Named Entities Recognition in Computer Field for Entity Attribute Semantic Knowledge Database

Honglin Wu\textsuperscript{1,a}, Ruoyi Zhou\textsuperscript{2,b} and Ke Wang\textsuperscript{3,c}

\textsuperscript{1}College of Computer Science and Engineering, Northeastern University, Shenyang, China
\textsuperscript{2}School of Information Engineering, Zhengzhou University, Zhengzhou, China
\textsuperscript{3}Research Center for Artificial Intelligence, Shenyang Linge Technology Co., Ltd., Shenyang, China

\textsuperscript{a}wuhl@mail.neu.edu.cn, \textsuperscript{b}zhouruo.yi@qq.com, \textsuperscript{c}flyingegg.ke@gmail.com

Keywords: Named entities recognition, Entity attribute, Semantic knowledge database.

Abstract. To construct the entity attribute semantic knowledge database in computer field, we need to achieve the relationship between the entities and attributes. That requires to identify the computer-named entities that present in the real text. Moreover, the verb collocation templates that describe the relationships would be achieved. In this paper, the necessary knowledge to recognize entities would be integrated into a generic framework by using entity-attribute concept. Thereby, the rules of entity recognition would be simplified. We transform the named entities recognition process of computer entities into an labeling process. For the given text to be processed, match the possible brand words or serial words driven by the brand attribute value and the series attribute value. Then the model sequence or the abstract entity suffix can be matched and marked in the text which successfully marked the brand or series. Finally, match the results of the annotation with the recognition rules, and output the marking sequence which accord with the rules as computer entity word. Proceed from the idea of entity-attribute framework, the name of an entity is the combination of the word mapping of the entity’s particular attribute value and the word mapping of the conceptual entity to which the entity belongs. By writing the specified entity naming rules in such way, it is possible to organically integrate the rules with the instantiation of supporting rules into the knowledge network centered on entities, instead of forming irrelevant dictionary knowledge that is only isolated for specific tasks only. Experimental result showed that the system achieved the F1 measure of 86.1%.

1. Introduction

To construct the entity attribute semantic knowledge database in computer field, we need to achieve the relationship between the entities and attributes. That requires to identify the computer-named entities that present in the real text. Moreover, the verb collocation templates that describe the relationships would be achieved.

The traditional methods to recognize named-entities, whether rule-based or statistics-based, organize and integrate the knowledge in different forms. But the knowledge isolates the entities. The rule-based method meticulously divides the entity-word units. But the word units are the external resources that are independent of the entities, just like dictionaries. The statistics-based methods only statistically integrate the probability models from the high quality of the marked training resources. The models deviate from the essence of language. In this paper, the necessary knowledge to recognize entities would be integrated into a generic framework by using entity-attribute concept. Thereby, the rules of entity recognition would be simplified.

2. The Process of Entity Recognition

The process of named entities recognition in computer field is shown in Fig.1. We transform the named entities recognition process of computer entities into a labeling process. For the given text to
be processed, match the possible brand words or serial words driven by the brand attribute value and the series attribute value. Then the model sequence or the abstract entity suffix can be matched and marked in the text which successfully marked the brand or series. Finally, match the results of the annotation with the recognition rules, and output the marking sequence which accord with the rules as computer entity word.

Figure 1. The process of named entities recognition in computer field.

In the next section, we will define the rules that can be used to recognize named entities in computer field. Those rules, together with the knowledge resources supporting those rules, will be integrated into the entity-attribute-framework knowledge database in computer field.

3. The Rules of Named Entities Recognition in Computer Field

The knowledge resources necessary to form a computer entity are the values of the attributes of computer, such as brand, serial and model. The words corresponding to these attribute values are finite set of known words that are relatively fixed, or strings that are easily triggered by the automaton. In the view of entity-attribute-framework, these knowledge resources are not isolated dictionary entries. They are related to computer conceptual entities and classified according to their semantic functions. Define attribute slots attr[Brand], attr[Series], and attr[Model], each representing the brand attributes, serial attributes and model attributes of the computer entity. And define the abstract entity word.
slot abs_entity, which is a node on the network of computer conceptual entities, i.e. "abs_entity ISA
computer". The rules for named entities recognition in computer field can be summed up as shown in
Table.1, consisting of attribute slots and abstract entity slots.

Table 1. The description of rules for named entities recognition in computer field.

| Rule   | Description                                      |
|--------|--------------------------------------------------|
| Rule 1# | Entity := attr[Brand] + attr[Series] + attr[Model] |
| Rule 2# | Entity := attr[Brand] + attr[Series]             |
| Rule 3# | Entity := attr[Brand] + attr[Model]              |
| Rule 4# | Entity := attr[Brand] + abs_entity               |
| Rule 5# | Entity := attr[Series] + attr[Model]             |
| Rule 6# | attr[Model] := ^[a-zA-Z0-9]+$                     |

In Table.1, Rule 1# to 5# are the rules of word formation in computer named entities. They are
arranged successively from the high to the low according to the reliability of the rule. Rule 1# has the
highest reliability, and it can identifies the most complete structure of a computer entity. Therefore, an
annotation sequence should be matched in accordance with the above rules in order to recognize the
relatively complete computer entity words first. Rule 6# is a regular expression for models. Since the
model is a string of numbers or letters with various values but fixed forms, it can be identified by
automaton.

Proceed from the idea of entity-attribute- framework, the name of an entity is the combination of
the word mapping of the entity's particular attribute value and the word mapping of the conceptual
entity to which the entity belongs. By writing the specified entity naming rules in such way, it is
possible to organically integrate the rules with the instantiation of supporting rules into the knowledge
network centered on entities, instead of forming irrelevant dictionary knowledge that is only isolated
for specific tasks only.

4. Experimental Result
In order to verify the proposed method of named entities recognition in computer field, we designed
the following experiment. We extracted a set of 100 articles in computer field from internet, which
included 471 named entities in computer field. The evaluation criteria for the experiment are
precision, recall and F-1 measure.

We applied the proposed method on the testing set and got 428 named entities. Among them, the
exact correct computer named entities are 387. That means that the system achieved the F1 measure of
86.1%. (The precision was 90.4%, and the recall was 82.2%.)

5. Conclusion
In this paper, the necessary knowledge to recognize entities would be integrated into a generic
framework by using entity-attribute concept. Thereby, the rules of entity recognition would be
simplified. We transform the named entities recognition process of computer entities into an labeling
process. For the given text to be processed, match the possible brand words or serial words driven by
the brand attribute value and the series attribute value. Then the model sequence or the abstract entity
suffix can be matched and marked in the text which successfully marked the brand or series. Finally,
match the results of the annotation with the recognition rules, and output the marking sequence which
accord with the rules as computer entity word. Proceed from the idea of entity-attribute- framework,
the name of an entity is the combination of the word mapping of the entity's particular attribute value
and the word mapping of the conceptual entity to which the entity belongs. By writing the specified
entity naming rules in such way, it is possible to organically integrate the rules with the instantiation
of supporting rules into the knowledge network centered on entities, instead of forming irrelevant
dictionary knowledge that is only isolated for specific tasks only.
Acknowledgements
This research was financially supported by the National Natural Science Foundation of China (61370155). We thank Li Zhang and Siwei Rao for the fundamental research on entity attribute knowledge base construction.

References
[1] Y. Yin, GDC: A robust tag recommendation algorithm, Journal of Computational Information Systems, vol. 22, pp. 8061-8069, 2015.
[2] F. Zhang, Z.M. Ma, J.W. Cheng, A survey on fuzzy ontology for the semantic web, Knowledge Engineering Review, vol. 3, pp. 1-44, 2016.
[3] H.L. Wu, R.Y. Zhou, K. Wang, Chinese language analyzing using entity attribute semantic knowledge base, Proceedings of the 2nd International Symposium on Advances in Electrical, Electronics and Computer Engineering, vol. 124, pp. 470-473, 2017.
[4] H.L. Wu, R.Y. Zhou, K. Wang, The structure of domain entity attribute semantic knowledge base for natural language processing, Proceedings of the International Conference on Computer, Electronics and Communication Engineering, 2017, unpublished.
[5] K. Wang, H.L. Wu, Research on neologism detection in entity attribute knowledge acquisition, Proceedings of the 5th International Conference on Machinery, Materials and Computing Technology, Advances in Engineering Research, vol. 126, pp. 690-693, 2017.
[6] H.L. Wu, R.Y. Zhou, K. Wang, Extracting attribute words for domain entity knowledge base construction, Proceedings of the International Conference on Electronics, Electrical Engineering and Information Science, 2017, unpublished.
[7] H.L. Wu, R.Y. Zhou, K. Wang, Retrieving collocation frameworks for entity attribute knowledge acquisition, Proceedings of the 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology, Advances in Engineering Research, vol.130, pp. 1550-1553, 2017.
[8] H.L. Wu, R.Y. Zhou, K. Wang, Template based attribute value words acquisition in entity attribute knowledge base construction, Proceedings of the International Conference on Computing Intelligence and Information System, 2017, unpublished.
[9] F. Zhang, Z.M. Ma, J.W. Cheng, Enhanced entity-relationship modeling with description logic, Knowledge-Based Systems, vol. 93, pp. 12-32, 2015.
[10] H.L. Wu, Y.Y. Liu, S.M. Liu, A sentence alignment model based on combined clues and kernel extensional matrix matching method, Proceedings of AASRI Conference on Computational Intelligence an Bioinformatics, pp. 468-473, 2012.
[11] H.L. Wu, Y.Y. Liu, and S.M. Liu, Machine translation based on direct matching between text and pattern, Proceedings of the 2012 International Conference on Computer Science and Service System, pp. 3751-3754, 2012.
[12] H.L. Wu, R.Y. Zhou, K. Wang, Knowledge representation of entity attribute frame for natural language understanding, Proceedings of the 2nd International Conference on Advances in Management Engineering and Information Technology, 2017, unpublished.
[13] H.L. Wu, R.Y. Zhou, K. Wang, Entity attribute framework for the compiling model of natural language understanding, Proceedings of the 2017 2nd International Seminar on Information Science, Computer Technology and Electrical Engineering, 2017, unpublished.