Overview of Natural Language Processing of Captions for Retrieving Multimedia Data

Eugene J. Guglielmo
Naval Weapons Center
Information Systems Department, Code 2724
China Lake, California 93555

Neil C. Rowe
Naval Postgraduate School
Computer Science Department, Code 52
Monterey, California 93943

Abstract

This paper briefly describes the current implementation status of an intelligent information retrieval system, MARIE, that employs natural language processing techniques. Descriptive captions are used to identify photographic images concerning various military projects. The captions are parsed to produce a logical form from which nouns and verbs are extracted to form the primary keywords. User queries are also specified in natural language. A two-phase search process employing coarse-grain and fine-grain match processes is used to find the captions that best match the query. A type hierarchy based on object-oriented programming constructs is used to represent the semantic knowledge base. This knowledge base contains knowledge of various military concepts and terminology with specifics from the Naval Weapons Center. *Methods* are used for creating the logical form during semantic analysis, generating the keywords to be used in the coarse-grain match process, and fine-grain matching between query and caption logical forms.

1 Introduction

Recent approaches to intelligent information retrieval have used natural language (NL) understanding methods instead of keywords and statistical methods. However, the best NL method is still unknown. This research studies a restricted form of information, the description associated with identifying multimedia data, i.e., natural language captions. The rationale and motivation for using captions was presented by Lum and Meyer-Wegener (1990). A prototype parser was developed for demonstrating how natural language queries could be used in conjunction with Structured Query Language (SQL) for specifying retrieval requests from a multimedia database.

Using these results at the Naval Postgraduate School, we have been able to design a more robust natural language processing and retrieval system for potential use at the Naval Weapons Center (NWC). The Center's Photo Lab maintains a database of over 100,000 photographs of project and historical data from the last 50 years. Both captions and supercaptions (caption about a set of captions) are used. The current search and retrieval strategy uses manually created keywords organized into a keyphrase - a head keyword and a string of descriptive nouns. Our strategy entails parsing the English captions to produce a logical form, then using the logical form as the basis of the retrieval. We have labeled this system MARIE (Epistemological Information Retrieval Applied to Multimedia).

2 Methodology

The information retrieval system we have developed is based on two stages: a coarse-grain match to reduce the list of possible information for a later fine-grain match (Rau 1987). Three tasks that we deemed essential for this system included the ability to represent and produce a logical form of the caption, the ability to generate keywords from the logical form, and the ability to load in previously stored caption logical forms for matching against the query logical form.

2.1 NL Parser

We have used an existing natural language processing program, the DBG Message Understanding System (Montgomery et al. 1989), as a starting point. This program was developed for understanding dialog conversations. To accommodate the existing captions at NWC, we had to make modifications to the grammar, functional parser, and template processor.

The grammar rules were changed to enable parsing of punctuation, descriptive noun phrases, dates, geographic locations, numeric and descriptive vehicle designations. Additional rules were introduced to handle theme-oriented phrases as opposed to agent-initiated sentences. The structure of functional parse output was altered to accommodate mapping into the type hierarchy. Specifically, tokens were introduced to allow linking together words based on syntactic relationships. The resulting output structure appears similar to slot-assertion notation.

In the original DBG system, the template processor produced frame structures for a semantic analysis of the sentence. This portion of the system was redone using an object-oriented programming methodology. We have created a single type hierarchy to hold both nouns and verbs. Producing the logical form is a matter of mapping the predicate expressions from the functional parse output into the type hierarchy. *Methods* are used to set inner cases for both nouns and verbs (e.g., theme, agent, location, etc.); set modifiers for nouns and verbs (e.g., adjectives and adverbs);

set correlations between classes (e.g., part_of, has_part, program_about, etc.); and generate the logical form from class instances and associated slot values.

2.2 Generating the Keywords

Keyword records to be used in the coarse-grain match are obtained from the type hierarchy directly rather than from the logical form output. An instance of a class uses the class name as the keyword. The keyword is based on *logically proper names*, not *definite descriptions* as described by Frixione et al. (1989). *Methods* are defined for caching keyword records containing the caption identifier and any case information to a keyword file for each class instance. Each class has a keyword file maintained in sorted order.

2.3 Matching

Once an English query is instantiated within the type hierarchy to reflect the query logical form, the instances indicate which class and subclass keyword files need to be examined in the coarse-grain match. The corresponding keyword files are read, and the keyword records are intersected using the caption-id as the unique identifier. In the future, case information will be used at query time for specifying the role for a word (e.g, initiator of an action as opposed to the recipient) and treated as a filter in selecting the appropriate case records within the keyword file. Caption-ids whose intersection score exceed a coarse-grain match threshold become eligible for fine-grain matching.

Fine-grain matching entails mapping the logical form for a stored parsed caption back into the type hierarchy and matching it against the query instances within the hierarchy. Figure 1 shows the appearance of the type hierarchy with the existence of both the query "missile on stand" and caption 262865, "Sidewinder AIM 9R missile on stand," within it.

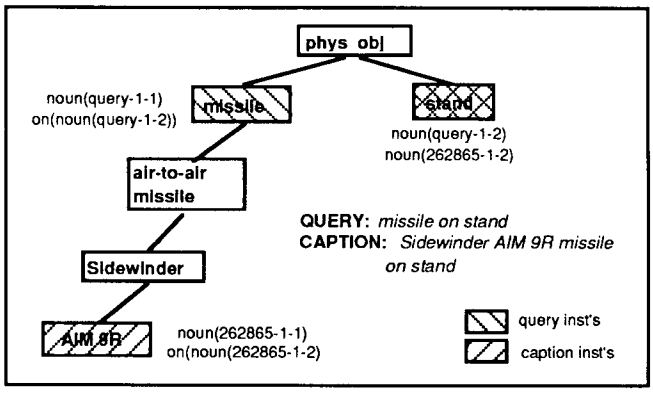

Figure 1 Fine-Grain Matching in Type Hierarchy

Instance matching is based on subtype matching. In Figure 1, the query instance for the class "missile" matches the query instance for the class "AIM-9R." Matching of relationships is currently based on exact matching. The matching process is being modified to allow relationship matching based on a predefined set of relationships. Caption-ids with match scores exceeding a fine-grain match threshold are presented to the user.

3 Implementation Status

The majority of the system is written in Quintus Prolog. with the type hierarchy being developed using the Elsa-Lap object-oriented Prolog tool. The system runs on Sun Sparcstations and was designed using a client-server relationship; the user search environment and key creation interface form the two clients and a server process handles the parsing of the natural language, generation of the keys, and the matching. The lexicon has over 1000 lexical items and the type hierarchy has over 200 classes. Further implementation and methodology details can be found in Guglielmo and Rowe (1991).

4 Future Research

The present system handles individual captions that describe an individual photograph. Future work will investigate supercaptions. For example, supercaptions that are used to represent all captions from the same chapter of a book or a supercaption that is used to represent all captions that pertain to a combat plan. All of the member captions share something in common, and the intersection of this common information forms the supercaption.

5 Conclusion

The ability to use natural language for query specification and retrieval holds the most promise over keyword and keyphrase approaches. We believe that the restricted use of natural language in captions for multimedia data retrieval is a less difficult task than full natural language fact retrieval. We feel that we have a system that can be demonstrated and built upon not only for retrieving images but also other forms of multimedia data as well.

References

Frixione, M., S. Gaglio, and G. Spinelli. 1989. Are There Individual Concepts? Proper Names and Individual Concepts in SI-Nets. *Intl. Journal Man-Machine Studies* 30:489-503.

Guglielmo, E.J. and Rowe, N.C. 1991. *Natural Language Processing of Captions for Retrieving Multimedia Data.* Tech Pub. TP-7203. Naval Weapons Center, Information Systems Department, China Lake, CA. December.

Lum, V. Y. and K. Meyer-Wegener. 1990. "An Architecture for a Multimedia Database Management System Supporting Content Search." In *Advances in Computing and Information, Proceedings of the International Conference on Computing and Information.* Niagra Falls, Canada, May 23-26.

Montgomery, C. A., J. Burge, H. Holmback, J. L. Kuhns, B. G. Stalls, R. Stumberger, and R. L. Russel Jr. 1989. The DBG Message Understanding System. In *Proceedings of the Annual AI Systems in Government Conference*, Washington, D.C., March 27-31.

Rau, L. 1987. "Knowledge Organization and Access in a Conceptual Information System." *Information Processing & Management* 23, no. 4:269-283.