Gillard L., de Loupy C. (2003), *"Coupling Named Entity Recognition, Vector-Space Model and Knowledge Bases for TREC-11 Question Answering Track"* in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Burger J. *et al*. (2000), "*Issues, tasks and program structures to roadmap research in question answering*", TREC-9, Gaithersburg.

Burger J.D., Ferro L., Greiff W., Henderson J., Mardis S., Morgan A. (2003), "*MITRE's Qanda at TREC-11*", in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Clarke C.L.A., Cormack G.V., Kemkes G., Laszlo M., Lynam T.R., Terra E.L., Tilker P.L. (2003), "*Statistical Selection of Exact Answers (MultiText Experiments for TREC 2002)*", in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Graesser A.C., Person N., Huber J., "*Mechanisms that Generate Questions*", chap. 9, Lawrence Erlbaum Associates, pp. 167-187, 1992.

Ittycheriah A. & Roukos S. (2003), "*IBM's Statistical Question Answering System – TREC11*", in Proceedings of the 11th-Text REtrieval Conference, NIST special publication SP 500-251, 2003.

Kuhn R. & de Mori R. (1996), *"The application of semantic decision trees to natural language understanding"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(5), pp. 449-460, 1996.

Lavenus K. & Lapalme G., *« Évaluation des systèmes de question réponse - Aspects méthodologiques »*, in Traitement Automatique des Langues, vol. 43, n°3/2002, p.181-208.

Li X., Roth D., "*Learning Question Classifiers*", in Proceeding of COLING'02, 2002.

Nyberg E. (2003), "PIQASso 2002", in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Plamondon L., Lapalme G., Kosseim L. (2003), "*The QUANTUM Question Answering System at TREC 11*", in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Pradhan S.S., Krugler V., Bethard S., Ward W., Jurafsky D., Martin J.H., Blair-Goldensohn S.J., Schlaikjer A.H., Filatova E., Duboue P.A., Yu H., Passonneau R.J., Hatzivassiloglou V., McKeown K.R., Illouz G. (2003), "*Building a Foundation System for Producing Short Answers to Factual Questions*", in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Roth D., Cumby C., Li X., Morie P., Nagarajan R., Rizzolo N., Small K., Yih W.-T. (2003), "*Question-Answering via Enhanced Understanding of Questions*", in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Sekine S., Sudo K., Nobata C. (2002), "*Extended Named Entity Hierarchy*". In Proceedings of the LREC-2002 Conference, pages 1818–1824.

Sutcliffe R.F.E. (2003), "*Question Answering Using the DLT System at TREC 2002*", in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.

Voorhees E. (2000), *"The TREC-8 Question Answering Track Report"* in Proceedings of the 8th Text Retrieval Conference (TREC-8), pp.77-82, Gaithersburg, Maryland, NIST Special Publication SP 500-246, 2000.

Voorhees E. (2001), *"Overview of the TREC-9 Question Answering Track"* in Proceedings of the 9th Text Retrieval Conference (TREC-9), pp.71-80, Gaithersburg, Maryland, NIST Special Publication SP 500-249, 2001.

Voorhees E. (2002), *"Overview of the TREC 2001 Question Answering Track"* in Proceedings of the 10th Text Retrieval Conference (TREC-10), pp.42-51, Gaithersburg, Maryland, NIST Special Publication SP 500-250, 2002.

Voorhees E. (2003), "*Overview of the TREC 2002 Question Answering Track"*, in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.
Yang H., Chua T.-S. (2003), "*The Integration of Lexical Knowledge and External Resources for Question Answering"*, in Proceedings of The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.