SpeechWeb: A Web of Natural-Language Speech Applications

R. A. Frost

Department of Computer Science, University of Windsor,
Windsor, Ontario, Canada N9B 3P4
richard@uwindsor.ca

Abstract

SpeechWeb consists of a collection of hyperlinked natural-language interfaces to applications which can be accessed through the Internet from speech browsers running on PCs. The applications contain hyperlinks which the browser uses to navigate SpeechWeb. The natural-language interfaces have been constructed as executable specifications of attribute grammars using a domain–specific programming language built for this purpose. The approach to natural-language processing is based on a new efficient compositional semantics that accommodates arbitrarily-nested quantification and negation. The user-independent speech browser is grammar based, and novel techniques have been developed to improve recognition accuracy.

SpeechWeb

SpeechWeb currently consists of several hyperlinked applications, including "solar man" who can answer questions about the solar system, "Judy" who knows some poems, and "Monty" who claims to be a student at the University of Windsor in Ontario. The following is an example of the spoken user-input component of a session with SpeechWeb.

```
can I talk to solar man?
what do you know?
which moons orbit earth or jupiter?
was every moon that orbits mars
                   discovered by Hall?
does something orbit no planet?
not every planet is orbited by phobos.
does every thing that orbits no planet
 and is not a person or a planet spin?
which moon that was discovered by hall
                  does not orbit mars?
can I talk to monty?

THE BROWSER IS REDIRECTED TO MONTY
hi monty
who is the president of the university
                            of windsor?
```


The Evolution of SpeechWeb

An earlier version of SpeechWeb was demonstrated at AAAI 1999 (Frost, 1999). Since then, SpeechWeb has been enhanced in the following ways: 1) The approach to natural-language processing has been extended to include arbitrarily-nested quantification and negation, thereby enhancing the expressive power of the interfaces, 2) techniques have been developed which significantly improve speech-recognition accuracy, and 3) SpeechWeb can now be accessed through wireless microphones communicating with PCs which are themselves connected through wireless to the Internet. Although this last enhancement is technically trivial, it provides true hands-free/eyes-free access to the web from remote locations.

In addition to access through a speech browser, the `solar man` application can also be accessed through an html web page, which contains a full listing of the executable specification and a paper describing the executable specification language.

```
www.cs.uwindsor.ca/users/r/
   richard/miranda/wage_demo.html
```

The Approach to Natural-Language Processing

The approach to natural-language processing which is used in SpeechWeb is based on a small first-order non-modal non-intentional subset of Montague Semantics (Montague, 1974) and (Dowty, Wall and Peters, 1981). To achieve computational tractability, Montague's characteristic functions are replaced by relations and operators of a higher-order relational algebra that includes partial application. A novel technique has been developed that allows arbitrarily-nested quantification and negation. Efficiency is achieved by representing the denotations of constructs involving negation by enumerating the members of complement sets.

This approach to negation, which is described in detail in Frost and Boulos (2002), is very simplistic compared to the comprehensive treatment of negation developed by other researchers (e.g. Iwanska, 1992), but is adequate for many questions to be answered with respect to a first-order knowledge source.

W/AGE: An Attribute-Grammar Programming Language

The natural-language interfaces are constructed as executable specifications of attribute grammars (Paaki, 1995) using the Windsor Attribute Grammar Programming Environment W/AGE (Frost, 1994). This language was built to enable the rapid construction of language processors that are based on compositional semantics.

The URL given earlier provides access to the W/AGE code for the "solar man" application. In summary, the program consists of a set of attribute type declarations, a dictionary defining the syntactic types and denotations of basic words (and the definition of other words as having the same meaning as a given phrase), a set of attribute-grammar productions in which syntax rules are augmented with semantic rules showing how the attributes of a compound phrase are computed from the attributes of the components of that phrase, and finally a set of definitions of semantic functions which provide links to the knowledge source.

Improving Speech-Recognition Accuracy

The SpeechWeb browser is constructed as a Java program linked to IBM's ViaVoice speech-recognition engine. This engine is grammar based in that it requires a "recognition grammar" to constrain the search space. The browser begins by downloading the recognition grammar from the remote location. The recognition engine is then specialized with this grammar. When an utterance is recognized, it is sent to the remote application for processing. If a hyperlink is returned, the browser is redirected to another location and a new grammar is downloaded from that site.

Recognition accuracy is improved if the recognition grammars define small languages. Rather than use conventional techniques for constraining the size of the input language (for example by restricting the vocabulary or range of syntactic constructs) we use an alternative technique which involves modifying the recognition grammar so that semantic constraints are "implemented" in the syntax. For example, the semantic constraint that a person cannot orbit anything (with respect to the solar-man database) is integrated into the syntax of the recognition grammar. This, and other similar restrictions on what can be recognized, has minimal impact on the user-friendliness of the interface, but results in significant improvement in recognition accuracy.

Future Research

Surprisingly little work has been done on the development of rules to assist in the design of hyperlinked speech applications. Constructing an application as a large collection of hyperlinked units, each having a small input language, can improve recognition accuracy if the user knows what s/he is allowed to say at each application. However, this approach complicates navigation, and makes it harder for the user to learn the scope of the input language. Guidelines need to be developed to help achieve a good compromise between recognition accuracy and ease of navigation. Also, very few guidelines are available which help in the design of recognition grammars.

In addition to the development of design guidelines to improve navigation and recognition accuracy, current work is directed at developing formal systems for the analysis and reformulation of queries, and for answer justification.

Acknowledgments

Many people at the University of Windsor have contribute to SpeechWeb. In particular, the author would like to thank Sanjay Chitte, Barbara Szydlowski, Tarek Haddad, Stephen Karamatos, Walid Mnaymneh, and Kunal Bhatia. The author acknowledges the support provided by NSERC in the form of an individual research grant.

References

Dowty, D. R., Wall, R. E. and Peters, S. 1981. *Introduction to Montague Semantics*. D. Reidel Publishing Company, Dordrecht, Boston, Lancaster, Tokyo.

Frost, R. A., and Boulos, P. 2002. An Efficient Compositional Semantics for Natural-Language Database Queries with Arbitrarily-Nested Quantification and Negation. Accepted for presentation at *the Fifteenth Canadian Conference on Artificial Intelligence*, AI'2002, to be held in Calgary in May 2002.

Frost, R. A. 1999. A Natural-Language Speech Interface Constructed Entirely as a Set of Executable Specifications. In *Proceedings of the Intelligent Systems Demonstrations, of the Eleventh National Conference on Artificial Intelligence, 908–909. AAAI'99* , Orlando, Florida: AAAI Press.

Frost, R. A. 1994. W/AGE The Windsor Attribute Grammar Programming Environment. *Schloss Dagstuhl International Workshop on Functional programming in the Real World*.

Iwanska, L. 1992 *A General Semantic Model of Negation in Natural Language: Representation and Inference*. Doctoral Thesis, Computer Science, University of Illinois at Urbana-Champaign.

Paaki, J. 1995. Attribute grammar paradigms — a high-level methodology in language implementation, *ACM Computing Surveys* 27(2) 196–255.