A Survey on Question Answering Technology from an Information Retrieval Perspective

Oleksandr Kolomiyets[1]

Language Intelligence & Information Retrieval, Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium

Marie-Francine Moens[1]

Language Intelligence & Information Retrieval, Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium

Abstract

This article provides a comprehensive and comparative overview of question answering technology. It presents the question answering task from an information retrieval perspective and emphasises the importance of retrieval models, i.e., representations of queries and information documents, and retrieval functions which are used for estimating the relevance between a query and an answer candidate. The survey suggests a general question answering architecture that steadily increases the complexity of the representation level of questions and information objects. On the one hand, natural language queries are reduced to keyword-based searches, on the other hand, knowledge bases are queried with structured or logical queries obtained from the natural language questions, and answers are obtained through reasoning. We discuss different levels of processing yielding bag-of-words-based and more complex representations integrating part-of-speech tags, classification of the expected answer type, semantic roles, discourse analysis, translation into a SQL-like language and logical representations.

Key words:
Question answering, information retrieval, natural language interfaces, retrieval and ranking models

1. Introduction

Question answering is a sophisticated form of information retrieval characterised by information needs that are at least partially expressed as natural language statements or questions, and is one of the most natural forms of human computer interaction. In contrast with classical information retrieval, where complete documents are considered relevant to the information request, in question answering, specific pieces of information are returned as an answer. The user of a question answering system is interested in a concise, comprehensible and correct answer, which may refer to a word, sentence, paragraph, image, audio fragment, or an entire document.

Question answering originated decades ago from research on natural language interfaces to access data- or knowledge bases. Today, these older technologies are seeing a revival. We note the growing power of querying in the context of multimedia data such as images, video, text, or audio with natural language expressions, and the relevance of inferring conclusions from content in text or any other media.

The amount of information stored in distributed or local systems and available publicly via the World Wide Web or privately via proprietary databases is constantly growing with a daily information consumption of around 34 GB/person[1]. Having a particular information need, expressed in a query, current search engines return long lists of potentially relevant documents without pinpointing the core of the result condensed to a short answer. Hence precise identification of a specific piece of information, i.e., receiving a direct answer to an information need, is becoming one of the most desirable functions for information consumers [83]. Returning precise answers demands well-expressed and precise formulations of the information need that go beyond a small set of loose terms, as it is usually the case for document search. Natural language statements or questions, in contrast, are a natural way to express a precise information need [121]. However, very little research is devoted to the information retrieval models that consider representations obtained from the natural language questions and natural

---

[1]According to the HMI Report of Global Infromation Industry Center, University of California, San Diego. Available at: http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf

language statements found in the documents. In this paper we systematically investigate possible representation formats and corresponding retrieval models that allow systems to effectively find direct answers in the documents and to integrate partial answers obtained from different parts of a document or from different documents.

Besides textual databases, repositories of multimedia information are becoming increasingly prominent. Content recognition in media such as images, video and audio is a major research domain, and traditionally there has been an interest in using text-based queries over multimedia resources, preferably in the form of natural language statements or questions. Another factor that will probably boost the demand for question answering technology is the increasing use of mobile devices, such as smartphones, to access information [69], for which traditional queries consisting of typed keywords are not very user-friendly. Speech interfaces, also called voice-enabled natural language interfaces, might very soon enable asking real questions or posting statements in spoken natural language [128, 147, 131].

There have been several surveys on question answering technologies in the past. Simmons reviewed the very first approaches to question answering in English [133]. Hirschman and Gaizauskas discussed background, motivation and general approaches to open domain question answering highly promoted by the Text Retrieval Conference (TREC) [50]. An update on the approaches used in open domain question answering was presented in [89] and [146]. The survey presented in this article is different from the previous ones in many ways. First, it does not make a distinction between open domain and restricted domain question answering, and it is not restricted to the approaches reported in the QA field only. It presents the QA task from an information retrieval perspective and emphasises the importance of the retrieval models, i.e., representation of queries and information documents, and the retrieval functions which are used for estimating the relevance between a query and an answer candidate. The main goal of this article is to provide a comprehensive and comparative overview of question answering technology. Our summary suggests that different approaches should be blended in order to build the next generation of question answering technology. We do not extensively compare the technologies in terms of their performance, because we lack their ground truth corpora and implementation details. However, we provide general assessments of state-of-the-art capabilities, shortcomings and computational complexity. Nor do we cover *interactive question answering*, where the system automatically proposes refinements of questions when there

3

are too many answers or expansions of questions, or when there are no answers at all. A good overview of interactive question answering can be found in [145]. In addition, we refrain from addressing *summarisation*, *clustering* and *visualisation techniques*, which would require us to review the vast literature on *human computer interaction* including interaction, displays, user modelling and user preferences. The techniques on *multilingual question answering* as a specific sub-field of question answering, when the question is posed in a different language than the information source, are not discussed here, but the general approaches can be found in [1].

We focus on the interactions between *information extraction*, *natural language processing* and *information retrieval research*. Information extraction techniques are necessary for analysing natural language questions and statements. These techniques often rely on natural language processing tools such as part-of-speech taggers, which determine the syntactic category of a word, and parsers that analyse a sentence and transform it into a dependency tree. Information retrieval provides the necessary functions for integrating and ranking possible answers that may meet the information need.

The article is organised as follows. First, we define the traditional types of question answering, followed by a short history of this technology. We present the high-level architecture of a typical question answering system and its components, followed by a short overview of evaluation metrics. The methodology section explains how the components are built and how they work together to obtain and rank the answers to an information need. Subsequently, we discuss bag-of-words approaches, morpho-syntactic analysis of natural language statements, classification of the question and expected answer type, finding the necessary discourse relationships, translation into and retrieval with a structured language, and translation into and reasoning with a logical representation. Sophisticated models can fuse information from different document components or even different documents. As an extension of the methodology section, we examine several pertinent research problems. Finally, we cite a number of novel question answering applications and conclude with pertinent research directions.

2. Definitions

Question answering is an information retrieval task constrained by an expression of all or a part of the information need as a set of natural language questions or statements. Examples are: *Who is the architect of the Hancock*

building in Boston?, or *What is the height of the Eiffel Tower?* Questions in natural language, on the one hand specify a well-defined information need, but on the other hand they carry more information than a simple list of search terms, as they represent syntactic and semantic relationships between the search terms.

Question is defined as a natural language sentence, which usually starts with an interrogative word and expresses some information need of the user. Sometimes a question has a form of an imperative construct and starts with a verb. In such a case the information request is called *statement*.

Question type is defined as a certain semantic category of questions characterised by some common properties. The major question types are: factoids, list, definition, hypothetical, causal, relationship, procedural, and confirmation questions. A *factoid question* is a question, which usually starts with a Wh-interrogated word (*What, When, Where, Who*) and requires as an answer a fact expressed in the text body. A *list question* is a question, which requires as an answer a list of entities or facts. A list question usually starts as: *List / Name [me] [all / at least NUMBER / some]*. A *definition question* is a question, which requires finding the definition of the term in the question and normally starts with *What is*. Related to the latter is the *descriptive question*, which asks for definitional information or for the description of an event, and the *opinion question* whose focus is the opinion about an entity or an event. A *hypothetical question* is a question, which requires information about a hypothetical event and has the form of *What would happen if.* A *causal question* is a question, which requires explanation of an event or artifact, like *Why*. A *relationship question* asks about a relation between two entities. A *procedural question* is a question, which requires as an answer a list of instructions for accomplishing the task mentioned in the question. A *confirmation question* is a question, which requires a Yes or No as an answer to an event expressed in the question.

Information source is defined as a collection of *information objects* (documents, video, audio, text, files or databases) available to the question answering system for extracting answers.

Today's question answering is not limited by the type of document or data repository - it can address both traditional databases and more advanced ones that contain text, images, audio and video. Structured and unstructured data collections can be considered as information sources in question answering [83]. Examples of *structured data* sources include relational databases (i), where the stored objects and their attributes have

5

well-defined semantics and relationships; knowledge bases of expert systems (ii), where in addition to the factual knowledge that is typical of data repositories, rules are added allowing reasoning and problem solving. Written text, speech, pictures, video and audio are all forms of *unstructured data*. *Unstructured* does not imply that the data is structurally incoherent, in which case it would simply be nonsense, but rather that its information is encoded in a form that is difficult for computers to interpret directly. Unstructured data allows querying of raw features (for example, words in a body of text), extracting information with clear semantics attached (structured information), or a combination of both. Related to this distinction between structured and unstructured data there is a traditional distinction between *restricted domain question answering*, or RDQA, and *open domain question answering* (ODQA). RDQA systems are designed to answer questions posed by users in a specific domain of competence, and usually rely on manually constructed data or knowledge sources. They often target a category of users who know and use the domain-specific terminology in their query formulation, as, for example, in the medical domain [156, 122, 88]. ODQA focuses on answering questions regardless of the subject domain. Extracting answers from a large corpus of textual documents is a typical example of an ODQA system [76]. Recently, we have witnessed an approach of question answering involving *semi-structured data*. These data often comprise text documents in which the structure of the document or certain extracted information is expressed by a markup. Such markups can be attributed manually (e.g., the structure of a document) and/or in an automatic way, e.g., markups for identified person and company names and their relationships in newspaper articles.

A *retrieval model* is defined by the form used to represent the information sources and the information need, and by the *retrieval function*, which estimates the relevance between the query and the document. In this context we can contrast data retrieval with information retrieval. *Data retrieval*, for example, from a relational database, assumes a data retrieval language (such as SQL) with a well-defined syntax and semantics, and data that similarly exhibit this syntax and semantics, allowing a deterministic match between an information need and the data. *Information retrieval*, on the other hand, typically handles queries and documents where the structure and semantics are to a large extent undefined, and different uncertain interpretations of the information need and document content are common. When retrieving answers from the document repository, often the match between query and information is non-deterministic, yielding a ranking of the information

according to *relevance*. In question answering technology, as no structural requirements on the data collection and natural language questions or statements are made, a non-deterministic ranking of the answers seems the most appropriate retrieval model.

In this article we address the processing of natural language queries and the retrieval or ranking functions used in question answering. We assume that content representations in the form of keywords or in the form of structured logical predicates are given. In the case of a textual document collection from which the answers are extracted, we explain the *natural language processing* techniques for building the content representations. Obtaining the necessary content representations from other media (for example, images, video, audio) is beyond the scope of this work, because this would entail an overview of the *signal processing* techniques involved.

3. A short history of question answering

In the literature the first cited question answering systems are BASE-BALL [42], built in 1961, and LUNAR [149], built in 1972, both of which interrogated a structured database using natural language questions. LU-NAR provided an interface to data from analyses of rock samples during the Apollo moon missions. The BASEBALL system answered questions about baseball games played over a period of one year. Both systems analysed questions based on a set of natural language patterns that were expected to occur in the input. Since the subject domain was restricted, an exhaustive set of analysis patterns embedded in a domain-specific vocabulary was built manually, so that the questions could be processed and successfully translated into a structured query form needed to interrogate the databases. BASEBALL and LUNAR were the first examples of *natural language interfaces to databases*, NLIDB. Today's research in NLIDB focuses on generic approaches to the detection of objects and their attributes and relationships in natural language questions, and on the translation of lexical items into the string tokens that are used to describe the database entries. An overview of NLIDB research can be found in [23, 8]. Recent work in this area includes [118, 103, 95].

In the 1980s and 1990s, knowledge base systems, usually built for restricted application domains, became very popular. This technology is well-suited to a question answering framework, in which the user is confronted with a certain problem and needs the answer. Access to the knowledge base

is usually organised through menus or a natural language interface. The system itself interactively asks the user additional questions in order to better understand the user's intent. To solve the problem, the system reasons using the knowledge available in the knowledge base and the additional information provided by the user. Such deductive question answering focuses on reasoning and provides explanations how a certain answer was obtained. Its early roots go back to early expert system technology, such as the MYCIN system [132], which was designed to offer explanations of medical concepts. Another example is the SHRDLU system which already in 1972 offered an interactive dialog interface to control robot interactions with toy blocks [148]. Today, natural language queries still attract substantial attention as a way to query knowledge bases (e.g., [139]).

The contemporary question answering era started in 1999. The annual Text Retrieval Conference (TREC)[2] has featured open domain question answering as one of its competition tracks [28]. The challenge is to provide a concise answer to a natural language question, given a large collection of textual documents. In recent evaluations two well-known collections of documents were used: AQUAINT with more than 1 million documents and AQUAINT2 with about 907K documents and 2.5 GB in size. The range of question topics is unrestricted, and often the analysis of question and of texts is rather shallow, and no advanced reasoning strategies are applied to find the answers. TREC had a major impact on interest in question answering and on the development of evaluation measures that compare the performance of difference systems [142]. This resulted in the first Web-based QA system, START, which was developed at the Massachusetts Institute of Technology in 2004 [56]. From 2002 onwards, the Initiative for the evaluation of XML retrieval (INEX) [38] launched a number of tasks focusing on exploiting the internal structure of marked documents for retrieving relevant information. Besides, the potential of semi-structured documents for answering questions has been discussed in [3, 53].

The Cross-Language Evaluation Forum or CLEF[3], established in 2000, promotes multilingual question answering, where the question is posed in a different language than the language of the documents in the repository [73]. This requires the translation of the question or of potential answer

---

[2]http://trec.nist.gov/
[3]http://www.clef-campaign.org/

sentences in order to match the question with information in the documents when finding the answer. Cross-language question answering may inspire cross-media question answering, where a question or statement in natural language interrogates image, video or audio archives. Recently, we witness the development of methods for cross-media alignment of content (e.g., across text and images), inspired by techniques of cross-lingual alignment [109]. Although to date querying multimedia databases has been implemented with a structured query language [33], a growing interest in using natural language is noticed [59, 34]. With the increasing power of analysis tools for processing questions or other natural language statements, we expect this interest to persist.

The NII[4] Test Collection for IR systems (NTCIR) Workshop is a series of evaluation workshops designed to enhance research in Information Access (IA) technologies including information retrieval, question answering, text summarisation, extraction, etc., and was launched in 1998. With an emphasis on large-scale test collections, this venue accommodated research groups working with Japanese and other Asian languages [126]. Apart from traditional and cross-lingual question answering NTCIR recently launched a geo-temporal QA track addressing geographic and temporal aspects in information retrieval, especially questions containing both kinds of constraints.

Throughout the evaluation campaigns we note a steady increase in question complexity, and as a result more advanced techniques for question answering were developed. Some of the key QA-techniques include: the incorporation of expected answer type as in [45, 84, 152], logic-based representations [85], syntactic and semantic structures for question answering [51, 7, 27, 49], temporal question answering [86, 43, 6, 124, 112], geographical question answering [40], semantic-role labelling for question answering [93, 97, 130], and discourse relationships and textual entailment for question answering [29, 44]. A global overview of research goals in question answering with respect to capabilities to the end-user is presented in [20].

4. The architecture of a typical question answering system

Despite their long history, question answering systems are characterised by a standard high-level architecture (Figure 1). The natural language question or statement and the content of the document are pre-processed and

---

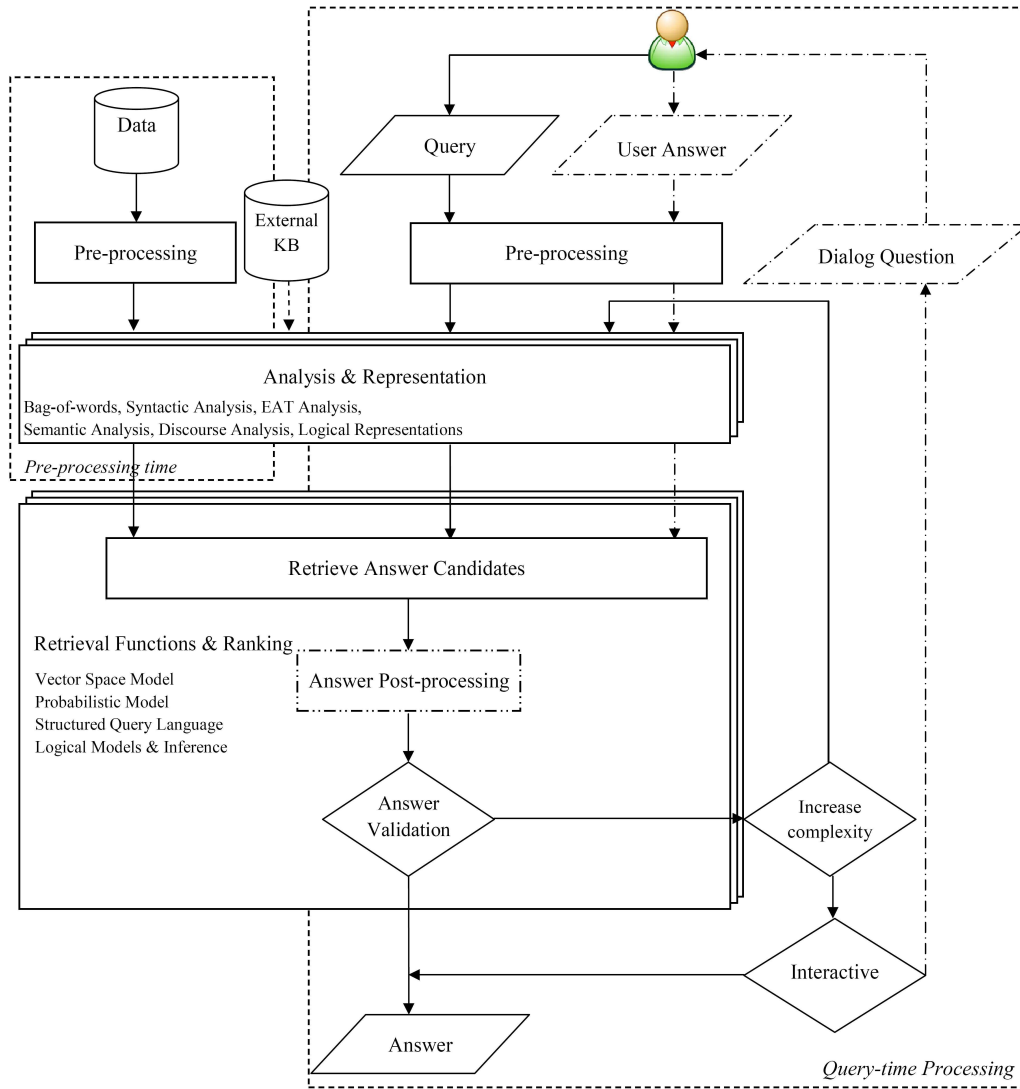[4]National Institute for Informatics, Tokyo

Figure 1: High-level architecture for question answering

translated into a format that is readable by a computer (for example, list
or bag of words, database tables, logical representation). Alternatively the
content of a document might be manually coded into a database or knowl-
edge base structure (Analysis and Representation). When confronted with
a textual document collection, the translation might involve different phases
increasing the level of sophistication. For instance, for a document repos-

itory more computationally expensive translations, (e.g., into a logical format), are only applied for the sentences that may potentially contain the answer. Such sentences are first filtered based on a simple representation, e.g., bag-of-words. In a subsequent step of the process the question and content representations undergo a deeper syntactic and semantic analysis, and are then compared by a retrieval function, which can range from a constrained matching process (for example, in a structured query language that matches database entries) to the use of sophisticated reasoning techniques that infer the answer. In a textual document, the answer is usually found as the resolution of an empty slot, in the context of matching a question representation with a sentence representation. In case of different possible answers or a partial fulfilment of the conditions expressed by the natural language question or statement, the retrieval function preferably yields a numerical result – that is, a ranking score – so that the answers can be ranked according to their relevance to the information need. Multiple identical answers might reinforce the ranking score, for example, for the question *When was Mozart born?*, if the same date of birth is found several times, then the chance that this is the right answer increases. Answers can be reformulated in certain cases: for instance, text-based answers can be translated into well-formed sentences, or video fragments can be merged into a single clip (Answer Post-processing). The different steps yield a pipelined architecture as shown in Figure 1. Question answering systems differ in their retrieval model: that is, they may use different approaches, as shown in the boxes, to both the question and content representation and the retrieval or ranking function. By contrast a typical document information retrieval system relies solely on a simplistic representation of queries and information objects, and an elementary retrieval function, which can quickly return a list of relevant objects, but does not perform any deep analysis of neither documents nor user queries, nor does it provide sophisticated retrieval models for answer finding and fusion.

5. The evaluation of a typical question answering system

When evaluating the results of a question answering system, several performance metrics have been proposed. When processing answers to factual questions, the *mean reciprocal rank* (MRR) calculates the average over a set of n queries, where different scores are attributed inversely proportional to the rank ($rank_i$) of the first correct answer in the answer list [143]. More

formally it is defined as:

$$MRR = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{rank_i} \qquad (1)$$

Since the MRR metric does not allow evaluating questions to which there is no answer in the information source, and does not credit the recall of the system, precision and recall measures are used to evaluate the performance of the system. For questions for which there is no answer in the document collection, the recall and precision of these NIL answers are computed, respectively as the percentage of NIL answers found and as the percentage of correct NIL answers in the found NIL answers. For factoid questions, accuracy is used as one of the major evaluation metrics, for which the answers are judged to be globally correct. Within TREC list questions are evaluated with *instance precision* (IP) and *instance recall* (IR) which are based on the complete list of known distinct instances of the answers. If S is the number of known answers, D is the number of correct, distinct responses returned by the system, and N is the total number of responses returned by the system, then $IP = D/N$ and $IR = D/S$. The F-score is then estimated giving equal weights to recall and precision ($\beta = 1$) using (2).

$$F(\beta) = \frac{(\beta^2 + 1) \times precision \times recall}{\beta^2 \times precision + recall} \qquad (2)$$

To make a distinction between systems that provide 'early' correct answers in the ranking, the *confidence score* has been introduced and presented in (3) as

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\text{number correct in first } i \text{ ranks}}{i} \qquad (3)$$

It has been noticed that if the evaluation requires only one answer per question, as, for example, for factoids, the average performance drops [72]. Also, a metric commonly used in text summarisation is sometimes applied: The Pyramid method [94] measures the overlap of words and phrases in the system-generated answer with the expert-generated answer, assuming a dictionary of equivalent paraphrases. It is computed proportionally to the number of votes, i.e., ground truth labels that contain the phrase. This metric is applied to the type of *Other*-questions introduced in TREC. Since

measuring the precision of *Other*-questions is difficult, as it requires all concept mentions defined by the expert, an *approximated precision value*, also known as *approximation to nugget precision*, is used and defined as (4):

$$1 - \frac{length - allowance}{length} \tag{4}$$

The final score for an *Other*-question is an F-score (2), with nugget recall weighted more heavily than nugget precision, e.g., with $\beta = 5$.

Since different types of questions are evaluated differently and may not reflect the overall performance of the system, evaluating over individual series of questions should provide more accurate scores. Each series is a mixture of different question types and a *weighted score* can be computed taking into account only the types of questions available in it (5). The final score is called *pre-series score* and is estimated as an average over all weighted scores.

$$WeightedScore = \mu \times factoid + \nu \times list + (1 - \mu - \nu) \times other \tag{5}$$

In the TREC QA 2007 μ and ν were equally chosen as $\frac{1}{3}$.

For illustration purposes we present the three best performing systems that participated in the TREC QA 2007 [28], QA@CLEF 2009 (monolingual English) [108] and NTCIR QA Track 2008 (English-Simplified Chinese and English-Japanese) [81], see Table 1.

System	TREC QA 2007			QA@CLEF 2009	NTCIR QA E-C	NTCIR QA E-J
	Factoid	*List*	*Other*	2009	E-C	E-J
	Accuracy	F-Score	$F(\beta = 3)$	Accuracy	$F(\beta = 3)$	$F(\beta = 3)$
1st best	0.706	0.479	0.329	0.61	0.2211	0.1627
2nd best	0.494	0.324	0.299	0.59	0.1930	0.1133
3rd best	0.289	0.147	0.281	0.57	0.1895	0.1133

Table 1: Evaluation scores of three best performing systems in TREC QA, QA@CLEF and NTCIR QA.

The evaluation metrics presented in this section are designed for performance comparison of question answering prototypes. In addition the usability of question answering and information retrieval systems is evaluated. A broad study on the *habitability* of question answering systems is presented in [100]. Habitability is used to describe how easily, naturally and effectively the users can express themselves given the constraints of a system language. This

study shows that the performance of a question answering system largely depends on the way questions are formulated, i.e., search terms, user experience and the awareness of the system's functions.

6. Main methods

When discussing the main methods and algorithms used in question answering, many different classifications can potentially be followed. Question answering is a form of information retrieval, and any retrieval model is defined by the representation of the information need, by the representation of the retrievable object, and by the retrieval or ranking function. The most instructive way to describe the methods starts from the simplest representations derived from the natural language statements and corresponding retrieval functions, and gradually moves to more advanced analysis techniques and retrieval functions that incorporate some degree of reasoning.

In the following sections we describe methods used in question answering to analyse the information needs expressed in natural language. The information objects can take any form (e.g., text, images, video, audio and multimedia). When the information content is expressed in the form of text, we point to valuable information extraction techniques that structure the information found in the text. When information is found in other media, we assume that its content has already been translated into a format that is easily interpretable by the computer, such as a database table or logical representation. Given the resulting representations of information need and information objects, an appropriate retrieval model is discussed. A retrieval function seeks to identify the correspondence between query and document representations. Throughout this article we illustrate the methods with the following example question: *Who killed Lee Harvey Oswald?*

6.1. Bag-of-words representations

Analysis and representation

The simplest approach is to consider the question as a bag-of-words, i.e., a set of words, possibly with stopwords or non-content bearing words removed (see Figure 2) without taking into account any structural characteristics that signal sentence or discourse grammaticality, or positional information of the words.
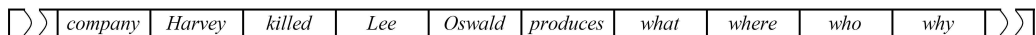
	company	Harvey	killed	Lee	Oswald	produces	what	where	who	why	

Figure 2: The vocabulary of a bag-of-words representation

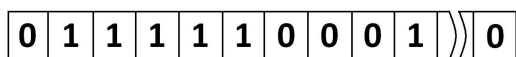| 0	1	1	1	1	1	0	0	0	1	0

Figure 3: Bag-of-words representation indicating the presence or absence of a word from the vocabulary

Retrieval function

The most straightforward case when information query and information objects are represented in a bag-of-words manner (Figure 3) uses a Boolean model that requires the Boolean expression built from the query to be matched by the document representation as shown in Figure 4. This retrieval function is used in the first iteration of a question answering system to find answer candidates to the posed query. Since this model requires a strict deterministic match, other retrieval models have been proposed, including the algebraic vector space model and various probabilistic models. Under the vector space model, an information object and natural language query are represented as a vector of terms in a p-dimensional space, where p is the number of terms in the vocabulary. The document and query vector are compared by computing their similarity (e.g., the cosine of the angle between the vectors) or distance. Compared to Boolean models, this approach returns answers to the questions even if the constraints posed by the question or query are only partially met, but at the cost of precision. Pointers to the above models can be found in [25].

The probabilistic models are more interesting to the question answering community, because they offer a natural way of integrating probabilistic relationships between terms into a model. One popular approach is the language model [26], in which a probabilistic content model is built from the information object (e.g., document or sentence), and possibly from the information query. To obtain a ranking function, we can compare information in the question using information models, such as the Kullback-Leibler divergence between two probability distributions [60]. In all models the replacement of words by their synonym terms found in lexico-semantic resources or thesauri (e.g., WordNet [79]) might improve the possibilities of finding correspondences between question and the information sought. In the algebraic and
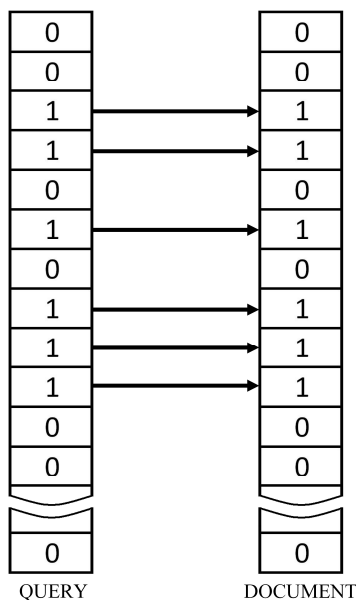
15

Figure 4: Matching mechanism between a document and a query given a bag-of-words representation.

probabilistic models *term frequencies* are usually taken into account, corrected by an *inverse document frequency* factor, which demotes words that occur in many documents.

Evaluation

Since we have taken a very simple approach to question analysis, and likewise to the analysis of documents and sentences, ignoring structural relations between words, the retrieved information may not be very precise. For example, for the query *A child with brown hair wearing a yellow sweater with blue jeans*, we might retrieve a video fragment showing a girl with yellow pants and a blue sweater. These models, although very popular in document retrieval and search technology, typically lack the precision needed for question answering, even for answering simple factual questions requiring the retrieval of an entity (e.g., person name, date) [84]. They have the advantage of low analysis complexity and low storage overhead for the indices, which are often in the form of term posting lists, so called *inverted files* [25]. However, in traditional question answering they have proven their usefulness for the initial rough filtering of documents and sentences that obviously have

no relevance to the information question at hand. It has been shown that correct answers can be found with a limited number of mapping rules to the queries by exploiting the redundancy of answers available [17].

6.2. Morpho-syntactic analysis of natural language statements

Analysis and representation

In a morphological analysis, stemming and lemmatisation can be useful, as they improve the chances of finding a good match of the query with an answer sentence. Stemming normalises words to their common root form. For example, *cat* and *cats* refer to the same stem *cat*. By contrast, lemmatisation identifies the lexeme *go* for *goes*, *went* and *gone*. However, these operations reduce the information contained in a term. In many languages inflectional forms of nouns and pronouns define cases, which indicate the syntactic or semantic function of a phrase, clause, or sentence constituent (e.g., identification of subject, object or circumstantial expressions) [96, 66].

A sentence statement (or a natural language query) exhibits a syntactic structure. It consists of content-bearing words (such as nouns, verbs, adjectives and adverbs,) which are connected with function words (such as determiners and prepositions), consistent with the syntax of language. The simplest approach for the sentence structure detection considers n-grams of words in the representation of a natural language statement [18], however, for a real syntactic analysis, a number of natural language processing tools can be used. They include part-of-speech (POS) tagging for detecting the syntactic word class (such as nouns or verbs), and phrase chunking for detecting base noun and verb phrases. Additional parse-based cues produced by breaking up a sentence into its constituents and building the dependency tree of a sentence between them can also be used [27] (see Figure 5).

Retrieval function

When the words of a natural language statement are tagged with part-of-speech (POS) labels, they contain additional disambiguating information, for example, the word *play* as a verb means something different from the word *play* as a noun. The matching process can incorporate this information. Likewise, when a sentence is parsed into its dependency tree, this information can function as additional constraints. Technically syntactic tree kernels are used to compute the similarity between two dependency parse trees of a sentence [22, 91, 19], where the word nodes are enriched with, for example, POS tags

```
                    killed
                  (root kill)

        N                      N Oswald
    (root who)          (root Lee Harvey Oswald)
    (rel subj)                (rel obj)
                            (sem PERSON)


                  U Lee            U Harvey
               (root Lee)        (root Harvey)
              (rel lex-mod)     (rel lex-mod)
```
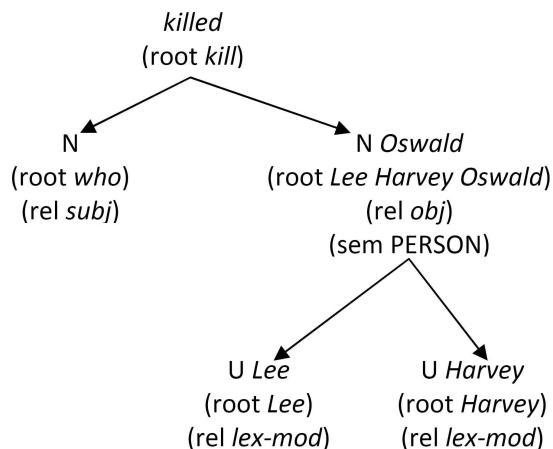
Figure 5: Dependency tree representation of a question

and possibly semantic information. Kernel functions can be efficiently computed by relying on dynamic programming techniques. Such kernel functions can be used to compute the ranked matching between the query representation and the answer candidate. Alternatively, syntactic rewriting rules are helpful to improve the finding of correspondences in question and information content. Tree edit models have been recently incorporated in question answering. They recognize textual entailments and paraphrases and can find semantic similarities between questions and answers via sequences of tree transformations [47].

Evaluation

Morpho-syntactic analysis of natural language expressions results in better capturing the structural relations between words in a sentence or question, resulting in an enriched meaning representation. Reliance on morpho-syntactic patterns has been successfully proven for automatically generating multiple-choice questions from texts [82], but its potential can be enlarged to general question answering. Phrasal nodes often represent a single unit of meaning. Using additional case information available for some languages already introduces semantic information to the retrieval process. Cui et al. [27] studied a fuzzy relation matching technique for factoid passage retrieval. Significant improvements over other retrieval systems were shown, namely up to 50-138% improvement in MRR and a precision level over 95% at rank 1. Li [63] showed that syntactic features contribute to the question answer-

ing performance increasing the MRR value up to 8.9%. On the downside, we increase the computational complexity of the processing, requiring that complex query analysis is performed in real-time and document analysis is carried out off-line at the expense of richer indexing representations, for instance, based on MapReduce architectures [67]. Time and space complexities will only increase with the advanced analysis methods discussed in the next sections.

6.3. Semantic classification of the expected answer type
Analysis and representation

A natural language question or query statement gives us additional information on the type of information that is expected as an answer. For instance, we might be looking for a person, the name of a company of interest, the place, the date, or even a picture of a person in a video. It has become common practice in question answering to automatically identify the semantic class of the expected answer in the question and the corresponding semantic class of the answer candidate in the information source.

Question classification goes back to the work of Lehnert [62], who defined 13 basic conceptual semantic question classes including *Causal antecedent*, *Goal orientation*, *Enablement*, *Causal consequent*, *Verification*, *Request* and others. Unfortunately, few studies have implemented this classification due to the need for a deep semantic analysis with the exception of the investigation of *Why* questions in QA research [141]. Motivated by early TREC evaluations and by the task settings, questions were categorized as factoid, list, definition, hypothetical, causal, procedural and confirmation queries. For factoid questions the question type classes correspond with an expected answer type (EAT), organised in *answer type taxonomies*, sometimes referred to as question ontologies. The following example illustrates how interrogative words can be used to identify the expected answer type:

what/which Aux(*be*) NP (*city - country - company*) *?* CITY - COUNTRY - COMPANY
who/whom ? PERSON
when ? TIME/DATE
where/what place ? LOCATION
what time (of day) ? TIME

A flat list taxonomy presented in [136] defines classes inspired by the Message Understanding Conference evaluation, like person, organisation,

`location`, `time`, `date`, `money` and `percent`. Other sub-types include: `duration`, `frequency`, `age`, `number`, `fraction`, `decimal`, `ordinal`, `math`, `equation`, `weight`, `length`, `temperature`, `angle`, `area`, `capacity`, `speed`, `rate`, `product`, `software`, `address`, `email`, `phone`, `fax`, `telex`, `www`, and `name` (default proper name). A flat set of 27 answer type categories, derived from the Microkosmos ontology [74] is proposed by Ogden *et al.* [99]. A typology of 140 text-surface patterns based on a manual analysis of 17,000 questions is presented in [52]. Other researchers proposed hierarchical classification schemes. Harabagiu *et al.* [45] presented a WordNet-based taxonomy of classes. This taxonomy contains 18 top categories (e.g., `Date`, `Time`, `Organisation`, `Location`) and 15 leaves of the top categories (e.g., `City`, `Country`, `Continent`, `Dimension`, `Duration`). This approach establishes many-to-many relationships between the various layer classes. An impressive hierarchical taxonomy is described by Suzuki *et al.* [138]. This taxonomy contains 150 categories, extracted from an analysis of 5,011 questions in Japanese. Another named-entity based taxonomy with 200 categories resulted from a manual analysis of a news wired corpus has been proposed in [129]. The most famous expected answer type taxonomy with regard to factoid questions is the one of Li and Roth [64]. Their two-layer taxonomy consists of a set of six coarse-grained categories and fifty fine-grained ones (e.g., `Abbreviation`, `Description`, `Entity`, `Human`, `Location` and `Numeric` as coarse classes; and `Expression`, `Manner`, `Color`, `City` etc. as fine-grained classes). As an illustration, the quantitative distribution of the coarse and fine-grained classes of 1,000 questions from TREC 10 and 11 is presented in Figure 6. An additional resource is the set of nearly 6,000 labelled questions written in English, which can be used to train question classifiers [152] besides an older collection of questions released by the University of Southern California [51].

The next problem is to identify the expected answer type of respectively the question and the information found in documents. Symbolic approaches use hand-crafted rules possibly combined in a grammar [99, 136, 52]. Features typically included in the rules are lexical elements such as words or phrases and syntactic information obtained through POS tagging and sentence parsing. More recently, supervised machine learning techniques became popular, which train a classification model from examples that are manually annotated (questions with their corresponding answer types). Creating a training set is a time-consuming process, but it does not require rule-writing skills. Classification of a new question is simply based on the patterns learned from

Class	#	Class	#
ABBREVIATION	18	term	19
abbreviation	2	vehicle	7
expression	16	word	0
DESCRIPTION	153	**HUMAN**	171
definition	126	group	24
description	13	individual	140
manner	7	title	4
reason	7	description	3
ENTITY	174	**LOCATION**	195
animal	27	city	44
body	5	country	21
color	12	mountain	5
creative	14	other	114
currency	8	state	11
disease/medicine	3	**NUMERIC**	289
event	6	code	1
food	7	count	22
instrument	1	date	146
lang	3	distance	38
letter	0	money	9
other	19	order	0
plant	7	other	24
product	9	period	18
religion	1	percent	7
sport	3	speed	9
substance	20	temp	7
symbol	2	vol.size	4
technique	1	weight	4

Figure 6: Distribution of 1,000 TREC questions according to the hierarchy of Li and Roth [64]

the training set and the same features as described above can be used. We can choose between a large number of techniques, such as rule and tree learners, support vector machines (SVM), maximum entropy classifiers and conditional random fields, but the context-dependent classification techniques that estimate parameters using a maximum entropy principle are often the most successful. The success factor of SVM lies in introducing a syntactic or semantic kernel function [152, 91]. The classification is quite successful with mean average precision values between 70-95.4% for the different questions with expected answer types trained and tested on different English datasets from TREC-8, 9 and 10. Comparative studies of the use of machine learning techniques for estimating the EAT are reported in [115, 152, 64]. The clas-

sification process often entails the more difficult task of segmentation of the EAT. For instance, in definitional or descriptive questions the answers might span one phrase and in other cases several sentences.

With respect to the information source, the analysis of answer candidates is done by employing shallow semantic analysers, such as named-entity recognizers, which provide semantic labels to the tokens in the candidate answer. The set of labels is limited to capabilities of the recognizer and adjusted to the chosen question taxonomy.

Retrieval function

Again, we would advocate a deterministic and a probabilistic approach. Under the deterministic approach, once the expected answer type has been identified, the answer candidates can be filtered by this expected semantic class. Under the probabilistic approach, the confidence score of the EAT identification or the scores of a few better hypotheses may be incorporated into a probabilistic retrieval model [83]. *List* questions especially require that information from different places in one or several documents is linked and merged in an understandable answer. *yes/no* questions need assessments whether the events or entity-relationships found in the question hold based on the information obtained from the documents. Another difficulty arises when questions are complex, entailing several question and answer types, possibly combined with conjunctions, disjunctions or conditional operators, often requiring the aggregation of pieces of information found. Approaches which utilise linguistic syntax-based patterns for questions and answers have also been proposed in [54, 135]. Such a pattern specifies a syntactic surface for a question and contains a number of pre-defined syntactic surfaces for answer bearing sentences, how the answer is to be found and what semantic type or string pattern it should have.

Evaluation

The importance of correctly identifying the expected answer type was studied in [84]. The author reports that 36.4 % of errors in question answering on TREC-8, 9 and 10 data are caused by incorrect EAT estimation. Because of the good recognition results of the type classes, it is worth investing in these technologies when building question answering systems. The problem is that many different question and answer type taxonomies exist in the form of flat lists or hierarchies. They form a kind of interlingua between questions and answers to facilitate their matching, but at the same time they are often

very specific for the application at hand. In the next section we discuss semantic classifications describing content at a more generic level, i.e., by means of semantic roles.

Identifying the expected answer type is very much linked to *real questions* whereas natural language queries often take other forms. For instance, *Show me a child with brown hair wearing a yellow sweater with blue jeans.* In addition, complex questions or query statements often contain many additional constraints that refine the information need. What type of information is more important than others is valuable when ranking incomplete answers. None of these problems have been studied in current question answering research.

6.4. Semantic classification of all the constituents of questions or natural language statements

Analysis and representation

Identifying the question type and expected answer type provides a semantic class label that can be locally estimated both in question (e.g., simply based on an interrogative question word) and candidate document sentences. Such an approach is too strict and may mislead the question answering process. A more complete semantic labelling of the full question and candidate answer sentence can improve performance as shown in [130]. The labels are assigned in the context of a situation, also referred to as a *case*, usually denoted or invoked by a main verb in the sentence. The sentence can be represented as a frame or predicate-argument structure that describes a general case of composing semantic roles: *who* did *what* to *whom*, *when* and *where*, *for which purpose*, *by what means*.

The history of the predicate-argument representation goes back to the very first implementations of question answering (see Section 6.7). Recently, state-of-the-art dependency parsers have been used to automatically represent queries as triplets and to disambiguate simple syntactic roles such as *Subject*, *Object* and indirect *Object* [96, 66]. More commonly, semantic role and frame classification systems yield a valuable structural decomposition of natural language expressions as shown in Figure 7 for the frame *killing*. On the left-hand side of the representation, the question constituents are presented, whereas *Who* is attached to an expected answer phrase pinpointing the semantic role that is desirable in the answer candidate. On the right-hand side are the major semantic roles of that frame. Each arrow denotes the role assignment for constituents with a confidence score as $P(R_{SR})$.
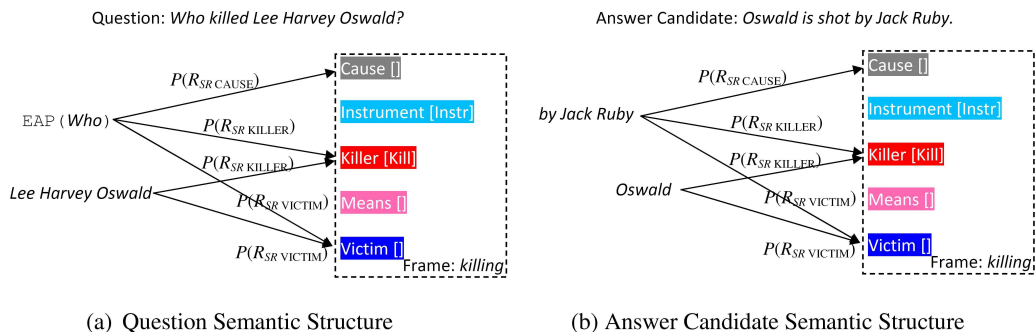
Figure 7: Representation as semantic role labelling in FrameNet

Temporal and spatial information are one of the challenges of QA for *When* and *Where* factoid questions [76]. Temporal question answering addresses the automated interpretation of questions by using temporal elements such as absolute or relative times, points, durations and extraction of answers with temporal aspects [6, 124, 43, 86, 125]. In addition, the temporal information serves as input for advanced reasoning techniques in question answering (see below). Spatial question answering yields answers to questions that involve spatial objects (for example, locations or regions), attributes (for example, size or shape), and relations (for example, above/below, inside/outside, or near/far), possibly requiring spatial inference [58].

There are two main resources for defining semantic roles and frames, namely FrameNet [10] and PropBank [105]. A frame in FrameNet describes all possible semantic roles that can be involved in a case, which can be invoked by any of several predefined lexical units for the frame. A frame in PropBank is organised around one main verb with simple mnemonics for generic role labels, such as `ARG0`, `ARG1`, `ARG2`, `ARGM-XXX` etc. The roles are defined globally throughout the corpus. By analogy with the FrameNet representation, our question example can be represented in terms of PropBank semantic roles (Figure 8), where `TARGET` is the label for the target (main) verb, `ARG0` is a mnemonic for a semantic *killer* role as defined in PropBank for this target verb and `ARG1` is a mnemonic for a semantic *corpse* role.

Besides major semantic roles there are other additional roles, like locative and temporal adjuncts (`ARGM-TMP` and `ARGM-LOC` for PropBank; and `Time[]` and `Place[Loc]` for FrameNet). Resources with temporal information include the TimeBank corpus annotated with the TimeML temporal markup language [111], [113]. Spatial roles have been annotated in the Generalized
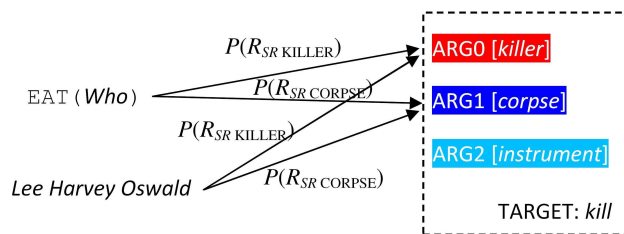
24

Figure 8: Question representation as semantic role labelling in PropBank

Upper Model (GUM) data set and the CLEF dataset [58].

The recognition of semantic roles and frames, and of the more specific temporal and spatial information and relations is commonly done by supervised machine learning techniques, and more specifically by means of context-dependent classification algorithms such as conditional random fields [77].

Retrieval function

In a retrieval context, taking into account semantics can constrain the mapping between the question and a potential answer sentence, i.e., not only should lexical items or words simply match, but also their semantic label. In addition, semantic labels can be used in the mapping without their lexical instantiations, thus broadening the search. The third possibility is to consider a logical retrieval model and infer the answer to a query (e.g., by means of temporal or spatial reasoning) [125]. This last possibility is discussed in section 6.7. A propagating algorithm with verb arguments has been presented in [97]. The method transforms step by step the structure preserving the original meaning of the sentence. The basic relations in WordNet (HYPERNYM, HYPONYM) along with additional ones (DERIVATION, ENTAILMENT, CAUSATION) are weighted and used for the transformations. Depending on the final score of the transformation, a new ranking is applied to the answer list.

Evaluation

To date, question answering based on semantic role labelling has resulted in the best matches between semantic representations of the query and answer candidates by filling empty slots in the question-representation with appropriate phrases of the same role from answer-candidate representations. Shen and Lapata [130] solve this step as a graph similarity problem that involves question and answer graphs (see Figure 7). The concept is that the more semantic roles the two graphs share, the more similar they are. The

proposed approach, which employs syntactic and semantic features, showed an improved performance level on a number of the TREC question collections, e.g., rising from 34.38% for syntactic match to 41.76% in terms of correct answers at rank 1, on the TREC 2005 dataset. Even more performance boost can be expected on questions, for which the semantic frames in FrameNet and corresponding annotations exist (from 13.16% for syntactic match versus 59.65% for syntactic and semantic match, in terms correct answers at rank 1). Narayanan and Harabagiu [93] reported a performance improvement for answer type recognition of complex questions as from 35% for the baseline up to 73.5% for the semantic topic-based approach. By contrast to the baseline with 8% of correct answers, probabilistic inferencing through interconnected semantic frames would enable answering 52% of complex questions considered in the paper. Ofoghi *et al.* [98] studied the impact of semantic analysis for question answering and showed that question answering highly benefits from advanced semantic processing of questions and documents. Recently Moschitti and Quarteroni [92] applied linguistic kernels to POS-tag sequences, syntactic parse trees and predicate-argument structures for answer re-ranking, which led to an 63% performance gain for answer candidate ranking in terms of MRR for the top five documents in comparison to a bag-of-words approach.

6.5. Identifying the necessary discourse relationships

Analysis and representation

In a more advanced retrieval model a limited form of discourse analysis for queries and documents is incorporated. Although beyond the scope of this paper, the information need might only be fully expressed using several questions – in some cases, subsequent questions might refine or broaden the original search. For instance, in an interactive setting, the user might give feedback on the answers by posing additional questions. In this case, it is important to detect how the data in the questions relate to each other. In the document collection, the complete answer to a question is not necessarily located, for instance, in one sentence or one image in the case of multimedia databases. Typically, answers to list questions are spread throughout the document collection. Here, also, it is important to identify how data are connected across documents and media. An important relationship is that of equivalence, where two or more mentions of an entity (a person or object, for example), an action or of an event refer to the same thing, in the context of the situation described in the discourse. When dealing with individual

entities, we often use the term *noun phrase co-reference resolution* to describe this phenomenon (e.g., use of certain pronouns to refer to the same person or object). The co-reference techniques in the field of textual question answering have been employed in [90, 153, 49, 127]. Besides equivalence, many other relations can be defined, including relations that refer to hypernymy and hyponymy, or temporal or spatial references. Within- and cross-document co-reference resolution is an area where substantial research is still needed in order to obtain accurate results. It is, however, a necessary condition in processing interactive question answering and for the fusion of answers across different documents. This is especially true when question answering is based on a structured or logical representation of questions and documents (see Section 6.7).

There are also a number of rhetorical relations that can be recognised in questions and data. For instance, Lehnert in [62] has suggested 13 conceptual categories for questions such as *Causal antecedent*, *Goal orientation*, *Enablement*, *Causal consequent*, *Verification*, *Request* etc. To the best of our knowledge very little research is reported on answering these types of questions. *Why* questions are studied by Verberne *et al.* [141]. *How* questions are investigated in [9]. On the recognition of rhetorical relations in expository texts, the work of [75] is very relevant. Recent advances in text-based argumentation detection [104] open up new research avenues for answering *Why* questions.

Retrieval function

We are not aware of any retrieval functions that incorporate knowledge of discourse relations, however this is an essential part of question answering with data sources of low redundancy.

Evaluation

Research is certainly needed with regard to answer finding for *How*, *Why* and other questions that require advanced forms of discourse understanding. In addition coreference resolution, content linking and answer aggregation are essential in the frame of question answering.

6.6. Translation into and retrieval with a structured language

Analysis and Representation

In the previous sections, we have presented the components (e.g., semantic role labelling) that allow the translation of a question posed in natural

language into a structured query language, and the translation of textual sentences into database entries. The most common method for querying relational databases is the *Structured Query Language*, or SQL, a standard developed by ANSI (American National Standard Institute) and ISO (International Standard Organisation) in 1986. SQL offers a simple syntax for data definition, formulation of the query and data updates. A typical SQL query is formulated as:

```
SELECT <attribute list>
FROM <list of tables>
WHERE <condition>
```

There exist many variations of SQL. In the context of information retrieval from documents, query languages such as XPath, XML Path Language and XQuery have been developed to manipulate data from XML (Extensible Markup Language) documents [38]. Example of the translation of the query into a structured language uses a set of symbolic rules [37, 101] and mappings [110].

The vast domain of information extraction offers technologies to translate textual data found in documents into structured database entries. Especially the approaches for entity-relation extraction are well developed (e.g., [24, 57, 150]). We refer to the sections on answer type recognition and semantic role labelling for a discussion of the classification algorithms.

Retrieval function

Once we have translated the question (possibly the natural language statements in a document) into a structured format, classical database models could be applied for retrieval that follow a deterministic model, where the data is retrieved only when it meets the conditions imposed by the query [33]. This approach lacks a framework for modelling uncertain translations of the query or uncertain representations of data. Nor does it provide answers or a ranked answer list when answers only partially fulfil the query.

Evaluation

Natural language interfaces to databases represent one the oldest forms of question answering technologies. The World Wide Web is often seen as a very large database with recent methods being capable of structuring and linking the information [150], offering enormous possibilities for the application of question answering technologies. Notwithstanding the steady interest

in QA that spans decades, we lack large benchmarking studies on the existing techniques.

6.7. Translation into and reasoning with a logical representation

Analysis and representation

As mentioned above questions and data can be represented in first-order logic [15]. A variant form, the Meaning Representation Language (MRL) was introduced to represent natural language statements [16]. It extends first-order predicate calculus by following a grammar formalism, which is based on case frames. MRL has been used to access relational databases via typical SQL queries (see previous section and [8]), but also serves as a basis for the translation of a query or document statement into a logical representation, more specifically into a predicate-argument structure.

Our example question, *Who killed Lee Harvey Oswald?* will have as MRL representation:

```
KILL(x, Lee Harvey Oswald)
```

where the argument at the first position denotes the killer, and at the second denotes the victim.

With x as variable, a logical program will look for an instance of the `KILL` predicate in the information source with a known value of the second argument as *Lee Harvey Oswald*. Along these lines is the work of Zettlemoyer and Collins [151], who map sentences to lambda calculus encodings of their meaning λx.person(x) kill(x, *Lee Harvey Oswald*).

The translation of natural language expressions into logical predicate-argument structures often integrates semantic role recognition (see Section 6.4). It relies on handcrafted symbolic knowledge [16, 93], although machine learning techniques gain in importance. As a complex predicate structure is recognized in the data, statistical relational learners are especially targeted [39]. Many existing logical formalisms integrate probabilistic statements, and probabilistic relational learners for recognising the logical forms become popular [117].

In addition to predicate-argument structures as logical representations, there have been several attempts to use direct natural language representations (which are more easier to extract at the expense of a generic usage) in question answering (e.g., [85]). Following this approach the question logical representation will get the form:

Question: *Who killed Lee Harvey Oswald?*
Question's logical representation: `person(x2) & kill(e1, x2, x6) &`
`Lee(x3) & Harvey(x4) & Oswald(x5) & nn_NNC(x6, x3, x4, x5)`

One example is the Logic Prover called COGEX [85]. The underlying idea in COGEX is to use syntactic-based relationships, such as syntactic objects, syntactic subjects, prepositional attachments, complex nominals and adverbial/adjectival adjuncts as logical representations of text.

Retrieval function

When the query and answer-candidate representations are represented using a logical formalism, the relevance of an answer to a query statement can be easily deduced by current theorem proving models [36]. Through knowledge base chaining, a true deduction of the answer can be achieved, even if several pieces are distributed across the knowledge base, which means that the answers are distributed in the different documents from which the knowledge base is built. In the information extraction and question answering fields the multi-document summarisation approach described in [78, 116] can provide insights into answer fusion from distributed answer candidates. External knowledge can supply a bridge between the question and its potential answer in case they do not match directly. As the theoretical framework allows us to introduce notions of confidence and probability, we can also use probabilistic approaches to compute the likelihood of an answer being correct. For instance, certain *deductions* can be relaxed, i.e., certain answers can still be proved even if all patterns do not match, resulting in a probabilistic ranking of the answers [85]. Multiple forms of reasoning are essential to respond to unanticipated questions, and to react when there are no answers in the known data sources (for example, by constraint relaxation). They can also help identify and explain false prepositions or various types of misunderstanding found in questions, generate summaries, or provide conditional responses.

To demonstrate the COGEX principle we use an answer-bearing sentence taken from the Wikipedia page about Lee Harvey Oswald[5]. The question's logical representation is as presented above.

Answer candidate: *Oswald is shot by Jack Ruby.*

---

[5]http://en.wikipedia.org/wiki/Lee_Harvey_Oswald

Answer's logical representation: `Oswald(x2) & nn_NNC(x3, x2) & shot(e1, x6, x2) & by(e1, x6) & Jack(x4) & Ruby(x5) & nn_NNC(x6, x4, x5) & person(x6)`

Supported by a number of linguistic and world knowledge axioms and after several relaxation steps, the prover will be able to establish the relevance between the question and the answer. The problem with the COGEX method lies in the variety of natural language expressions that express similar content and verbal ambiguity. When documents are represented as predicates and arguments, which are the actual words within a body of text, this seriously hampers the possibilities and the accuracy of the inference.

As mentioned above, the retrieval model can be deterministically oriented, with the semantic information acting as extra constraints or filters. Alternatively, the semantic recognition probability can be incorporated into a probabilistic retrieval model. In addition, semantic role labelling also includes the recognition of temporal and spatial expressions in the query. Recognition of temporal information allows reasoning with contextually underspecified temporal expressions (temporal expressions such as *last week* or *two weeks ago*). Moreover, it supports delayed interpretation and reasoning with persistence of events: that is, how long an event or its outcome lasts. When geo-spatial information is also recognised or added to the retrievable documents, the retrieval model can integrate geo-spatial reasoning, however this domain is still lacking in experimental results. Benamara and Saint-Dizier [13] argue in favour of advanced reasoning for QA to infer answers or to provide explanations or justifications of the reasoning patterns and knowledge involved.

In addition, semantic information can also be represented as a graph. Graph representations offer several possibilities for ranking potential answers that may be relevant to a query. If the graphs have undirected edges, techniques of (possibly relaxed) subgraph matching may be applied [87]. Alternatively, we can use algorithms for spreading activation in a graph such as those described by Narayanan and Harabagiu [93]. When the edges of the graph are directed, the most intuitive ranking method is to incorporate the query and potential answer graph into a Bayesian network and thereby make probabilistic inferences based on this network [107]. A famous retrieval model is the Bayesian network model [140], usually called the *inference network model*, which as far as we know has not been used for question answering to date.

Evaluation

The above methods have a large potential for future question answering, though most of the studies regard isolated question answering systems. The described methods are the most advanced in understanding questions and potential answers, at the expense of a large computational complexity. The reported performance gain of the logic-based model in the scope of TREC evaluation improves on average by 30% of correctly answered questions, although we miss here benchmarking studies that pinpoint the merits and bottlenecks. The many different ontological ways by which predicates and arguments can be defined, might be a serious threat to the further development of these approaches as will be seen in the following sections.

6.8. Conclusive findings regarding the methods

The lack of a unique standard framework where all the systems can be compared, the difficulty of system and experiment replication, the number of different ontological ways by which semantic descriptors of question and answers can be defined, and finally the errors introduced by the natural language processing tasks make evaluation, comparison and error analysis of the existing methods difficult. Nevertheless we list a number of conclusive findings.

Estimation of the expected answer type and the semantic classes of information in the question and candidate answers are two essential parts of any question answering system [84]. Deeper semantic analysis of the information source, such as semantic role labelling and discourse analysis, can improve the performance level of question answering, but it usually is a time consuming task. For factual open domain question answering, where the information redundancy online can be exploited, such an analysis might not pay off the effort. On the other hand, for more specific and complex questions a deeper semantic analysis is a key feature of the system.

Nowadays a perfect question answering system should employ not only one or two retrieval models, but rather a set of models that depends on the system requirements, the type of question posed, the type of interrogated data, the type of interface and other criteria. The employment of the models and techniques for question answering should be considered very carefully. The following aspects have to be taken into account when selecting models and techniques for question answering:

- **Type of question answering** (open domain, restricted domain), or

How can information redundancy be exploited?

If a question answering system targets factoid questions, a set of simple mapping rules can be used generating queries to search engines with predefined patterns for the answers. In this case the bag-of-words model is appropriate and delivers the fastest response [17, 14].

- **Types of questions** (factoids, list, definitional vs. causal, relations, procedural), or How deep should the semantic analysis be?

  If the question answering system targets complex questions or data sources with low redundancy, a deeper semantic analysis is required (broader coverage of EATs, semantic role labelling and discourse analysis). For inferencing and answer fusion, the system should deal with logical representations of natural language [86, 13]. In this case, the bag-of-words representation as a pre-filtering step might even hinder the performance.

- **Types of data that is interrogated**, or How deep should the semantic analysis of the information source be?

  When questioning documents that contain essentially factual information, representations of questions and potential answers that capture rather shallow semantics based on expected answer types seem to be sufficient, where also the redundancy of the answer might improve the performance. On the other hand, if the documents contain procedural content or explain why certain events happen, event and discourse structures are important to recognize in the documents, so that *How*- and *Why*-questions can be answered. If the documents contain salient temporal and spatial information, it is likely that questions involving temporal and spatial constraints will be posed, again demanding appropriate semantic processing [125, 40]. In these cases logical representations of natural language and logical retrieval models seem the most promising.

- **Value of wrong answers and no answers**, or What are the consequences of wrong answers?

  Depending on the datasource, it may happen that there is no correct answer in it, although the selected retrieval model may provide some answers. In this case one should decide whether or not a wrong answer or partial answer is better than no answer at all.

- **Response time**, or How fast should the system provide the answer?

Response time is one of the crucial factors of the question-answering performance. Online question answering systems should guarantee short response times. Taking into account the complexity of the question, the system should find a trade-off between the retrieval models, indexing techniques and the time for answer finding. In this case one may consider MapReduce [67] and 'just-in-time' semantic representations [134].

- **Type of users and usability criteria**: lexicon, response time, system familiarity and functional awareness of users, expected response time, or What are the typical scenarios of interaction with the system?

As mentioned at the beginning of the article, open domain systems are developed for a broad group of users whose information need has to be satisfied in a very short time. By contrast, restricted domain QA users may use a specific lexicon or specific semantic representations, and such users are ready for longer response times. Depending on the system requirements, the question-answering technology can be used for investigation purposes when the user is interested not only in the answer, but also in the system's confidence and partial answers, providing multiple answers or tailoring them according to a user's profile [100, 114].

7. Prominent issues

In spite of the variety of information retrieval models, it is possible that the question answering system fails to return an appropriate set of answer candidates at an acceptable level of precision and recall. If the system generates too many incorrect answers, we may wish to improve the precision of the results. In case the system fails to find any answers or generate too few correct answers, we are interested in improving the recall of question answering. QA can be tackled by relaxing the query and eliminating some of the keywords when too few candidates are returned, or by strengthening the query by adding keywords when too many answer-bearing instances are found. Such an approach was incorporated in [106].

In previous sections, we have explored various degrees of processing of the natural language query. We have examined the corresponding processing of the natural language of documents or information extracted from other

media. Starting from a bag of words approach, we have moved to advanced relationships. In general, information extraction techniques can serve a dual purpose. First, the semantic or even syntactic dependency relations in a sentence can enforce additional constraints, all of which must be fulfilled in the retrieval of information. This can make the information more precise. Second, if the assigned classifications of entities and their relations in the query statements and in documents or indexing descriptions of the information objects are fine-grained, then they may serve as substitutes for the words or other low level features of multimedia objects. We can thereby improve the recall of information retrieval results, because the search is more general and low level features (such as certain words in a body of text) can be ignored as they are replaced by more abstract concepts.

In information retrieval, there is a tradition of expanding or replacing the query with related terms or synonyms. The goal is to strengthen the query with additional related information and thus to improve the precision of the search. In this way, we can also cope with varying formulations of similar content in order to improve the recall of the retrieval. In general information retrieval, queries are often expanded by adding keywords or terms, such as those obtained via topic signatures as proposed in [65] that group all terms related to a topic based on statistical association techniques. One example might be the words *table*, *menu*, *waiter*, *order*, *eat*, *pay*, *tip* related to a restaurant visit scenario. By contrast, in question answering we see a tendency to really paraphrase a query. *Paraphrasing* has been a research area for the information retrieval community over the past two decades [71, 154, 120]. For example, the synset *make*, *produce*, *create* has the genus *manufacture*, defined in the gloss of its homomorphic nominalisation *for sale*. Accordingly, the document containing a string *Hallmark remains the largest maker of greeting cards.* is considered as answer-bearing for the natural query *What company sells the most greeting cards?* [106]. Morphological alterations allow us to answer the question *Who invented the paper clip?* where the verb *invented* is mapped to its nominalisation *inventor*, which, in turn, is a sub-category of the estimated answer category `Person`. Lexico-semantic rephrasing patterns can also be learned from corpora, as in [155].

To demonstrate how paraphrasing works, we modify the example question to *How did Lee Harvey Oswald die?* so that no trivial token-based answer can be found. As no (or very few) answer candidates are found, an additional paraphrasing step will allow to broaden the search space and thereby find more answer candidates. For our example we use the principle of lexical

chaining to rephrase the query. By analogy with [106], we use WordNet and look up the word *die*, as one of the tokens in the query. In WordNet the verb *die* is specified as having derivationally related forms containing the noun *death*. A further search results in the entry for *death*, for which in the synset (sense number 7) the word is linked with the hypernyms *killing*, *kill*, and *putting to death*. A new paraphrased query posed against the document collection will deliver a larger set of potential answers.

Another problem that arises with complex pipelined question answering systems, where advanced semantic content representations form the basis for advanced retrieval functions, is that each of the components might not work properly and introduces errors which propagate. Empirically investigating the effect and severeness of these errors is an interesting research question.

8. Applications of question answering

Over more than 50 years of question answering, there have been an enormous number of applications, including both academic prototypes and industrial implementations. In this section, we mention a few example applications that have notable characteristics in terms of domain application, architecture, performance, or research impact.

With the broad use of the World Wide Web, search engines have become critical online services. When conventional search techniques failed to provide concise answers to users' queries, web-based question answering systems were developed. Starting with AskMSR [18] and START [55], many other online question answering services appeared, such as AnswerBus[6], AskJeeves (recently renamed to Ask.com[7]), Answer.com[8], Wondir[9] and others.

Domain-specific applications of question answering are usually not available online, and their development can be traced only by scientific articles and technical reports. The following list of references is by no means complete and simply cites some interesting applications in restricted domain question answering. Diekema *et al.* [31] presented a question answering system aimed at effective information retrieval in the domain of Reusable Launch Vehicles (aerospace engineering). The application is designed for a collaborative

---

[6]http://www.answerbus.com/index.shtml
[7]http://www.ask.com
[8]http://www.answer.com
[9]http://www.wondir.com

learning environment with real users and live questions. Rinaldi *et al.* [119] developed a question answering system that creates a domain-specific knowledge base from a collection of electronic documents and interrogates it with users' queries. This work differs in that the information source is first pre-processed into a logically coherent domain-specific structure which is used for answering questions posed by users. WEBCOOP [12], developed for the tourist domain, utilises a logic-based representation of the data and facilitates its interrogation with natural language queries. A system for automatic railway inquiry over mobile devices, such as mobile phones, was presented in [5]. It shows that the restricted domain question answering technology addresses a broader target user group than domain experts. A comprehensive overview of restricted domain question answering with a list of applications is available in [88].

As domain-specific information is usually very structured and well organised, novel approaches with specific knowledge representation techniques have become more important. In this domain, academic prototypes deal with small-scale knowledge bases that represent the domain and interrogate them using natural language queries. In this context we note the following work: Frank *et al.* [37] with a QA system for structured knowledge sources; Demner-Fushman and Lin [30] with a clinical question answering system; Basili *et al.* [11] on the question answering approach for ontological resources; Ou *et al.* [102] with an approach for ontology-based question answering in the context of *Movies and Cinemas*; and an Ontology-Portable Question Answering System for the Semantic Web (AquaLog) as presented in [70].

Some interesting applications have been recently presented in the scope of the Cross-Language Evaluation Forum (CLEF) and NTCIR Workshop including question answering using Wikipedia [53, 2, 137, 35], question answering over European legislation [41, 4], and question answering with geographical reasoning GikiCLEF [123, 32, 61].

9. The Future of Question Answering

In this survey we considered question answering as an information retrieval task, in which users receive direct answers extracted from documents to their natural language queries instead of the relevant documents. We demonstrated certain compatibility between the old approaches, where question answering was mainly seen as providing a natural language interface to

data- and knowledge bases, and the more recent QA approaches promoted by the TREC, CLEF and NTCIR that focus on open domain QA. Such an approach was motivated by recent advances in natural language understanding of questions, sentences and query statements, to advancements in information retrieval from semi-structured collections, and by recent interest in probabilistic reasoning with content models. In this section we elaborate on promising evolutions of question answering research and anticipate their future impact. We go deeper into the analysis of the questions and textual information, the retrieval models, evaluation of future question answering systems and promising applications.

Starting with simple representations, in this survey we have gradually increased the complexity of the representations of queries and documents. The syntactic and semantic analyses of the question and of document sentences in a textual database have two purposes. First, they provide additional constraints in the matching of information need and candidate answer sentences in order to potentially increase the precision of the resulting answers. The syntactic structure detected in the sentences or the disambiguation of phrases by assigning them semantic labels are examples of such constraints. We also classify expected answer type finding under this category, as this also has the potential to make the search for the answer more precise. Second, the analysis may translate the content into a more general form, which might improve the possibilities of matching and thus can improve the retrieval recall. Natural language analyses eventually lead to a translation of the statements into a structured form such as a structured query or sentences, or into first-order logic representations.

Questions or query statements in natural language represent the detailed information needs of a user. A high degree of the natural language expressiveness allows us to structure the entities (e.g., object, persons) and their relationships into sentences. In order to accurately answer such questions, rich representations of queries and document content need to be built that contain the necessary entity relations. This is confirmed by current research that sees the Web as a knowledge base [150]. First, current natural language understanding is far from perfect. We can extract explicit factual information from text, but the extraction of conceptual information that also requires discourse understanding remains a distant goal. In such a setting linking of information within and across different documents is very important. We need ways of accurately identifying co-referring information (within and across documents, within and across media). Current latent class models

that learn the different phrasings of a relation between entities from a large corpus with no or little human supervision or annotation seem promising. A first step in this direction is made by Lita and Carbonell [68], and Chali and Joty [21]. We need ways to more correctly deal with temporal and spatial information.

Increasingly we will have to deal with spoken questions. Their transcriptions might contain noise, if the speech is intertwined with background, the speaker has a different accent, etc. This poses additional challenges to the analysis of the questions. Often the speech transcription offers several probable transcriptions, i.e., several valid translations of the query into a structured or logical language, opening the way for probabilistic logical representations. We do not have experience in analysing spoken queries, but their analysis is an important future research topic. A good start in spoken language question answering has been presented in [128] as well as QA on mobile devices [80].

The next generation of question answering systems will have to take into account today's multimedia data. There exists a mixture of natural language text, images, video, audio, user added tags, and metadata. On the question side, users may express their queries using a variety of modalities. Such a situation requires a QA system to semantically label and analyse content, generate and integrate representations for all types of content, fuse the recognised content (possibly combined with user-generated metadata), and maybe reason with the content and compute a relevance score for the candidate answers. The semantic analysis might be performed in a cross-lingual or cross-media setting, respectively referring to cross-lingual and cross-media question answering. In addition, many current (multimedia) repositories contain controlled metadata added at creation time (e.g., time pointers to video frames). Future retrieval functions have to integrate these different sources of information.

Such an evolution perfectly maps onto the domain of *semantic search*. Semantic search is the collective name for methods that look at information that is available on the Web or within organisations, beyond the level of individual words or phrases. Understanding content at the semantic level promises to facilitate the tasks of existing information retrieval engines, but it also allows us to aggregate, compare and reason with content as structured data.

There are, however, many research questions that are yet to be addressed. We need to investigate how structured document collections obtained through information extraction techniques can help us to select efficient query pro-

cessing methods. We need ways to reason with the representations efficiently, taking into account uncertain representations. Current graph-based models, whether they represent undirected graphs (e.g., Markov networks) or directed graphs (e.g., Bayesian networks), may help guide this work. In this survey, we have explained different query processing techniques of various complexity. However, very little research has been published to date regarding efficient and scalable indices and search structures for question answering retrieval.

In addition, as in any information retrieval task we need to deal with context identification, user-adapted answers and effective interfaces in order to improve the quality of the given answers and the usability of the QA systems. Questions might be short or ambiguous lacking the necessary context. Answers might be complex and aggregated from different sources needing advanced navigation, exploration and visualisation tools. In this respect a 2003 Roadmap for question answering by Burger et al. [20] is still up to date.

Despite ten years of system evaluations in the TREC QA track for specific kinds of questions and answers [144], the community is still searching for good evaluation procedures that accurately measure progress. Moreover, the past evaluation campaigns showed how challenging the organisation of evaluation set-ups (datasets, question lists, answer lists etc) is. Current metrics ignore dimensions, like breadth of coverage of a system, accuracy and confidence in an answer, speed, and domain adaptability. Largely relying on natural language techniques and external resources, the real impact of such techniques and resources cannot be estimated since the experiments are non-replicable. Another issue that hampers the transparent evaluations of question answering systems is the standard ground truth corpora. Due to legal issues, expensive human effort in cleaning up the data, and the composing the list of the questions with a corresponding list of answers, the available data for question answering evaluations are quite limited. After a long series of evaluations the questions and related answers in the data are already known and can be successfully utilised [46]. A new emerging research field of automated *question generation* may help in producing question lists to new textual datasets [48]. The issues of efficiency, effectiveness and system scalability of question answering also remain neglected, which may be the reason for the little attention to QA as a technology for search engines.

Question answering is a complex task needing effective improvements of different research areas, including information retrieval, natural language processing, database technologies, Semantic Web technologies, human computer interaction, speech processing and computer vision.

10. Conclusions

This paper gave a summary of important methods in question answering over the last decades. It defined question answering as an information retrieval problem where the question is posed in natural language. The queried data can be of any type or medium having in mind current heterogeneous databases consisting of text, images, video and audio possibly augmented with classical table-form database information. We have explained simple, computationally light approaches and gradually introduced advanced forms of analysis of the question (and of documents in case of textual information), yielding advanced content representations. We discussed different levels of processing yielding bag-of-words and more complex representations integrating POS, classification of the expected answer type, semantic roles, discourse analysis, translation of the question into a SQL-like language or logical representation. On one extreme we have keyword-based searches, on the other sophisticated knowledge bases queried with structured or logical queries obtained from the natural language questions, and answers obtained through reasoning. This evolution shows that the old distinction between open domain and restricted domain question answering becomes less pronounced. Also in open domain question answering, questions and content in natural language are translated into more structured forms apart from low level bag-of-words representations, allowing a more effective retrieval of the requested information. The description of this evolution and of the integration of components originating in different disciplines (e.g., information retrieval, databases and computational linguistics) constitutes the main contribution of this paper. Additionally, the paper pinpoints many research questions to be studied in the years to come.

11. Acknowledgments

References

[1] *Working Notes for the CLEF 2009 Workshop.* 2009. ISBN 978-88-88506-84-5.

[2] Sisay Fissaha Adafre, Valentin Jijkoun, and Maarten de Rijke. The University of Amsterdam at WiQA 2006. In *CLEF 2006 Working Notes*, 2006.

[3] Eugene Agichtein, Chris Burges, and Eric Brill. Question answering over implicitly structured Web content. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 18–25, Washington, DC, USA, 2007. IEEE Computer Society.

[4] Eneko Agirre, Arantxa Otegi, and Hugo Zaragoza. Using semantic relatedness and word sense disambiguation for (CL) IR. In Carol Peters, Giorgio Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peas, and Giovanna Roda, editors, *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, volume 6241 of *Lecture Notes in Computer Science*, pages 166–173. Springer Berlin / Heidelberg, 2010.

[5] Avinash J. Agrawal. Using domain specific question answering technique for automatic railways inquiry on mobile phone. In *Proceedings of the Fifth International Conference on Information Technology: New Generations (ITNG '08)*, pages 1111–1116, Washington, DC, USA, 2008. IEEE Computer Society.

[6] David Ahn, Steven Schockaert, Martine De Cock, and Etienne Kerre. Supporting temporal question answering: Strategies for offline data collection. In *Proceedings of the 5th International Workshop on Inference in Computational Semantics*, pages 127–132, 2006.

[7] Enrique Alfonseca, Marco De Boni, Jos-Luis Jara-Valencia, and Suresh Manandhar. A prototype question answering system using syntactic and semantic information for answer retrieval. In *Proceedings of the 10th Text Retrieval Conference (TREC-10)*, pages 680–686, 2001.

[8] Ion Androutsopoulos. Natural language interfaces to databases - an introduction. *Natural Language Engineering*, 1:29–81, 1995.

[9] Farida Aouladomar. Some foundational linguistic elements for QA systems: An application to e-government services. In *Proceeding of the 2005 Conference on Legal Knowledge and Information Systems: JURIX 2005: The Eighteenth Annual Conference*, pages 81–90, 2005.

[10] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 86–90, Stroudsburg, PA, USA, 1998. ACL.

[11] Roberto Basili, Dorte H. Hansen, Patrizia Paggio, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. Ontological resources and question answering. In *Proceedings of the Workshop on Pragmatics of Question Answering*, pages 78–84, Boston, Massachusetts, USA, 2004. ACL.

[12] Farah Benamara. Cooperative question answering in restricted domains: The WEBCOOP experiment. In *Workshop on Question Answering in Restricted Domains. 42nd Annual Meeting of the Association for Computational Linguistics (ACL-2004)*, pages 31–38, Barcelona, Spain, 2004. ACL.

[13] Farah Benamara and Patrick Saint-Dizier. Advanced relaxation for cooperative question answering. In Mark T. Maybury, editor, *New Directions in Question Answering*, pages 263–274. AAAI Press, 2004.

[14] Matthew W. Bilotti, Paul Ogilvie, Jamie Callan, and Eric Nyberg. Structured retrieval for question answering. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 351–358, New York, NY, USA, 2007. ACM.

[15] Fischer Black. Deductive question answering system. In M. Minsky, editor, *Semantic Information Processing*, pages 354–402. MIT Press, Cambridge, Massachusetts, 1968.

[16] Patrick Blackburn and Johan Bos. *Representation and Inference for Natural Language. A First Course in Computational Semantics.* Center for the Study of Language and Information - Lecture Notes, 2005.

[17] Eric Brill, Jimmy Lin, Michele Banko, Susan Dumais, and Andrew Ng. Data-intensive question answering. In *Proceedings of the Tenth Text Retrieval Conference (TREC)*, pages 393–400, 2001.

[18] Eric Brill, Susan Dumais, and Michele Banko. An analysis of the AskMSR question-answering system. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP '02)*, pages 257–264, Morristown, NJ, USA, 2002. ACL.

[19] Razvan C. Bunescu and Raymond J. Mooney. Subsequence kernels for relation extraction. In *Advances in Neural Information Processing Systems*, pages 171–178, Cambridge, MA, 2005. MIT Press.

[20] John Burger, Claire Cardie, Vinay Chaudhri, Robert Gaizauskas, Sanda Harabagiu, David Israel, Christian Jacquemin, Chin-Yew Lin, Steve Maiorano, George Miller, Dan Moldovan, Bill Ogden, John Prager, Ellen Riloff, Amit Singhal, Rohini Shrihari, Tomek Strzalkowski, Ellen Voorhees, and Ralph Weischedel. Issues, tasks and program structures to roadmap research in question & answering (Q&A). Technical report, NIST, 2001.

[21] Yllias Chali and Shafiq R. Joty. Improving the performance of the random walk model for answering complex questions. In *HLT '08: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies*, pages 9–12, Morristown, NJ, USA, 2008. ACL.

[22] Michael Collins and Nigel Duffy. Convolution kernels for natural language. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, pages 625–632, Cambridge, MA, 2001. MIT Press.

[23] Ann Copestake and Karen Sparck Jones. Natural language interfaces to databases. *Knowledge Engineering Review*, 5:225–249, 1990.

[24] Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom M. Mitchell, Kamal Nigam, and Seán Slattery. Learning to construct knowledge bases from the World Wide Web. *Artificial Intelligence*, 118(1-2):69–113, 2000.

[25] Bruce Croft, Donal Metzler, and Trevor Stroghman. *Search Engines: Information Retrieval in Practice*. Pearson, 2009.

[26] Bruce W. Croft and John Lafferty. *Language Modeling for Information Retrieval*. Kluwer Academic Publishers, Norwell, MA, USA, 2003.

[27] Hang Cui, Renxu Sun, Keya Li, Min-Yen Kan, and Tat-Seng Chua. Question answering passage retrieval using dependency relations. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 400–407, New York, NY, USA, 2005. ACM.

[28] Hoa Trang Dang, Diane Kelly, and Jimmy J. Lin. Overview of the TREC 2007 question answering track. In *Proceedings of The Sixteenth Text REtrieval Conference (TREC 2007)*, 2007.

[29] Rodrigo de Salvo Braz, Roxana Girju, Vasin Punyakanok, Dan Roth, and Mark Sammons. Knowledge representation for semantic entailment and question-answering. In *IJCAI-05 Workshop on Knowledge and Reasoning for Answering Questions*, 2005.

[30] Dina Demner-Fushman and Jimmy J. Lin. Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*, 33(1):63–103, 2007.

[31] Anne R. Diekema, Ozgur Yilmazel, Jiangping Chen, Sarah Harwell, Elizabeth D. Liddy, and Lan He. What do you mean? Finding answers to complex questions. In *Proceedings of the AAAI Spring Symposium: New Directions in Question Answering*, pages 87–93, 2003.

[32] Iustin Dornescu. Semantic QA for encyclopaedic questions: EQUAL in GikiCLEF. In Carol Peters, Giorgio Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peas, and Giovanna Roda, editors, *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, volume 6241 of *Lecture Notes in Computer Science*, pages 326–333. Springer Berlin / Heidelberg, 2010.

[33] Ramez Elmasri and Shamkant B. Navathe. *Fundamentals of Database Systems (5th Edition)*. Addison Wesley, 2006.

[34] Guzen Erozel, Nihan Kesim Cicekli, and Ilyas Cicekli. Natural language querying for video databases. *Information Sciences*, 178(12):2534 – 2552, 2008.

[35] Sergio Ferrández, Antonio Toral, Óscar Ferrández, Antonio Ferrández, and Rafael Muñoz. Exploiting Wikipedia and EuroWordNet to solve

cross-lingual question answering. *Information Sciences*, 179(20):3473 – 3488, 2009.

[36] Melvin Fitting. *First-order Logic and Automated Theorem Proving (2nd Edition)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1996.

[37] Anette Frank, Hans-Ulrich Krieger, Feiyu Xu, Hans Uszkoreit, Berthold Crysmann, Brigitte Jörg, and Ulrich Schäfer. Question answering from structured knowledge sources. *Journal of Applied Logic*, 5(1):20–48, 2007.

[38] Norbert Fuhr, Jaap Kamps, Mounia Lalmas, and Andrew Trotman, editors. *Focused Access to XML Documents, 6th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2007, Selected Papers*, volume 4862 of *Lecture Notes in Computer Science*, 2008. Springer.

[39] Lise Getoor and Ben Taskar. *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2007.

[40] Fredric Gey, Ray Larson, Mark Sanderson, Hideo Joho, Paul Clough, and Vivien Petras. GeoCLEF: The CLEF 2005 cross-language geographic information retrieval track overview. In Carol Peters, Fredric Gey, Julio Gonzalo, Henning M uller, Gareth Jones, Michael Kluck, Bernardo Magnini, Maarten de Rijke, and Danilo Giampiccolo, editors, *Accessing Multilingual Information Repositories*, volume 4022 of *Lecture Notes in Computer Science*, pages 908–919. Springer Berlin / Heidelberg, 2006.

[41] Ingo Gloeckner and Bjoern Pelzer. Extending a logic-based question answering system for administrative texts. In Carol Peters, Giorgio Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peñas, and Giovanna Roda, editors, *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, volume 6241 of *Lecture Notes in Computer Science*, pages 265–272. Springer Berlin / Heidelberg, 2010.

[42] B. F. Green, A. K. Wolf, C. Chomsky, and K. Laughery. BASEBALL: An automatic question answerer. In *Proceedings of Western Computing Conference*, volume 19, pages 219–224, 1961.

[43] Sanda Harabagiu and Cosmin Adrian Bejan. Question answering based on temporal inference. In *Proceedings of the AAAI-2005 Workshop on Inference for Textual Question Answering*, 2005.

[44] Sanda Harabagiu and Andrew Hickl. Methods for using textual entailment in open-domain question answering. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 905–912, Stroudsburg, PA, USA, 2006. ACL.

[45] Sanda Harabagiu, Dan Moldovan, Rada Mihalcea, Mihai Surdeanu, and Vasile Rus. Falcon: Boosting knowledge for answer engines. In *Proceedings of Text REtrieval Conference (TREC-9)*, pages 479–488, 2000.

[46] Matthias H. Heie, Edward W. D. Whittaker, and Sadaoki Furui. Optimizing question answering accuracy by maximizing Log-Likelihood. In *Proceedings of the ACL 2010*, pages 236–240. ACL, 2010.

[47] Michael Heilman and Noah A. Smith. Tree edit models for recognizing textual entailments, paraphrases, and answers to questions. In *uman Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1011–1019, Stroudsburg, PA, USA, 2010. ACL.

[48] Michael Heilman and Noah A. Smith. Good question! Statistical ranking for question generation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, pages 609–617, Stroudsburg, PA, USA, 2010. ACL.

[49] Andrew Hickl, John Williams, Jeremy Bensley, Kirk Roberts, Ying Shi, and Bryan Rink. Question answering with LCC's CHAUCER at TREC 2006. In *Proceedings of Text REtrieval Conference TREC 2006*, 2006.

[50] L. Hirschman and R. Gaizauskas. Natural language question answering: The view from here. *Natural Language Engineering*, 7(4):275–300, 2001.

[51] Eduard Hovy, Laurie Gerber, Ulf Hermjakob, Deepak Ravichandran, Chin-Yew Lin, and Deepak Ravich. Toward semantics-based answer pinpointing. In *Proceedings of the First International Conference on Human Language Technology Research*, pages 1–7, Stroudsburg, PA, USA, 2001. ACL.

[52] Eduard Hovy, Ulf Hermjakob, and Deepak Ravichandran. A question/answer typology with surface text patterns. In *Proceedings of the second international conference on Human Language Technology Research*, pages 247–251, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[53] Valentin Jijkoun and Maarten de Rijke. Overview of WiQA 2006. In *Evaluation of Multilingual and Multi-modal Information Retrieval*, pages 265–274, Berlin, Heidelberg, 2006. Springer-Verlag.

[54] Michael Kaisser and Tilman Becker. Question answering by searching large corpora with linguistic methods. In *Proceedings of the 13th Text REtrieval Conference (TREC 2004)*, Gaithersburg, 2004.

[55] Boris Katz, Jimmy J. Lin, and Sue Felshin. The START multimedia information system: Current technology and future directions. In *Proceedings of the International Workshop on Multimedia Information Systems*, pages 117–123, 2002.

[56] Boris Katz, Sue Felshin, Jimmy J. Lin, and Gregory Marton. Viewing the Web as a virtual database for question answering. In Mark T. Maybury, editor, *New Directions in Question Answering*, pages 215–226. AAAI Press, 2004.

[57] Boris Katz, Gary Borchardt, and Sue Felshin. Syntactic and semantic decomposition strategies for question answering from multiple sources. In *Proceedings of the AAAI 2005 Workshop on Inference for Textual Question Answering*, pages 35–41, 2005.

[58] Parisa Kordjamshidi, Martijn van Otterlo, and Marie-Francine Moens. From language towards formal spatial calculi. In *Proceedings of Computational Models of Spatial Language Interpretation (COSLI)*, pages 17–24, 2010.

[59] Onur Kucuktunc, Ugur Gudukbay, and Ozgur Ulusoy. A natural language-based interface for querying a video database. *IEEE MultiMedia*, 14(1):83–89, 2007.

[60] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[61] Ray R. Larson. Interactive probabilistic search for GikiCLEF. In Carol Peters, Giorgio Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peas, and Giovanna Roda, editors, *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, volume 6241 of *Lecture Notes in Computer Science*, pages 334–341. Springer Berlin / Heidelberg, 2010.

[62] Wendy Lehnert. A conceptual theory of question answering. In *Proceedings of the Workshop on Reading Comprehension Texts as Evaluation for Computer-Based Language Understanding Systems*, pages 651–657, Seattle, 1977. ACL.

[63] Xiaoyan Li. Syntactic features in question answering. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, pages 383–384, New York, NY, USA, 2003. ACM.

[64] Xin Li and Dan Roth. Learning question classifiers: The role of semantic information. *Natural Language Engineering*, 12(3):229–249, 2006.

[65] Chin-Yew Lin and Eduard H. Hovy. The automated acquisition of topic signatures for text summarization. In *Proceedings of the 18th Conference on Computational Linguistics*, pages 495–501, Morristown, NJ, USA, 2000. ACL.

[66] Dekang Lin. Dependency-based evaluation of MINIPAR. In *Proceedings of the Workshop on the Evaluation of Parsing Systems*, Granada, 1998.

[67] Jimmy Lin and Chris Dyer. *Data-Intensive Text Processing with MapReduce*. Morgan & Claypool Publishers, 2010.

[68] Lucian Vlad Lita and Jaime Carbonell. Unsupervised question answering data acquisition from local corpora. In *CIKM '04: Proceedings of the Thirteenth ACM International Conference on Information and*

Knowledge Management, pages 607–614, New York, NY, USA, 2004. ACM.

[69] Xu Liu and David Doermann. Mobile Retriever: Access to digital documents from their physical source. *International Journal on Document Analysis and Recognition*, 11(1):19–27, September 2008.

[70] Vanessa Lopez, Michele Pasin, and Enrico Motta. AquaLog: An ontology-portable question answering system for the semantic Web. In *Proceedings of the 2nd European Semantic Web Conference (ESWC)*, pages 546–562, 2005.

[71] Barry G. T. Lowden and Anne N. De Roeck. The REMIT system for paraphrasing relational query expressions into natural language. In *VLDB '86: Proceedings of the 12th International Conference on Very Large Data Bases*, pages 365–371, San Francisco, CA, USA, 1986. Morgan Kaufmann Publishers Inc.

[72] Bernardo Magnini, Alessandro Vallin, Christelle Ayache, Gregor Erbach, Anselmo Peñas, Maarten de Rijke, Paulo Rocha, Kiril Simov, and Richard Sutcliffe. Overview of the CLEF 2004 multilingual question answering track. In Carol Peters, Paul Clough, Julio Gonzalo, Gareth J. F. Jones, Michael Kluck, and Bernardo Magnini, editors, *Multilingual Information Access for Text, Speech and Images*, volume 3491 of *Lecture Notes in Computer Science*, pages 371–391. Springer Berlin / Heidelberg, 2005.

[73] Bernardo Magnini, Danilo Giampiccolo, Pamela Forner, Christelle Ayache, Valentin Jijkoun, Petya Osenova, Anselmo Peñas, Paulo Rocha, Bogdan Sacaleanu, and Richard Sutcliffe. Overview of the CLEF 2006 multilingual question answering track. In Carol Peters, Paul Clough, Fredric Gey, Jussi Karlgren, Bernardo Magnini, Douglas Oard, Maarten de Rijke, and Maximilian Stempfhuber, editors, *Evaluation of Multilingual and Multi-modal Information Retrieval*, volume 4730 of *Lecture Notes in Computer Science*, pages 223–256. Springer Berlin / Heidelberg, 2007.

[74] Kavi Mahesh and Sergei Nirenburg. A situated ontology for practical NLP. In *Proceedings of the Workshop on Basic Ontological Issues in*

Knowledge Sharing, International Joint Conference on Artificial Intelligence (IJCAI-95), 1995.

[75] Daniel Marcu. The rhetorical parsing of unrestricted texts: a surface-based approach. *Computational Linguistics*, 26(3):395–448, 2000.

[76] Mark T. Maybury, editor. *New Directions in Question Answering*, 2004. AAAI Press.

[77] Andrew McCallum, Karl Schultz, and Sameer Singh. Factorie: Probabilistic programming via imperatively defined factor graphs. In *Advances on Neural Information Processing Systems*, pages 1249–1257, 2009.

[78] Kathleen R. McKeown, Vasileios Hatzivassiloglou, Regina Barzilay, Barry Schiffman, David Evans, and Simone Teufel. Columbia multi-document summarization: Approach and evaluation. In *Proceedings of the Document Understanding Conference (DUC 2001)*, pages 43–64, 2001.

[79] George A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38:39–41, November 1995.

[80] Taniya Mishra and Srinivas Bangalore. Qme!: A speech-based question-answering system on mobile devices. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, pages 53–63. ACL, 2010.

[81] Teruko Mitamura, Eric Nyberg, Hideki Shima, Tsuneaki Kato, Tatsunori Mori, Chin-Yew Lin, Ruihua Song, Chuan-Jie Lin, Tetsuya Sakai, Donghong Ji, and Donghong Ji. Overview of the NTCIR-7 ACLIA tasks: Advanced cross-lingual information access. In *Proceedings of NTCIR-7 Workshop Meeting*, pages 77–102, 2008.

[82] Ruslan Mitkov, Le An Ha, and Nikiforos Karamanis. A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*, 12(2):177–194, 2006.

[83] Marie-Francine Moens. *Information Extraction: Algorithms and Prospects in a Retrieval Context (The Information Retrieval Series 21)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[84] Dan Moldovan, Marius Pasca, Sanda Harabagiu, and Mihai Surdeanu. Performance issues and error analysis in an open-domain question answering system. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 33–40, Stroudsburg, PA, USA, 2002. ACL.

[85] Dan Moldovan, Christine Clark, Sanda Harabagiu, and Steve Maiorano. Cogex: A logic prover for question answering. In *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 87–93, Morristown, NJ, USA, 2003. ACL.

[86] Dan Moldovan, Christine Clark, and Sanda Harabagiu. Temporal context representation and reasoning. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 1099–1104, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.

[87] Diego Mollá. Learning of graph-based question answering rules. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, pages 37–44, Stroudsburg, PA, USA, 2006. ACL.

[88] Diego Mollá and José Luis Vicedo. Question answering in restricted domains: An overview. *Computational Linguistics*, 33(1):41–61, 2007.

[89] Diego Mollá-Aliod and José-Luis Vicedo. Question answering. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing, Second Edition*. CRC Press, Taylor and Francis Group, Boca Raton, FL, 2010.

[90] Thomas S. Morton. Using coreference for question answering. In *Proceedings of the Workshop on Coreference and its Applications*, CorefApp '99, pages 85–89, Morristown, NJ, USA, 1999. ACL.

[91] Alessandro Moschitti. Syntactic and semantic kernels for short text pair categorization. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 576–584, Stroudsburg, PA, USA, 2009. ACL.

[92] Alessandro Moschitti and Silvia Quarteroni. Linguistic kernels for answer re-ranking in question answering systems. *Information Processing & Management*, 2010.

[93] Srini Narayanan and Sanda Harabagiu. Question answering based on semantic structures. In *Proceedings of the 20th International Conference on Computational Linguistics*, Morristown, NJ, USA, 2004. ACL.

[94] Ani Nenkova, Rebecca Passonneau, and Kathleen McKeown. The Pyramid Method: Incorporating human content selection variation in summarization evaluation. *ACM Transactions on Speech and Language Processing*, 4(2):4, 2007.

[95] Anh Kim Nguyen and Huong Thanh Le. Natural language interface construction using semantic grammars. In *Proceedings of the Thenth Pacific Rim International Conference on Artificial Intelligence (PRICAI-08)*, pages 728–739, 2008.

[96] Joakim Nivre, Johan Hall, and Jens Nilsson. Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, pages 2216–2219, 2006.

[97] Adrian Novischi and Dan I. Moldovan. Question answering with lexical chains propagating verb arguments. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 897–904, Stroudsburg, PA, USA, 2006. ACL.

[98] Bahadorreza Ofoghi, John Yearwood, and Liping Ma. The impact of semantic class identification and semantic role labeling on natural language answer extraction. In *Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval*, pages 430–437, Berlin, Heidelberg, 2008. Springer-Verlag.

[99] Bill Ogden, Jim Cowie, Eugene Ludovik, Hugo Molina-Salgado, Sergei Nirenburg, Nigel Sharples, and Svetlana Sheremtyeva. CRLs TREC-8 systems cross-lingual IR, and Q&A. In *Proceedings of the 8th Text REtrieval Conference (TREC-8)*, pages 513–523, 1999.

[100] William Ogden, James Mcdonald, Philip Bernick, and Roger Chadwick. Habitability in question-answering systems. In Tomek Strzalkowski and Sanda M. Harabagiu, editors, *Text, Speech and Language Technology*, pages 457–473. Springer Netherlands, New York, 2006.

[101] Óscar Ferrández, Rubén Izquierdo, Sergio Ferrández, and José Luis Vicedo. Addressing ontology-based question answering with collections of user queries. *Information Processing & Management*, 45(2):175–188, 2009.

[102] Shiyan Ou, Constatntin Orasan, Dalila Mekhaldi, and Laura Hasler. Automatic question pattern generation for ontology-based question answering. In *Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, pages 183–188, Cocnut Grove, Florida, USA, 2008.

[103] Majdi Owda, Zuhair Bandar, and Keeley Crockett. Conversation-based natural language interface to relational databases. In *Proceedings of IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops*, pages 363–367, Los Alamitos, CA, USA, November 2007.

[104] Raquel Mochales Palau and Marie-Francine Moens. Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the International Conference on Artificial Intelligence and Law*, pages 98–107, New York, NY, USA, 2009. ACM.

[105] Martha Palmer, Paul Kingsbury, and Daniel Gildea. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106, March 2005.

[106] Marius Pasca and Sanda M. Harabagiu. High performance question/answering. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 366–374, New York, NY, USA, 2001. ACM.

[107] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[108] A. Peñas, P. Forner, R. Sutcliffe, A. Rodrigo, C. Forascu, I. Alegria, D. Giampiccolo, N. Moreau, and P. Osenova. Overview of ResPubliQA 2009: Question answering evaluation over european legislation. In *Proceedings of the 10th Cross-Language Evaluation Forum Conference*

on *Multilingual Information Access Evaluation: Text Retrieval Experiments*, CLEF'09, pages 174–196, Berlin, Heidelberg, 2009. Springer-Verlag.

[109] Phi The Pham, Marie-Francine Moens, and Tinne Tuytelaars. Cross-media alignment of names and faces. *IEEE Transactions on Multimedia*, 12(1):13–27, January 2010.

[110] Ana-Maria Popescu, Oren Etzioni, and Henry Kautz. Towards a theory of natural language interfaces to databases. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, pages 149–157, New York, NY, USA, 2003. ACM.

[111] James Pustejovsky, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo. The TimeBank corpus. In *Corpus Linguistics*, pages 647–656, 2003.

[112] James Pustejovsky, Roser Saurí, José M. Castaño, Dragomir R. Radev, Robert J. Gaizauskas, Andrea Setzer, Beth Sundheim, and Graham Katz. Representing temporal and event knowledge for QA systems. In Mark T. Maybury, editor, *New Directions in Question Answering*, pages 99–112. AAAI Press, 2004.

[113] James Pustejovsky, José Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. TimeML: Robust specification of event and temporal expressions in text. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, Tilburg, January 2003.

[114] Silvia Quarteroni and Suresh Manandhar. Designing an interactive open-domain question answering system. *Natural Language Engineering*, 15(1):73–95, January 2009.

[115] Dragomir Radev, Weiguo Fan, Hong Qi, Harris Wu, and Amardeep Grewal. Probabilistic question answering on the Web. In *WWW '02: Proceedings of the 11th International Conference on World Wide Web*, pages 408–419, New York, NY, USA, 2002. ACM.

[116] Dragomir R. Radev. Experiments in single and multidocument summarization using MEAD. In *Proceedings of the Document Understanding Conference (DUC 2001)*, 2001.

[117] Luc De Raedt and Kristian Kersting. Probabilistic inductive logic programming. In *Probabilistic Inductive Logic Programming*, pages 1–27, 2008.

[118] Vinitha Reddy, Kyle Neumeier, Joshua McFarlane, Jackson Cothren, and Craig W. Thompson. Extending a natural language interface with geospatial queries. *IEEE Internet Computing*, 11:82–85, 2007.

[119] Fabio Rinaldi, James Dowdall, Michael Hess, Diego Mollá-Aliod, Rolf Schwitter, and Kaarel Kaljurand. Knowledge-based question answering. In *Proceedings of the Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems*, volume 2773 of *Lecture Notes in Computer Science*, pages 785–792. Springer Berlin / Heidelberg, 2003.

[120] Fabio Rinaldi, James Dowdall, Kaarel Kaljurand, Michael Hess, and Diego Mollá. Exploiting paraphrases in a question answering system. In *Proceedings of the Second International Workshop on Paraphrasing*, pages 25–32, Morristown, NJ, USA, 2003. ACL.

[121] Dmitri Roussinov, Weiguo Fan, and José Robles-Flores. Beyond keywords: Automated question answering on the Web. *Communications of the ACM*, 51(9):60–65, 2008.

[122] Erik Tjong Kim Sang, Gosse Bouma, and Maarten de Rijke. Developing offline strategies for answering medical questions. In *Proceedings of the AAAI-05 Workshop on Question Answering in Restricted Domains*, pages 41–45. Pittsburgh, PA, USA, 2005.

[123] Diana Santos and Lus Miguel Cabral. GikiCLEF: Crosscultural issues in an international setting: Asking non-English-centered questions to Wikipedia. In *CLEF 2009 Working Notes*, 2009.

[124] Estela Saquete, José Luis Vicedo González, Patricio Martínez-Barco, Rafael Muñoz, and Fernando Llopis. Evaluation of complex temporal questions in CLEF-QA. In Carol Peters, Paul Clough, Julio Gonzalo,

Gareth J. F. Jones, Michael Kluck, and Bernardo Magnini, editors, *Multilingual Information Access for Text, Speech and Images*, volume 3491 of *Lecture Notes in Computer Science*, pages 591–596. Springer Berlin / Heidelberg, 2005.

[125] Estela Saquete, José Luis Vicedo González, Patricio Martínez-Barco, Rafael Muñoz, and Hector Llorens. Enhancing QA systems with complex temporal question processing capabilities. *Journal of Artificial Intelligence Research*, 35:775–811, 2009.

[126] Yutaka Sasaki, Hsin-Hsi Chen, Kuang hua Chen, and Chuan-Jie Lin. Overview of the NTCIR-5 cross-lingual question answering task (CLQA1). In *Proceedings of NTCIR-5 Workshop Meeting*, pages 230–235, 2005.

[127] Barry Schiffman, Kathleen McKeown, Ralph Grishman, and James Allan. Question answering using integrated information retrieval and information extraction. In *Proceedings of the Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, pages 532–539, Rochester, New York, April 2007. ACL.

[128] Edward Schofield and Zhiping Zheng. A speech interface for open-domain question-answering. In *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 177–180, Morristown, NJ, USA, 2003. ACL.

[129] Satoshi Sekine and Chikashi Nobata. Definition, dictionary and tagger for extended named entities. In *Proceedings of International Conference on Language Resources and Evaluation (LREC'04)*, pages 1977–1980, 2004.

[130] Dan Shen and Mirella Lapata. Using semantic role to improve question answering. In *Proceedings of Empirical Methods in Natural Language Processing*, pages 12–21, Prague, 2007. ACL.

[131] J. Sherwani, Dong Yu, Tim Paek, Mary Czerwinski, Y. C. Ju, and Alex Acero. Voicepedia: Towards speech-based access to unstructured information. In *Proceedings of the 8th Annual Conference of the International Communication Association (Interspeech 2007)*, pages 146–149. International Speech Communication Association, 2007.

[132] Edward H. Shortliffe. *Computer Based Medical Consultations: MYCIN*. American Elsevier, 1976.

[133] R. F. Simmons. Answering English questions by computer: a survey. *Communications of the ACM*, 8(1):53–70, 1965.

[134] Sharon Small and Tomek Strzalkowski. HITIQA: High-quality intelligence through interactive question answering. *Natural Language Engineering*, 15(1):31–54, 2009.

[135] M. M. Soubbotin and S.M. Soubbotin. Patterns of potential answer expressions as clues to the right answers. In *Proceedings of the Tenth Text REtrieval Conference (TREC)*, pages 293–302, 2001.

[136] Rohini Srihari and Wei Li. A question answering system supported by information extraction. In *Proceedings of the Sixth Conference on Applied Natural Language Processing*, pages 166–172, Stroudsburg, PA, USA, 2000. ACL.

[137] Richard F. E. Sutcliffe, Josef Steinberger, Udo Kruschwitz, Mijail Alexandrov-Kabadjov, and Massimo Poesio. Identifying novel information using latent semantic analysis in the WiQA task at CLEF 2006. In Carol Peters, Paul Clough, Fredric Gey, Jussi Karlgren, Bernardo Magnini, Douglas Oard, Maarten de Rijke, and Maximilian Stempfhuber, editors, *Evaluation of Multilingual and Multi-modal Information Retrieval*, volume 4730 of *Lecture Notes in Computer Science*, pages 541–549. Springer Berlin / Heidelberg, 2007.

[138] Jun Suzuki, Hirotoshi Taira, Yutaka Sasaki, and Eisaku Maeda. Question classification using HDAG kernel. In *Proceedings of the Workshop on Multilingual Summarization and Question Answering*, pages 61–68, Stroudsburg, PA, USA, 2003. ACL.

[139] Valentin Tablan, Danica Damljanovic, and Kalina Bontcheva. A natural language query interface to structured information. In *Proceedings of the 5th European semantic web conference on The semantic web: research and applications*, ESWC'08, pages 361–375, Berlin, Heidelberg, 2008. Springer-Verlag.

[140] Howard R. Turtle and W. Bruce Croft. Inference networks for document retrieval. In *Proceedings of the 13th Annual International ACM*

SIGIR Conference on Research and Development in Information Retrieval, pages 1–24, New York, NY, USA, 1990. ACM.

[141] Suzan Verberne, Lou Boves, Nelleke Oostdijk, and Peter-Arno Coppen. Using syntactic information for improving Why-question answering. In *Proceedings of the International Conference on Computational Linguistics*, COLING '08, pages 953–960, Stroudsburg, PA, USA, 2008. ACL.

[142] Ellen Voorhees and Donna Harman. TREC*: Experiment and Evaluation in Information Retrieval.* The MIT Press, Cambridge, MA, 2005.

[143] Ellen M. Voorhees. The TREC-8 question answering track report. In *Proceedings of TREC-8*, pages 77–82, 1999.

[144] Ellen M. Voorhees. TREC: Continuing information retrieval's tradition of experimentation. *Communications of the ACM*, 50(11):51–54, 2007.

[145] Nick Webb and Bonnie Webber, editors. *Special Issue on Interactive Question Answering.* Cambridge University Press, 2008.

[146] Bonnie Webber and Nick Webb. Question answering. In *The Handbook of Computational Linguistics and Natural Language Processing*, pages 630–654. Wiley-Blackwell, 2010.

[147] E. W. D. Whittaker, J. Mrozinski, and S. Furui. Factoid question answering with web, mobile and speech interfaces. In *Proceedings of the NAACL-HLT(2006) Companion volume: demonstrations*, pages 288–291, Stroudsburg, PA, USA, 2006. ACL.

[148] Terry Winograd. Procedures as a representation for data in a computer program for understanding natural language. *Cognitive Psychology*, 3 (1):1–191, 1972.

[149] W. A. Woods, R. A. Kaplan, and B. Nash-Webber. The lunar sciences natural language information system: Final report: BBN Report #2378. Technical report, Bolt Beranek and Newman Inc.,Cambridge, MA., June 1972.

[150] Alexander Yates. Extracting world knowledge from the Web. *IEEE Computer*, 42(6):94–97, 2009.

[151] Luke S. Zettlemoyer and Michael Collins. Learning context-dependent mappings from sentences to logical form. In *ACL-IJCNLP '09: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2*, pages 976–984, Morristown, NJ, USA, 2009. ACL.

[152] Dell Zhang and Wee Sun Lee. Question classification using support vector machines. In *SIGIR '03: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 26–32, New York, NY, USA, 2003. ACM.

[153] Zhiping Zheng. AnswerBus question answering system. In *Proceedings of the Conference on Human Language Technology*, pages 399–404, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[154] Ingrid Zukerman and Bhavani Raskutti. Lexical query paraphrasing for document retrieval. In *Proceedings of the 19th International Conference on Computational Linguistics*, pages 1–7, Morristown, NJ, USA, 2002. ACL.

[155] Ingrid Zukerman, Bhavani Raskutti, and Yingying Wen. Query expansion and query reduction in document retrieval. In *ICTAI '03: Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence*, pages 552–559, Washington, DC, USA, 2003. IEEE Computer Society.

[156] Pierre Zweigenbaum. Question answering in biomedicine. In *Proceedings of the EACL2003 Workshop on NLP for Question Answering*, pages 1–4, Budapest, 2003.