Query-Based Summarization using Rhetorical Structure Theory

Wauter Bosma

Human Media Interaction, University of Twente

Abstract

Research on Question Answering is focused mainly on classifying the question type and finding the answer. Presenting the answer in a way that suits the user's needs has received little attention. This paper shows how existing question answering systems—which aim at finding precise answers to questions—can be improved by exploiting summarization techniques to extract more than just the answer from the document in which the answer resides. This is done using a graph search algorithm which searches for relevant sentences in the discourse structure, which is represented as a graph. The Rhetorical Structure Theory (RST) is used to create a graph representation of a text document. The output is an extensive answer, which not only answers the question, but also gives the user an opportunity to assess the accuracy of the answer (is this what I am looking for?), and to find additional information that is related to the question, and which may satisfy an information need. This has been implemented in a working multimodal question answering system where it operates with two independently developed question answering modules.

1 Introduction

A question answering (QA) system pinpoints an *answer* to a given question in a set of documents. A *response* is then generated for this answer, and presented to the user (c.f. Hirschman and Gaizauskas 2001). Discussion of the task of pinpointing the answer is beyond the scope of this paper. I will assume that the sentence which best matches the question, the *answer sentence*, is located by a QA system in a corpus of text documents. What remains is the task of generating an appropriate response and present it to the user.

Question answering systems traditionally try to find an 'exact answer'. An exact answer is a "text string consisting of a complete answer and nothing else" (Voorhees 2003). Strings that contain a correct answer with additional text are considered 'inexact'. Finding exact answers is also the focus of large-scale question answering evaluation programs such as TREC (Voorhees and Tice 2000).

Studies have shown, however, that users appreciate receiving more information than *only* the exact answer (Burger et al. 2000). Consulting a question answering system is only part of a user's attempt to fulfill an information need: it's not the end point, but some steps along what has been called a 'berry picking' process, where each answer/result returned by the system may motivate a follow-up step (Bates 1990). The user may not only be interested in the answer to the question, but also in related information. The 'exact answer approach' fails to show leads to related information that might also be of interest to the user. Lin et al. (2003) show that when searching for information, increasing the amount of text returned to users significantly decreases the number of queries that they pose to the system, suggesting that users utilize related information from supporting text.

In both commercial and academic QA systems, the response to a question tends to be more than the exact answer, but the sophistication of their responses varies from system to system. There are three degrees of sophistication in response generation.

Exact answer. The most basic form of answer presentation is to present only an exact answer. For instance, an exact answer to the question "what is the cause of RSI?" could be:

the movement always involves contraction of the same muscles

Answer plus context. If only an exact answer is provided, users have great difficulty assessing the accuracy of the answer, and thus whether the answer is correct. If the user is provided with more context (i.e. surrounding text), she will exploit this in order to find out whether the answer is indeed an answer to the question (Lin et al. 2003). Most of the current QA systems follow this approach, and return not only the answer but also part of the surrounding text, in which the answer itself may be highlighted. This can be a few lines of text, or only the single sentence in which the answer occurs. For instance, the response to the question about RSI causes could consist of the answer sentence, the preceding sentence and the sentence following the answer sentence:

Despite fewer working hours, the same quantity of work had to be finished. A possible explanation of the development of RSI as a result of frequently repeated movements which are performed with low exertion is that **the movement always involves contraction of the same muscles***. This happens for instance when working with a display device.*

Extensive answer. Lin et al. (2003) have shown that users prefer to receive more information than only an exact answer, but simply returning to the user a particular quantity of surrounding text is likely to produce incoherent results. Furthermore, the surrounding text may include irrelevant information or unnecessary details. Although—similarly to an answer plus context—an extensive answer includes more information than just the exact answer, the difference is that the extensive answer approach specifically aims at producing a coherent response that includes, apart from the answer, also related information which might interest the user. For instance, an extensive answer to the question about RSI causes could be:

A possible explanation of the development of RSI as a result of frequently repeated movements which are performed with low exertion is that **the movement always involves contraction of the same muscles***. This happens for instance when working with a display device. Eventually they can cease to function and the muscle will lose strength.*

This paper presents a method to produce extensive answers by extracting the sentences which are most salient with respect to the question, from the document which contains the answer. This is very similar to creating an extractive summarization: in both cases, the goal is to extract the most salient sentences from a document. In case of summarization, the result should reflect the communicative intent conveyed by the original document, i.e. the summarization contains the most salient parts of the original document. In question answering, what is relevant depends on the user's question rather than on the intention of the writer of the document which happens to contain the answer. In other words, the output of the summarization process is adapted to suit the user's declared information need (i.e. the question). This branch of summarization has been called *query-based summarization* (c.f. Chali 2002).

The method proposed here uses a pointer to the (exact) answer as a summarization parameter. The sentences which are most closely related to the answer sentence are extracted and the resulting extensive answer is presented to the user. This answer includes the answer sentence itself. For this type of summarization, determining the salience of a sentence as done in generic summarization no longer suffices. Instead of using a static notion of salience, the strength of the relation between the answer and each sentence is used for summarization. Rhetorical Structure Theory is used to find those relations.

In short, the following method is proposed. The rhetorical (RST) structure of the document to be summarized is transformed into a weighted graph, in which each vertex represents a sentence. The weight of an edge represents the distance between the two sentences. Given that a sentence a is relevant to the answer, the weight of a path from sentence a to another sentence b represents the level of relevance of sentence b to the answer. Given an appropriate assignment of weights in the graph, such a graph can be used to determine which sentences are the most relevant to the answer.

This paper is structured as follows. First, background knowledge about coherence, Rhetorical Structure Theory and summarization is provided in section 2. Section 3 discusses the proposal to answer extension and section 4 discusses its application in a real system. This paper concludes with a discussion and possible follow-ups on this research in section 5. Although this work is aimed at the Dutch language, all examples have been translated to English. This is possible because all methods presented in this paper are language independent.

2 Background

2.1 Coherence in Discourse

What makes discourse different from just any list of sentences, is that sentences in discourse are somehow related to each other, i.e. by means of coreference, substitution, ellipsis, conjunction and lexical cohesion (Halliday and Hasan 1976). All these phenomena account for relations between words or groups of words sentences in discourse. Such relations are called *cohesive* relations (Mani, Bloedorn and Gates 1998).

However, it is argued that there is more to discourse than only cohesion. Several theories have been developed to model the structure of discourse, most notably the intentional structure of Grosz and Sidner (1986) and the rhetorical structure (RST) of Mann and Thompson (1987). Both theories state that discourse can be segmented into non-overlapping spans of texts, that an intentional relation holds between those segments, and that a segment may in turn be further segmented into smaller segments which are also subject to an intentional relation.

The main difference between theories of text organization is the number of relation types that can be identified. Some argue that any coherence relation between two spans of text can be classified as one of a finite number (usually in the order of tenths) of rhetorical relation types (c.f. Mann and Thompson 1988). Others state that the number of possible rhetorical relations is ultimately infinite, so it makes no sense trying to classify relations or to define a definite relation set (c.f. Grosz and Sidner 1986). Instead, Grosz and Sidner (1986) restrict themselves to only two relations—DOMINANCE and SATISFACTION-PRECEDENCE.

2.2 Rhetorical Structure Theory

For the purpose of text summarization, RST has theoretical and pragmatic advantages over other theories. Good levels of agreement have been measured between human annotators of RST, which indicates that RST is well defined (Mann and Thompson 1988, den Ouden 2004). Furthermore, a corpus of RST-annotated English news articles is publically available, which can be used for training and evaluating RST-based summarization algorithms (Carlson, Marcu and Okurowski 2002). Another advantage of RST is that RST defines coherence relations very formally and elaborately, which makes computational applications easier to develop.

According to RST, a rhetorical relation typically holds between two contiguous spans, of which one span (the *nucleus*) is more central to the writer's intention that the other (the *satellite*), whose sole purpose is to increase the reader's understanding or belief of what is said in the nucleus. Sometimes, two related spans are of equal importance, in which case there is a *multinuclear* relation between them. The related spans form a new span, which can in turn participate in a relation with another span. The smallest units of discourse are *elementary discourse units* or *edus*.

The idea behind RST is that all rhetorical relations that can possibly occur in a text can be categorized into a finite set of relation types. The Rhetorical Structure Theory is primarily a method of text analysis. Mann and Thompson (1988) define a set of discourse relations that commonly occur in English texts, but RST has also been applied with other relation sets (such as in Carlson and Marcu 2001). The optimal relation set may depend on the genre and the application (Marcu and Echihabi 2002, André and Rist 1995)

2.3 Query-based Summarization

There are several flavors of summarization:

Abstractive vs. extractive. A feature of an extractive summarization is that each sentence of the summarization is literally copied from the source document. Abstracting involves *rewriting* a text in fewer words, rather than *extracting* the most salient portions of a text.

Multi-document vs. single-document. A multi-document summarization contains the most relevant information from a set of documents, while in single-document summarization, only a single document is used.

Query-based vs. generic. A query-based summarization is tailored to suit the user's declared information need, while a generic summarization reflects the writer's communicative intent as conveyed by the source document.

This paper discusses query-based single-document extracts—the summarization will not contain any sentences that are not present in the original document. The query is a question posed by the user. Because the answer is already pinpointed in a document by a question answering engine, a pointer to the answer can be used as a summarization parameter.

While creating an extract for a particular answer, a candidate sentence can only be included if something is known about the relation between the candidate sentence and the answer sentence. Indications of a strong relation between two sentences include statistical measures of text similarity, such as the number of denotations of mutually used concepts. This paper focuses on the use of rhetorical relations. More in particular, RST.

RST has proven to be very useful to facilitate summarization (Marcu 1997). In his summarization effort, Marcu used the nuclearity of relations in the rhetorical structure to determine which sentence is more salient, but he also explored other features as additional indicators of importance, such as sentence length (Marcu 1997, Marcu 1998).

The elementary discourse units of the RST analyses used for summarization are sentences. RST can be used to make a more detailed analysis of discourse, including relations between clauses, but for making an extractive summarization, using a finer granularity than sentences is not necessary. If more detailed analyses were used, the extract could also contain parts of sentences, but this would require rewriting the extracted text into a grammatical whole.

Query-based summarization has been applied in information retrieval (c.f. Chali 2002, Saggion, Bontcheva and Cunningham 2003), but also in multi-document summarization (Mani and Bloedorn 1997). In multi-document summarization—like in question answering—the source documents of the summarization are not written to satisfy the information need expressed by the query at hand.

Mani and Bloedorn (1997) used graphs to formalize relations between sentences inside a document for multi-document summarization. A spreading activation algorithm is then used to perform a query-based summarization, given a starting node that is selected for the query. Although Mani and Bloedorn (1997) aim at multi-document summarization, a similar graph-based algorithm to perform query-based summarization can also be applied in single-document summarization, as demonstrated by this paper.

3 An Approach to Query-Based Summarization Using RST

This section describes a two-step approach to query-based summarization. First, the relations between sentences are defined in a discourse graph. Then, this graph is used to perform the summarization. During the first step, the rhetorical structure is transformed into a graph representation. The second step exploits a graph search algorithm in order to extract the most salient sentences from the graph. The starting node of the search is the node representing the answer sentence.

The summarization should consist of the most salient sentences, given the starting node. This can be realized by determining the *distance* between the answer sentence and each of the other sentences. The sentences which are most closely related to the answer sentence are included in the summarization.

A simple measure of distance between two sentences would be the *linear distance*, i.e. the number of sentences in between the two sentences, given the linear order of the sentences in the text. For instance, a summarization could consist of the answer sentence and a number of successive (and/or preceding) sentences. However, experience shows that summarizations are often incoherent if they are based on solely this measure of distance between sentences. At paragraph boundaries, for instance, two contiguous sentences can be rhetorically very distant.

The distance between sentences can also be measured by their distance in the RST graph, which I call the *rhetorical distance*. The Rhetorical Structure Theory defines relations between two spans of text, which can be used to derive the distance from one sentence to another. The graph which is created from the rhetorical structure can be used as a computational model for summarization.

The most nuclear sentence of an RST analysis is the sentence which is most central to the writer's purpose. The graph ensures that, similarly to Marcu's approach, a nucleus is preferred over a satellite: in both summarization approaches, a satellite cannot be included in a generic summarization without its nucleus. The consequence is that in the specific case that the entry point of the summarization—the answer sentence—is the most nuclear sentence in the RST analysis, the result resembles the result of the summarization approach by Marcu (1997). However, the graph-based approach is more general in the sense that the summarization can start from any specific sentence rather than only the most nuclear sentence of the analysis.

RST analyses as weighted graphs

It is relatively straightforward to derive a graph from a rhetorical structure. While RST is not designed as a computational framework, graph theory is very suited for this purpose. A RST tree can be converted to a discourse graph by means of the following steps.

1. For each elementary discourse unit in the RST tree, create a vertex associated with it.

2. For each directed relation, create an edge from the nuclear sentences of the nucleus to the nuclear sentences of the satellite of the relation.

Figure 1: Rhetorical structure examples.

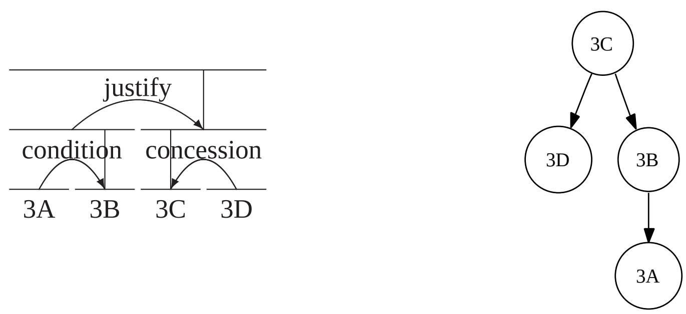

Figure 2: Rhetorical structure example and a discourse graph created for this rhetorical structure.

A sentence is a nuclear sentence of a text span if it is not part of any sub span (of the text span) which participates as a satellite in a directed relation with any other sub span. A text span can have multiple nuclear sentences if multinuclear relations are involved. For instance, in the RST diagram on the left in Figure 1, the set of nuclear sentences of the entire document (denoted as 1A:1D) contains only sentence 1C. The right diagram shows a rhetorical structure in which the set of nuclear sentences of 2A:2D consists of sentences 2C and 2D.

The result of the transformation is an a-cyclic directed graph of which the vertices correspond to elementary discourse units, and the edges define relations between them. Figure 2 shows an example of a rhetorical structure and a discourse graph that was created as described above. During the transformation from RST to graph, part of the structural information is lost because sentences of the graph are directly connected to other sentences, while in RST, one end of a relation can also span more than one discourse unit. If in RST one sentence was related to a text span of two sentences, it is related to the nucleus of the two sentences in the discourse graph. In practice, this means that if the inclusion of a sentence in a summarization was justified by a rhetorical relation, the nucleus of that relation must be included in the summarization as well. This is in line with Mann and Thompson's (1988) definition of directedness

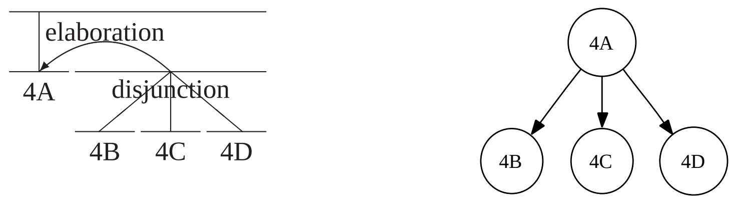

Figure 3: Rhetorical structure containing a multinuclear relation and the corresponding discourse graph.

of relations, which states that a nucleus of a directed relation has meaning without the satellite, but not the other way round.

If a multinuclear relation is involved, as in Figure 3, each of the sentences participating in the multinuclear relation (in the example: sentences 4B, 4C and 4D) is connected with the nucleus of the multinuclear span. That is, in the example, sentence 4A is connected to each of the sentences 4B, 4C and 4D, but sentences 4B–4D are not directly mutually connected. The reason for this is that in terms of RST, there is a mutual (multinuclear) relation between the sentences 4B–4D, but only in the context of this relation. They are mutually independent: if we know that 4B contains relevant information in a particular context, there is no way to be sure that, to any extent, 4C is relevant as well, based on the relevance of 4B.

Now we have a discourse graph T, we assume that given two sentences $a, b \in T$ for which there is a path from a to b, we can say that they are related and therefore if a is relevant to the answer, b is also relevant to the answer. If a path contains more than one edge, the sentences are related only indirectly and an indirect relation is weaker than a direct relation between two sentences.

The strength of a relation between two sentences could be calculated by just counting the number of edges in the path between the vertices of the sentences. However, it may be the case that there is more than one sentence with an equally long path to the starting point of the summarization. This means that during a summarization, the two sentences are equally likely to be included in the summarization, although there may be other indications of one sentence being better suited for inclusion in the summarization than the other.

In order to remedy this situation, we can assign weights to vertices and edges in the discourse graph. A greater distance is reflected by a greater weight. A low weight of the path from a to b indicates a high probability that b is relevant given, that a is relevant. The total weight of the path from a to b is denoted as $weight(a, b)$. The weight of a path between two sentences is defined as long as if there is a path that connects them. The weight of a path is the sum of the weights of its edges and vertices.

Given the entry point of the summarization (the answer sentence), the shortest path from the this sentence to any other sentence defines the relevance of the topic of the other sentence to the final answer. All we have to do now in order to be able to extract an answer, is to determine the weights.

3.1 Determining Weights

Weights of edges in the discourse graph can be determined by using features of the rhetorical structure from which the graph was created, such as features of the text spans on either side of the relation for which the edge was created as well as features of the relation itself. Also vertices can be weighted. The weight of a vertex depends on features of the sentence it corresponds to. The only constraint is that all weights of edges and vertices are non-negative.

The rhetorical structure has many features that may be relevant for determining weights to edges or vertices. Currently, only three features are considered when assigning weights. For these features, there is at least some evidence that they can contribute to the quality of a summarization. Further research may motivate the use of other features as well. For instance, the algorithm does not differentiate between relation types because there is not sufficiently specific evidence to support this. The following features are considered, in order of relative importance.

1. Each edge has a basic weight, which is the same for all edges in the graph. This makes the distinction between directly and indirectly related sentences explicit. Two sentences are less closely related if the path that connects them consists of more edges.

2. For each edge, a weight is added depending on the number of sentences in the satellite of the corresponding rhetorical relation. If a particular satellite contains more sentences than another satellite of the same nucleus, the author apparently spent more words on it, which may indicate that the author finds this topic more important than a shorter one, although they both are a satellite of the same nucleus.

3. For each vertex, a weight is added depending on the number of words in the sentence. According to Marcu (1998), this is a good measure for the amount of new information contained in the sentence.

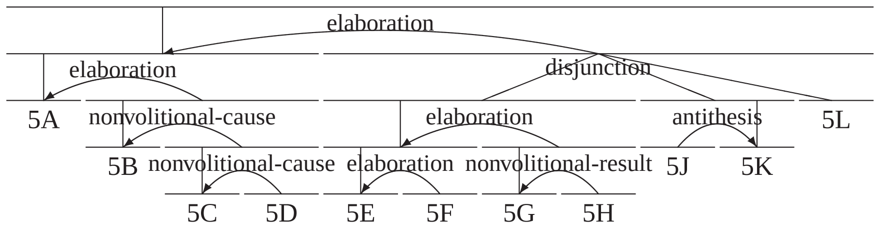

Figure 4: Rhetorical structure tree of the text fragment.

The weights of edges and vertices are calculated as follows.

$weight(e) =$

> $a + b \cdot \frac{1}{sentences(sat(r))}$, if e is the edge that was created for the relation r, where $sat(r)$ is the satellite of r, and $sentences(s)$ is defined as the number of sentences of a span s, a is the basic weight, and b is a constant factor of the 'satellite size' component of the edge weight;

$weight(v) =$

> $c \cdot \frac{1}{words(s)}$, if v is the vertex that was created for the sentence s, where $words(s)$ is the number of words in s, and c is a constant.

The constants a, b and c are used to balance the three factors of the distance between two sentences: the number of edges (represented by a) is more important than the number of sentences in the satellite (represented by b), and the number of sentences in the satellite is more important than the number of words in the sentence (represented by c).

Example 1: Extraction

This example shows how three sentences can be extracted from a text, based on its RST analysis, and given the entry point of the summarization. In a QA context, the entry point would be the answer sentence. Two of the extracted sentences are direct or indirect satellites of the answer sentence, the third is the answer sentence itself. The RST analysis of the following (segmented) text is shown in Figure 4. The entry point for the extraction is sentence 5E.

[A high pressure of workload, stress and repeatedly carrying out the same operation for a long period of time are the most important factors causing RSI to develop.]^{5A}

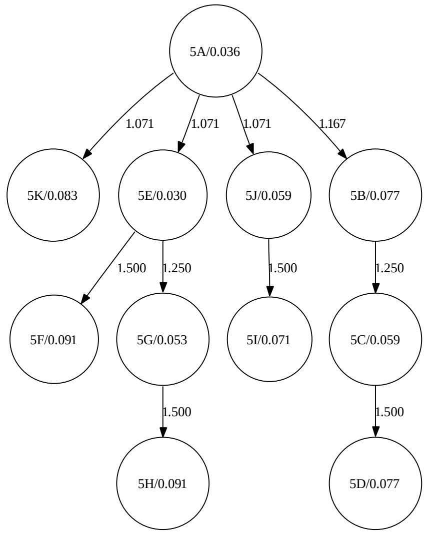

Figure 5: Weighted rhetorical structure graph of a text fragment. The vertices are labeled *sentence*/*weight*, in which *sentence* refers to the sentence corresponding to the vertex. The edges are labeled by their weights.

[In the Netherlands the work pressure increased with approximately 1.5% per year.]^{5B} [This is the result of shorter working hours in the eighties and nineties of the twentieth century.]^{5C} [Despite fewer working hours, the same quantity of work had to be finished.]^{5D} [A possible explanation of the development of RSI as a result of frequently repeated movements which are performed with low exertion is that the movement always involves contraction of the same muscles.]^{5E} [This happens for instance when working with a display device.]^{5F} [The motorial entities can be damaged because of oxygen lack and the impossibility of removing waste products.]^{5G} [Eventually they can cease to function and the muscle will lose strength.]^{5H} [There are however also indications that the complaints do not arise from damaged muscles.]^{5J} [Instead, they supposedly arise from abnormalities in the response of the brain to signals from the muscles.]^{5K} [Another possibility is that psychological factors can lead to symptoms of RSI.]^{5L}

First, a discourse graph is created from an RST analysis (as shown in Figure 5). The graph contains weighted edges and vertices. For this graph, the total weight of the paths from sentence 5E to each sentence in the graph is calculated using Dijkstra's shortest paths algorithm (Dijkstra 1959). A path in a graph is an alternating sequence of vertices and edges, beginning and ending with a vertex. For instance, in the graph of Figure 5, there is a path over three vertices and two edges from 5E to 5H. The weight of this path is the sum of the weights of all of its edges and vertices. In the

	5A	5B	5C	5D	5E	5F	5G	5H	5J	5K	5L
			...					...			
5E	—	—	—	—	0.030	1.621	1.333	2.924	—	—	—
			...					...			

Table 1: Weight table showing the total weight of the path from 5E to each sentence in the rhetorical structure graph of Figure 5.

Figure 6: Extraction graph of the three sentences selected for inclusion in the summary, and the corresponding structure in RST notation, which is derived from the original RST analysis.

case of the path from 5E to 5H, this is $0.03 + 1.25 + 0.053 + 1.5 + 0.091 = 2.924$.

The weights of the paths originating from 5E are shown in Table 1. Only four sentences are reachable from 5E. Since the selection of sentences is based on the weight of their path from 5E, a sentence which is associated with an unreachable vertex cannot be included in the extract.

From this table, the sentences with the cheapest path from the entry point 5E are selected. The selected sentences are filtered out, resulting in the discourse graph on the left in Figure 6. For the sentences in this graph, the rhetorical structure can be derived using the original RST analysis in Figure 4. The result is the rhetorical structure in Figure 6. This rhetorical structure may be used for further processing, for example for the purpose of speech synthesis (den Ouden 2004). The output of the extraction process would be the following text. The answer sentence is highlighted.

> **A possible explanation of the development of RSI as a result of frequently repeated movements which are performed with low exertion is that the movement always involves contraction of the same muscles.** This happens for instance when working with a display device. The motorial entities can be damaged because of oxygen lack and the impossibility of removing waste products.

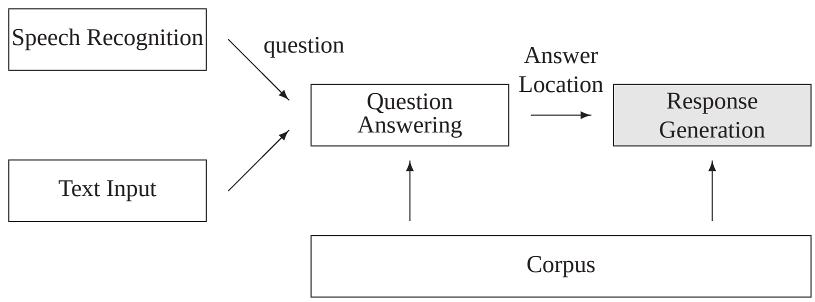

Figure 7: Simplified architecture of the IMIX system. The work in this paper is implemented in the 'response generation' module.

4 Answer Extraction in IMIX

The approach to query-based summarization is implemented as part of a working multimodal question answering system, which has been developed within the context of IMIX. IMIX is the Interactive Multimodal Information Extraction program of the Netherlands Organization for Scientific Research (NWO), with the objective of building a fully multimodal question answering dialog system (i.e. multimodal input and output). Currently, there is a first version of the IMIX system which is capable of answering typed and spoken questions in Dutch about medical issues. The answer is presented using speech, and an HTML page with text and images. Other IMIX modules are responsible for question answering, speech recognition, speech synthesis and the graphical user interface.

A simplified model of the architecture of the IMIX system is depicted in Figure 7. The Question Answering module receives a spoken or typed question from Speech Recognition or Text Input. The output of Question Answering is a pointer to a single sentence in a corpus, which is shared between Question Answering and Response Generation. This paper describes the 'Response Generation' module, which takes the question answering result (the answer sentence) as input for producing a coherent response. The Response Generation module has access to the QA corpus. Therefore, it has access to not only the sentence that was found by QA, but also to its context, i.e. to the entire document in which the answer sentence resides.

The response generation module in IMIX uses the summarization method described in this paper. Because in IMIX the system's response to questions has to be brief, the size of the responses is limited to a maximum of three sentences. The generated responses have not yet been formally evaluated, but information evaluations show that the responses are generally coherent, and that additional sentences (beyond the answer sentence) contain information which is strongly related to the question. The following are examples of responses that were generated for questions by the IMIX system.

Question: *What is RSI?*

Answer: **RSI is a name for a large number of diseases which affect the neck, shoulders, arms and hands.** *Repetitively making the same movements may cause complaints.*

Question: *What is the cause of RSI?*

Answer: **A possible explanation of the development of RSI as a result of frequently repeated movements which are performed with low exertion is that the movement always involves contraction of the same muscles**. *This happens for instance when working with a display device. Eventually they can cease to function and the muscle will lose strength.*

Although automated RST analysis can be performed on English texts (Marcu and Echihabi 2002), this is not yet the case for Dutch. Because Dutch is the interaction language of IMIX, the RST analyses used for extraction still have to be created manually. Because this is very time-consuming, at present, the RST-analyzed corpus is only a subset of the QA corpus. In cases where an RST analysis is missing, the response generation module falls back to giving only the answer sentence instead of a multi-sentence extract.

5 Discussion and Future Work

Question answering systems can benefit from responding with more extensive answers by means of query-based summarization. The presented approach to query-based summarization consists of two steps. First, the rhetorical structure tree is used to build distance graphs which determine the distances between individual sentences. Then, these graphs are used to decide which sentences are most relevant to the answer. The result is an answer that is more informative than an 'exact answer' (as returned by traditional QA systems), and more concise than a full document (as returned by IR systems)—a compromise between question answering and information retrieval, taking the best features from both.

The advantage of the separation between formalization (graph construction) and extraction (graph search and sentence extraction) is that the latter is fairly generic: it can also be applied to discourse graphs that are not RST-based. Mani and Bloedorn (1997) experimented with summarization based on conceptual similarity relations between sentences. The conceptual graphs could be integrated with the RST-based graphs, in order to exploit all available indications of relevance.

The extraction method has been tested with promising results on a limited scale in the IMIX question answering system, but more thorough experiments are required in order to test both the performance of the approach and the validity of the more general case of extending answers using the source document.

Future versions of the IMIX system will be capable of participating in more complex dialogs than just answering isolated questions. Because the summarizer is aware of coherence relations, its output is also RST-annotated text. Being able to reason about its output is very useful for a dialog system in order to parse and reply to subsequent utterances of the user. For instance, it would be useful for a dialog system to know that part of its output participates in an 'evidence' relation with another portion of the output. RST can also be used to improve speech synthesis (den Ouden 2004).

Another challenge is to investigate how query-based summarization methods apply to multimodal documents. Rhetorical Structure Theory itself applies to multimodal documents without any extensive modifications (c.f. André 1994), but this direction of RST has to be further explored, and further tools have to be developed for the generation of multimedia responses including pictures and animations.

Acknowledgements

This work is funded by the Interactive Multimodal Information Extraction (IMIX) program of the Netherlands Organization for Scientific Research (NWO).

References

André, E. (1994), *Ein plan-basierter Ansatz zur Generierung multimedialer Präsentationen*, PhD thesis, Universität des Saarlandes, Saarbrücken.

André, E. and Rist, T. (1995), Generating coherent presentations employing textual and visual material, *Artificial Intelligence Review, Special Volume on the Integration of Natural Language and Vision Processing* **9**(2–3), 147–165.

Bates, M. J. (1990), The berry-picking search: user interface design, *in* H. Thimbleby (ed.), *User Interface Design*, Addison-Wesley.

Burger, J., Cardie, C., Chaudhri, V., Gaizauskas, R., Harabagiu, S., Israel, D., Jacquemin, C., Lin, C.-Y., Maiorano, S., Miller, G., Moldovan, D., Ogden, B., Prager, J., Riloff, E., Singhal, A., Shrihari, R., Strzalkowski, T., Voorhees, E. and Weishedel, R. (2000), Issues, tasks, and program structures to roadmap research in question & answering (q&a), NIST DUC Vision and Roadmap Documents.

Carlson, L. and Marcu, D. (2001), Discourse tagging manual, *ISI Tech Report ISI-TR-545*, Information Sciences Institute.

Carlson, L., Marcu, D. and Okurowski, M. E. (2002), Building a discourse-tagged corpus in the framework of rhetorical structure theory, *in* J. van Kuppevelt and R. Smith (eds), *Current Directions in Discourse and Dialogue*.

Chali, Y. (2002), Generic and query-based text summarization using lexical cohesion, *in* R. Cohen and B. Spencer (eds), *Advances in Artificial Intelligence: 15th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2002*, Calgary, Canada, pp. 293–302.

den Ouden, H. (2004), *Prosodic realizations of text structure*, PhD thesis, University of Tilburg.

Dijkstra, E. (1959), A note on two problems in connection with graphs, *Numerische Mathematik* **1**, 269–271.

Grosz, B. J. and Sidner, C. L. (1986), Attention, intentions and the structure of discourse, *Computational Linguistics* **12**(3), 175–204.

Halliday, M. A. and Hasan, R. (1976), *Cohesion in English*, Longman, London.

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

Lin, J., Quan, D., Sinha, V., Bakshi, K., Huynh, D., Katz, B., and Karger, D. R. (2003), What makes a good answer? the role of context in question answering, *Proceedings of the Ninth IFIP TC13 International Conference on Human-Computer Interaction*, Zürich, Switzerland.

Mani, I. and Bloedorn, E. (1997), Multi-document summarization by graph search and matching, *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI'97)*, pp. 622–628.

Mani, I., Bloedorn, E. and Gates, B. (1998), Using cohesion and coherence models for text summarization, *Proceedings of AAAI Spring Symposium on Intelligent Text Summarization*, Stanford.

Mann, W. C. and Thompson, S. A. (1987), Rhetorical structure theory: A theory of text organization, *Technical Report ISI/RS-87-190, NTIS Identifying Number ADA 183038*, University of Southern California, Information Sciences Institute, Marina del Rey, CA, United States.

Mann, W. C. and Thompson, S. A. (1988), Rhetorical structure theory: Toward a functional theory of text organization, *Text* **8**(3), 243–281.

Marcu, D. (1997), From discourse structures to text summaries, *The Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization*, Madrid, Spain, pp. 82–88.

Marcu, D. (1998), To build text summaries of high quality, nuclearity is not sufficient, *The Working Notes of the the AAAI-98 Spring Symposium on Intelligent Text Summarization*, AAAI, Stanford, CA, pp. 1–8.

Marcu, D. and Echihabi, A. (2002), An unsupervised approach to recognizing discourse relations, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, Philadelphia, PA.

Saggion, H., Bontcheva, K. and Cunningham, H. (2003), Robust generic and query-based summarization, *10th Conference of the European Chapter of the Association for Computational Linguistics*, *EACL-2003*, Hungary.

Voorhees, E. M. (2003), Overview of the trec 2003 question answering track, *Proceedings of the Twelfth Text REtrieval Conference (TREC 2003)*.

Voorhees, E. M. and Tice, D. M. (2000), Building a question answering test collection, *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, ACM Press, Athens, Greece, pp. 200–207.