

CHINESE CONCEPT MAP GENERATION BASED ON OPEN INFORMATION EXTRACTION

Lung-Hao Lee & Yuen-Hsien Tseng
National Taiwan Normal University, Taiwan
Email: samtseng@ntnu.edu.tw

Abstract. Creating a concept map automatically from Chinese texts is challenging, because 1) Chinese words require no delimiters in between (like those spaces in English), which makes word, phrase, or concept identification difficult; 2) Chinese is discourse-oriented, meaning there is no strict grammar imposed on a sentence, which makes relation extraction difficult. For example, the three Chinese sentences “蘋果營養豐富” (‘Apple nutritious’), “蘋果是營養豐富的” (‘Apple are nutritious’) and “蘋果富含營養” (‘Apple are rich in nutrition’) are semantically synonymous sentences, but the first one, which lacks an overt verb, is used more often than the other two in real-world situations.

In this study, we present an Extended Chinese Open Relation Extraction (ECORE) approach that is able to extract entity-relation triples (i.e. a relation that links two concepts) from Chinese free texts based on a series of natural language processing techniques, *i.e.*, word segmentation, POS tagging, syntactic parsing, and linguistic rules. Take the sentence as an example: “蘿蔔是蔬菜，也是根莖類植物” (‘Radish is vegetable, and is a rootstock plant.’), our ECORE software can correctly extract two triples: <蘿蔔(‘Radish’), 是(‘is’), 蔬菜(‘vegetable’)> and <蘿蔔(‘Radish’), 是(‘is’), 根莖類植物(‘rootstock plant’)>, although the subject in the second clause is elided. Another example is a paragraph describing the 吳郭魚 (‘Tilapia’) in a textbook of elementary school. Figure 1 shows the original texts, 8 extracted triples and a generated concept map.

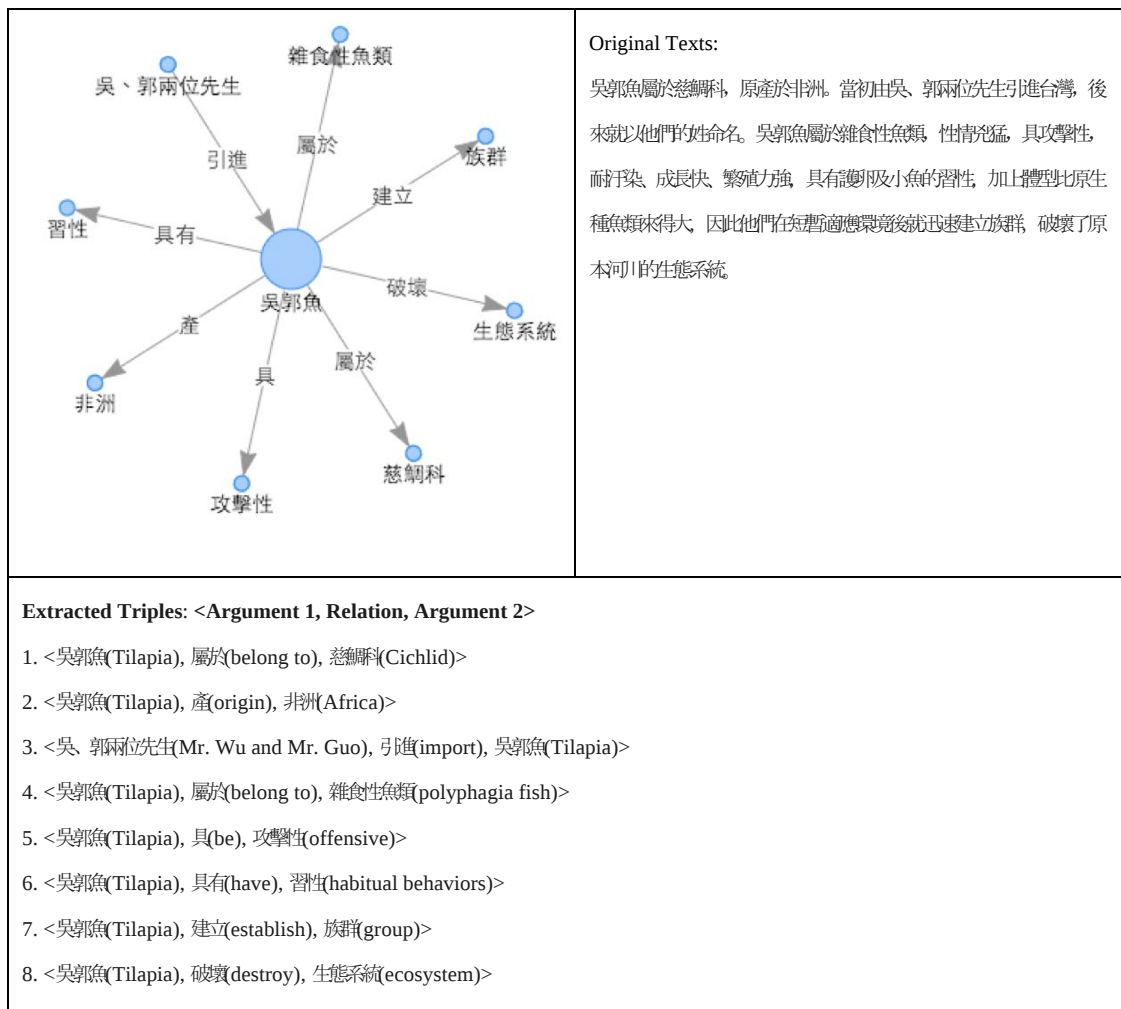


Figure 1: Example sentences, extracted triples, and a generated concept map.

To illustrate the performance of our ECORE concept mapping software, we randomly selected paragraphs from the science textbooks of elementary schools, and then manually annotated the relation triples for each Chinese sentences. In total, 141 entity-relation triples were obtained as gold standard. Performance evaluation of our ECORE software was conducted based on: 1) exact match; and 2) relation-only match. For exact match, each component of the extracted triple must be identical with the gold standard. For relation-only match, the extracted triple is regarded as a correct case if an extracted relation agreed with the relation

of the gold standard. We compared ECORE with the first Chinese open relation extraction method ever built: CORE (Tseng et al., 2014). Table 1 shows the experimental results. Our ECORE system obviously performs better than CORE no matter which metrics (exact match or relation only) were concerned.

Chinese Open IE		Precision	Recall	F-Score
Exact Match	ECORE	0.3065	0.4326	0.3588
	CORE	0.1649	0.2199	0.1884
Relation Only	ECORE	0.6784	0.9574	0.7941
	CORE	0.3936	0.5248	0.4498

Table 1: Experimental Results for performance evaluation

In conclusion, this study proposes a concept map generation method based on open information extraction techniques. Experimental results show that our ECORE software is effective for entity-relation extraction, and demonstrates its feasibility for Chinese concept map generation. The resulting software could be used to automatically generate concept maps from science textbooks for elementary schools to facilitate the teaching and learning based on the concept map approach.

Acknowledgement:

This work was partially supported by the Ministry of Science and Technology, under the grant MOST 103-2221-E-003-013-MY3 and the “Aim for the Top University Project” and “Center of Learning Technology for Chinese” of National Taiwan Normal University, sponsored by the Ministry of Education, Taiwan.

References

- Yuen-Hsien Tseng, Lung-Hao Lee, Shu-Yen Lin, Bo-Shun Liao, Mei-Jun Liu, Hsin-Hsi Chen, Oren Etzioni, & Anthony Fader (2014). Chinese Open Relation Extraction for Knowledge Acquisition, Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, pp 12-16, Gothenburg, Sweden, 26–30.