

# Factoid Question Answering with Web, Mobile and Speech Interfaces

E.W.D. Whittaker      J. Mrozinski      S. Furui

Dept. of Computer Science  
Tokyo Institute of Technology  
2-12-1, Ookayama, Meguro-ku  
Tokyo 152-8552 Japan

{edw,mrozinsk,furui}@furui.cs.titech.ac.jp

## Abstract

In this paper we describe the web and mobile-phone interfaces to our multi-language factoid question answering (QA) system together with a prototype speech interface to our English-language QA system. Using a statistical, data-driven approach to factoid question answering has allowed us to develop QA systems in five languages in a matter of months. In the web-based system, which is accessible at <http://asked.jp>, we have combined the QA system output with standard search-engine-like results by integrating it with an open-source web search engine. The prototype speech interface is based around a VoiceXML application running on the Voxeo developer platform. Recognition of the user's question is performed on a separate speech recognition server dedicated to recognizing questions. An adapted version of the Sphinx-4 recognizer is used for this purpose. Once the question has been recognized correctly it is passed to the QA system and the resulting answers read back to the user by speech synthesis. Our approach is modular and makes extensive use of open-source software. Consequently, each component can be easily and independently improved and easily extended to other languages.

## 1 Introduction

The approach to factoid question answering (QA) that we adopt was first described in (Whittaker et al., 2005b) where the details of the mathematical model and how it was trained for English were given. The approach has been successfully evaluated in the 2005 text retrieval conference (TREC) question answering track evaluations (Voorhees and Trang Dang, 2005) where our group placed eleventh out of thirty participants (Whittaker et al., 2005a). Although the TREC QA task is substantially different to web-based QA this evaluation showed that our approach works and provides an objective assessment of its quality. Similarly, for our Japanese language system we have evaluated the performance of our approach on the NTCIR-3 QAC-1 task (Whittaker et al., 2005c). Although our Japanese experiments were applied retrospectively, the results would have placed us in the mid-range of participating systems. In (Whittaker et al., 2006b) we described how our approach could be used for the rapid development of web-based QA systems in five very different languages. It was shown that a developer proficient with the tools, and with access to suitable training data, could build a system in a new language in around 10 hours. In (Whittaker et al., 2006a) we evaluated the performance of the systems for four of our five languages. We give a brief summary of our approach to QA in Section 2.

In this paper we introduce our web-based QA system which is publicly accessible at <http://asked.jp>, permitting questions in English, Japanese, Chinese, Russian and Swedish and

is discussed in Section 3. Since answers in factoid QA are inherently well-suited to display on small screens we have also made a mobile-phone interface which is accessible at the same address when using an HTML browser from a mobile phone. This is discussed in Section 4. There are several other QA systems on the web including Brainboost (Brainboost, 2005) and Lexxe (Lexxe, 2005) but they only try to answer questions in English and do not have convenient mobile interfaces.

Entering whole questions rather than just keywords is tedious especially on a mobile-phone so we have also begun to look at speech interfaces. In this paper we describe a prototype speech interface to our English-language QA system. This prototype is currently intended primarily as a platform for further research into speech recognition and answering of questions from an acoustic modelling point-of-view (e.g. low-bandwidth, low-quality VoIP channel), from a language modelling perspective (e.g. irregular word order in questions vs. text, and very large out-of-vocabulary problem) and also in terms of dialog modelling. There have been several attempts at speech interfaces to QA systems in the literature e.g. (Schofield and Zheng, 2003) but as far as we know ours is the only system that is publicly accessible. We discuss this interface in Section 5.

## 2 Statistical pattern classification approach to QA

The answer to a question depends primarily on the question itself but also on many other factors such as the person asking the question, the location of the person, what questions the person has asked before, and so on. For simplicity, we choose to consider only the dependence of an answer  $A$  on the question  $Q$ . In particular, we hypothesize that the answer  $A$  depends on two sets of features extracted from  $Q$ :  $W = \mathcal{W}(Q)$  and  $X = \mathcal{X}(Q)$  as follows:

$$P(A | Q) = P(A | W, X), \quad (1)$$

where  $W$  can be thought of as a set of  $l_W$  features describing the “question-type” part of  $Q$  such as *who*, *when*, *where*, *which*, etc. and  $X$  is a set of features comprising the “information-bearing” part of  $Q$  i.e. what the question is actually about and what

it refers to. For example, in the questions, *Where is Mount Everest?* and *How high is Mount Everest?* the information-bearing component is identical in both cases whereas the question-type component is different.

Finding the best answer  $\hat{A}$  involves a search over all  $A$  for the one which maximizes the probability of the above model:

$$\hat{A} = \arg \max_A P(A | W, X). \quad (2)$$

This is guaranteed to give us the optimal answer in a maximum likelihood sense if the probability distribution is the correct one. Making various conditional independence assumptions to simplify modelling we obtain the final optimisation criterion:

$$\arg \max_A \underbrace{P(A | X)}_{\text{retrieval model}} \cdot \underbrace{P(W | A)}_{\text{filter model}}. \quad (3)$$

The  $P(A | X)$  model is essentially a language model which models the probability of an answer sequence  $A$  given a set of information-bearing features  $X$ . It models the proximity of  $A$  to features in  $X$ . This model is referred to as the *retrieval model*.

The  $P(W | A)$  model matches an answer  $A$  with features in the question-type set  $W$ . Roughly speaking this model relates ways of asking a question with classes of valid answers. For example, it associates names of people or companies with *who*-type questions. In general, there are many valid and equiprobable  $A$  for a given  $W$  so this component can only re-rank candidate answers retrieved by the retrieval model. Consequently, we call it the *filter model*.

## 3 Web interface

The web-based interface to our QA systems has been accessible at <http://asked.jp> since December 2005 and although still primarily a research system and not widely advertised it attracts around five unique users a day. Currently we do not perform language detection for an input question so the user must first select a language-specific system before inputting a question in a language other than English.

In Figure 1 we show the results page for the question “*How high is Mount Everest?*”. As can be seen

the left-hand side of the page contains the familiar title, link and summaries of pages relevant to the query that is common to most of today's web search engines. These results are produced by an open-source web search engine which is run locally and currently contains about 100 million web-pages in its database. Down the right-hand side of the results page we present the answers that were found by our QA system. These answers are presented in order of probability as determined by Equation (3). When the mouse is rolled over an answer a Java-script pop-up box is displayed that shows more context for a given answer. This allows the user to determine more rapidly the validity of an answer and also partially compensates for inaccurate answer identification by the system. Each answer can also be clicked on whereupon the user is redirected to the page from which the answer was taken. This re-direction is effected through a redirect via our own web-server so that for a given question we can see which answers were clicked on. Eventually, it is hoped this could be used for unsupervised system adaptation.

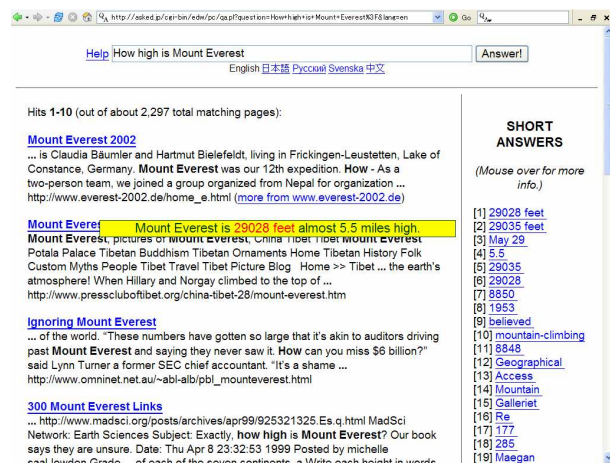


Figure 1: Results page for “How high is Mount Everest?”.

The same basic layout and design is repeated for each of the five language-specific systems. In Figure 2 we show the results page for the Japanese question of “What plant do Pandas eat?”.

The average response time to present the full results page for a question in each language is currently around 10 seconds. The web-search and QA systems are run in parallel and the outputs combined when both are complete.

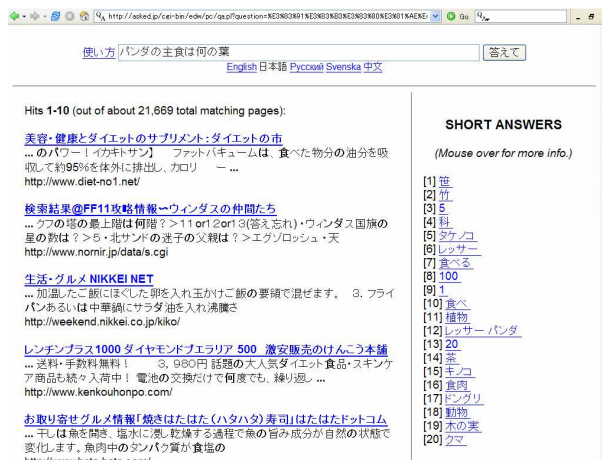


Figure 2: Results page for “What plant do Pandas eat?” in Japanese.

## 4 Mobile-phone interface

Since the priorities with a mobile-phone interface revolve around speed, display size and cost to the user, the interface is basically a whittled down version of the web-based interface described in the previous section. The only requirement for being able to use the mobile phone interface is that the phone must contain an HTML browser. In countries like Japan this has been fairly standard for many years but it is expected that this will become more common worldwide in the near future with the continued roll-out of 3G mobile phone services.

For the mobile-phone interface the standard web-search results section has been removed entirely from the results section and instead only the top 20 short answers are displayed without pop-up boxes or corresponding context. Such a strategy minimizes the number of bytes transmitted and ensures that most answers are adequately displayed on most mobile-phone interfaces with a minimum amount of scrolling. Although not yet implemented we aim to allow users to click on an answer and be taken to the part of the page that contains the answer rather than loading a whole page which could sometimes be several megabytes in size.

## 5 Speech interface

Our implementation of the speech interface to the QA system was greatly simplified by the avail-

ability of the Voxeo developer platform<sup>1</sup> which provides free access, for development purposes, to a VoiceXML browser running our application. The application can be accessed through: (i) a U.S. telephone number at (800) 289-5570 then PIN:9991423955; (ii) SIP VoIP clients at (SIP:9991423955sip.voxeo.net); (iii) Free World Dialup at (\*\*86919991423955); and (iv) SkypeOut at (+99000936 9991423955).

Since most VoiceXML applications are designed for use with small vocabulary, rule-based grammars we only use VoiceXML and Voxeo's browser to handle negotiation of the questions and answers with the user through simple dialogs. The recognition of the question itself is performed using a dedicated large-vocabulary speech recognizer with a language model (LM) trained on English-language questions. The speech recognizer we use is the open-source Sphinx-4 recognizer (Walker et al., 2004) which runs in a server mode and has been adapted to use more complex LMs than those permitted by the default ARPA format word  $N$ -gram LMs. Currently we use a linear interpolation of a word and class-based trigram LM each of which were trained on a large corpus of English-language questions (Hallmarks, 2002)—the same data used to train the English-language QA system (Whittaker et al., 2005b).

## 6 Conclusion and Further work

Having recapped a basic overview of our statistical approach to question answering (QA), in this paper we have described the web and mobile-phone interfaces to our multi-language QA system and how they can be accessed. In addition, we have described our first attempt at a prototype speech interface which will be used as a platform for future research. Eventually our aim is to make the QA performance of the speech interface the same as that obtained through the web and mobile-phone interfaces. This will be achieved through a combination of acoustic, language and dialog model adaptation on the speech side, and making the QA system more robust to underspecified and errorful questions on the QA side. We think these demonstration systems show significant progress has already been made and

give a hint of how information access to QA systems might be achieved in the near future.

## 7 Acknowledgments

This research was supported by JSPS and the Japanese government 21st century COE programme. The authors also wish to thank Dietrich Klakow for all his contributions.

## References

- Brainboost. 2005. <http://www.brainboost.com>.
- Academic Hallmarks. 2002. Knowledge Master Educational Software. PO Box 998, Durango, CO 81302 <http://www.greatauk.com/>.
- Lexxe. 2005. <http://www.lexxe.com>.
- E. Schofield and Z. Zheng. 2003. A Speech Interface for Open-domain Question-answering. In *Proceedings of the 41st Annual Meeting of the ACL*, Sapporo, Japan, July.
- E.M. Voorhees and H. Trang Dang. 2005. Overview of the TREC 2005 Question Answering Track. In *Proceedings of the 14th Text Retrieval Conference*.
- W. Walker et al. 2004. Sphinx-4: A Flexible Open Source Framework for Speech Recognition. Technical report, Sun Microsystems Inc.
- E.W.D. Whittaker, P. Chatain, S. Furui, and D. Klakow. 2005a. TREC2005 Question Answering Experiments at Tokyo Institute of Technology. In *Proceedings of the 14th Text Retrieval Conference*.
- E.W.D. Whittaker, S. Furui, and D. Klakow. 2005b. A Statistical Pattern Recognition Approach to Question Answering using Web Data. In *Proceedings of Cyberworlds*.
- E.W.D. Whittaker, J. Hamonic, and S. Furui. 2005c. A Unified Approach to Japanese and English Question Answering. In *Proceedings of NTCIR-5*.
- E.W.D. Whittaker, J. Hamonic, T. Klingberg, Y. Dong, and S. Furui. 2006a. Monolingual Web-based Factoid Question Answering in Chinese, Swedish, English and Japanese. In *Proceedings of the Workshop on Multilingual Question Answering, EACL*.
- E.W.D. Whittaker, J. Hamonic, T. Klingberg, Y. Dong, and S. Furui. 2006b. Rapid Development of Web-based Monolingual Question Answering Systems. In *Proceedings of ECIR2006*.

---

<sup>1</sup><http://www.voxeo.com/developers>