Versatile question answering systems: seeing in synthesis

Sparsh Mittal, Ankush Mittal

Versatile question answering systems: seeing in synthesis

Sparsh Mittal*

Department of Electrical and Computer Engineering,
Iowa State University,
Ames, IA-50011, USA
E-mail: sparsh@iastate.edu
*Corresponding author

Ankush Mittal

Computer Science and Engineering,
College of Engineering Roorkee,
Uttarakhand, 247667, India
E-mail: dr.ankush.mittal@gmail.com

Abstract: Recent advances in massive information storage and growth of internet have led to increased necessity of tools such as search engines and QA systems to get meaningful answers from the vast amount of information. The complex questions appearing in real-life generally require multiple techniques or sub-division of question at various levels. Several practical systems have been developed to explore this. This paper reviews state-of-art QA systems which provide enhanced and versatile functionality at various levels in their architecture and then makes the following contributions:

1 it provides useful overview of research trends and recent developments in the area of QA
2 the paper introduces and defines basic design parameters for any QA system
3 it resolves semantic heterogeneity by clarifying the meaning and context in which different terms are used by different researchers
4 it classifies the existing QA systems in broad groups, based on the point at which versatility (multiplicity) is introduced/handled in the system
5 by developing a unified view of existing frontiers of QA systems, it provides clear directions for future research and development in the area.

Keywords: question answering system; QAS; versatile QA systems; survey; QA design parameters.

Reference to this paper should be made as follows: Mittal, S. and Mittal, A. (2011) 'Versatile question answering systems seeing in synthesis', *Int. J. Intelligent Information and Database Systems*, Vol. 5, No. 2, pp.119–142.

Biographical notes: Sparsh Mittal received his BTech in Electronics and Communications Engineering from the Indian Institute of Technology, Roorkee, India. He was the graduating topper of his batch and his major project was awarded Institute Silver Medal. At present, he is pursuing his PhD in Electrical and Computer Engineering at Iowa State University, USA. He has been awarded scholarship and fellowship from IIT Roorkee and ISU. His research interests include natural language processing, data mining and computer architecture.

Ankush Mittal received both his BTech and MS (Research) degrees from Indian Institute of Technology Delhi, and PhD from The National University of Singapore (NUS). He served as a Faculty in NUS and Indian Institute of Technology Roorkee from 2001 to 2009. At present, he is the Dean (Research) and a Professor in the Department of Computer Science and Engineering, College of Engineering Roorkee, India. He has more than 60 international reputed journal papers. He was conferred with several awards, including IBM Faculty Award 2008, Outstanding Teacher Award 2008 (IIT Roorkee), Young Scientist Award 2006 (The National academy of Sciences), and Young Scientist Award 2008 (Indian Science Congress association). His research interests include video processing, e-learning, multicore computing and AI.

1 Introduction

For human-computer interaction natural language is the best information access mechanism for humans. Hence, question answering systems (QAS) have special significance and advantages over search engines and are considered to be the ultimate goal of semantic web research for user's information needs (Trotman et al., 2007). Development of QA systems has been motivated by various workshops and research forums such as TREC, CLEF, and NTCIR and in recent times, many directions of enhancing capabilities of QA systems have been explored.

Traditional QA systems generally implement a single processing stream or technique that performs three standard steps sequentially: question analysis, search, and answer selection (Hovy et al., 2000; Jijkoun et al., 2004). These systems cover a limited number of corpus or other resources and utilise a limited number of alternative techniques. Thus, it is difficult to properly answer a variety of complex questions posed by users. Thus, systems which use multiple techniques or resources etc. for answering are required.

Recently, many practical systems have been developed, which employ additional processing module or provide multimedia content-rich answers and so on. Thus, these QA systems approach the problem of answering the question in a multi-pronged manner. We call these systems 'versatile' QA systems. Compared to traditional 'strict pipeline architecture' QAS, these systems either provide enhanced answer (e.g., multimedia QAS) or enable the system to answer those questions, which were previously impossible to answer (e.g., multi-focus QAS, multi-sentence QAS, multi-document QAS, etc.). Each of the system has its own advantages and addresses and resolves a particular bottleneck of traditional systems. Taken together, these systems complement each other and serve as definite steps in development of next-generation 'all-in-one' QA systems having all these functionalities together. The approach of using multiple techniques to augment a task is very effective and has been successfully employed in several technical fields (Khaitan et al., 2008).

The contributions of our work are as follows:

1 this paper reviews several QA systems with special focus on QA systems providing versatile features

2 it provides useful overview of research trends and recent developments in the area of QA

3 the paper introduces and defines basic design parameters for any QA system

4 the paper resolves semantic heterogeneity by clarifying the meaning and context in which different terms are used by different researchers

5 by developing an over-all view of QA systems, it provides clear directions for future research and development in the area.

The significant advantages of this review and development of unified view of apparently diverse QA systems are as follows:

1 it provides a basis of comparison and identifying similarities between different classes of the state-of-art QA systems

2 on the basis of above, developing single generic architecture along with specific modules for different systems would lead to functionality-reuse, portability and savings in resources and development cost.

3 it enables the experts in one field to benefit from the advancement in another field in designing better systems

4 this classification gives an insight as to how the QA systems actually answer a question, thus, helps to understand the task at fundamental level

5 identification of these parameters also facilitates development of existing systems.

The advances and research carried out in the field of QA systems is rather extensive to do full justice in terms of evaluation within this paper and any study would at best give a bird-eye view. In this paper, we attempt to give a good, but obviously limited overview of the field of QA systems. The rest of the paper is organised as follows. In Section 2, we briefly review several versatile QAS. For the sake of brevity, we take a representative system to define the scope/task of these QAS, and briefly describe their architecture, research trends and limitations (wherever required). Section 3 clarifies the differences between terms (name of systems) which are often confused. Section 4 introduces and defines basic QAS design parameters and provides explanation of their use. Finally, Section 5 concludes the work and provides directions for future research.

2 Versatile QA systems

Due to obvious complexity in natural-languages and complicated needs of human-users, the questions posed in natural language present special difficulty and seek intricate answers. Such difficulty can occur in properly analysing the question, in using right technique to retrieve the proper answer or in presenting answer in a manner/language suitable to the user. Recently, various QAS have been developed each of which primarily addresses one or few of these issues. In this section, we survey these systems.

2.1 Multimedia QAS

Medium is a means of conveying a representation (to a human), e.g., a diagram or a text. Multimedia QAS refers to such QA systems which provide textual and multimedia contents (such as images, sounds, video, etc.) as answers to user queries. The input to the system, however, is provided only through text. Yang et al. (2003) refer to their system as ‘news video QA system’, while Huang et al. (2007) call their system as ‘video QA system’.

The well-known start multimedia information system (Katz et al., 2002b) uses natural language annotations (Katz and Lin, 2002), which are machine-parseable sentences and phrases that describe the content of various information segments. These serve as metadata (data of data) describing the types of questions that a particular piece of knowledge is capable of answering. User’s query is compared with the annotations stored in the knowledge-based, using various NLP tools. The effectiveness of START has been further enhanced by development of a ‘virtual’ database called Omnibase (Katz et al., 2002a) that provides a uniform abstraction layer over multiple web knowledge sources. Omnibase serves as a natural language interface to heterogeneous data on the World Wide Web. Natural language annotations serve as the enabling technology that allows the integration of start and Omnibase. With the help from Omnibase, start translates user queries into a structured request.

The VideoQA (Yang et al., 2003) is a video-based QA system which provides precise video answers (news summaries) to simple factoid questions posed over the news video collection. The semantic contents of video come from multiple sources including the inherent content features, accompanying speech or closed captioned text, metadata and external resources such as the web-based news articles. VideoQA architecture is composed of several stages. During the pre-processing (or preparation) stage, VideoQA performs video story segmentation and classification, as well as video transcript generation and correction. The resulting video story may contain shots of different genre types. Users interact with VideoQA using short natural language questions with implicit constraints on contents, context, duration, and genre of expected videos. In the question answering stage, they first perform question analysis to extract the key terms in the question; the type of questions and its likely answer targets; the type of video genre, and the implicit duration constraint. The error correction to deal with the problem of recognition errors in news transcript improves the performance of the system. The answer of the system consists of news video summaries and news transcripts sentences.

Huang et al. (2007) present “A mobile video question answering system for e-learning”. It is designed to integrate with multimedia content question answering, retrieval, and annotation in the wireless and mobile environment. A learner can ask questions through wireless connection. After video question answering process, a list of ranked answers together with text and videos are represented. Learners can directly annotate the returned answers through clicking and browsing. The system subsequently sends the annotated results as email to learners. In this way, the retrieved multimedia content could be further browsed and traced in desktop PCs. The system brings together the benefits of mobile learning, learning with multimedia content and judicious use of portable systems (such as PDA) and fixed systems (such as desktop) in over-all learning process.

Multimedia QA systems enable the users to access heterogeneous web sources in a unified manner, which systems such as in Kirk et al. (1995) and Knoblock et al. (2001) have been unable to do, because of the need of formulating the query in SQL, Datalog, or some similarly formal language. Some of the challenges in multimedia systems are difficulty in accessing heterogeneous data, and its coherent and contextual presentation; crucial dependence of system performance on natural language annotations which affects knowledge coverage; inadequacy of manual annotation and issue of scaling it to web (or large repository of knowledge), need of higher processing power to provide real-time response, etc. In today's world, multimedia systems are being employed in several applications (Pande and Zambreno, 2009) such as entertainment (mobile phones, internet video websites), defence (video-surveillance and tracking) and public-domain (telemedicine, remote and distant learning, traffic monitoring and management) and several researchers are working to provide real-time support for such applications (Pande and Zambreno, 2008a; Sliwko and Nguyen, 2007). Multimedia information retrieval in these areas, (such as internet) will increasingly gain popularity. Some of the efforts made towards addressing the above mentioned challenges include using automatically built annotations, utilising robust NLP techniques, and facilitating data integration process through well-designed authoring tools, etc.

2.2 Multimodal QAS

In a communication act, modality refers to the sensory or perceptual experience (e.g., visual, tactile, etc.) (Anastopoulou et al., 2001). Multimodal QAS are those which employ multiple modalities at both question (input) and answer (output) level. Sanchis et al. (2006) use the term 'spoken QA' for their QUASAR QA system.

Sanchis et al. (2006) describe their QUASAR system, which is based on passage retrieval based approach. It works with real speech input; using a language model learnt using the questions of the test. This enhances word accuracy as generally questions are constructed from a small vocabulary. They have also introduced measures to ensure that the errors in speech recognition minimally affect passage retrieval and answer extraction phase.

Schofield and Zheng (2003) demonstrate a multimodal interface for asking questions and retrieving a set of likely answers. They describe spoken question answering using a commercial dictation engine with language models customised to questions, a web-based text-prediction interface allowing quick correction of errors, and an open-domain question-answering system, AnswerBus available on the web. Such an interface is particularly appropriate for mobile networked devices (such as Palm and Pocket PC devices) with screens that are too small to display general web pages and documents.

Yeh and Darrell (2008) design a multimodal QA system, where the users can pose the question through both text and picture (photo) modalities. Such a system is very useful to query about any special object/event, which is hard to describe in text but are easy to distinguish visually. They propose system architecture, and described an algorithm for searching a multimodal query database. On the server's side, image matching techniques are applied in conjunction with the existing question search/matching mechanism to lookup relevant information in the multimodal query database.

Design of multimodal systems, poses more challenges than posed by multimedia systems. At level of input inaccuracies arise out of recognition error, which are very high for non-text input such as voice or photo. Even assuming an excellent speech recognition system, ambiguity in intention/meaning, use of homograph words, additional computational overhead of decoding the speech or image matching, etc. come up, which are absent while using text input. Much remains to be done to fully harness the immense potential of multimodal systems and ensure a more tightened integration in design of both QA system and the input recognition and interface system. Despite these, multimodal QAS have great potential.

2.3 Multi-strategy QAS

Multi-strategy QAS refers to systems which utilise multiple strategies to extract the answer to any question.

Nyberg et al. (2003) describe their multi-strategy QA system named JAVELIN. During extraction of candidate answers from relevant documents, it employs a variety of strategies, ranging from simple finite state transducers to classifiers. The 'light' information extraction strategy implements specific strategies for definition questions, relationship questions, and person biography questions. Another strategy is based on using statistical features, train SVM and KNN classifiers to separate correct answers from incorrect ones for each specific answer type. The final strategy implements a finite state transducer. It is most appropriate for question types where answers can be extracted using a handful of simple patterns. These strategies are implemented in an object-oriented architecture, so that different strategies can be tested individually and then integrated into the system in a straightforward manner.

2.4 Multi-source QA

Multi-source question answering refers to systems which utilise multiple kinds of databases as corpus for finding the answers. Hildebr et al. (2004) use the term 'multiple knowledge source' QA for such a system. The term 'multi-resource' QA also appears in literature (Katz et al., 2005). Thus, such systems provide uniform access to diverse knowledge resources.

Clarke et al. (2002) employ techniques for utilising both unstructured text and structured databases for question answering. The work by Clarke et al. (2002) uses structured data collection consisting of tables containing answers of frequently occurring questions. The structured data consists of data gathered from web. The passages are retrieved from TREC corpus, local web corpus and another small 'trivia' corpus (treated as unstructured corpus).

Each of the knowledge sources has its own advantages and disadvantages. Prager et al. (2001) have shown that question answering in structured text improves accuracy. However, it may not be able to answer complex questions. On the other hand, despite providing comparatively low accuracy, unstructured texts have been used in majority of existing QA systems, as a large body of information on the web is available in form of unstructured text.

2.5 Multi-stream (or agent) QAS

A stream is a small question answering system on its own. A multi-stream QA system employs multiple streams for finding the answer of the question. Chu-Carroll et al. (2003b) refer each stream as one agent. Each of the multiple streams can possibly use different answering strategy or knowledge sources and produces a ranked list of answer candidates. For example, an answering agent/stream may employ statistical methods for extracting answers to questions from a large corpus, while another answering agent may transform select natural language questions into logical forms and query structured knowledge sources for answers (Chu-Carroll et al., 2003b).

Chu-Carroll et al. (2003b) describe PIQUANT, which they also call a multi-strategy and multi-source question answering system. The system integrates multiple answering agents that utilise both structured and unstructured knowledge sources for answering the questions. For answering from structured corpus, PIQUANT uses two text-based answering agents, one utilising a primarily knowledge-driven approach (Prager et al., 2000) and the other adopting statistical methods (Ittycheriah et al., 2001). For answering from unstructured texts, several knowledge sources such as public databases such as the US geological survey, websites with data in formatted tables from websites, public domain lexicons such as WordNet, and the Cyc knowledge-based have been used. A question is converted into a query suitable for these knowledge sources, and final answer retrieved is validated with Cyc Sanity checker for relevance.

Jijkoun and De Rijke (2004) present a multi-stream QA system named Quartz. The streams used in Quartz (Jijkoun and De Rijke, 2004) are table lookup, collection patterns, web patterns, collection ngrams, web ngrams and tequesta. Each of the streams finds correct answers that are not found by other streams, thus, each stream contributes to the system's performance, but the significance of the contribution depends on the question type. For selection of final answer from a pool of answer candidates, the scores of answers are normalised and re-ranked. Then similar answers across the pools of answer candidates are identified. Finally, six pools are merged and the answer with the highest confidence is selected. Multi-stream architectures enable easy modification, maintenance, and testing of the different subsystems as well as easy integration of multiple source of information. The challenges in such system are similar to those presented in previous two systems. Tllez-Valero et al., (2008) present answer validation method for leading an ensemble of QA systems for providing precise answers, better than those provided by the individual components. The system employs different QA systems to extract candidate answers in parallel, along with corresponding support text. Then using answer validation methods (which employ statistical laws and NLP techniques), it takes the decision to accept or reject the answer; seeing whether it is supported and correct.

A fundamental research issue in multi-stream QAS (and also in multi-strategy, multi-source QA systems) is 'final answer selection or resolution or validation'. This pertains to selecting or from or combining answer candidates (presented by different agent or from different corpora) to construct the final answer to be presented to the user. Clearly, simply presenting all the answers will be insufficient, as the answers from different components may have different degree of utility or relevance. For answer resolution, PIQUANT uses answer feedback mechanism, where answers obtained from unstructured text is fed back into search process, to retrieve passages which are used to justify the answers. Thus, all candidate answers are generated in a uniform fashion, simplifying the answer resolution process.

2.6 Multi-lingual QAS and cross-lingual QAS

Multi-lingual QAS refers to the QA systems which can work on and interact with user in multiple languages. Cross-lingual QA is concerned with providing an answer in one language (the target language) to a question posed in a different language (the source language). Multi-lingual QA has many advantages, such as it allows users to interact with machines in their native languages, contributing to easier, faster, and more equal information access. Second, cross-lingual capabilities enable QA systems to access information stored only in language-specific text collections.

There has been a good amount of research in this area motivated by both, its potential applications and existence of QA tracks in several workshops and evaluation campaigns. Since, a wide variety of techniques have been applied into practice, it will be worthwhile to review the tasks and challenges presented in workshops such as cross-language evaluation forum (CLEF) (http://clef-qa.itc.it/)or MLQA organised by European Chapter of Association of Computational Linguistics (EACL) (http://nlp.uned.es/MLQA06/) CLEF offers several tracks designed to evaluate different aspects of the systems. For example, the task on mono-, bi- and multi-lingual textual document retrieval on news collections (ad hoc), multiple language question answering (QA@CLEF), multi-lingual retrieval of web documents (WebCLEF) interactive cross-language retrieval (iCLEF), cross-language retrieval in image collections (ImageCLEF), etc.

Most systems tackle the cross-lingual problem by translating the question or query posed in the source language in the target language, and then uses a QA system developed for the target language for retrieving an answer. At the three QA@CLEF campaigns (2003–2005) the participants performed 20 different cross lingual tasks (Peters, 2003, 2005; Peters and Borri, 2004), using off-the-shelf translation software to translate the question or the keywords of the query of the source language into the target language, followed by processing the translated question using a QA system designed for the target language. Bos and Nissim (2006) propose a different approach to cross-lingual QA by translating the answer into the target language. Such programs have drawn attention in different parts of the world and systems made in numerous languages have been developed.

Some of the challenges in designing these systems are difficulty in translation due to lack of one-to-one correspondence between different languages, inherent difficulty in translating some words likes creative works, numeric expressions, etc.

2.7 Multiple choice QAS

This refers to the QA systems where the system has to answer questions, from a set of multiple choices provided as possible answers. It is a simplified but nevertheless challenging area in question answering research. The presence of possible answer itself alleviates many problems/issues present in other systems, however, the constraint of selecting one among the present choices and ‘rejecting’ the other choices (which might be quite close) impose major challenges.

The approach used in the multiple-choice QA (MCQA) systems focus on answer selection and validation. Apart from the local given corpus, the web is employed as corpus, either for extracting answers or for learning lexical patterns which are then used to improve the system itself. Studies suggest that the resulting data redundancy provides more reliable answer extraction (Clarke et al., 2001). Different approaches to improve system performance exist, such as using probabilistic algorithms to learn the best question paraphrase (Radev et al., 2001) or training a QA system to find possible sentence-length answers (Mann, 2001). When several potential answers are retrieved, answer validation techniques rank them, selecting the most probable answer. This is the basic approach in MCQA. Techniques to answer validation range from purely statistical methods based on web search to the use of semantic techniques (Harabagiu and Maiorano, 1999).

Rauber and Awadallah (2001) compare different answer selection techniques within a multiple-choice QA setting. The multiple choice questions have been taken from both the

English and the Arabic versions of the TV show 'Who wants to be a Millionaire?', as well as from the TREC-2002 questions. The procedure to answer such a question, consists of taking appropriate information from both the question and each of the answers, using web [e.g., Google search engine in work by Rauber and Awadallah (2001)] to extract candidate answers. The relevance of each answer is assessed using different answer selection techniques, such as simple hit counts, or full-edged analysis of the documents retrieved by the search engine. The experiments reveal suitability of different answer selection techniques for both the languages.

2.8 Multiple sentence QAS

Single-sentence questions are defined as the questions composed of one sentence. Multiple-sentence questions are defined as the questions composed of two or more sentences: For example, "My computer reboots as soon as it gets started. OS is Windows XP. Is there any homepage that tells why it happens?" (Tamura et al., 2005). The task in multiple sentence QAS is to answer such questions. Bilotti and Nyberg (2006), refer such questions as multi-part questions. A multi-sentence query often contains contents that are not directly used for question type identification, such as greetings or apologies. Such sentences create noise in process of automatic answering of the questions.

Takechi et al. (2007) describe a technique of question segmentation and type identification for multi-sentence queries in open domain question answering in Japanese. They use chunking based identification method. Chunking is a process of identifying chunks that indicate some sort of visual or semantic unit. Here, the target unit is question segments. The goal is to extract question segments in a query and identify their question types. For question type identification, they divide a question article into sentences, carry out chunking with respect to each article and then extract question segments labelled with their question types. Their experimental observations suggest that simple word features cannot be used to accurately classify sentences. It is concluded that errors often occur in the boundaries of adjacent questions, and when chunks contained more than one question. They suggest separating segmentation from question type identification as a possible improvement.

Tamura et al. (2005) discuss classification of multiple sentence questions using SVM classifier. They use five sets of features: word unigrams, word bigrams, semantic categories of nouns, question focuses, and semantic categories of question focuses. The core sentence extraction component in their system, extracts the most important sentence for question classification, thus, helping to remove the noisy features present in multiple sentence questions. They then classify the question using the only information in the sentence. The experiments conducted from questions collected from online QAS verify their approach.

2.9 Multi-layer QAS

A multi-layer QAS refers to the QA system where a question is iteratively decomposed into a set of simpler questions that can be directly answered using the document corpus, and final answer is constructed using answers generated by layers in bottom-to-top fashion. Here, a question is decomposed into various sub-questions, which may be further decomposed at a different layer. The question at a layer can be either directly answered or to answer a question at an upper layer, answer to the decomposed question at a lower layer is required. The last layer (called level 1) has questions, which can be answered using the document corpus. Thus, it uses query reformulation to provide a conceptual framework to resolve contradictory answers (Blake, 2003). The concept of multi-layer QAS can be understood from the following example [taken from Blake (2003)] for the question 'What is the relationship between alcohol consumption and breast cancer?'. To answer such a question, the questions generated at the next lower layer could be, 'Do people with breast cancer consume significantly more alcohol than people in a similar population?' etc. To answer such a question, the question at next lower level could be, 'How much alcohol do people with breast cancer consume?' and 'How much alcohol do people without breast cancer consume?' Proceeding in this way would lead to simple questions, which can be answered based on a study of alcohol consumption of a group of people.

The approach taken by Blake (2003) is called information synthesis (IS). It uses meta-analysis to unify contradictory findings from multiple findings (Ingelfinger et al., 19994). This approach is based on the premise that the person posing the question has knowledge of the information required and integration methods required to answer the question. Information synthesis is appropriate for the class of questions that users can reformulate as a question with a quantitative answer. An important advantage of information synthesis is to reduce the effect of publication bias (Begg and Berlin, 1988). Moreover, the authors in (Blake, 2003) claim that multiple answers, obtained directly from a document or through information synthesis could be the only accurate answer for a diverse and technical domain like biomedicine, etc. However, they have not automated the conceptual reformulation of a question. Such automation is the main challenge in building a multi-layer QA system.

2.10 Multi-perspective QAS

Multi-perspective question answering (MPQA), targets opinion based questions of the following sort:

1. How Bush's decision not to ratify the Kyoto Protocol is looked upon by Japan and other US allies?
2. How do the Chinese regard the human rights record of the USA?

An interest in methods for automatic classification of opinions, sentiments and emotions has recently been growing (Dave et al., 2003). Stoyanov et al. (2005) investigate the use of machine learning and rule-based subjectivity and opinion source filters to guide MPQA systems and argue that the use of these tools may substantially improve the performance of an end-to-end MPQA system. The solution proposed by Stoyanov et al. (2005) for MPQA systems is based on using special corpus, which is manually annotated with phrase-level opinion information, e.g., OpQA is corpus of opinion questions and answers, which consists of documents from the MPQA corpus. However, these systems are still far from a fully functional question answering system. The utility of MPQA system lies in presenting the various positions on a topic currently being expressed in the world press or by groups of experts, to help the user answer the question for himself or herself.

2.11 Multi-focus QAS

In literature, 'question focus' is defined as 'the question concept that embodies the information expectations expressed by the question'. Multi-focus questions are defined as questions containing multiple foci. A multi-focus question can be decomposed into several subquestions and for such questions, 'question focus' can be re-defined as the set of the question foci of its subquestions. A QAS which answers such questions and works on identifying question focus set is defined as multi-focus QAS.

An example of multi-focus question is as follows:

Q_1 Give me the capitals, national flowers, and national trees of all the countries.

The question foci of Q1 are 'capital', 'national flower', 'national tree', and 'country'. The focus degree (*FD*) is defined as the cardinality of the question focus set. Therefore, the focus degree of question Q_1 is 4. In terms of focus degree, a 'multi-focus question' is a question with $FD > 1$. On the other hand, a question with a $FD = 1$ is called a single-focus question.

Lin and Liu (2008) advocate the need of identifying such questions and developing systems for answering them. For simulation they manually decomposed the complex questions into multiple single focus questions, and submitted individual questions to the QA system START. The individual answers can be accumulated and presented as final answer in a table format to the user. However, this task has not been automated.

They have identified four kinds of relations between different focus terms existing in a single question, which reflect the inter-dependence of different sub-questions. Sub-questions which are independent of each other can be answer independently, however, for a question such as Q_2 'What are the five largest cities and their populations?', where the sub-questions $Q_{2.1}$ 'What are the five largest cities?' and $Q_{2.2}$ 'What is the population of (ans$Q_{2.1}$)' are dependent and can only be answered serially with first sub-question answered before the second one. Moreover, the correctness of answer to second sub-question crucially depends on the answer to the first sub-question. Despite the high complexity in automatic identification of the sub-questions, such systems are promising and closely model the way humans pose their complex questions.

2.12 Multi-Parser QAS

The QAS, where different types of parsers are employed on different types of query data in multi-Parser architecture (MPA) are called multi-Parser QAS.

Due to the complexity of natural language, the traditional way of using a single grammar for a single language parser leads to an inefficient, fragile, and often very large language processing system. Xu et al. (2001) investigate the effect of using multiple parsers on different query data. They employ two styles of parsers, having different characteristics. The first is the Earley parser, where the prediction step is precompiled. The other is the GLR parser. Their results show that grammar partitioning with composition of Earley parsers can speed up processing substantially for both semantic and syntactic grammars when compared to the unpartitioned grammars. Such advancements are expected to help the question answering domain in the long run. The MPA proposed by Luk et al. (2000) alleviate many parsing issues simultaneously by partitioning a single grammar into multiple subgrammars and composing subparsers for the subgrammars. Ruland et al. (1998) developed multi-Parser multi-strategy architecture for noisy input. It uses a full parser first, and then a partial parser when the full parser does not return a result. CMUs Janus parser (Woszczyna et al., 1994) uses a similar strategy.

2.13 Multi-document comparison QAS

Multi-document QAS constructs and presents its single answer for a question by taking parts of answers from multiple documents. For complex questions where the answer to a question may not be completely present in one document aggregating information from multiple locations and presenting all the information as a whole is required. Intuitively, comparison seeking questions is one such class of questions, where multi-document QAS can prove to be exclusively effective. For example, to answer questions such as 'What is the difference between glycoprotein and lipoprotein?', it is very unlikely to have a direct answer available in a single document which a QAS can present to the user. In such a case, generating answers from multiple documents is highly useful.

Mittal et al. (2008) present a biomedical multi-document QA system named 'BioinQA'. It constructs a single answer for the comparison type questions from multiple documents. The system searches the answers from multiple locations that can be either from a same document or multiple documents. To meaningfully answer comparison based questions, they have used the technique of 'entity cluster matching-based passage sieving'. As expected, different answers explaining same aspect of entities being compared should be presented as single answer. For example, for a question such as 'What is the difference between OSI model and TCP model?', the entities being compared are OSI model and TCP model. A meaningful answer to this comparison would be to describe and contrast the same aspect such as number of layers; type of connection; functionality of corresponding layer, etc. Thus, while it makes perfect sense to expect 'OSI model has seven layers. TCP model has four layers'; an answer such as 'OSI model has seven layers. In TCP model, the protocols have been invented before models, so the functionalities are perfectly described' would not make much sense.

Their approach of constructing single answer from multiple documents is general enough and can be applied to technical questions other than comparison seeking questions, such as "What is the similarity between polymorphic wavelet transform and secure wavelet transform?" (Pande and Zambreno, 2008b, 2010) A characteristic of such questions is that the correlation or interaction between sub-questions generated must be clearly identified and taken into consideration. Effective integration of multiple pieces of data is required to present the answer in proper context and answering the exact need of the user.

2.14 Question based multi-document summarisation system

Multi-document summarisation concerns with the task of providing summary from multiple documents in response to a question. The challenge in such systems is to provide a user with only the 'right' amount of the interesting information in coherent and concise form, omitting all the redundant and 'uninteresting' material. The quality of the summary depends strongly on the user's need – a summary that focuses on one of several topics contained in the material may prove to be either very useful or completely useless depending on what the user's' interests are (Leuskiet al., 2003). Thus, generating summary in response to a question is effective, as it appropriately reflects user's information need.

Afantenos et al. (2005) provide an excellent survey of multi-document summarisation systems. It categorises different summarisation systems based on the techniques and factors such as input (number of documents, language, text or multimedia), purpose (generic v/s user-oriented, etc.), output quality (extracts v/s abstracts, etc.). The authors have also compared similar systems based on these parameters and presented criterion for evaluation. Shi et al. (2007) introduce BioSquash, a question-oriented extractive summarisation system on biomedical multi-documents that are relevant to a question. The system was based upon a general-purpose summariser, Squash. The BioSquash system has four main components: the annotator, concept similarity, extractor, and editor modules. In the annotator module, the system annotates the documents and the question text with syntactic and shallow semantic information. In the concept similarity module, actual semantic meanings of both general and biomedical concepts and the ontological relations among these concepts are obtained. Using information thus obtained, the summariser in extractor stage primarily does the task of content selection and that in editor stage focuses on linguistic readability. As of now, the human capability to combine information from many documents into a succinct summarising statement is still beyond the capability of the automated systems. Some of the major challenges in this QAS are rewording the sentences for regeneration of text, identification and removal of redundancy and sub-sentential modification (Barzilay and McKeown, 2005; Knight and Marcu, 2000).

2.15 Multi-dimensional mark-up QAS

Multi-dimensional markup QAS deals with answering questions from corpora with multiple types of annotations. Annotation is extra information asserted with a particular point in a document or other piece of information (Wikipedia). The challenge in such systems lies in integration of different layers of text annotation.

The work by Ogilvie (2004) outlines the possibility of using multi-dimensional markup for question answering, with no system or experimental results. Jijkoun et al. (2005) describe initial experiments with XQuesta, a QAS based on multi-dimensional markup on English and Dutch Newspaper Corpora. It makes use of the multi-dimensional approach to linguistic annotation embodied in XIRAF. The system analyses an incoming question to determine the required answer type and keyword queries for retrieving relevant snippets from the corpus. From these snippets, candidate answers are extracted, ranked, and returned.

2.16 Multi-grained QAS

The word granularity denotes the depth of detail that a system collects. A multi-grained QAS detects the right granularity of answers and return answers of different granularity. Different purposes require different levels of granularity.

The work by Sorg (2008) proposes a multi-grained QAS. Traditionally, information retrieval (IR) and QA systems return answers of only a specific granularity. IR systems typically return whole documents as answers where as QA systems try to find an exact answer to the question. Traditional QASs are not able to answer complex questions, e.g., questions based on a large context or questions where the answer is not explicitly stated in the text but must be inferred. Thus, a multi-grained QAS can fill this gap and will provide answers of different granularity ranging from whole documents to exact answers in a flexible way. To handle the trade-off between completeness of the answer and its length, methods need to be developed to identify the right granularity of the answers given a query and the information sources.

2.17 Multi-search engine QAS

A multi-search engine QAS refers to a QA system which utilises existing search engines for answer extraction. Echihabi et al. (2003) use the term 'multiple-engine question answering' for their system TextMap. A promising feature of such system is utilisation of existing search engine with only some extra module to properly utilise these search engines as underlying retrieval tools.

Agichtein et al. (2001) present the Tritus system that automatically learns to transform natural language questions into queries containing terms and phrases expected to appear in documents containing answers to the questions. It automatically learns multiple query transformations, optimised specifically for each search engine. It attempts to maximise the probability of an IR system returning documents that contain answers to a given question. Various search engines do not perform uniformly well across different question types, or even across different subtypes of the basic question type. Then intelligently deciding weighting functions for the answers returned by different systems can provide best possible answer for multiple types of questions.

2.18 Multi-channel QAS

A multi-channel QAS provides multiple channels of communication or presentation of answer to the user. Such system focuses on integrating multi-channel delivery strategies into a single system.

In Terrassa, Catalonia, Spain ISAC system (which stands for 'Interactive Citizen Information and Attention Service') has been employed (Wilson and Blakemore, 2008) for providing a common validated source for answering user questions. Alternative service channels include face-to-face, telephone, and web. In the context of e-government, such a system will help to improve citizen access, make services available to more people than before. In the context of information systems as public utility, such systems have great promise to provide several channels for allowing people to select the most appropriate channel to suit the time of day, usage context and personal circumstances.

2.19 Multi-participants QAS (community QAS)

Multi-participant QAS refers to QA systems where information seekers can obtain specific answers to their questions by posting questions for other participants to answer. This is quite different from other QAS as here the task of the system is to facilitate proper interaction among multiple users and not to get an answer itself. However, it is still useful to classify such QA systems in this list due to their increasing utility.

Presently, question answering communities such as Naver (http://www.naver.com/) and Yahoo! Answers (http://answers.yahoo.com/) have emerged as popular, and often effective, means of information seeking on the web. In addition to using general-purpose web search engines, users now have an option to post their often complex and specific questions on community QA sites, and get them answered by other users. These sites are growing rapidly and millions of questions and answers have already been posted in just two years since introduction of Yahoo! Answers.

The satisfaction of information seeker is of prime importance in such collaborative QA communities and (Liu et al., 2008) introduce methods for predicting it. They present a general prediction model, and develop a variety of content, structure, and community-focused features for this task. They also explored state-of-the-art classification techniques to implement their models and the experimental results show the effectiveness of their approach.

3 Clarifying the differences

Due to non-standard usage of terminologies in different contexts by different researchers and close similarity between few pairs of systems, confusion is sometimes created between them. In what follows, we clarify the differences between the usages of different terminologies.

3.1 Multimodal v/s multimedia

We use the terminology defined by Turk (2000) to differentiate between multimodal and multimedia systems. The distinction between multimedia and multimodal user interfaces is based on the system's input and output capabilities. A multimodal user interface supports multiple computer input and output, e.g., using speech together with pen-based gestures. A multimedia user interface supports multiple outputs only, e.g., text with audio or tactile information provided to the user. As a result, multimedia research is a subset of multimodal research (Turk, 2000). In other words, multimodal systems offer multiple modalities at both input and output level, where as multimedia systems offer unimodality at input and multiple modality at the output. According to this definition, a system which provides multiple modalities at only input level will be classified as multimodal (and not multimedia system).

3.2 Multi-modal v/s multi-channel

Both of these work with multiple modalities and inside the QAS, the task of taking question input, processing it to generate answer and presenting it to the user can be same for both. However, the difference lies in the context or application in which they are used. Multi-channel QAS is generally used where the users are remotely situated and query a central information system through different channels of communication such as phone or internet. In response, the multi-channel QAS returns the answer in user's preferred channel of communication. In a real-life setting, such channel of communication can include web or special technologies for people with disability (e.g., accessible web browser, screen reader, deaf interpreter, etc.). On the other hand, in multi-modal system the users are situated locally and use multiple modalities (such as text, audio, video, etc.), to pose a question or to receive the answer.

3.3 Multi-strategy v/s multi-source

Although the strategies to extract answers from different types of sources are expected to be different and hence, it is tempting to conclude that a multi-source QAS will automatically be a multi-strategy QAS too, however, these two features are defined separately in literature (Chu-Carroll et al., 2003b), moreover, these terms give a minimal description of the system characteristics. Hence, as the name implies, these systems use multiple strategies and sources respectively.

3.4 Multi-stream v/s multi-strategy or multi-source

A multi-stream QA system uses multiple streams, each of which can possibly use different answering strategy or knowledge source. However, in multi-strategy QAS the main feature is only to employ multiple strategies and in a multi-source QAS, the main feature is only to extract answers from multiple knowledge sources.

3.5 Multi-stream v/s multi-search engine

These are almost similar. However, the multi-stream QAS uses multiple streams (different strategies for retrieving or re-ranking answers or even different sources of knowledge), here, each of the stream is merely a component of the system itself, and possibly lacks an independent existence due to lack of interface (input, output functionality) and a few other important components. However, multi-search engine uses different search engines, each of which is independent, a fully functional system in itself, possibly using different strategies for searching, retrieving and presentation of answer and different source.

3.6 Multi-source v/s multimedia or multimodal

Multi-source QA is generally referred to systems where all the corpuses are primarily text corpuses. However, in multimedia/multimodal systems, answers are extracted from sources which differ in modality such as text, video, audio.

3.7 Multi-document v/s multi-source

Multi-document QAS finds answer passages from different documents and presents it as ‘single’ answer to the user where as multi-source QA system simply finds answers from different sources and presents them as ‘separate’ or ‘multiple’ mutually independent answers.

3.8 Multiple sentence v/s multi-focus

A multiple sentence question may have one or more questions, each of which may be expressed in one or more sentences. To answer such questions, one needs to identify different questions and answer them (in parallel). These questions are independent and may at best require reference resolution of pronoun, etc. Moreover, such resolution can be done just seeing the structure of the question, as different questions are contained explicitly in different sentences. In a multiple focus question, more than one questions are asked (usually) within a same sentence. However, identification of these questions cannot be done just by seeing the question, but requires answering one of them, using the answer to construct the second (or third) question and so on. Thus in general, different questions can only be answered serially, not in parallel.

3.9 Multi-focus v/s multi-layer

Both of these QASs are used for complex questions, require conceptual reformulation of questions to create sub-questions (or new questions for lower layer) and for both systems the process of generating sub-questions have not been automated. Despite these similarities, differences do exist. Most importantly, multi-focus represents a characteristic of a question, while as multilayer shows the approach used to generate the answer to a question. Multi-focus questions can be answered in a multi-layer architecture. In fact, because of its generality, several kinds of question can be grouped to fall within the purview of multi-layer QAS, however, such coarse classification does not provide useful insight into the nature of problem and because of lack of automated implementation, it does not practically resolve the task.

3.10 Multi-search engine v/s multi-participant

The former employs search engines which are all automatic, while as the answers in multi-participant question answering are also provided by the users or participants in the community.

4 Design parameters or characteristic features

Despite the apparent diversity between different systems, there have several common features. We have identified a few of such features which characterise each of the system, and bring them to a common denominator. Just as interconnection networks are characterised and classified according to basic and useful properties such as diameter, bisection width, etc.; these fundamental properties also forms a basis of classification for QAS. These can be used as design parameters in study and development of existing and new systems. Formally, a QAS has the following parameters:

- *Qsize* (κ): The *Qsize* of a QAS, denoted by κ is defined as the number of fragments or parallel (but different forms) sub-questions (generated by decomposition or transformation or translation of language or translation of sentence structure) employed to extract the answer.
- *MDimension* (λ): The *Mdimension* of a QAS, denoted by λ is defined as the number of different kind of knowledge sources or modalities that a system can work in, at input (question), processing or output (answer presentation) level.
- *TCount* (Γ): *TCount* of a QAS, denoted by Γ refers to the number of techniques/strategies used for extracting answers from a single knowledge source/modality.

- *AnswerCardinality* (Φ): *AnswerCardinality* of a QAS, denoted by Φ, is defined as the number of answers it is expected to produce from a single knowledge source or modality. Multiple-choice QAS provides only one answer. The other systems (such as question-based summarisation, etc.) can provide many answers (some of them may be repeated, some may be incorrect and others may be candidate answers), and leave it to the user to select the best/correct or the desired one. For example, in summarisation task, redundancy is avoided within a single summary (answer), however, the definition of the system does not constrain its providing several summaries as possible answers. Many QAS usually return top five or ten answers which is their Φ value. If need be, this parameter can be easily controlled by user.
- *AnswerSpan* (Λ): The *AnswerSpan* of the QAS, denoted by Λ is defined as the magnitude of answer it is expected to provide. We refrain from using words 'longest' or 'shortest' answer, as from point of view of theoretical studies, it is more important to provide a coarse measure of answer length, than measuring exact number of words in an answer. Thus, it is more important to report figures such as one-word answer or one paragraph answer than to specify '50' word or '60' word answers. Moreover, it is also interesting to mention whether AnswerSpan is fixed, controllable by user (at the time of question input, to get answer of desired length) or system (at run time, seeing nature of question) etc. Based on these parameters, we have constructed the Table 1.

Table 1 shows typical values of the parameters for different systems. The symbol K in column for κ stands for a value larger than one. The K values in two rows (i.e., for two systems) need not be same, same symbol is used only to show multiplicity. Similar is true for M, T, N in other columns. A '-' in the table represents that definition of the system does not specify or constrain it and thus, they can be designed for a desired value. A brief justification for the values in the table is as follows.

What we discuss here is valid for simple QA systems and terms are used in the sense of 'as the name suggests'. For example, a user can request multimedia information through a multi-focus question, however, the term 'multi-focus question' in itself does not specify nature of or modality of retrieved answer, hence, the values of different parameters given in the table for different systems refer to that in their simplest application or representative system. Exceptions do exist, however, such as when visual modality is used for expressing certainty or uncertainty to reflect user's trust in the answer, the system becomes multimodal, and however, the number of knowledge sources still remains one. The reason is that the extra modality is used for augmenting the question answering process and to add extra feature/function. To keep the discussion simple and generic, we have deliberately excluded such system from consideration (Marsi and Van Rooden, 2007). Moreover, these values do not put an upper limit or restriction and systems with higher values of these parameters can (and should) be designed, e.g., a multiple-grained question answering system can draw its answers of different granularity from different types of knowledge sources, however, the value of 'one' for λ (MDimension) parameter merely suggests minimum such value required to fulfil the criterion of definition.

Table 1 Versatile QA systems and their design parameters

	κ	λ	Γ	Φ	Λ
Multimedia QA	1	M	1	N	-
Multimodal QA	K	M*	1	N	-
Multi-strategy QA	1	1	T	N	-
Multi-source QA	1	M	1	N	-
Multi-stream QA	1	M	T	N	-
Multi/cross lingual QA	K	M*	1	N	-
Multi-choice QA	1	1	1	1	Fix
Multi-sentence QA	≥ 1	1	1	N	-
Multi-layer QA	K	1	1	N	-
Multi-perspective QA	K	1	1	N	-
Multi-focus QA	K1	1	1	N	-
Multi-Parser QA	1	1	1	N	-
Multi-Doc comparison QA	K	1	1	N	-
Multi-Doc summarisation QA	1	M	1	N	-
Multi-dimensional markup QA	≥ 1	1	T	N	-
Multi-search engine QA	K	M	T	N	-
Multi-grained QA	1	1	T	N	Var
Multi-channel QA	K	M*	1	N	-
Multi-participant QA	1	≥ 1	≥ 1	N	-

Notes: 1 here, K = focus degree of question (FD); *here, K = M; Fix = fixed;
Var = varied by users and system.

Based on above table, these QA systems can be roughly divided into broad categories, depending upon the point where versatility (multiplicity) is employed in the architecture. Systems with $\kappa > 1$ handle or generate multiple sub-questions and form the group 1. Systems with $\lambda > 1$ need capability of searching in multiple corpora or knowledge sources and form group 2. Systems with $\Gamma > 1$ employ multiple techniques within single knowledge source and form group 3. This classification can be further extended.

Finally, there are a few more terms in this context, though used rarely. Chu-Carroll et al. (2003a) present a multi-level answer resolution algorithm, in the context of their multi-strategy multi-source question answering system. As two answering agents in a system with the same pipeline architecture, can have multiple points where intermediate results can be combined, such as the question analysis, passage retrieval, and answer selection phases. In PIQUANT, the knowledge-based agent may accept input from the statistical agent after each of these three phases.

5 Conclusions and research directions

This survey presented an integrated view of diverse question answering research with emphasis on versatile QA systems. We reviewed several state-of-arts QA systems, explaining the minute differences in terminology used in literature. Finally, we also introduced and defined basic design parameters and classified the systems based on them.

The list of systems and parameters presented here is by no means exhaustive and we hope addition of more by development of novel systems. For example, a 'multi-target user' QAS can be designed to address the needs of users of different background. The system would take some information (apart from question) from the user to assess his/her level, background, interests or preferences. Then, it will generate possibly different answers for the same question for different users. Such as system is feasible and useful where target user-groups with non-overlapping (specific) information needs can be identified. For example, in the field of biomedicine, the expected (and favoured) answer for the question, 'How does HIV increase the chances of getting TB?' may be completely different for a researcher, a medical practitioner, a college going student (in e-learning set up) and a layman. A researcher would prefer technical answer with scientific jargon, equation, etc. from a scientific journal, a medical practitioner may favour an answer with experience of other practitioners along with description of symptoms and precautions and a layman would like plain to-the-point answer and so on. Thus, such a system could be customised for personal use. Use of a QAS which ranks the answers merely based on statistical features such as number of occurrences, part of speech features would be counter-productive here as it totally neglects the context of the user. Then one can define term 'UserCoverage' to denote the number of kinds (broad categories) of users that a QAS can adapt to, in terms of providing user-specific answers to their queries.

Future efforts should focus on developing generic architectures and utilising them for developing more than one of these systems. We also encourage research to increase the value of parameters such as Γ, κ within a single system. To satisfactorily answer the complex questions generated in real-life scenarios, the next generation 'all-in-one' system will consist of several of such systems; and research should be done to smartly and efficiently integrate such components in a single system.

References

Afantenos, S., Karkaletsis, V. and Stamatopoulos, P. (2005) 'Summarization from medical documents: a survey', *Artificial Intelligence in Medicine*, February, Vol. 33, No. 2, pp.157–177.

Agichtein, E., Lawrence, S. and Gravano, L. (2001) 'Learning search engine specific query transformations for question answering', in *WWW '01: Proceedings of the 10th International Conference on World Wide Web*, pp.169–178, ACM, New York, NY, USA.

Anastopoulou, S., Baber, C. and Sharples, M. (2001) 'Multimedia and multimodal systems: commonalities and differences', available at http://www.informatics.susx.ac.uk/research/hct/hctw2001/papers/anastopoulou.pdf.

Barzilay, R. and McKeown, K.R. (2005) 'Sentence fusion for multidocument news summarization', *Comput. Linguist.*, Vol. 31, No. 3, pp.297–328.

Begg, C. and Berlin, J. (1988) 'Publication bias: a problem in interpreting medical data', *Journal of the Royal Statistical Society*, No. 151, pp.419–63.

Bilotti, M.W. and Nyberg, E. (2006) 'Evaluation for scenario question answering systems', in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2006)*, (Poster.).

Blake, C. (2003) 'A technique to resolve contradictory answers', in *AAAI Spring Symposium on New Directions in Question Answering*.

Bos, J. and Nissim, M. (2006) 'Cross-lingual question answering by answer translation', in Peters, C. (Ed.): *Cross Language Evaluation Forum: Working Notes for the CLEF 2006 Workshop (CLEF 2006)*.

Chu-Carroll, J., Czuba, K., Prager, J. and Ittycheriah, A. (2003a) 'In question answering, two heads are better than one', in *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pp.24–31, Association for Computational Linguistics, Morristown, NJ, USA.

Chu-Carroll, J., Prager, J., Welty, C., Czuba, K. and Ferrucci, D. (2003b) 'A multi-strategy and multi-source approach to question answering', in *TREC 2002*, pp.281–288.

Clarke, C.L.A., Cormack, G.V. and Lynam, T.R. (2001) 'Exploiting redundancy in question answering', in *Proceedings of the 24th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*, ACM Press, pp.358–365.

Clarke, C.L.A., Cormack, G.V., Kemkes, G., Laszlo, M., Lynam, T.R., Terra, E.L. and Tilker, P.L. (2002) 'Statistical selection of exact answers (multitext experiments for TREC 2002)', in *TREC*.

Dave, K., Lawrence, S. and Pennock, D.M. (2003) 'Mining the peanut gallery: opinion extraction and semantic classification of product reviews', in *WWW '03: Proceedings of the 12th International Conference on World Wide Web*, pp.519–528, ACM, New York, NY, USA.

Echihabi, A., et al. (2003) 'Multiple-engine question answering in TextMap', *Proceedings of the 12th Text Retrieval Conference (TREC-2003)*, pp.772–781.

Harabagiu, S. and Maiorano, S. (1999) 'Finding answers in large collections of texts: paragraph indexing + abductive inference', in *Proc. of the AAAI Fall Symposium on QA Systems*.

Hildebr, W., Katz, B. and Lin, J. (2004) 'Answering definition questions using multiple knowledge sources', in *Proceedings of HLT-NAACL 2004*, HLT, pp.49–56.

Hovy, E., Gerber, L., Hermjakob, U., Junk, M. and Lin, C.Y. (2000) 'Question answering in webclopedia', in *Proceedings of the Ninth Text REtrieval Conference (TREC-9)*, pp.655–664.

Huang, Y-T., Chung, C-I., Tsai, C-C., Shen, C-H., Wu, Y-C. and Yang, J-C. (2007) 'A mobile video question answering system for e-learning', *Seventh IEEE International Conference on Advanced Learning Technologies, 2007. ICALT 2007*, July, pp.74–78.

Ingelfinger, J., et al. (1994) *Biostatistics in Clinical Medicine*, McGraw-Hill Health Professions D.

Ittycheriah, A., Franz, M. and Roukos, S. (2001) 'IBM's statistical question answering system – TREC-10', in *TREC*.

Jijkoun, V. and De Rijke, M. (2004) 'Answer selection in a multi-stream open domain question answering system', in *Proceedings 26th European Conference on Information Retrieval (ECIR04)*, LNCS, Vol. 2997, pp.99–111, Springer.

Jijkoun, V., et al. (2005) 'The University of Amsterdam at QA@CLEF 2005', in Peters, C. (Ed.): *Results of the CLEF 2005 Cross-Language System Evaluation Campaign. Working Notes for the CLEF 2005 Workshop*.

Jijkoun, V., Mishne, G., Monz, C., Rijke, M., Schlobach, S. and Tsur, O. (2004) 'The University of Amsterdam at the TREC 2003 question answering track', in *Proceedings TREC 2003*, pp.586–593.

Katz, B. and Lin, J. (2002) 'Annotating the semantic web using natural language', in *Proceedings of the 2nd Workshop on NLP and XML*, pp.1–8.

Katz, B., Borchardt, G. and Felshin, S. (2005) 'Syntactic and semantic decomposition strategies for question answering from multiple resources', in *Proceedings of the AAAI 2005 Workshop on Inference for Textual Question Answering*.

Katz, B., Felshin, S., Yuret, D., Ibrahim, A., Lin, J. and Mcfarland, M.A.J. (2002a) 'Omnibase: uniform access to heterogeneous data for question answering', in *Proceedings of the 7th International Workshop on Applications of Natural Language to Information Systems*.

Katz, B., Lin, J.J. and Felshin, S. (2002b) 'The START multimedia information system: current technology and future directions', in *Proceedings of the International Workshop on Multimedia Information Systems, 2002*.

Khaitan, S.K., McCalley, J.D. and Chen, Q. (2008) 'Multifrontal solver for online power system time-domain simulation', *IEEE Transactions on Power Systems*, November, Vol. 23, No. 4, pp.1727–1737.

Kirk, T., Levy, A.Y., Sagiv, Y. and Srivastava, D. (1995) 'The information manifold', in *Proceedings of the AAAI 1995 Spring Symp. on Information Gathering from Heterogeneous, Distributed Environments*, pp.85–91.

Knight, K. and Marcu, D. (2000) 'Statistics-based summarization – step one: sentence compression', in *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, AAAI Press/The MIT Press, pp.703–710.

Knoblock, C.A., Minton, S., Ambite, J.L., Ashish, N., Muslea, I., Philpot, A. and Tejada, S. (2001) 'The Ariadne approach to web-based information integration', *International Journal of Cooperative Information Systems*, Vol. 10, Nos. 1–2, pp.145–169.

Leuski, A., Lin, C-Y. and Hovy, E. (2003) 'iNeATS: interactive multi-document summarization', in *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pp.125–128, Association for Computational Linguistics, Morristown, NJ, USA.

Lin, C.J. and Liu, R.R. (2008) 'An analysis of multi-focus questions', in *Proceedings of the SIGIR 2008 Workshop on Focused Retrieval*, p.3036.

Liu, Y., Bian, J. and Agichtein, E. (2008) 'Predicting information seeker satisfaction in community question answering', in *SIGIR '08: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.483–490, ACM, New York, NY, USA.

Luk, P.C., Meng, H. and Weng, F. (2000) 'Grammar partitioning and Parser composition for natural language understanding', in *Proceedings of ICSLP*.

Mann, G.S. (2001) 'A statistical method for short answer extraction', in *Proc. of the 39th Annual Meeting of the Association for Computational Linguistics*.

Marsi, E.C. and Van Rooden, F. (2007) 'Expressing uncertainty with a talking head in a multimodal question answering system', in *Proceedings of the Workshop on Multimodal Output Generation (MOG 2007)CTIT Workshop Proceedings WP 07-01*, 25–26 January, pp.105–116.

Mittal, S., Gupta, S., Mittal, A. and Bhatia, S. (2008) 'BioinQA: addressing bottlenecks of biomedical domain through biomedical question answering system', in *The International Conference on Systemics, Cybernetics and Informatics (ICSCI-2008)*, pp.98–103.

Nyberg, E., et al. (2003) 'The JAVELIN question answering system at TREC 2003: a multi-strategy approach with dynamic planning', in *Proceedings of TREC 12*.

Ogilvie, P. (2004) 'Retrieval using structure for question answering', in *TDM, CTIT Workshop Proceedings Series*, pp.17–25, Centre for Telematics and Information Technology (CTIT), University of Twente, Enschede, The Netherlands.

Pande, A. and Zambreno, J. (2008a) 'Design and analysis of efficient reconfigurable wavelet filters', in *Proceedings of IEEE International Conference on Electro/Information Technology (IEEE EIT 2008)*, IEEE, pp.337–342.

Pande, A. and Zambreno, J. (2008b) 'Polymorphic wavelet architecture over reconfigurable hardware', in *Proceedings of Field Programmable Logic and Applications (FPL 2008)*, IEEE, pp.471–474.

Pande, A. and Zambreno, J. (2009) 'An efficient hardware architecture for multimedia encryption and authentication using discrete wavelet transform', in *Proceedings of IEEE International Symposium on VLSI 2009*, IEEE CS, pp.85–90.

Pande, A. and Zambreno, J. (2010) 'A reconfigurable architecture for secure multimedia delivery', in *23rd Intl. Conf. VLSI Design*, pp.258–263.

Peters, C. (Ed.) (2003) *Results of the CLEF 2003 Cross-Language System Evaluation Campaign*, 21–22 August, ISTI-CNR Pisa, Trondheim, Norway.

Peters, C. (Ed.) (2005) *Results of the CLEF 2005 Cross-Language System Evaluation Campaign. Working Notes for the CLEF 2005 Workshop*, 21–23 September, ISTI-CNR, Vienna, Austria.

Peters, C. and Borri, F. (Eds.) (2004) *Cross Language Evaluation Forum: Working Notes for the CLEF 2004 Workshop (CLEF 2004)*, 15–17 September, ISTI-CNR, Pisa, Italy.

Prager, J., Brown, E., Coden, A. and Radev, D. (2000) 'Question-answering by predictive annotation', in *SIGIR '00: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.184–191, ACM, New York, NY, USA.

Prager, J., Radev, D. and Czuba, K. (2001) 'Answering what-is questions by virtual annotation', in *HLT '01: Proceedings of the First International Conference on Human Language Technology Research*, pp.1–5, Association for Computational Linguistics, Morristown, NJ, USA.

Radev, D.R., Qi, H., Zheng, Z., Blair-Goldensohn, S., Fan, W. and Prager, J. (2001) 'Mining the web for answers to natural language questions', in *Proceedings of the Tenth International Conference on Information and Knowledge Management*, ACM Press, pp.143–150.

Rauber, A. and Awadallah, R. (2001) 'Web-based multiple choice question answering for English and Arabic questions', in *Proceedings of the 28th European Conference on Information Retrieval (ECIR 2006)*, 10–12 April, pp.515–518.

Ruland, T., Spilker, J., Weber, H. and Worm, K.L. (1998) 'Making the most of multiplicity: a multi-parser multi-strategy architecture for the robust processing of spoken language', in *Proc. of the 1998 International Conference on Spoken Language Processing (ICSLP 98)*, pp.1163–1166.

Sanchis, E., Buscaldi, D., Grau, S., Hurtado, L. and Griol, D. (2006) 'Spoken QA based on a passage retrieval engine', *Spoken Language Technology Workshop, 2006*, IEEE, December, pp.62–65.

Schofield, E. and Zheng, Z. (2003) 'A speech interface for open-domain question-answering', in *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pp.177–180, Association for Computational Linguistics, Morristown, NJ, USA.

Shi, Z., Melli, G., Wang, Y., Liu, Y., Gu, B., Kashani, M.M., Sarkar, A. and Popowich, F. (2007) 'Question answering summarization of multiple biomedical documents', in *CAI '07: Proceedings of the 20th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence*, pp.284–295, Springer-Verlag, Berlin, Heidelberg.

Sliwko, L. and Nguyen, N.T. (2007) 'Using multi-agent systems and consensus methods for information retrieval in internet', *International Journal of Intelligent Information and Database Systems*, Vol. 1, No. 2, pp.181–198.

Sorg, P. (2008) 'Learning methods in multi-grained query answering', in *International Semantic Web Conference, Lecture Notes in Computer Science*, Vol. 5318, pp.926–931, Springer.

Stoyanov, V., Cardie, C. and Wiebe, J. (2005) 'Multi-perspective question answering using the OpQA corpus', in *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp.923–930, Association for Computational Linguistics, Morristown, NJ, USA.

Takechi, M., Tokunaga, T. and Matsumoto, Y. (2007) 'Chunking-based question type Identification for multi-sentence queries', in *Proceedings of the SIGIR 2007 Workshop on Focused Retrieval*, pp.41–48.

Tamura, A., Takamura, H. and Okumura, M. (2005) 'Classification of multiple-sentence questions', in *IJCNLP, Lecture Notes in Computer Science*, Vol. 3651, pp.426–437, Springer.

Tllez-Valero, A., Montes y Gmez, M., Villaseor-Pineda, L. and Peas, A. (2008) 'Improving question answering by combining multiple systems via answer validation', *Computational Linguistics and Intelligent Text Processing*, Vol. 4919, pp.544–554.

Trotman, A., Geva, S. and Kamps, J. (2007) 'Report on the SIGIR 2007 workshop on focused retrieval', *SIGIR Forum*, Vol. 41, No. 2, pp.97–103.

Turk, M. (2000) 'Perceptual user interfaces', *Communications of the ACM*, Vol. 43, pp.33–34.

Wilson, F. and Blakemore, M. (2008) 'ISAC multichannel inclusive e-government using natural language processing', August, available at http://www.mcegov.eu/media/488/080616-gironafinal.doc.

Woszczyna, M., et al. (1994) 'Janus 93: towards spontaneous speech translation', in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-94*, Vol. 1, pp.I/345–I/348.

Xu, K., Weng, F., Meng, H.M. and Luk, P.C. (2001) 'Multi-parser architecture for query processing', *Eurospeech 2001*, pp.1077–1080.

Yang, H., Chaisorn, L., Zhao, Y., Neo, S-Y. and Chua, T-S. (2003) 'VideoQA: question answering on news video', in *MULTIMEDIA '03: Proceedings of the Eleventh ACM International Conference on Multimedia*, pp.632–641, ACM, New York, NY, USA.

Yeh, T. and Darrell, T. (2008) 'Multimodal question answering for mobile devices', in *IUI '08: Proceedings of the 13th International Conference on Intelligent User Interfaces*, pp.405–408, ACM, New York, NY, USA.