Using Scenario Knowledge in Automatic Question Answering

Sanda Harabagiu and Andrew Hickl
Language Computer Corporation
1701 North Collins Boulevard
Richardson, Texas 75080 USA
`sanda@languagecomputer.com`

Abstract

This paper describes a novel framework for using scenario knowledge in open-domain Question Answering (Q/A) applications that uses a state-of-the-art textual entailment system (Hickl et al., 2006b) in order to discover textual information relevant to the set of topics associated with a scenario description. An intrinsic and an extrinsic evaluation of this method is presented in the context of an automatic Q/A system and results from several user scenarios are discussed.

1 Introduction

Users of today's automatic question-answering (Q/A) systems generally have complex information needs that cannot be satisfied by asking single questions in isolation. When users interact with Q/A systems, they often formulate sets of queries that they believe will help them gather the information that needed to perform one or more specific tasks. While human users are generally able to identify their information needs independently, the information needs of organizations are often presented in the form of short prose descriptions – known as *scenarios* – which outline the range of knowledge sought by a customer in order to achieve a specific outcome or to accomplish a particular task. (An example of one scenario is presented in Figure 1.)

Recent work in Q/A has sought to use information derived from these kinds of scenarios in order to retrieve sets of answers that are more relevant – and responsive – to a customer's information needs. While (Harabagiu et al., 2005) used *topic signatures* (Lin and Hovy, 2000;

Scenario Description
The customer has commissioned a research project looking at the impact of the outsourcing of American jobs on the United States' relationship with India. After conducting research on U.S. companies currently doing business in India, the customer wants to know why American corporations have sought to outsource jobs to India, the types of economic advantages that American companies could gain from relocating to India, and the kinds of economic or political inducements that India has offered to American companies looking to outsource jobs there. The customer is not interested in demographic information on Indian employees of American firms.

Table 1: Example of a User Scenario.

Harabagiu, 2004) computed automatically from collections of documents relevant to a scenario in order to approximate the semantic content of a scenario, (Narayanan and Harabagiu, 2004) employed formal models of the interrelated events, actions, states, and relations implicit to a scenario in order to produce fine-grained, context-sensitive inferences that could be used to answer questions. Scenario knowledge was also included in the form of axiomatic logic transformation developed in (Moldovan et al., 2003). Under this approach, information extracted from the scenario narrative is converted to logical axioms that can used in conjunction with a logic prover in order justify answers returned for questions.

In this paper, we propose that scenario-relevant passages in natural language texts can be identified by recognizing a semantic relation, known as *contextual entailment* (CE), that exists between a text passage and one of a set of subquestions that are conventionally implied by a scenario. Under this model, we expect that a scenario S can be considered to contextually entail a passage t, when there exists at least one subquestion q derived from S that textually entails the passage t. We show that by using a state-of-the-art textual entailment system (Hickl et al., 2006b), we can provide Q/A systems with another mechanism for approximating the inference between questions and relevant answers. We show how each of these cases of con-

textual entailment can be computed and how it can be used in the intrinsic and extrinsic evaluation of a Q/A system.

The remainder of the paper is organized in the following way. Section 2 introduces our notion of contextual entailment and provides a framework for recognizing instances of CE between scenarios and both questions and answers. Section 3 describes the textual entailment system used at the core of our CE system. Sections 4 and 5 describe separate frameworks for intrinsically and extrinsically evaluating the impact of CE on current Q/A systems. Section 6 presents results from our evaluations, and Section 7 summarizes our conclusions

2 Recognizing Contextual Entailment

We define *contextual entailment* (CE) as a directional relation that exists between a text passage t and one of a set of implicit subquestions q that can be derived from a user's interpretation of a scenario. Informally, we consider that a scenario S *contextually entails* a passage t when there exists at least one subquestion q implied by S that can be considered to entail t.

We expect that the meaning of an information-seeking scenario S can be represented as a question under discussion (QUD) Q_S, which denotes a partially-ordered set of subquestions ($q \in Q_S$) that represent the entire set of questions that could potentially be asked in order to gather information relevant to S. Taken together, we expect these subquestions to represent the widest possible construal of a user's information need given S.
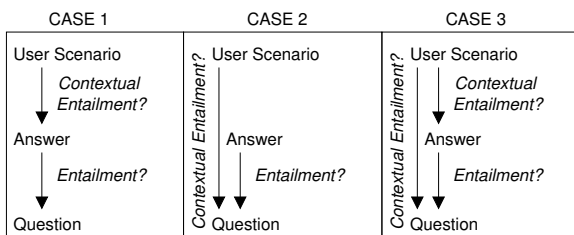

Figure 1: Three types of Contextual Entailment

We believe the set of subquestions implied by Q_S can be used to test whether a text passage is relevant to S. Since the formal answerhood relation between a question and its answer(s) can be cast in terms of (logical) entailment (Groenendijk, 1999; Lewis, 1988), we believe that systems for recognizing *textual entailment* (Dagan et al., 2005) could be used in order to identify those text passages that should be considered when gath-

ering information related to a scenario. Based on these assumptions, we expect that the set of text passages that are textually entailed by subquestions derived from a scenario represent information that is more likely to be relevant to the overall topic of the scenario as a whole.

We expect that there are three types of contextual entailment relationships that could prove useful for automatic Q/A systems. First, as illustrated in Case 1 in 1, CE could exist between a scenario and one of the set of answers returned by a Q/A system in response to a user's question. Second, as in Case 2, CE could be established directly between a scenario and the question asked by the user. Finally, as in Case 3, CE could be established both between a scenario and a user's question as well as between a scenario and one of the answers returned by the Q/A system for that question.

Figure 2 provides examples of each of these three types of contextual entailment using the scenario presented in Figure 1.
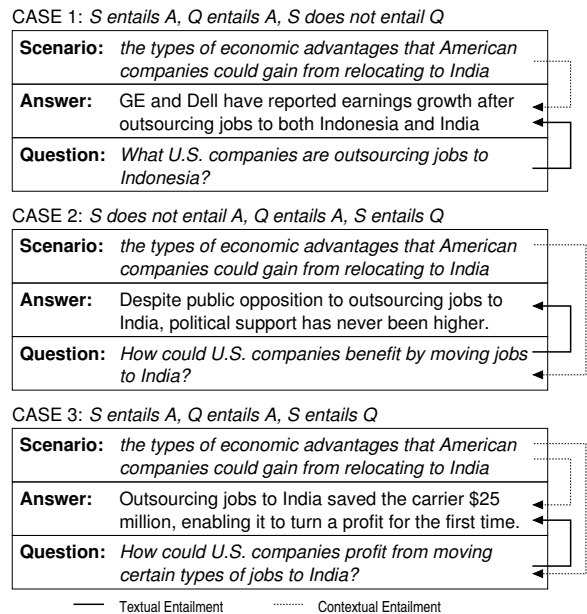

Figure 2: Examples of Contextual Entailment.

In Case 1, the scenario textually entails the meaning of the answer passage, as *earnings growth* from outsourcing necessarily represents one of the types of *economic advantages* that can be derived from outsourcing. However, the scenario cannot be seen as entailing the user's question, as the user's interest in job outsourcing in Indonesia cannot be interpreted as being part of the topics that are associated with the scenario. In this case, recognition of contextual entailment would allow systems to be sensitive to the types of

scenario-relevant information that is encountered – even when the user asks questions that are not entailed by the scenario itself. We expect that this type of contextual entailment would allow systems to identify scenario-relevant knowledge throughout a user's interaction with a system, regardless of topic of a user's last query.

In Case 2, the user's question is entailed by the scenario, but no corresponding entailment relationship can be established between the scenario and the answer passage identified by the Q/A system as an answer to the question. While *political support* may be interpretable as one of the benefits realized by companies that outsource, it cannot be understood as one of the *economic advantages* of outsourcing. Here, recognizing that contextual entailment could not be established between the scenario and the answer – but could be established between the scenario and the question – could be used to signal the Q/A system to consider additional answers before moving on to the user's next question. By identifying contextual entailment relationships between answers and elements in a scenario, systems could perform valuable forms of answer validation that could be used to select only the most relevant answers for a user's consideration.

Finally, in Case 3, entailment relationships exist between the scenario and both the user's question and the returned answer, as *saving $25 million* can be considered to be both an *economic advantage* and one of the ways that companies *profit* from outsourcing. In this case, the establishment of contextual entailment could be used to inform topic models that could be used to identify and extract other similarly relevant passages for the user.

In order to capture these three types of CE relationships, we developed the architecture for recognizing contextual entailment illustrated in Figure 3.

This architecture includes three basic types of modules: (1) a *Context Discovery* module, which identifies passages relevant to the concepts mentioned in a scenario, (2) a *Textual Entailment* module, which recognizes implicational relationships between passages, and (3) a *Entailment Merging* module, which ranks relevant passages according to their relevance to the scenario itself. In *Context Discovery*, document retrieval queries are first extracted from each sentence found in a scenario. Once a set of documents has been as-
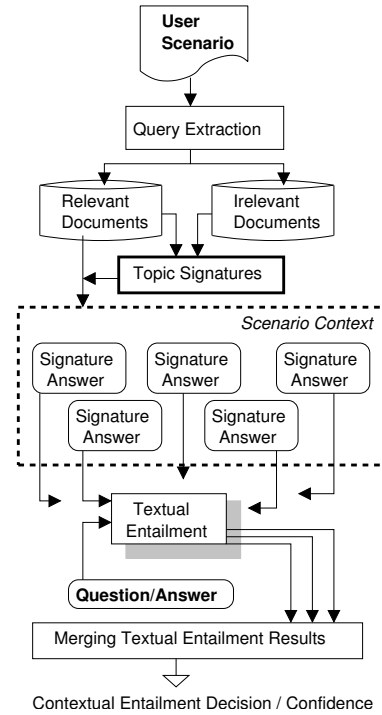

Figure 3: Contextual Entailment Architecture.

sembled, *topic signatures* (Lin and Hovy, 2000; Harabagiu 2004) are computed which identify the set of topic-relevant concepts – and relations between concepts – that are found in the relevant set of documents. Weights associated with each set of topic signatures are then used to extract a set of relevant sentences – referred to as *topic answers* – from each relevant document. Once a set of topic answers have been identified, each topic answer is paired with a question submitted by a user and sent to the *Textual Entailment* system described in Section 2. Topic answers that are deemed to be positive entailments of the user question are assigned a confidence value by the TE system and are then sent to a *Entailment Merging* module, which uses logistic regression in order to rank passages according to their expected relevance to the user scenario. Here, logistic regression is used to find a set of coefficients b_j (where $0 \leq j \leq p$) in order to fit a variable x to a logistic transformation of a probability q.

$$\text{logit}(q) = \log \frac{q}{1-q} = b_0 + \sum_{j=1}^{p} b_j x_j + e$$

We believe that since logistic regression uses a maximum likelihood method, it is a suitable technique for normalizing across range of confidence values output by the TE system.
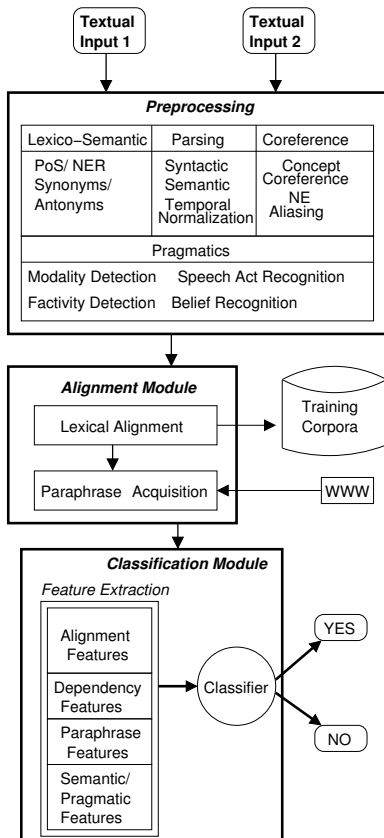
Figure 4: Textual Entailment Architecture.

3 Recognizing Textual Entailment

Recent work in computational semantics (Haghighi et al., 2005; Hickl et al., 2006b; MacCartney et al., 2006) has demonstrated the viability of supervised machine learning-based approaches to the recognition of textual entailment (TE). While these approaches have not incorporated the forms of structured world knowledge featured in many logic-based TE systems, classification-based approaches have been consistently among the top-performing systems in both the 2005 and 2006 PASCAL Recognizing Textual Entailment (RTE) Challenges (Dagan et al., 2005), with the best systems (such as (Hickl et al., 2006b)) correctly identifying instances of textual entailment more than 75% of the time.

The architecture of our TE system is presented in Figure 4.[1] Pairs of texts are initially sent to a *Preprocessing Module*, which performs syntactic and semantic parsing of each sentence, resolves coreference, and annotates entities and predicates with a wide range of lexico-semantic and prag-

---

[1]For more information on the TE system described in this section, please see (Hickl et al., 2006b) and (Harabagiu and Hickl, 2006).

matic information, including named entity information, synonymy and antonymy information, and polarity and modality information.

Once preprocessing is complete, texts are then sent to an *Alignment Module*, which uses *lexical alignment* module in conjunction with a *paraphrase acquisition module* in order to determine the likelihood that pairs of elements selected from each sentence contain corresponding information that could be used in recognizing textual entailment. *Lexical Alignment* is performed using a Maximum Entropy-based classifier which computes an alignment probability $p(a)$ equal to the likelihood that a term selected from one text corresponds to an element selected from another text. Once these pairs of corresponding elements are identified, alignment information is then used in order to extract portions of texts that could be related via one or more phrase-level alternations or "paraphrases". In order to acquire these alternations, the most likely pairs of aligned elements were then sent to a *Paraphrase Acquisition* module, which extracts sentences that contain instances of both aligned elements from the World Wide Web.

Output from these two modules are then combined in a final *Classification Module*, which uses features derived from (1) lexico-semantic properties, (2) semantic dependencies, (3) predicate-based features (including polarity and modality), (4) lexical alignment, and (5) paraphrase acquisition in order learn a decision tree classifier capable of determining whether an entailment relationship exists for a pair of texts.

4 Intrinsic Evaluation of Contextual Entailment

Since we believe CE is intrinsic to the Q/A task, we have evaluated the impact of contextual entailment on our Q/A system in two ways. First, we compared the quality of the answers produced, with and without contextual entailment. Second, we evaluated the quality of the ranked list of paragraphs against the list of entailed paragraphs identified by the CE system and the set of relevant answers identified by the Q/A system. This comparison was performed for each of the three cases of entailment presented in Figure 2.

We have explored the impact of knowledge derived from the user scenario through different forms of contextual entailment by comparing the
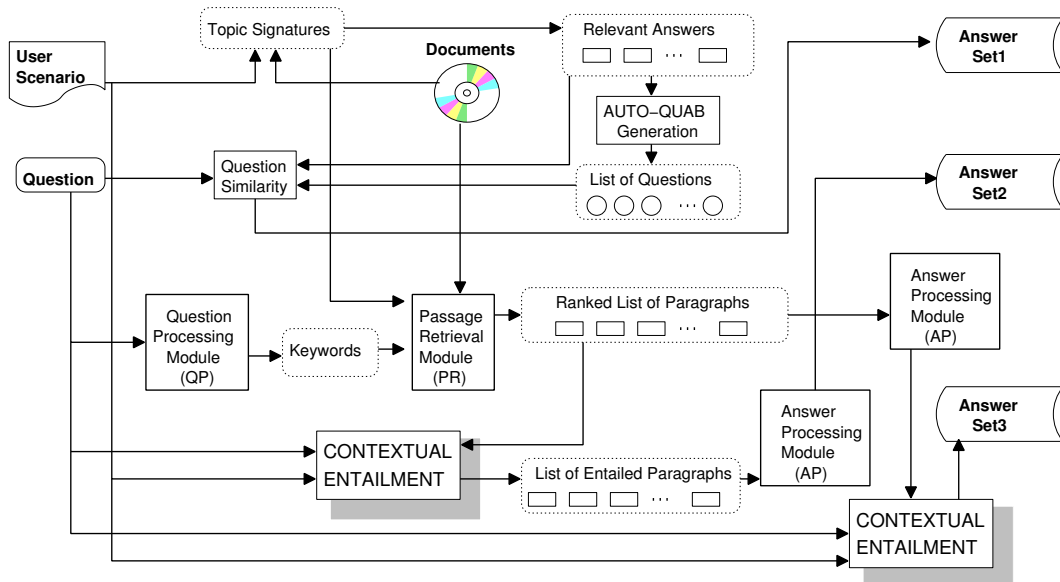
Figure 5: Framework for Intrinsic Evaluation of Contextual Entailment in Q/A.

results of such knowledge integration in a Q/A system against the usage of scenario knowledge reported in (Harabagiu et al., 2005).

Topic signatures, derived from the user scenario and from documents are used to establish text passages that are relevant to the scenario, and thus constitute relevant answers. For each such answer, one or multiple questions were built automatically with the method reported in (Harabagiu et al., 2005). When a new question was asked, its similarity to any of the questions generated based on the knowledge of the scenario is computed, and its corresponding answer is provided as an answer for the current question as well. Since the questions are ranked by similarity to the current question, the answers are also ranked and produce the Answer Set$_1$ illustrated in Figure 5.

When a Q/A system is used for answering the question, the scenario knowledge can be used in two ways. First, the keywords extracted by the Question Processing module can be enhanced with concepts from the topic signatures to produce a ranked list of paragraphs, resulting from the Passage Retrieval Module. These passages together with the question and the user scenario are used in one of the contextual entailment configurations to derive a list of entailed paragraphs from which the Answer Processing module can extract the answer set 2 illustrated in Figure 5. In another way, the ranked list of paragraphs is passed to the Answer Processing module, which provides a set of ranked answers to the contextual entailment configurations to derive a list of entailed answers, rep-

resented as answer set 3 in Figure 5. We evaluate the quality of each set of answers, and for the answer set 2 and 3, we produce separate evaluation for each configuration for the contextual entailment.

5 Extrinsic Evaluation of Contextual Entailment

Questions asked in response to a user scenario tend to be complex. Following work in (Hickl et al., 2004), we believe complex questions can be answered in one of two ways: either by (1) using techniques (similar to the ones proposed in (Harabagiu et al., 2006)) for automatically decomposing complex questions into sets of informationally-simpler questions, or by (2) using a multi-document summarization (MDS) system (such as the one described in (Lacatusu et al., 2006)) in order to assemble a ranked list of passages which contain information that is potentially relevant to the user's question.

First, we expect that contextual entailment can be used to select the decompositions of a complex question that are most closely related to a scenario. By assigning more confidence to the decompositions that are contextually entailed by a scenario, systems can select a set of answers that are relevant to both the user scenario – and the user's question. In contrast, contextual entailment can be used in conjunction with the output of a MDS system: once a summary has been constructed from the passages retrieved for a query, contextual en-
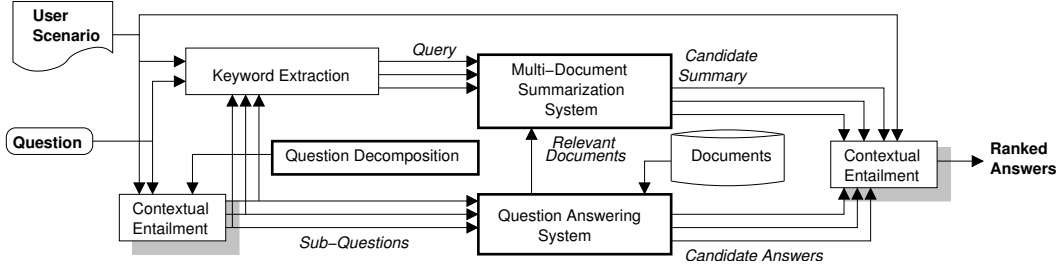
Figure 6: Framework for Extrinsic Evaluation of Contextual Entailment in Q/A.

tailment can be used to select the most relevant sentences from the summary.

The architecture of this proposed system is illustrated in Figure 6.

When using contextual entailment for selecting question decompositions, we rely on the method reported in (Harabagiu et al., 2006) which generates questions by using a random walk on a bipartite graph of salient relations and answers. In this case, the recognition of entailment between questions operates as a filter, forcing questions that are not entailed by any of the signature answers derived from the scenario context (see Figure 3) to be dropped from consideration.

When entailment information is used for re-ranking candidate answers, the summary is added to the scenario context, each summary sentence being treated akin to a signature answer. We believe that the summary contains the most informative information from both the question and the scenario, since the queries that produced it originated both in the question and in the scenario. By adding summary sentences to the scenario context, we have introduced a new dimension to the processing of the scenario. The contextual entailment is based on the textual entailments between any of the texts from the scenario context and any of the candidate answers.

6 Experimental Results

In this section, we present preliminary results from four sets of experiments which show how forms of textual – and contextual – entailment can enhance the quality of answers returned by an automatic Q/A system.

Questions used in these experiments were gathered from human interactions with the interactive Q/A system described in (Hickl et al., 2006a). A total of 6 users were asked to spend approximately 90 minutes gathering information related to three different information-gathering scenarios similar

to the one in Table 1. Each user researched two different scenarios, resulting in a total of 12 total research sessions. Once all research sessions were completed, linguistically well-formed questions were extracted from the system logs for each session for use in our experiments; ungrammatical questions or keyword-style queries were not used in our experiments. Table 2 presents a breakdown of the total number of questions collected for each of the 6 scenarios.

Scenario Name	Users	Total Qs	Avg. Q/Session	σ^2
S_1. India Outsourcing	4	45	11.25	2.50
S_2. Chinese WMD Proliferation	4	38	9.50	6.45
S_3. Libyan Bioweapons Programs	4	63	15.75	2.22
Total	12	146	12.17	1.23

Table 2: Questions Collected from User Experiments.

In order to evaluate the performance of our Q/A system under each of the experimental conditions described below, questions were re-submitted to the Q/A system and the top 10 answers were retrieved. Two annotators were then tasked with judging the correctness – or "relevance" – of each returned answer to the original question. If the answer could be considered to provide either a complete or partial answer to the original question, it was marked as *correct*; if the answer contained information that could not be construed as an answer to the original question, it was marked as *incorrect*.

6.1 Textual Entailment

Following (Harabagiu and Hickl, 2006), we used TE information in order to filter answers identified by the Q/A system that were not entailed by the user's original question. After filtering, the top-ranked entailed answer (as determined by the Q/A system) was returned as the system's answer to the question. Results from both a baseline version and a TE-enhanced version of our Q/A system are presented in Table 4.

Although no information from the scenario was used in this experiment, performance of the Q/A

		S$_1$	S$_2$	S$_3$	Total
# of Questions		45	38	63	146
baseline	top 1	8 (17.78%)	6 (15.79%)	11 (17.46%)	25 (17.12%)
TE	top 1	10 (22.22%)	8 (21.05%)	16 (25.40%)	34 (23.29%)
baseline	top 5	17 (37.78%)	16 (42.11%)	27 (42.86%)	60 (41.10%)
TE	top 5	20 (44.44%)	17 (44.74%)	32 (50.79%)	69 (47.26%)

Table 3: Impact of Textual Entailment on Q/A.

system increased by more than 6% over the baseline system for each of the three scenarios. These results suggest that TE can be used effectively in order to boost the percentage of relevant answers found in the top answers returned by a system: by focusing only on answers that are entailed by a user's question, we feel that systems can better identify passages that might contain information relevant to a user's information need.

6.2 Contextual Entailment

In order to evaluate the performance of our contextual entailment system directly, two annotators were tasked with identifying instances of CE amongst the passages and answers returned by our Q/A system. Annotators were instructed to mark a passage as being *contextually entailed* by a scenario only when the passage could be reasonably expected to be associated with one of the subtopics they believed to be entailed by the complex scenario. If the passage could not be associated with the extension of any subtopic they believed to be entailed by the scenario, annotators were instructed to mark the passage as not being contextually entailed by the scenario. For evaluation purposes, only examples that were marked by both annotators were considered as valid examples of CE.

Annotators were tasked with evaluating three types of output from our Q/A system: (1) the ranked list of passages retrieved by our system's *Passage Retrieval* module, (2) the list of passages identified as being CE by the scenario, and (3) the set of answers marked as being CE by the scenario (AnsSet$_3$). Results from the annotation of these passages are presented in Table 4.

	S$_1$		S$_2$		S$_3$		Total	
	#	%Rel	#	%Rel	#	%Rel	#	%Rel
Ranked Paragraphs	450	40.4%	380	31.3%	630	42.5%	1460	39.3%
Entailed Paragraphs	112	46.5%	87	44.8%	149	52.4%	348	48.6%
Answer Set 3	304	44.4%	188	39.9%	322	49.1%	814	45.2%

Table 4: Distribution of CE.

Annotators marked 39.3% of retrieved passages as being CE by one of the three scenarios. This number increased substantially when only passages identified by the CE system were considered, as annotators judged 48.6% of CE passages

and 45.2% of CE-filtered answers to be valid instances of contextual entailment.

6.3 Intrinsic Evaluation

In order to evaluate the impact of CE on a Q/A system, we compared the quality of answers produced (1) when no CE information was used (AnsSet$_1$), (2) when CE information was used to select a list of entailed paragraphs that were submitted to an *Answer Processing* module (AnsSet$_2$), and (3) when CE information was used directly to select answers (AnsSet$_3$). Results from these three experiments are presented in Table 5.

		S$_1$	S$_2$	S$_3$	Total
# of Questions		45	38	63	146
AnsSet$_1$	top 1	12 (26.67%)	9 (23.68%)	19 (30.16%)	40 (27.39%)
AnsSet$_2$	top 1	16 (35.56%)	11 (28.95%)	26 (41.27%)	53 (36.30%)
AnsSet$_3$	top 1	14 (31.11%)	15 (39.47%)	31 (49.21%)	60 (41.09%)
AnsSet$_1$	top 5	21 (46.67%)	17 (44.74%)	30 (47.62%)	68 (46.58%)
AnsSet$_2$	top 5	24 (53.33%)	18 (47.37%)	35 (55.55%)	77 (52.74%)
AnsSet$_3$	top 5	29 (64.44%)	20 (52.63%)	39 (61.90%)	88 (60.27%)

Table 5: Intrinsic Evaluation of CE on Q/A Performance.

As with the TE-based experiments described in Section 7.1, we found that the Q/A system was more likely to return at least one relevant answer among the top-ranked answers when contextual entailment information was used to either rank or select answers. When CE was used to rank passages for Answer Processing (AnsSet$_2$), accuracy increased by nearly 9% over the baseline (AnsSet$_1$), while accuracy increased by almost 14% overall when CE was used to select answers directly (AnsSet$_3$).

6.4 Extrinsic Evaluation

In order to evaluate the performance of the framework illustrated in Figure 6, we compared the performance of a question-focused MDS system (first described in (Lacatusu et al., 2006)) that did not use CE with a similar system that used CE to rank passages for a summary answer.

When CE was not used, sentences identified by the system's Q/A and MDS system for each question were combined and ranked based on number of question keywords found in each sentence. In the CE-enabled system (analogous to the system depicted in Figure 6), only the sentences that were contextually entailed by the scenario were considered; sentences were then ranked using the real-valued *entailment confidence* computed by the CE system for each sentence. Results from this system are presented in Table 6.

Although the CE-enabled system was more likely to return a scenario-relevant sentence in top

	S_1	S_2	S_3	Total
# of Questions	45	38	63	146
Without CE top 1	14 (31.11%)	15 (39.47%)	31 (49.21%)	60 (41.09%)
With CE top 1	20 (44.44%)	16 (42.11%)	32 (50.79%)	68 (48.23%)
Without CE top 5	29 (64.44%)	20 (52.63%)	39 (61.90%)	88 (60.27%)
With CE top 5	29 (64.44%)	21 (55.26%)	40 (63.49%)	90 (61.64%)

Table 6: Extrinsic Evaluation.

position (48.23%) than the system that did not use CE (41.09%), differences between the systems were much less apparent when the top 5 answers returned by each system were compared.

7   Conclusions

This paper introduced a new form of textual entailment, known as contextual entailment, which can be used to recognize scenario-relevant information in both the questions users ask and in the answers that automatic Q/A systems return. In addition to outlining a framework for recognizing contextual entailment in texts, we showed that contextual entailment information can significantly enhance the quality of answers returned by a Q/A system in response to users' questions about a particular scenario. In our evaluations, we found that using contextual entailment allowed Q/A systems to improve their accuracy by more than 10% overall.

8   Acknowledgments

References

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2005. The PASCAL Recognizing Textual Entailment Challenge. In *Proceedings of the PASCAL Challenges Workshop*.

Jeroen Groenendijk. 1999. The logic of interrogation: Classical version. In *Proceedings of the Ninth Semantics and Linguistics Theory Conference (SALT IX)*, Ithaca, NY.

Aria Haghighi, Andrew Ng, and Christopher Manning. 2005. Robust textual inference via graph matching. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, Vancouver, British Columbia, Canada, October.

Sanda Harabagiu and Andrew Hickl. 2006. Methods for using textual entailment in open-domain question-answering. In *Proceedings of the Joint International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics (COLING-ACL 2006)*, Sydney, Australia.

Sanda Harabagiu, Andrew Hickl, John Lehmann, and Dan Moldovan. 2005. Experiments with Interactive Question-Answering. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*.

Sanda Harabagiu, Finley Lacatusu, and Andrew Hickl. 2006. Answering complex questions with random walk models. In *2006 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, Seattle, WA.

Sanda Harabagiu. 2004. Incremental Topic Representations. In *Proceedings of the 20th COLING Conference*, Geneva, Switzerland.

Andrew Hickl, John Lehmann, John Williams, and Sanda Harabagiu. 2004. Experiments with Interactive Question-Answering in Complex Scenarios. In *Proceedings of the Workshop on the Pragmatics of Question Answering at HLT-NAACL 2004*, Boston, MA.

Andrew Hickl, Patrick Wang, John Lehmann, and Sanda Harabagiu. 2006a. FERRET: Interactive Question-Answering for Real-World Environments. In *Proceedings of the Joint International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics (COLING-ACL 2006) Interactive Presentations Program*, Sydney, Australia.

Andrew Hickl, John Williams, Jeremy Bensley, Kirk Roberts, Bryan Rink, and Ying Shi. 2006b. Recognizing Textual Entailment with LCC's Groundhog System. In *Proceedings of the Second PASCAL Challenges Workshop*, Sydney, Australia.

Finley Lacatusu, Andrew Hickl, Kirk Roberts, Ying Shi, Jeremy Bensley, Bryan Rink, Patrick Wang, and Lara Taylor. 2006. LCC's GISTexter at DUC 2006: Multi-Strategy Multi-Document Summarization. In *Proceedings of the 2006 Document Understanding Conference (DUC 2006)*, New York, New York.

David Lewis. 1988. Relevant Implication. *Theoria*, 54(3):161–174.

Chin-Yew Lin and Eduard Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *Proceedings of the 18th COLING Conference*, Saarbrücken, Germany.

Bill MacCartney, Trond Grenager, Marie-Catherine de Marneffe, Daniel Cer, and Christopher D. Manning. 2006. Learning to recognize features of valid textual entailments. In *Proceedings of the Joint Human Language Technology Conference and Annual Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2006)*, New York, New York.

Dan Moldovan, Christine Clark, Sanda Harabagiu, and Steve Maiorano. 2003. COGEX: A Logic Prover for Question Answering. In *Proceedings of HLT/NAACL-2003*.

Srini Narayanan and Sanda Harabagiu. 2004. Question Answering based on Semantic Structures. In *Proceedings of COLING-2004*.