

FISER: An Effective Method for Detecting Interactions between Topic Persons

Yung-Chun Chang^{1,2}, Pi-Hua Chuang¹, Chien Chin Chen¹ and Wen-Lian Hsu²

¹ Department of Information Management, National Taiwan University
No. 1, Sec. 4, Roosevelt Rd., Taipei City 10617, Taiwan (R.O.C)

² Institute of Information Science, Academia Sinica
No. 128, Sec. 2, Academia Rd., Taipei City 11529, Taiwan (R.O.C)
changyc@iis.sinica.edu.tw, r99725008@ntu.edu.tw, paton@im.ntu.edu.tw, hsu@iis.sinica.edu.tw

Abstract. Discovering the interactions between the persons mentioned in a set of topic documents can help readers construct the background of the topic and facilitate document comprehension. To discover person interactions, we need a detection method that can identify text segments containing information about the interactions. Information extraction algorithms then analyze the segments to extract interaction tuples and construct an interaction network of topic persons. In this paper, we define interaction detection as a classification problem. The proposed interaction detection method, called FISER, exploits nineteen features covering syntactic, context-dependent, and semantic information in text to detect interactive segments in topic documents. Empirical evaluations demonstrate the efficacy of FISER, and show that it significantly outperforms many well-known Open IE methods.

Keywords: Topic Person Interaction, Information Extraction, Interactive Segment.

1 Introduction

The Web has become an abundant source of information because of the prevalence of Web2.0, and Internet users can express their opinions about topics easily through various collaborative tools, such as weblogs. Published documents provide a comprehensive view of a topic, but readers are often overwhelmed by large number of topic documents. To help readers comprehend numerous topic documents, several topic mining methods have been proposed. For instance, Chen and Chen [4] summarized the incidents of a topic timeline to help readers understand the story of a topic quickly. Basically, a topic is associated with specific times, places, and persons [11]. Discovering the interactions between the persons can help readers construct the background of the topic and facilitate document comprehension. According to [12], interaction is a kind of human behavior that makes people take each other into account or have a reciprocal influence on each other. Examples of person interactions include compliment, criticism, collaboration, and competition. The discovery of topic person interactions involves two key tasks, namely *interaction detection* and *interaction extraction*. Interaction detection first partitions topic documents into

segments and identifies the segments that convey possible interactions between persons. Then, interaction extraction applies an information extraction algorithm to extract interaction tuples from the interactive segments. In this paper, we investigate interaction detection. In contrast to Open IE research [1], which focuses on discovering static and permanent relations (e.g., *capital-of*) between entities, the interaction relations we investigate are dynamic and topic-dependent. To identify dynamic interactions between persons, we define interaction detection as a classification problem. We also propose an effective interaction detection method, called FISER (Feature-based Interactive SEgment Recognizer), which employs nineteen features cover syntactic, context-dependent, and semantic information in text to detect interactive segments in topic documents. Our experiment results show that FISER can identify interactive segments accurately and the proposed features outperform those of well-known Open IE systems dramatically.

The remainder of this article is organized as follows. In Section 2, we discuss Open IE and explain how it differs from our research. We describe the proposed FISER method in Section 3, and evaluate its performance in Section 4. Then, in Section 5, we present our conclusions.

2 Related Work

Our research is closely related to Open IE, which is a novel information extraction paradigm proposed by Banko et al. [1]. In Open IE, the objective is to recognize the relations between entities without providing any relation-specific human input. Like our approach, Open IE involves two tasks, namely, *relation detection* and *relation extraction* [5, 9]. In [9], Li et al. demonstrated that relation detection is critical to outputting reliable relation tuples. However, our survey of Open IE literature revealed that most Open IE approaches omit relation detection, or they exploit simple heuristics to detect relation segments. For example, TEXTRUNNER [1] employs six syntactic features to detect relation segments in a text corpus. The drawback with the approaches is that they do not consider text semantics, so they may not perform well in terms of relation detection [5]. In [2, 14], the authors view relation extraction as a sequence labeling problem, and employ conditional random fields (CRFs) [8] to recognize relation expressions. Because the models are trained and tested with relation segments, in practice, a relation detection component is needed to achieve a good relation extraction performance [5, 9]. Zhu et al. [14] proposed a statistical framework, called StatSnowball, to conduct both traditional information extraction and Open IE. The framework employs discriminative Markov logic networks (MLNs) to learn the weights of relation extraction patterns, which are generally linguistic-structure rules and keyword-matching rules.

Our method differs from existing Open IE approaches in a number of respects. First, to the best of our knowledge, all existing Open IE approaches detect static and permanent relations. By contrast, our method detects interactive segments and the interactions between persons are dynamic and topic-dependent. Second, in addition to syntactic features, we devise useful context-dependent and semantic features to detect interactive segments effectively. Finally, most Open IE approaches analyze the text

between entities. Our method further considers the contexts before and after person names to enhance the relation detection performance.

3 FISER System

The process of FISER is comprised of three key components, namely, *candidate segment generation*, *feature extraction*, and *interactive segment recognition*. At present, FISER is designed for Chinese topics. Candidate segment generation extracts important person names from a set of topic documents, and then partitions the documents into candidate segments that may contain information about the interactions between the topic persons. Next, the feature extraction component extracts representative text features from each candidate segment. The features are used by the interactive segment recognition component to classify interactive segments. We discuss each component in detail in the following sub-sections.

3.1 Candidate Segment Generation

Given a topic document d , we first apply the Chinese language parser CKIP AutoTag¹ to decompose the document into a sequence of sentences $S = \{s_1, \dots, s_k\}$. The parser also breaks a sentence into tokens and tags their parts-of-speech. It also labels the tokens that represent a person's name. In our experiment, we observed that many of the labeled person names rarely occurred in the topic documents and the rank-frequency distribution of person names followed Zipf's law [10]. To discover the interactions between important topic persons, the low frequency person names are excluded and $P = \{p_1, \dots, p_e\}$ denotes the set of important topic person names. In Chinese, the main constituent of a sentence is a simple phrase [13]. Therefore, two or more consecutive sentences may express a coherent discourse; and an interactive segment may include a number of sentences. In our candidate segment generation algorithm, we consider two types of candidate segments: an *intra-sentential* segment in which the person names appear in the same sentence, and an *inter-sentential* segment in which the person names are distributed among consecutive sentences. Given a topic person name pair (p_i, p_j) , the algorithm processes document sentences one by one and considers a sentence as the initial sentence of a candidate segment if it contains person name p_i (p_j). Then, it examines the initial sentence and subsequent sentences until it reaches an end sentence that contains person name p_j (p_i). If the initial sentence is identical to the end sentence, the algorithm generates an intra-sentential candidate segment; otherwise, it generates an inter-sentential candidate segment. However, if a period appears in between the initial and end sentences, we drop the segment because a period indicates the end of a discourse in Chinese. In addition, if p_i (p_j) appears more than once in a candidate segment, we truncate all the sentences before the last p_i (p_j) to make the candidate segment concise. By running all person name pairs over the topic documents, we obtain a candidate segment set $CS = \{cs_1, \dots, cs_m\}$.

¹ <http://ckipsvr.iis.sinica.edu.tw/>

```

Candidate Segment Generation
INPUT:  $(p_i, p_j)$  – a topic person name pair;  $S = \{s_1, \dots, s_k\}$  – a sequence of sentences from a topic document  $d$ .
BEGIN
   $inCandidate = false$ 
   $cs = \{\}$ 
  FOR  $l = 1$  TO  $l = k$ 
    IF  $s_l$  contains  $p_i(p_j)$  &&  $inCandidate == false$ 
      add  $s_l$  into  $cs$  and set  $inCandidate = true$ 
    ELSE IF  $s_l$  contains  $p_i(p_j)$  &&  $inCandidate == true$ 
       $cs = \{\}$  and add  $s_l$  into  $cs_n$ 
    ELSE IF  $s_l$  contains  $p_j(p_i)$  &&  $inCandidate == true$ 
      add  $s_l$  into  $cs$ , save  $cs$  into candidate segment set  $CS$ 
      set  $inCandidate = false$  and  $cs_n = \{\}$ 
    ELSE IF  $inCandidate == true$  &&  $s_l$  has a period
       $cs = \{\}$  and set  $inCandidate = false$ 
  END FOR
END

```

Fig. 1. Candidate segment generation algorithm

3.2 Interactive Segment Recognizer and Feature Extraction

To recognize interactive segments in CS , we treat interaction detection as a binary classification problem. In this work, we utilize the maximum entropy (ME) classification method [3], which is a logistic regression-based statistical model. Let IS denote that a segment is interactive. ME classifies a candidate segment in terms of the following conditional probability:

$$P(IS | cs_l) = \frac{1}{Z(cs_l)} \exp\left(\sum_j w_j * f_j(IS, cs_l)\right) \quad (1)$$

$$Z(cs_l) = \exp\left(\sum_j w_j * f_j(IS, cs_l) + \sum_k w_k * f_k(\neg IS, cs_l)\right), \quad (2)$$

where f_j is a feature function and w_j is its weight. A feature function indicates a specific condition between IS and cs_l . $Z(cs_l)$ is a smoothing factor that is used to normalize $P(IS|cs_l)$ within the range $[0,1]$. Given a training dataset, the weights of the feature functions can be derived appropriately by the conditional maximum likelihood estimation method. The learned weights are then used by Eq. 1 to detect interactive segments. Generally, interactions between entities are described by verbs [7], but not all verbs express interactions. For instance, the candidate segment cs_1 shown below contains more than one verb; however, the interaction between the given person names 胡錦濤(Hu Jintao) and 歐巴馬(Barack Obama) is not described by a verb. While the verb 審問(interrogated) in cs_2 indicates repulsion between the given person names 馬英九(Ma Ying-jeou) and 蔡英文(Tsai Ing-wen), the segment contains another important topic person name, 宋楚瑜(James Soong), who is irrelevant to the interaction. Because detecting interactions is so difficult, in addition to syntactic properties, the semantic and context information should be considered to ensure that interaction detection is successful. The following presents the proposed features including syntactic, context-dependent, and semantic information in text.

[*cs*₁] 中國(Nc)_{China} 國家主席(Na)_{paramount leader} 胡錦濤(Nb)_{Hu Jintao} 即將(D)_{will} 訪問(VC)_{visit} 美國(Nc)_{United States} , (COMMACATEGORY) 將(D)_{will} 有與(V_2)_{have} 美國(Nc)_{United States} 總統(Na)_{president} 歐巴馬(Nb)_{Barack Obama} 見面(VA)_{meet} 的(DE) 機會(Na)_{chance} , (COMMACATEGORY)
(Hu Jintao, the current Paramount Leader of the People's Republic of China, will visit the U.S., and have a chance to meet Barack Obama, the President of the United States.)

[*cs*₂] 馬英九(Nb)_{Ma Ying-jeou} 總統(Na)_{president} 在(P)_{during} 辯論會(Na)_{debate} , (COMMACATEGORY) 不但(Cbb)_{not only} 反駁(VC)_{retorted} 宋楚瑜(Nb)_{James Soong's} 的(DE) 質疑(VE)_{questions} , (COMMACATEGORY) 且(Cbb)_{but also} 審問(VC)_{interrogated} 宇昌案爭議(Nb)_{Yu-Chang controversies} , (COMMACATEGORY) 指責(VC)_{challenge} 蔡英文(Nb)_{Tsai Ing-wen} 的(DE) 道德標準(Na)_{moral standards} 。(PERIODCATEGORY)
(Taiwan President Ma Ying-Jeou not only retorted James Soong's questions during the debate, but also interrogated Tsai Ing-Wen on the Yu-Chang controversy, challenging her moral standards.)

Syntactic feature set:

- **VERB RATIO (*vr*):** The ratio of transitive verbs to intransitive verbs for the given person names in a candidate segment.
- **VERB COUNT (*vc*):** The number of verbs in a candidate segment.
- **VERB COUNT BETWEEN TOPIC PERSONS (*vcp*):** The number of verbs for the given person names in a candidate segment.
- **SEGMENT LENGTH (*sl*):** The length of a candidate segment (i.e., the number of tokens).
- **VERB DENSITY (*vd*):** The ratio of verbs to the length of a candidate segment.
- **SPECIFIC PUNCTUATION (*sp*):** It is equal to 1 if the punctuation { ; , \ } appears in a candidate segment; otherwise, it is 0.
- **DISTANCE OF TOPIC PERSONS (*dp*):** The number of tokens in the given person names of a candidate segment; that is, the distance of the given person names.
- **MIDDLE TOPIC PERSON (*mp*):** It is equal to 1 if person names other than the given person names occur in a candidate segment. For instance, *mp* is 1 for *cs*₂.
- **INTRA-SENTENTIAL SEGMENT (*iss*):** It is equal to 1 if a candidate segment is intra-sentential; otherwise, it is 0. For instance, *iss* is 0 for *cs*₂.
- **FIRST POSITION (*fp*):** The first position of the given person names in a candidate segment. For instance, *fp* is 1 for *cs*₂.
- **LAST POSITION (*lp*):** The last position of the given person names in a candidate segment. For instance, *lp* is 16 for *cs*₂.

Context-dependent feature set:

- **TRI-WINDOW COUNT (*tc*):** The number of verbs in the tri-window (i.e., three consecutive tokens) before and after the given person names. For instance, *tc* is 1 for *cs*₂.
- **INTERACTIVE VERB (*iv*):** It is equal to 1 if a candidate segment contains a verb on an interactive verb list; otherwise, it is 0. The verb list is compiled by using the log likelihood ratio (LLR) [10], which is an effective feature selection method. Given a training dataset comprised of interactive and non-interactive segments, LLR calculates the likelihood that the occurrence of a verb in the interactive segments is not random. A verb with a large LLR value is closely associated with the interactive segments. We rank the verbs in the training dataset in terms of their LLR values, and select the top 150 verbs to compile the interactive verb list.
- **INTERACTIVE BIGRAM (*ib*):** It is equal to 1 if a candidate segment contains a bigram of an interactive bigram list; otherwise, it is 0. The bigram list is compiled in a similar way to the verb list by selecting the top 150 bigrams in the training dataset based on their LLR values.

Semantic feature set:

- **SENTIMENT VERB COUNT** (*svc*): This is the number of sentiment verbs in a candidate segment. Intuitively, interactions can occur with positive or negative semantics. For instance, the verb 審問(interrogated) in cs_2 describes criticism between the given person names, and it is a sentiment verb with negative semantics. Here, we employ the NTU Chinese Sentiment Dictionary (NTUS)², which contains 2812 positive and 8276 negative Chinese sentiment verbs compiled by linguistic experts.

- **NEGATIVE ADVERB COUNT** (*nac*): The number of negative adverbs (e.g., 未曾(have never)) in a candidate segment.

- **INTERACTIVE SEMEME** (*is*): It is equal to 1 if a sememe of a verb in a candidate segment is on an interactive sememe list; otherwise, it is 0. A sememe is a semantic primitive of a word defined by E-HowNet [6], which is a Chinese lexicon compiled by Chinese linguistic experts. Basically, an interaction can be described by different synonyms. By considering the sememes of the verbs in a candidate segment, we may increase the chances of detecting interactions. For each sememe in E-HowNet, we compute its information gain [10] in discriminating the interactive and non-interactive segments of the training dataset. However, a sememe with a high information gain can be an indicator of non-interactive segments. Therefore, we process sememes one by one according to the order of their information gains. We compute the frequency that a sememe occurs in the interactive segments. If the sememe tends to occur in the interactive segments, we regard it as an interactive sememe; otherwise, it is a non-interactive sememe. We compile the interactive sememe list by selecting the first 150 interaction sememes.

- **NON-INTERACTIVE SEMEME** (*ns*): It is equal to 1 if a sememe of the verbs in a candidate segment is on a non-interactive sememe list; otherwise, it is 0. Similar to *is* the non-interactive sememe list is compiled by selecting the first 150 non-interactive sememes in the training dataset.

- **FREQUENT SEMEME** (*fs*): It is equal to 1 if a sememe of the verbs in a candidate segment is on a frequent sememe list; otherwise, it is 0. We rank the sememes of verbs according to their occurrences in the interactive segments of the training dataset. The frequent sememe list is compiled by selecting the first 150 frequent sememes.

4 Performance Evaluation

In this section, we present the data corpus used for the performance evaluation; examine the effects of the proposed features; and compare the proposed feature set with those of well-known Open IE methods.

4.1 Data Collection

In information extraction, evaluations are normally based on official corpora. Most previous information extraction studies used the Automatic Context Extraction (ACE) datasets³ to evaluate system performance. However, the relations (e.g., *capital of*) defined in the datasets are static and therefore irrelevant to interactions between persons. To the best of our knowledge, there is no official corpus for interaction

² <http://nlg18.csie.ntu.edu.tw:8080/opinion/pub1.html>

³ <http://www.itl.nist.gov/iad/mig/tests/ace/>

detection; therefore, we compiled our own data corpus for the performance evaluations. Table 1 shows the statistics of the data corpus, which comprises ten political topics in Taiwan from 2004 to 2010. Each topic consists of 50 news documents (all longer than 250 words) downloaded from Google News. As mentioned in Sec. 3, many of the person names labeled by CKIP rarely occur in topic documents. Hence, we selected the first frequent person names whose accumulated frequencies reached 70% of the total person name frequency count in the topic documents. We extracted 1747 candidate segments from the topic documents by using the candidate segment generation algorithm. Then, three experts labeled 455 segments as interactive, and the Kappa statistic of the labeling process was 0.615. The statistic indicates that our annotated data corpus is substantial. It is noteworthy that approximately 45% of the interactive segments are inter-sentential. In other words, expressions of person interactions often cross sentences. Meanwhile, 77% of the intra-sentential segments are non-interactive. Since interaction expressions are rare and sentence-crossing, detecting interactions is difficult.

Table 1. The statistics of data corpus

# of topics	10
# of topic documents	500
# of tagged person names	436
# of evaluated person names	85
# of person name pairs	432
# of interactive segments (intra)	266
# of interactive segments (inter)	189
# of non-interactive segments (intra)	905
# of non-interactive segment (inter)	387

4.2 Effects of the Features

We use 10-fold cross validation to examine the effects of the proposed features. For each evaluation run, a topic and the corresponding candidate segments are selected as test data, and the remaining topics are used to train FISER. The results of the 10 evaluation runs are averaged to obtain the global performance. The evaluation metrics are the precision, recall, and F1-score. We use F1 to determine the superiority of the features because it balances the precision and recall scores.

Table 2 shows the performance of the syntactic, context-dependent, and semantic features, denoted as $FISER_{\text{syntactic}}$, $FISER_{\text{context-dependent}}$, and $FISER_{\text{semantic}}$ respectively. As shown in the table, the syntactic features cannot detect interactive segments correctly. This is because they are incapable of discriminating between interactive segments. Since both interactive and non-interactive segments are comprised of grammatical sentences, they have similar syntactic feature values. For instance, the averages of VERB RATIO (vr) for the interactive and non-interactive segments are 2.11 and 2.33 respectively; and there is no significant difference in terms of t-testing with a 99% confidence level. Besides, as mentioned in Sec. 4.1, both intra-sentential and inter-sentential segments can be non-interactive. Therefore, the syntactic INTRA-SENTENTIAL SEGMENT (iss) and SPECIFIC PUNCTUATION (sp) features,

which are used to judge inter-sentential segments, are indiscriminative. By contrast, the context-dependent features detect interactive segments successfully. We observe that the compiled interactive verb list and interactive bigram list are closely associated with person interactions, so the INTERACTIVE VERB (*iv*) and INTERACTIVE BIGRAM (*ib*) features discriminate interactive segments effectively. Meanwhile, the verbs used to describe person interactions tend to occur immediately before or after the given person names. Thus, the context-dependent TRI-WINDOW COUNT (*tc*) feature is useful for filtering out non-interactive segments. It is noteworthy that the semantic features produce a high precision rate, but a low recall rate. Our analysis of the experimental data showed that segments containing positive or negative verbs generally reveal person interactions; hence, the semantic feature SENTIMENT VERB COUNT (*svc*) yields high detection precision. However, a significant proportion of the interactive segments do not have sentimental semantics, so the feature cannot increase the detection recall rate. While the semantic features INTERACTIVE SEMEME (*is*), NON-INTERACTIVE SEMEME (*ns*), and FREQUENT SEMEME (*fs*) try to increase the detection of interactive segments by considering the sememes of verbs, the expert-compiled E-HowNet is not comprehensive enough to identify various person interactions. Notably, FISER achieves its best performance when all the features are applied together (denoted as FISER_{all}). In other words, the context-dependent features and semantic features do not conflict with each other.

Table2. Experimental result of each feature category

<i>Features Category</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
FISER _{syntactic}	19.4%	1.1%	2.0%
FISER _{context-dependent}	66.3%	41.5%	49.3%
FISER _{semantic}	66.7%	16.9%	26.2%
FISER _{all}	70.5%	51.9%	58.6%

4.3 Comparison with Open IE Methods

Since interaction detection is an innovative research issue, we hardly find systems for comparisons. Nevertheless, our research is closely related to the relation detection task of Open IE, which focuses on detecting text segments conveying static and permanent relations between entities. Hence, we compare FISER with three well-known Open IE methods, namely, TEXTRUNNER [1], O-CRF [2], and StatSnowball [14]. TEXTRUNNER, the first Open IE system, employs six syntactic features to extract the relations between entities. O-CRF considers syntactic features, including POS tags, conjunctions, and regular expressions of syntax. It also uses context words to identify relation keywords between entities. StatSnowball also adopts syntactic features to identify relation keywords between entities. The selected features include POS tags and occurrences of non-stop words. Notably, O-CRF and StatSnowball, which are designed for relation extraction, extract interaction keywords from a candidate segment in our experiment. Hence, a candidate segment is classified as non-interactive if no interactive keyword is extracted from it. Additionally, to examine the effectiveness of the proposed features in a fair manner, we also train a ME classifier using the features of each compared method (denoted as TEXTRUNNER_F, O-CRF_F,

and, StatSnowball_F respectively) and employ the 10-fold cross validation to obtain its global performance.

Table3. The interaction detection result of compared methods

<i>Features</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
TEXTRUNNER	32.7%	2.2%	4.0%
O-CRF	42.1%	8.8%	14.6%
StatSnowball	48.1%	5.5%	9.9%
TEXTRUNNER _F	48.8%	34.3%	38.9 %
O-CRF _F	53.2%	39.8%	43.5%
StatSnowball _F	52.6%	25.2%	32.1%
FISER _{all}	70.5%	51.9%	58.6%

As shown in the Table 3, FISER outperforms all the compared methods and feature sets. As the compared methods and feature sets simply use syntactic features, they cannot sense the semantics of person interactions in candidate segments successfully. By contrast, FISER incorporates semantic and context-dependent features, and thus achieves the best precision, recall, and F1 score. O-CRF outperforms StatSnowball and TEXTRUNNER because its feature set considers the context information of a candidate segment. It is interesting to note that O-CRF and StatSnowball are inferior to O-CRF_F and StatSnowball_F, and the recall rates of O-CRF and StatSnowball are very low. Basically, O-CRF and StatSnowball employ the CRF model to learn the extraction patterns of interaction keywords. Since the non-interactive segments have no interaction keywords, only the interactive segments of the training data are useful for pattern learning. As shown in Table 1, most of the candidate segments are non-interactive. Thus, the learned extraction patterns cannot detect interactive segments completely, and the recall rates of the methods deteriorate. The outcome corresponds well with the observation in [9] that detecting relation segments is necessary to ensure that extractions of relation keywords are reliable. Based on the experimental results, we conclude that syntactic features cannot detect interactive segments correctly. Existing Open IE studies focus on discovering static and permanent relations between entities. In [2], Banko and Etzioni claim that 86% of relation expressions are in the given entities. However, according to our data corpus, only 56% of the interaction expressions are in the given person names in Chinese. Therefore, the compared methods are inferior in terms of detecting interactive segments.

5 Conclusion and Feature Works

A topic is associated with specific times, places, and persons. Discovering the interactions between the persons would help readers construct the background of the topic and facilitate document comprehension. In this paper, we have proposed an interaction detection method called FISER, which employs nineteen features covering syntactic, context-dependent, and semantic information in text to detect interactive segments in topic documents. Our experiment results demonstrate the efficacy of FISER and show that it outperforms well-known Open IE methods.

In our future work, we will employ sophisticated syntactic features, such as the dependency tree of a sentence, to enhance FISER's syntactic features. We will also investigate using information extraction algorithms to extract interaction tuples from the detected interactive segments and construct an interaction network of topic persons.

References

1. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M., Etzioni, O.: Open information extraction from the web. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence, pp. 2670-2676, (2007)
2. Banko M., Etzioni, O.: The tradeoffs between open and traditional relation extraction. In: Proceedings of the 46th Annual Meeting on Association for Computational Linguistics on Human Language Technologies, pp. 28-36, (2008)
3. Berger, A.L., Pietra, V.J.D., Pietra, S.A.D.: A maximum entropy approach to natural language processing. Computational Linguistics, vol. 22, pp. 39-71, (1996)
4. Chen, C.C., Chen M.C.: TSCAN: A content anatomy approach to temporal topic summarization. IEEE Transactions on Knowledge and Data Engineering, vol. 24, pp. 170-183, (2012)
5. Hirano, T., Asano, H., Matsuo, Y., Kikui, G.: Recognizing relation expression between named entities based on inherent and context-dependent features of relational words. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 409-417, (2010)
6. Huang, S.L., Chung, Y.S., Chen, K.J.: E-HowNet: the expansion of HowNet. In: Proceedings of the 1st National HowNet Workshop, pp. 10-22, (2008)
7. Kim, M.Y.: Detection of gene interactions based on syntactic relations. Journal of Biomedicine and Biotechnology, vol. 2008, pp. 371-380, (2008)
8. Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning, pp. 282-289, (2001)
9. Li, W., Zhang, P., Wei, F., Hou, Y., Lu, Q.: A novel feature-based approach to Chinese entity relation extraction. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies, pp. 89-92, (2008)
10. Manning, C.D., Schütze, H.: Foundations of statistical natural language processing: MIT Press, Cambridge, Massachusetts, 1st edn., (1999)
11. Nallapati, R., Feng, A., Peng, F., Allan, J.: Event threading within news topics. In: Proceedings of the 13th ACM International Conference on Information and Knowledge Management, pp. 446-453, (2004)
12. Vernon, G.M.: Human interaction: an introduction to sociology: Ronald Press Co., New York, 1st edn., (1965)
13. Wang, Y.K., Chen, Y.S., Hsu, W.L.: Empirical study of Mandarin Chinese discourse analysis: an event-based approach. In: Proceedings of 10th IEEE International Conference on Tools with Artificial Intelligence, pp. 466-473, (1998)
14. Zhu, J., Nie, Z., Liu, X., Zhang, B., Wen, J.R.: StatSnowball: a statistical approach to extracting entity relationships. In: Proceedings of the 18th International Conference on World Wide Web, pp. 101-110, (2009)