Generation of natural responses through syntactic patterns

Glenda B. Anaya and Leila Kosseim
CLaC Laboratory - Concordia University
1400 de Maisonneuve West, Montréal, Canada H3G 1M8
{anayaca,kosseim}@cs.concordia.ca

Mots-clefs – Keywords

Système de question-réponse, Génération de texte, Patrons syntaxiques, Paraphrases
Question-answering, Natural Language Generation, Syntactic Patterns, Paraphrases

Résumé - Abstract

Le but des systèmes de question-réponse est de trouver des réponses exactes et factuelles à des questions exprimées en langue naturelle en recherchant dans une grande collection de documents. Notre recherche vise plutôt à générer des réponses complètes, sous forme de phrases, étant donnée la réponse exacte. La génération de telles phrases-réponses est une tâche importante car ces phrases peuvent être employées par un système de question-réponse pour améliorer la recherche de réponses exactes ou bien, pour améliorer l'interface entre le système et l'utilisateur en fournissant des réponses plus naturelles. Suite à une étude de corpus de phrases-réponses, nous avons développé un ensemble de patrons syntaxiques de réponses correspondant à chaque patron syntaxique de question.

The goal of Question-Answering (QA) systems is to find short and factual answers to open-domain questions by searching a large collection of documents. The subject of this research is to formulate *complete* and *natural* answer-sentences to questions, given the short answer. The answer-sentences are meant to be self-sufficient; that is, they should contain enough context to be understood without needing the original question. Generating such sentences is important in question-answering as they can be used to enhance existing QA systems to provide answers to the user in a more *natural way* and to provide a pattern to actually extract the answer from the document collection.

1 Introduction

The goal of Question-Answering (QA) systems is to find short and factual answers to open-domain questions by searching a large collection of documents. The subject of our research is to formulate *complete* and *natural* answer-sentences to questions, given the short answer. The answer-sentences are meant to be self-sufficient; that is, they should contain enough context to be understood without needing the original question. Figure 1 shows an example of a question,

Q:	*What two US biochemists won the Nobel Prize in medicine in 1992? (Edwin Krebs Edmond Fischer)*
A1:	*Edwin Krebs and Edmond Fischer are the two US biochemists who won the Nobel Prize in medicine in 1992.*
A2:	*The two US biochemists who won the Nobel Prize in medicine in 1992 are Edwin Krebs and Edmond Fischer.*
A3:	*In 1992, Edwin Krebs and Edmond Fischer were the two US biochemists who have won the Nobel Prize in medicine.*

Figure 1: Example of a question and its associated answer-sentences

its exact answer and three possible answer-sentences that we wish to generate. Generating such sentences is important in question-answering as they can be used to enhance existing QA systems to provide answers to the user in a more natural way and to provide a pattern to actually extract the answer from the document collection.

2   Previous Work in Answer Formulation

To date, most work in QA has been involved in *answer extraction*; that is, locating the answer in a document collection. In contrast, the problem of *answer formulation* has not received much attention. Answer formulation performed to improve interaction with the user aims for high precision by producing only a few formulations of good linguistic quality. However, to our knowledge, little research has yet addressed this issue.

To improve answer extraction, for example, when looking for the answer to *What country is the biggest producer of tungsten?* and knowing that the answer could have the form *The biggest producer of tungsten is* <LOCATION> or <LOCATION> *is the biggest producer of tungsten*, the QA system can search for these formulations in the document collection and instantiate <LOCATION> with the matching noun phrase. This technique is now used in several QA systems such as (de Chalendar *et al.*, 2002; Soubbotin & Soubbotin, 2001; Plamondon *et al.*, 2002; Duclaye *et al.*, 2002). In the work of (Brill *et al.*, 2001; Brill *et al.*, 2002), the system searches the web for a list of possible answer formulations generated by permuting the words of the questions and in (Agichtein & Gravano, 2000; Lawrence & Giles, 1998), answer formulations are produced specifically to improve the retrieval of documents, not the retrieval of exact answers.

Most work on question reformulation has worked at the word level; whether using simple word permutations or including lexical variations, such as synonyms. However, reformulations that take into account syntactic information, has, to our knowledge not been investigated. This is probably due to the fact that most work has not been performed for the generation of self-sufficient answer-sentences.

3   Corpus Analysis

To capture and analyze the variety of answer formulations for various types of questions (*What*, *Which*, *When*, *Where*, *Who*, *Name*, *How* and *Why*) and their corresponding answer-sentence; we designed a survey composed of 150 questions and exact answers, taken from the TREC-8,

Generation of natural responses in question-answering

For how long is an elephant pregnant? (22 months)
67%   An elephant is pregnant for 22 months.
20%   An elephant's pregnancy lasts for 22 months.
10%   The pregnancy of an elephant lasts 22 months.
5%   An elephant has a 22 months pregnancy.

Figure 2: Excerpt of the compilation of answers

Type	Question Pattern	%	Answer Pattern
Who	WP VBD NP PP	61%	AA VBD NP PP
	Who released the Internet worm in the late 1980s?		*AA released the Internet worm in the late 1980s.*
		14%	PP AA VBD NP
			In the late 1980s AA released the Internet worm.
		10%	NP +tobeAux VBD PP +by AA.
			The Internet worm was released in the late 1980s by AA.
What	WP [VBD|VBZ] NP	78%	NP [VBD|VBZ] AA
	What is an atom?		*An atom is AA.*
		22%	AA [VBD|VBZ] NP
			AA is an atom.

Table 1: Examples of question and answer patterns

TREC-9 and TREC-10 competitions (Voorhees & Harman, 1999; Voorhees & Harman, 2000; Voorhees & Harman, 2001) and asked 40 people to formulate a complete and natural sentence-long answer that would best answer each of these questions. The answers obtained from the survey were compiled and classified according to their syntactic form. Figure 2 shows an example of this compilation. We found a great variety of answer formulations for questions of type *What*, *Name* and *Why* and these tended to have a more complex structure including one or more prepositional phrases. The other types of questions had fewer variations; the answers were more stereotypical.

Once we compiled the answers for each question, both the questions and the answers were tagged using the Brill part-of-speech Tagger (Brill, 1995) and two main syntactic constituents: noun phrases (NP) using NPExtractor (Bergler & Knoll, 1996) and prepositional phrases (PP). In addition answer patterns include an extra tag (denoted AA) to indicate the position of the exact answer, and also include supplemental grammatical words and verb tense information needed to create well-formed sentences. For example, to answer the question *What year did WWII begin?*, a possible answer would be *WWII began (past tense) +in AA*. The result of the analysis is a set of answer patterns for each possible question pattern. Table 1 shows results of this process for questions of type *Who* and *What*.

Due to the variety of questions, the number of question patterns found is considerable (for 150 questions, 92 patterns were found), which makes us consider that the evaluation of fewer types of questions would have been preferable to obtain a more representative number of patterns for each type of question. Table 2 shows the distribution of patterns for each type of questions. Questions of type *Name*, *Which*, *Why* and *How* were rather few and have a great difference in their formulations.

Nb. of questions	Type of question	Nb. Question Patterns
74	What	43
22	How	20
28	Who	10
9	When	5
3	Which	3
2	Why	2
6	Where	3
6	Name	6
150		92

Table 2: Distribution of the question patterns in the training corpus

Type of question	Nb of questions	% Coverage (recall)	% Correct Formulation (precision)
What	64	48%	77%
How	16	37%	100%
Who	15	60%	66%
When	8	88%	100%
Which	7	0%	-
Why	3	0%	-
Where	7	49%	100%
Name	0	-	-
	120	47%	88%

Table 3: Results of the evaluation with the test corpus

4 Implementation and Evaluation

To test our model, we implemented the patterns in a system called AnsForm. AnsForm proceeds to automatically extract the tags and the words of a given tagged question from the test corpus, creates a pattern and match it with the question patterns identified from our training set. If a counterpart is found, the system checks the answer patterns associated with the matched question and automatically generates an answer-sentence as the output.

4.1 Evaluation

To evaluate the performance of our approach, we ran the system with a test corpus made up of 120 questions (and their exact answers) taken from the TREC-8, TREC-9 and TREC-10 question corpus. Considering recall as a measure of coverage (quantity); how many questions from the test set did match question patterns identified from our training set, and precision as a measure of accuracy (quality); how many answer formulation were grammatically correct, we found that the system has only 47% of recall and 88% of precision. This is shown in Table 3. The low coverage of our system is due to the large variety of types of question formulations (8) that we analyzed and the inherent difficulties in gathering a representative training corpus. Because our training set is about the same size as our test set, we do not have a more representative number of patterns for each type of question.

Questions of type *How*, where the system performed poorly (37% of recall), is an example of the direct relation between the variety of type of question formulation, question patterns and the

coverage of a question. Although we have a large number of patterns for this type of question (see Table 2), the variation of this type into several categories such as *How long*, *How many*, *How much*, *How tall* ... makes the 20 patterns too few to cover different formulations. It is interesting to note, however, that questions of type *Where* were always well-reformulated even if we have only three (3) patterns for this type. We believe that this is because *Where* questions have a more stereotypical structure.

4.2 Error Analysis

As shown in table 3, on average, 88% of the answer-sentences that were generated were grammatically correct. Some of the errors can be blamed on the tagging process: Brill's tagger has not been adapted to tag questions, and our own noun phrase extractor does not tag correctly certain types of noun phrases like conjunctive NPs. However, other errors were due directly to our approach. We assumed that if two questions share the same syntactic pattern, then they should also share identical syntactic patterns for their answer-sentences. In general, this assumption holds. For example, the questions:*What/WP persons head/NP is/VBZ on a dime/PP?/.* and *What/WP soviet seaport/NP is/VBZ on the Black Sea/PP ?/* share the same question pattern (WP NP VBZ PP) and can both be answered by the following patterns:

Pattern	Example
AA VBZ PP	*AA is on a dime.*
	AA is on the Black Sea.
+the NP +that VBZ PP +is AA	*The person's head that is on a dime is AA.*
	The soviet seaport that is on the Black Sea is AA.
AA VBZ +the NP PP	*AA is the person's head on a dime.*
	AA is the soviet seaport on the Black Sea.
+the NP PP VBZ AA	*The person's head on a dime is AA.*
	The soviet seaport on the Black Sea is AA.

However, in some cases, different questions may share the same grammatical pattern, but cannot use the same answer pattern. For example: *Who/WP was/VBD Galileo/NP ?/* and *Who/WP discovered/VBD radium/NP ?/* both share the pattern (WP VBD NP); but while the second question can be answered by NP +was VBD +by AA: *Radium +was discovered +by AA*, the first cannot use the same answer pattern as:*Galileo +was was +by AA* is not grammatically correct. A solution to this problem would be to assign different tags to different types of verbs in order to produce distinct patterns for each case.

5 Conclusion and Future Work

The answer formulation approach presented in this paper shows that patterns that include syntactic information can effectively produce answer-sentences for most of the matched questions. Our evaluation shows that with the small training corpus that we used (150 questions) only about a half of the test corpus could be covered. We believe that with a larger training corpus, the recall rate could be increased significantly. However, our goal is not to have high coverage, but have high grammatical precision. We are more interested in generating answer-sentences for human interface reasons, than for extraction purposes. Our work only considered self-suffcient answer-sentences, but if the question is also given to the user, we should consider the treatment of language phenomena such as ellipsis and anaphora. Also to improve the AnsForm system itself, it would be useful to train the part-of-speech tagger and the noun phrase extractor on

questions, so as to reduce tagging errors and to concentrate on a smaller set of question types. To improve the approach in general we believe that using only part-of-speech tags and a few syntactic tags are not sufficient; semantic information must be taken into account. For example, we should distinguish verb classes and identify semantic types of prepositional phrases (ex. temporal, locative, . . .) which cannot be placed in the same syntactic positions to produce answer-sentences.

6   Acknowledgments

References

AGICHTEIN E. & GRAVANO L. (2000). Snowball: Extracting Relations from Large Plain-Text Collections. In *Proceedings of the 5th ACM International Conference on Digital Libraries*.

BERGLER S. & KNOLL S. (1996). Coreference Patterns in the Wall Street Journal. In C. PERCY, C. MEYER & I. LANCASHIRE, Eds., *Synchronic corpus linguistics. Papers from the Sixteenth International Conference on English Language Research on Computerized Corpora*, p. 85–96.

BRILL E. (1995). Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging. *Computational Linguistics*, **21**(4), 543–565.

BRILL E., DUMAIS S. & BANKO M. (2002). An Analysis of the AskMSR Question-Answering System. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP-2002)*, Philadelphia.

BRILL E., LIN J., BANKO M., DUMAIS S. & NG A. (2001). Data-Intensive Question Answering. In *Proceedings of The Tenth Text Retrieval Conference (TREC-X)*, p. 393–400, Gaithersburg, Maryland.

DE CHALENDAR G., DALMAS T., F. E.-G., FERRET, O. GRAU B., HURAULT-PLANTET M., ILLOUZ G., MONCEAUX L., ROBBA I. & VILNAT A. (2002). The Question Answering System QALC at LIMSI: Experiments in Using Web and WordNet. In *Notebook Proceedings of TREC-11*, p. 457–465, Gaithersburg, Maryland.

DUCLAYE F., YVON F. & COLLIN O. (2002). Using the Web as a Linguistic Resource for Learning Reformulations Automatically. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, p. 390–396, Las Palmas, Spain.

LAWRENCE S. & GILES C. L. (1998). Context and Page Analysis for Improved Web Search. *IEEE Internet Computing*, **2**(4), 38–46.

PLAMONDON L., LAPALME G. & KOSSEIM L. (2002). The QUANTUM Question Answering System at TREC-11. In *Notebook Proceedings of TREC-11*, p. 157–165, Gaithersburg, Maryland.

SOUBBOTIN M. & SOUBBOTIN S. (2001). Patterns of Potential Answer Expressions as Clues to the Right Answers. In *Proceedings of TREC-10*, p. 175–182, Gaithersburg, Maryland.

E. M. VOORHEES & D. K. HARMAN, Eds. (1999). *Proceedings of The Eight Text REtrieval Conference (TREC-8)*, Gaithersburg, Maryland. NIST. available at http://trec.nist.gov/pubs/trec8/t8_proceedings.html.

E. M. VOORHEES & D. K. HARMAN, Eds. (2000). *Proceedings of The Ninth Text REtrieval Conference (TREC-9)*, Gaithersburg, Maryland. NIST. available at http://trec.nist.gov/pubs/trec9/t9_proceedings.html.

E. M. VOORHEES & D. K. HARMAN, Eds. (2001). *Proceedings of The Tenth Text REtrieval Conference (TREC-X)*, Gaithersburg, Maryland. NIST. available at http://trec.nist.gov/pubs/trec10/t10_proceedings.html.