Extracting Chinese Entity Names and Their Relations

Yimin ZHANG
NLP/IR, Intel China Research Center
6F1, North Kerry Center
100020 Beijing, China,
yimin.zhang@intel.com

Joe F ZHOU
NLP/IR, Intel China Research Center
6F1, North Kerry Center
100020 Beijing, China,
joe.f.zhou@intel.com

Abstract

We built an entity extraction tool that uses machine-learned rules to locate and determine entity names in natural running text, such as people, companies, products, and places. It also uses machine-learned rules to identify relations between those Entities, e.g., the relation between a product and a company, a person and a company, etc. The key techniques used are the identification of Chinese NP chunks and machine learning methods adapted from dictionary induction systems for information extraction, such as Crystal and Rapier.

The Chinese NP chunking is based on cascaded finite state automata combined with dependency relations extracted from corpus. In many cases, structural ambiguities inherent in the Chinese language are correctly handled using dependency relations between words.

The extraction of entity names and relations is rule-based. The rules are learned using machine learning techniques. Coreference resolution is introduced to identify more implicit relations. For example, from "该公司的总裁 (CEO of this company)", we first resolve what "该公司 (this company)" is referred to, then use the rules to extract the relation between the people and the company.

References

Califf M. and Mooney R. (1997) *Relational learning of pattern-match rules for information extraction.* Working Papers of the ACL-97 Workshop in Natural Language Learning, pp9-15.

Huffman S. (1995) *Learning information extraction patterns from examples.* IJCAI-95 Workshop on new approaches to learning for natural language processing, pp127-142.

Soderland S., Fisher D., Aseltine, J. and Lehnert W. (1995) *Crystal: Inducing a conceptual dictionary.* Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95), pp1314-1319.

Dekang Lin. (1997) *Using syntactic dependency as local context to resolve word sense ambiguity.* In Proceedings of ACL/EACL-97, Madrid, Spain, pp64-71.

A sample screen shot: