A System for Answering Non-Factoid Japanese Questions by Using Passage Retrieval Weighted Based on Type of Answer

Masaki Murata, Sachiyo Tsukawaki, Toshiyuki Kanamaru, Qing Ma, and Hitoshi Isahara
National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan
{murata, mutiyama, isahara}@nict.go.jp

Abstract

We constructed a system for answering non-factoid Japanese questions. We used passage retrieval methods for the system. We extracted paragraphs based on terms from an input question and output them as the desired answer. We classified the non-factoid questions into six categories. We used a particular method for each category. For example, we increased the scores of paragraphs including the word "reason" for questions including the word "why." We performed experiments using the NTCIR-6 QAC-4 data collection and tested the effectiveness of our methods.

1 Introduction

A question-answering system is an application designed to produce the correct answer to a question given as input. For example, when "What is the capital of Japan?" is given as input, a question-answering system may retrieve text containing sentences like "Tokyo is Japan's capital and the country's largest and most important city. Tokyo is also one of Japan's 47 prefectures." from websites, newspaper articles, or encyclopedias. The system then outputs "Tokyo" as the correct answer. We believe question-answering systems will become a more convenient alternative to other systems designed for information retrieval and a basic component of future artificial intelligence systems. Numerous researchers have recently been attracted to this important topic. These researchers have produced many interesting studies on question-answering systems [8, 7, 4, 5, 10, 12]. Evaluation conferences and contests on question-answering systems have also been held. In particular, the U.S.A. has held the Text REtrieval Conferences (TREC) [30], and Japan has hosted the Question-Answering Challenges (QAC) [24]. These conferences and contests aim to improve question-answering systems. Researchers who participate in them make question-answering systems that they use to answer the same questions, and each system's performance is then examined to glean possible improvements. We have investigated the potential of question-answering systems [17, 18, 21, 22] and studied their construction by participating in the QAC 1, 2, and 3 [24, 25, 26] at NTCIR (NII Test Collection for IR Systems) 3, 4, and 5 [19, 20, 16].

In NTCIR-6 QAC-4, we addressed non-factoid question answering. For example, when the question is "Why are the people opposed to the Private Information Protection Law?" the system retrieves sentences based on terms appearing in the question and outputs the answer using the retrieved sentences. Numerous studies have addressed the issues that surround answering non-factoid questions [2, 3, 31, 29, 6, 13, 9, 1].

For QAC-4, we constructed a system for answering non-factoid Japanese questions. We used passage retrieval methods for the system. We extracted paragraphs based on terms from an input question and output them as the desired answer. We classified the non-factoid questions into six categories. We used a particular method for each category. For example, we increased the scores of paragraphs including the word "reason" for questions including the word "why." We performed experiments using the NTCIR-6 QAC-4 data collection and tested the effectiveness of our methods.

2 Categories of Non-Factoid Questions

In this study, we used the following six categories of non-factoid questions.

1. Definition-oriented questions (Questions that require a definition be given in response.)

   e.g., *K-1 to wa nandesuka?* (What is K-1?)

2. Reason-oriented questions (Questions that require a reason be given in response.)

   e.g., *kojin jouhou hokogou ni hantai shiteiru hito wa doushite hantai shiteiru no desuka?* (Why are the people opposed to the Private Information Protection Law?)

3. Method-oriented questions (Questions that require an explanation of a method be given in response.)

   e.g., *sekai isan wa donoyouni shite kimeru no desuka?*" (How is a World Heritage determined?)

4. Degree-oriented questions (Questions that require an explanation of the degree of something be given in response.)

5. Change-oriented questions (Questions that require a description of things that change be given in response.)

   e.g., *shounen hou wa dou kawari mashitaka?* (How was the juvenile law changed?)

6. Detail-oriented questions (Questions that require a description of the particulars or details surrounding a sequence of events be given in response.)

   e.g., *donoyouna keii de ryuukyuu oukoku wa nihon no ichibu ni natta no desuka?* (How did Ryukyu come to belong to Japan?)

3 Question-answering Systems in this Study

The system has three basic components:

1. Prediction of type of answer The system predicts the answer to be a particular type of expression based on whether the input question is indicated by an interrogative pronoun, an adjective, or an adverb. For example, if the input question is "Why are the people opposed to the Private Information Protection Law?", the word "why" suggests that the answer will be an expression that describes a reason.

2. Document retrieval The system extracts terms from the input question and retrieves documents by using these terms. Documents that are likely to contain the correct answer are thus gathered during the retrieval process. For example, for the input question "Why are the people opposed to the Private Information Protection Law?", the system extracts "people," "opposed," "Private," "Information," "Protection," and "Law" as terms and retrieves the appropriate documents based on them.

3. Answer detection The system separates the retrieved documents into paragraphs and retrieves those that contain terms from the input question and a clue expression (e.g., "to wa" (copula sentence) for the definition sentence). The system outputs the retrieved paragraphs as the desired answer.

3.1 Prediction of type of answer

We used the following rules for predicting the type of answer.

1. Definition-oriented questions Questions including expressions such as "*to wa nani*," "*donna*," "*douiu*," "*douitta*," "*nanimono*," "*donoyouna mono*," "*donna mono*," and "*douiu koto*" (which all mean "what is") are recognized by the system as being definition-oriented questions.

2. Reason-oriented questions Questions including expressions such as "*naze*" (why), "*naniyue*" (why), "*doushite*" (why), "*nani ga riyuu de*" (what is the reason), and "*donna riyuu de*" (what reason), are recognized by the system as being reason-oriented questions.

3. Method-oriented questions Questions including expressions such as "*dou*," "*dousureba*," "*douyatte*," "*dono youni shite*," "*ikani shite*," "*ikani*," and "*donnna houhou de*" (which all mean "how") are recognized by the system as being method-oriented questions.

4. Degree-oriented questions Questions including expressions such as "*dorekurai*" (how much), "*dorekurai no*" (to what extent), and "*dono teido*" (to what extent), are recognized by the system as being degree-oriented questions.

5. Change-oriented questions Questions including expressions such as "*naniga chigau*" (What is different), "*donoyuni kawaru*" (How is ... changed), and "*dokoga kotonaru*" (What is different), are recognized by the system as being change-oriented questions.

6. Detail-oriented questionsQuestions including expressions such as "*dono you na keii*," "*dono you na ikisatsu*," and "*dono you na nariyuki*" (which all mean "how was") are recognized by the system as being detail-oriented questions.

3.2 Document retrieval

Our system extracts terms from a question by using the morphological analyzer, ChaSen [11]. The analyzer first eliminates prepositions, articles, and similar parts of speech. It then retrieves documents by using the extracted terms.

The documents are retrieved as follows:

We first retrieve the top k_{dr1} documents with the highest scores calculated from the equation

$$Score(d)$$

$$= \sum_{\text{term } t} \left(\frac{tf(d,t)}{tf(d,t) + k_t \frac{length(d) + k_+}{\Delta + k_+}} \times log \frac{N}{df(t)} \right), \quad (1)$$

where d is a document, t is a term extracted from a question, $tf(d,t)$ is the frequency of t occurring in d, $df(t)$ is the number of documents in which t appears, N is the total number of documents, $length(d)$ is the length of d, and Δ is the average length of all documents. Constants k_t and k_+ are defined based on experimental results. We based this equation on Robertson's equation [27, 28]. This approach is very effective, and we have used it extensively for information retrieval [15, 23, 14]. The question-answering system uses a large number for k_t.

We extracted the top 300 documents and used them in the next procedure.

3.3 Answer detection

In detecting answers, our system first generates candidate expressions for them from the extracted documents. We use two methods for extracting candidate expressions. Method 1 uses a paragraph as a candidate expression. Method 2 uses a paragraph, two continuous paragraphs, or three continuous paragraphs as candidate expressions.

We give each candidate expression the following score.

$$\begin{aligned} & Score(d) \\ = \; & -min_{t1 \in T} log \prod_{t2 \in T3} (2dist(t1,t2) \frac{df(t2)}{N}) \\ + \; & 0.00000001 \times length(d) \\ = \; & max_{t1 \in T} \sum_{t2 \in T3} log \frac{N}{2dist(t1,t2) * df(t2)} \\ + \; & 0.00000001 \times length(d) \end{aligned} \quad (2)$$

$$T3 = \{t | t \in T, 2dist(t1,t) \frac{df(t)}{N} \leq 1\}, \quad (3)$$

where d is a candidate expression, T is the set of terms in the question, $dist(t1,t2)$ is the distance between $t1$ and $t2$ (defined as the number of characters between them with $dist(t1,t2)$ = 0.5 when $t1 = t2$), and $length(d)$ is the number of characters in a candidate expression. The numerical term, $0.00000001 \times length(d)$, is used for increasing the scores of long paragraphs.

For reason-oriented questions, our system uses some reason terms such as "riyuu" (reason), "gen'in" (cause), and "nazenara" (because) as terms for Equation 2 in addition to terms from the input question. This is because we would like to increase the score of a document that includes reason terms for reason-oriented questions.

For method-oriented questions, our system uses some method terms such as "houhou" (method), "tejun" (procedure), and "kotoniyori" (by doing) as terms for a second document retrieval (re-ranking) in addition to terms from the input question.

For detail-oriented questions, our system uses some method terms such as "keii" (a detail, or a sequence of events), "haikei" (background), and "rekishi" (history) as terms for a second document retrieval (re-ranking) in addition to terms from the input question.

For degree-oriented questions, when candidate paragraphs include numerical expressions, the score ($Score(d)$) is multiplied by 1.1.

For definition-oriented questions, the system first extracts focus expressions. When the question includes expressions such as "*X-wa*", "*X-towa*", "*X-toiunowa*", and "*X-tte*", X is extracted as a focus expression. The system multiplies the score ($Score(d)$) of the candidate expression by 1.1. When the candidate expression includes focus expressions having modifiers (including modifier clauses and modifier phrases), the modifiers are used as candidate expressions, and the scores of the candidate expressions are multiplied by 1.1.

We show an example of a candidate expression which is a modifier clause in a sentence as follows.

> Question sentence:
>
> *sekai isan jouyaku to wa dono youna jouyaku desu ka?*
>
> (What is Convention concerning the Protection of the World Cultural and Natural Heritage?)
>
> Sentence including answers:
>
> *1972 nen no dai 17 kai yunesuko soukai de saitaku sareta sekai isan jouyaku ....*
>
> (Convention concerning the Protection of the World Cultural and Natural Heritage, which is adopted in 1972 in the 17th general assembly meeting of UN Educational, Scientific and Cultural Organization.)

Finally, our system extracts candidate expressions having high scores ($Score(d)$s) as the desired output. Our system extracts candidate expressions having scores that are no less than the highest score multiplied by 0.9 as the desired output.

4 Experiments

The experimental results are listed in Table 1. One hundred non-factoid questions were used in the experiment. The questions, which were generated by the

Table 1. Results

Method	A		B		C		D	
Method 1	0.072	(25/345)	0.188	(65/345)	0.041	(14/345)	0.699	(241/345)
Method 2	0.016	(6/366)	0.328	(120/366)	0.066	(24/366)	0.590	(216/366)

QAC-4 organizers, are natural and were not generated by using target documents. The QAC-4 organizers checked four or fewer outputs for each question. Methods 1 and 2 determine what we use as answer candidate expressions (Method 1 uses one paragraph as a candidate answer. Method 2 uses one paragraph, two paragraphs, or three paragraphs as candidate answers.).

"A," "B," "C," and "D" are the evaluation criteria. "A" indicates the output that describes the same content as that of the answer. Even if there is a supplementary expression in the output, which does not change the content, the output is judged to be "A." "B" indicates the output that contains some content similar to that of the answer but contains different overall content. "C" indicates the output contains part of the same content as that of the answer. "D" indicates the output does not contain any of the same content as that of the answer.

We made the following findings.

- Method 1 obtained higher scores in Evaluation A than Method 2. This indicates that Method 1 can extract an completely relevant answer more accurately than Method 2.
- Method 2 obtained higher scores in Evaluations A, B, and C than Method 1. This indicates that Method 2 can extract more partly relevant answers than Method 1. When we would like to extract an completely relevant answer, we should use Method 1. When we would like to extract more answers, including partly relevant answers, we should use Method 2.

We show an example of the output of Method 1, which is judged to be "A."

Question sentence:

jusei ran shindan wa douiu baai ni okonawareru noka?

(When is fertilized ovum diagnosis performed?)

System output:

omoi idenbyou no kodono ga umareru no wo fusegu.

(Preventing the birth of a child with a serious genetic disorder)

Examples of answers given by organizers:

omoi idenbyou

(a serious genetic disorder)

omoi idenbyou no kodomo ga umareru kanousei ga takai baai

(In cases when the plausibility of the birth of a child with a serious genetic disorder is high.)

We show an example of the output of Method 2, which is judged to be "A."

Question sentence:

houshasen no hibaku ga jintai ni ataeru eikyou wa donoyouna mono desuka.

(What effect does radiogenic exposure have on the human body?)

System output:

housha sen wa jintai no saibou ya idenshi ni chokusetsu sayou shi, kinou wo sokonattari hakai suru. kono eikyou wa ketsueki wo tsukuru kotsuzui saibou, seishoku saibou nado saibou bunretsu no sakana bubun hodo ukeyasuku, shoujou wa hibaku ryou ni hirei shite omokunaru. ippanjin no 1 nen kan no hibaku kyoyou senryou wa hourei de 1 miri shi-beru to sareteiru. ippoude, iryou koui deno hibaku ni seigen wa naku, i no ekkusu sen shuudan kenshin de wa 1 kai de abiru housha sen ga saidai de 4 miri shi-beruto ni naru. tadashi, kono teido no ryou dewa kenkou he no eikyou wa nai to iu. housha sen eikyou wa kojin sa ga aru ga, ippanni 250 miri shi-beruto wo koeru housha sen wo ittokini abiruto, kentaikan nado no shoujou ga araware, 1 shi-beruto (1000 miri shi-beruto) de hakike ya geri nado no kyuusei houshasen shoujou wo okosu.

(Radiation directly affects the cells and genes of the human body by causing their functions to deteriorate. It can also kill the body's cells and genes. Bone marrow cells, which make blood, and reproductive cells, which make cells divide rapidly,

are easily affected by radiation. The damage is proportional to the amount of exposure. The permissible radiogenic exposure for an ordinary person by law in Japan is 1 mSv per year. The same law does not regulate exposure during medical treatment. The exposure to radiation during a gastric group X-ray examination is at most 4 mSv. It is said this amount does not affect a person's health. Typically, exposure to more that 250 mSv of radiation at one time causes fatigue. Exposure to 1 Sv (1000 mSv) causes acute radiation syndrome, the symptoms of which are ailments such as nausea and diarrhea. It should be remembered that exposure to radiation affects individuals differently.)

Examples of answers given by organizers:

datsuryoku kan wo okosu

(developing weakness)

geri nado wo okosu

(having diarrhea)

outo

(vomiting)

5 Conclusion

We constructed a system for answering non-factoid Japanese questions. An example of a non-factoid question is "Why are the people opposed to the Private Information Protection Law?" We used passage retrieval methods for the system. We extracted paragraphs based on terms from an input question and output them as the desired answer. We classified the non-factoid questions into six categories. We used a particular method for each category. For example, we increased the scores of paragraphs including the word "reason" for questions including the word "why." We performed experiments using the NTCIR-6 QAC-4 data collection and tested the effectiveness of our methods.

Acknowledgements

We are grateful to all the organizers of NTCIR 6 who gave us a chance to participate in the NTCIR 6 contest to examine and improve our question-answering system. We greatly appreciate the kindness of all those who helped us.

References

[1] Y. Asada. Processing of definition type questions in a question answering system. *the master dissertation in Yokohama National University*, 2006. (in Japanese).

[2] A. Berger, R. Caruana, D. Cohn, D. Freitag, and V. Mittal. Bridging the lexical chasm: Statistical approaches to answer-finding. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR-2000)*, pages 192–199, 2000.

[3] S. Blair-Goldensohn, K. R. McKeown, and A. H. Schlaikjer. A hybrid approach for qa track definitional questions. In *Proceedings of the 12th Text Retrieval Conference (TREC-2003)*, pages 185–192, 2003.

[4] C. L. A. Clarke, G. V. Cormack, and T. R. Lynam. Exploiting redundancy in question answering. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2001.

[5] S. Dumis, M. Banko, E. Brill, J. Lin, and A. Ng. Web question answering: Is more always better? In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2002.

[6] K.-S. Han, Y.-I. Song, S.-B. Kim, and H.-C. Rim. Phrase-based definitional question answering using definition terminology. In *Lecture Note in Computer Science 3689*, pages 246–259, 2005.

[7] A. Ittycheriah, M. Franz, W.-J. Zhu, and A. Ratnaparkhi. IBM's Statistical Question Answering System. In *TREC-9 Proceedings*, 2001.

[8] J. Kupiec. MURAX: A robust linguistic approach for question answering using an on-line encyclopedia. In *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1993.

[9] H. Maehara, J. Fukumoto, and N. Kando. A be-based automated evaluation for question-answering system. *IEICE-WGNLC2005-109*, pages 19–24, 2006. (in Japnese).

[10] B. Magnini, M. Negri, R. Prevete, and H. Tanev. Is it the right answer? Exploiting web redundancy for answer validation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 2002.

[11] Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, and M. Asahara. Japanese morphological analysis system ChaSen version 2.0 manual 2nd edition. 1999.

[12] D. Moldovan, M. Pasca, S. Harabagiu, and M. Surdeanu. Performance issues and error analysis in an open-domain question answering system. *ACM Transactions on Information Systems*, 21(2):133–154, 2003.

[13] K. Morooka and J. Fukumoto. Answer extraction method for why-type question answering system. *IEICE-WGNLC2005-107*, pages 7–12, 2006. (in Japnese).

[14] M. Murata, Q. Ma, and H. Isahara. High performance information retrieval using many characteristics and many techniques. *Proceedings of the Third NTCIR Workshop (CLIR)*, 2002.

[15] M. Murata, K. Uchimoto, H. Ozaku, Q. Ma, M. Utiyama, and H. Isahara. Japanese probabilistic information retrieval using location and category information. *The Fifth International Workshop on Information Retrieval with Asian Languages*, pages 81–88, 2000.

[16] M. Murata, M. Utiyama, and H. Isahara. Japanese question-answering system using decreased adding with multiple answers.

[17] M. Murata, M. Utiyama, and H. Isahara. Question answering system using syntactic information. 1999. http://xxx.lanl.gov/abs/cs.CL/9911006.

[18] M. Murata, M. Utiyama, and H. Isahara. Question answering system using similarity-guided reasoning. *Information Processing Society of Japan, WGNL 2000-NL-135*, pages 181–188, 2000. (in Japanese).

[19] M. Murata, M. Utiyama, and H. Isahara. A question-answering system using unit estimation and probabilistic near-terms IR. *Proceedings of the Third NTCIR Workshop (QAC)*, 2002.

[20] M. Murata, M. Utiyama, and H. Isahara. Japanese question-answering system using decreased adding with multiple answers. *Proceedings of the NTCIR Workshop 4 (QAC)*, 2004.

[21] M. Murata, M. Utiyama, and H. Isahara. Use of multiple documents as evidence with decreased adding in a Japanese question-answering system. *Journal of Natural Language Processing*, 12(2), 2005.

[22] M. Murata, M. Utiyama, and H. Isahara. Japanese question-answering system for contextual questions using simple connection method, decreased adding with multiple answers, and selection by ratio. *Asia Information Retrieval Symposium (AIRS)*, pages 601–607, 2006.

[23] M. Murata, M. Utiyama, Q. Ma, H. Ozaku, and H. Isahara. CRL at NTCIR2. *Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization*, pages 5–21–5–31, 2001.

[24] National Institute of Informatics. *Proceedings of the Third NTCIR Workshop (QAC)*. 2002.

[25] National Institute of Informatics. *Proceedings of the Fourth NTCIR Workshop (QAC)*. 2004.

[26] National Institute of Informatics. *Proceedings of the Fifth NTCIR Workshop (QAC)*. 2005.

[27] S. E. Robertson and S. Walker. Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. In *Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1994.

[28] S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In *TREC-3*, 1994.

[29] R. Soricut and E. Brill. Automatic question answering: Beyond the factoid. In *In Proceedings of the Human Language Technology and Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL-2004)*, pages 57–64, 2004.

[30] TREC-10 committee. The tenth text retrieval conference. 2001. http://trec.nist.gov/pubs/trec10/t10_proceedings.html.

[31] J. Xu, A. Licuanan, and R. Weischedel. Trec 2003 qa at bbn: Answering definitional questions. In *Proceedings of the 12th Text Retrieval Conference (TREC-2003)*, pages 98–106, 2003.