ICIC Express Letters
Part B: Applications
Volume **9**, Number **8**, August **2018**

ICIC International ©2018 ISSN 2185-2766
pp. **747–756**

WHY BIOMEDICAL RELATION EXTRACTION IS AN OPEN ISSUE?

Farzane Ghamami and Mohammadreza Keyvanpour

Department of Computer Engineering
Alzahra University
Deh-e-Vanak St., Tehran 1993893973, Iran
f.ghamami@student.alzahra.ac.ir; keyvanpour@alzahra.ac.ir

Received January 2018; accepted April 2018

Abstract. *Biomedical information has been receiving rapid expansion in recent years. Articles on biomedical area are available in free-form texts in libraries and databases. Many articles published on biomedicine field contain up-to-date information. Besides, the new findings of biomedical projects are widespread, such as the names of genes, proteins, drugs, treatments, and diseases. Considering the rapid progress of bioscience, manual exploration of a huge volume of information will be time-consuming and costly. One of the techniques for using this information includes relation extractions. Biomedical relation extraction methods gather valuable, reliable, and affordable information from free-form texts. However, in biomedical relation extraction, to manage and discover new and valuable information is a very challenging task. This paper first describes biomedical relation extraction methods. Second, we proposed a classification for biomedical relation extraction challenges. Finally, the methods were assessed by addressing the challenges. This paper prepares a good platform to compare the biomedical relation extraction methods by evaluating the methods in the light of the challenges. An appropriate comparison leads to better understanding of the methods. Consequently, perfect understanding of the methods will help researchers to eliminate the obstacles and achieve the methods development and improvement.*

Keywords: Text mining, Biomedical information, Relation extraction, Challenges

1. **Introduction.** Biomedical relation extraction is an active area in biomedical information extraction and management. In recent years biomedical information has progressed. The high volume of information is one of the outcomes of the rapid advancement [1]. The use and management of biomedical information and methods to stay updated and informed are important issues [2]. One of the sources for accessing biomedical information is scientific literature databases and libraries [3]. Biomedical relation extraction is an information extraction subtask that is able to organize, manage, and discover new information from biomedical texts [4]. Information obtained from biomedical relation extraction is used in biomedical databases and as a part of the other tasks inputs, such as event extraction [4, 5]. Furthermore, biomedical relation extraction outputs are used in question-answering systems [6], diagnosis categorization [7], clinical decision support [8], and ontology population, and learning [9]. This study proposed a classification for challenges, according to the wide variety of applications of relation extraction from biomedical texts and the particular importance of understanding and resolving the challenges. This classification introduces and organizes the challenges of the methods for better cognition and understanding, and then assesses the methods to address the challenges. This classification can play a significant role in the comparing and selecting of the biomedical relation methods and, improving them also. Surveying previous works brought light and understanding into this study, and made highlight the motivation of this study also. Few previous related works were reviewed. So far, many significant works have been published

DOI: 10.24507/icicelb.09.08.747

in the area. Generally, in previous studies, biomedical relation extraction methods are two main approaches: pattern-based and supervision-based methods. Some of the most comprehensive classifications are briefly introduced as follows: Bedmar [10] suggested three main approaches to relation extraction from biomedical texts. These approaches include linguistic-based, pattern-based, and machine-learning. The linguistic-based approach focuses on the natural language process (NLP), and the detection of relations depends on NLP preprocessing and tasks. The pattern-based approach includes a set of methods that recognizes relations via a predefined pattern set that can be both manual and automatic. The machine learning approach refers to the method that does not need manual encoding for extraction and classification of biomedical relations [10]. In most categories, this approach consists of two methods: feature-based and kernel-based. Katukuri et al. [11] examine and categorize these methods by considering new aspects of relation extraction methods. They divided biomedical relation extraction into three classes: rule-based, graphical, and discriminative models. The rule-based methods are the same as the pattern-based methods that are conditional and bounded. In Katukuri et al.'s [11] study, graphical models extract relations through the help of the grammar structure of sentences feed to a graph algorithm like hidden Markov model (HMM). Later, several studies such as the one by Aggarwal and Zhai [6], Bui [12], Huang [13], and Song et al. [14], presented similar categories consisting of three main approaches: co-occurrence, rule-based, and classification. The co-occurrence approach extracts relations by considering the frequency rate of entities mentioned together in computation using statistical means. The classification approach includes every method that can detect and classify the relation type of two or multi-classes. Due to the fact that biomedical relation extraction lacks comparative platform, this study offers a classification for the challenges in this field. Thus, the deficiency of standardization in the biomedical relation extraction techniques makes quantitative comparison impossible [15]. In this field of study, it was realized that obstacles and challenges can be removed to improve and compare these methods qualitatively by addressing the challenges. This study was meant to accurately survey techniques by preparing a qualitative assessment. The qualitative assessment of techniques to address the challenges led to a powerful tool for improving the techniques. Also, this paper presents the distinct insight of the researchers for the deep understanding of various aspect of biomedical relation extraction problem. It also tries to highlight the challenges of biomedical relation extraction. Consequently, it can help researchers to practically choose the appropriate method and efficient resolution and overcome challenges in future works.

This paper is organized as follows. Section 2 introduces biomedical relation extraction. Section 3 describes the proposed classification for biomedical relation extraction challenges. Section 4 presents a comparative assessment of the methods. Section 5 discusses the advantages of this paper. Section 6 concludes the paper.

2. **Relation Extraction in Biomedical Domain.**

Information extraction subtasks can detect structured information from free-form texts [6]. Biomedical relation extraction is defined as the process of exploring biomedical texts according to the syntactic and semantic roles and natural language processing tasks, in order to detect the relationships between existing entities (such as genes, proteins, and drugs) [16, 17]. Generally, biomedical texts are highly dimensional, relational, and complex [13]. Consequently, to get acceptable results, need considerable attention and accuracy [18]. As regards to the fast growth of biomedical information, biomedical relation extraction plays a key role in using organization of updated information. Figure 1 shows the general biomedical relation extraction workflow architecture. The extraction of relation depends on the number and type of desired relations. For example, if there is preference specific relation between two entities like the drug-target interaction called binary relation extraction (well as seen in

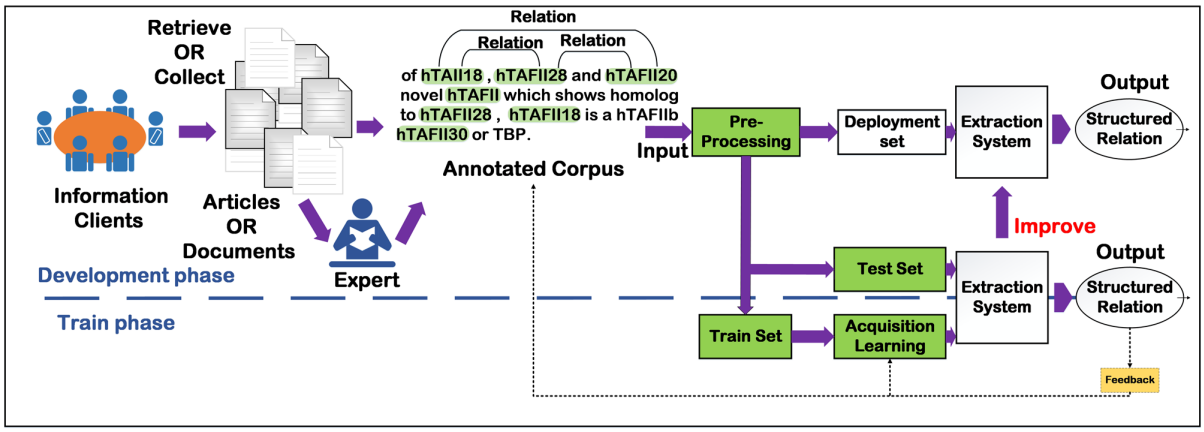

FIGURE 1. Biomedical relation extraction workflow adapted from [16, 19]

Bisebromoamide$_{\text{Drug}}$ inhibits **ERK**$_{\text{Genes and Genes Products(GGP)}}$ phosphorylation.
Bisebromoamide **Interact with:** ERK
Biomedical Relation Type: Drug-target Interaction

FIGURE 2. An example of binary relation extraction [20]

Expression of the **Sigma(K)**$_{\text{E1}}$-dependent on **cw1H**$_{\text{E2}}$ gene depended on **gerE**$_{\text{E3}}$.
Sigma(K) **Interact with:** cwlH
gerE **Interact with:** cwlH
Biomedical Relation Type: protein-gene Interaction

FIGURE 3. An example of complex relation [21]

Figure 2). If more than two entities have a relation in a sentence, it is called complex relation extraction (Figure 3). Often, in the biomedical domain, there are complex relations. Formal biomedical relation extraction, generally independent of any approach, is as follows (Note that relations are considered as binary.):

$$R = \{r_1, r_2, \ldots, r_n\} \tag{1}$$

$$\forall s_i =\in D : s = \{w_1, \ldots, w_i, e_1, w_i + 1, \ldots, w_i + j, \ldots, e_2, w_i + j + 1, \ldots, w_n\} \tag{2}$$

$$y_i = f(s_i) \tag{3}$$

$$\forall y_i \subset D = \begin{cases} \text{true}, & \exists y_i \in R : y_i = \langle \{e_1, e_2, r\} \rangle \in r_n \\ \text{false}, & \text{otherwise} \end{cases} \tag{4}$$

where $n = 1, 2, 3$ has r_n relation types as shown in Equation (1); s is the ith sentence from the document of the dataset including entities and n words as shown in Equation (2); $f(.)$ is the mapping function; y is the mapped set (Features, kernels, patterns, rules, etc.) under $f(.)$ corresponding to s_i as shown in Equation (3). Extracting the biomedical relation is successful if the extracted relation r from the mapped sample y_i is a member of a predefined and desire set of R, as is well shown in Equation (4). According to this study, generally, biomedical relation extraction has two main approaches and they are self-direction and set-dependent. This classification is as shown in Figure 4.

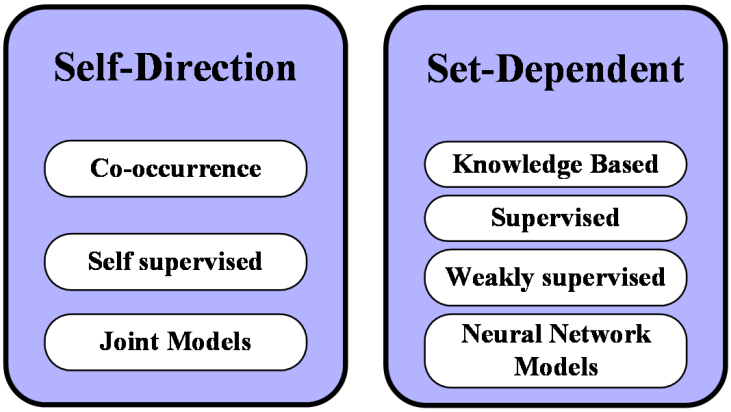

FIGURE 4. Approaches classification for biomedical relation extraction

- Self-Direction. This approach consists of techniques that use statistical means for hidden relation discovery and relation extraction from biomedical texts. This approach learns relations based on similarity and relies on occurrence. Self-direction includes co-occurrence, self-supervised, and joint models. In this category, the number of relations for extraction is infinite. *Co-occurrence*: The co-occurrence approach is the simplest way to discover relations [12, 22]. *Self-supervised*: This approach refers to methods that generate small set of relations, and then discover the remaining relations [23]. *Joint models*: This approach enables relation to be discovered and recognized by joints and linkages between entities recognition, coreference resolution, and relation cooperative [24].
- Set-dependent: This approach is divided into techniques that focus on a predefined set to find and extract biomedical relations. This approach consists of knowledge-based and supervised techniques. *Knowledge-based*: These methods highly depend on the domain of input text. Also, in this category relation extraction relies on the specific predefined templates based on input text in the form of pattern-set or rule-set [22]. This class of methods consists of pattern-based and rule-based methods [12, 13]. *Supervised or classification approach*: This category of biomedical relation extraction methods identifies relations between biomedical entities using machine learning methods [6]. Significant methods of supervised approaches are feature and kernel methods. Feature-based methods extract important information from the text by pruning.

$$s(t.F) = \textbf{Feature Space}\{X_1, X_2, \ldots, X_n\} \tag{5}$$

$$\{w_1, \ldots, e_1, w_j, \ldots, e_2, \ldots, w_n\} \widehat{=} \{X_1, X_2, \ldots, X_n\} \tag{6}$$

where t is the elements of a sample chosen as feature Equation (5); finally, a sample (such as a sentence) turns to feature vector Equation (6). Kernel methods transfer inputs to kernel space implicitly based on the similarity of the samples [1].

$$\forall s \in D, \quad \Phi : s \to \eta \tag{7}$$

$$K(X_i, X_j) = \langle \Phi(s_i),\ \Phi(s_j) \rangle \tag{8}$$

For all samples (The sample metrics is the sentence) that belong to dataset Equation (7), there is a mapping function ϕ that is based on the similarity of transferring the samples Equation (7) to the K kernel space by an inner product Equation (8) [26]. *Weakly supervised*: This method is also known as bootstrapping [22, 23]. *Neural Network models*: This method identifies relations by neural networks in raw texts [25].

3. **Biomedical Relation Extraction Challenges.** As biomedical information is rapidly increasing, it is very important for researchers around the world to do coherent and aligned research to have up-to-date information. Furthermore, there is lack of comprehensive and standard methodology in biomedical relation extraction. The present study proposed a classification of the existing challenges of biomedical text relation extraction to improve the direction of methods and eliminate challenges as shown in Figure 5. However, the manual relation extraction from biomedical texts is almost impossible [1]. A fully automatic biomedical relation extraction is a challenging task and an open issue [28]. This proposed classification tries to identify the challenges to reduce human role and costs, and improve automatic biomedical relation extraction techniques. The following section describes briefly the challenges in order depth-first.

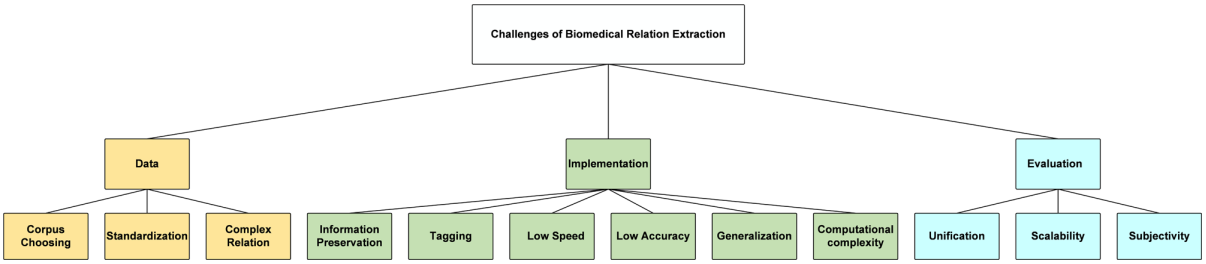

FIGURE 5. The proposed classification of biomedical relation extraction challenges

- *Data challenges*: This class of challenges refers to those posed by biomedical text nature and informative dataset collection method obstacles. The biomedical text is complex [13]. The biomedical text has long and related sentences. Indeed, biomedical texts are not grammatical [12]. This class of challenges is divided into three categories: corpus choosing, standardization, and complex relation.
  - Corpus choosing: This involves the selection of the most informative text part as corpus problems. According to the research goals, each section of a biomedical, article is useful [27]. Due to short lengths, simple structures, and proximity to relevant biomedical cues, sentences are considered as corpus in most researches [28]. On the other hand, a whole document probe makes implicit the associations between entities discovered [29]. Although the selection of sentences in most cases increases the accuracy; it also increases the dimensions of a dataset [15]. Choosing the informative part of the biomedical text, remains a challenge.
  - Standardization: Data standardization involves two issues: first, the challenges posed by the non-availability of proper standardized biomedical abbreviations, synonyms, and terminology [1, 13]; second, the lack of standard in sample generation [15].
  - Complex relation: A characteristic of biomedical literature is ungrammatical sentences [30]. Uninformative words in complex sentences along with high dimension, make sparse space for relation extract [10, 31].
- *Implementation challenges*: This class of challenges is related to the difficulty of automating biomedical relation extraction, technique performance, and achieving the correct result costs. Implementation challenges include information preservation, tagging, low speed, low accuracy, generalization, and computational complexity.
  - Information preservation: According to Joachims [32], biomedical documents can be represented from different levels. As the level of representation increases, valuable information can be preserved [31, 33]. The point is that all algorithms and techniques cannot be applied to all levels of representation.
  - Tagging: Due to lack of a proper standard for naming and tagging biomedical entities [6] and its high cost [34], tagging is an existing challenge.

  - Low speed: According to the large volume of biomedical information, researchers require techniques that perform within acceptable time [35, 36, 38]. Besides, some applications of biomedical relation extraction like question-answer system need to generate the output of extraction process [37].
  - Low accuracy: Biomedical literature characterizes decrease in accuracy of performance. The same techniques' performance in non-biomedical texts provides accurate and more satisfactory results [1].
  - Generalization: Often, the biomedical train set and the test set are not independent [15]. Consequently, this method is well-adapted only to a specific dataset and cannot be widely used.
  - Computational complexity: Discovery of correct relations from the raw text with minimum parameter settings, preprocessing, and requirement is one important challenge [15, 34, 39].
- *Evaluation challenges*: Researchers in the biomedical area usually use their own datasets and evaluate with precision, recall, and f-measure metrics Equation (9) [10]. The assessment metric determines whether the used method is reliable, and which method is more appropriate and efficient [10]. Hence, metrics and evaluation in the biomedical area are essential challenges [15].

$$\begin{aligned} Precision &= \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN}, \\ F\text{-}Measure &= \frac{(1+\beta)Precision \times Precision}{\beta^2 Precision + Recall} \end{aligned} \tag{9}$$

  - Unification: The biomedical relation extraction concept depends greatly on identic relation definition [15]. Lack of a standard definition of identical relations is an obstacle to generalization and comparison of the methods. On the other hand, the correctness of the output of the methods is limited to the comparison with the gold standard annotation, which requires development and maturity [10].
  - Scalability: As a result of the huge dimension of biomedical exploration space, the large volume of biomedical documents and their importance and maintaining the reliability of relation extraction methods are challenges faced by the researchers [40, 41]. The remarkable point is that methods with acceptable performance in the biomedical field increase in data scale, experience performance decrease [12].
  - Subjectivity: Different datasets, occasional parameter setting, and arbitrary initialization lead to biased techniques [10, 15]. Accordingly, local results usually gain in the biomedical relation extraction field [34].

4. **The Assessment of the Techniques.** This section organizes the techniques by addressing the challenges shown in Table 1. Also, by qualitative assessment, it allows a deeper understanding of techniques, resolution of challenges, the selection of appropriate techniques, and their improvement. As shown in Table 1, when the techniques resolved the challenge completely, it means the (✓) mark has been used. Indeed, the (✓) shows the methods, that handle specific challenges. For example, co-occurrence method is very fast. However, most of the significant studies in biomedical field gain low accuracy. This assessment is based on qualitative comparison to almost all the efficient existing methods. (↑) mark shows that the techniques obtained a relative improvement. Relative improvement refers to the acceptable result in facing challenges. Methods that perform more better than the marked method exist. For example, joint models can detect complex relation. The results of kernel-based methods and neural networks models are more accurate than joint models. In general, these methods work to face the afore-mentioned challenges medium and can still improve them. Finally, (×) mark means the challenge exists as an

TABLE 1. The biomedical relation extraction techniques handled regular challenges

		Data			Implementation						Evaluation		
		Corpus Choosing	Standardization	Complex Relation	Information Preservation	Tagging	Low Speed	Low Accuracy	Generalization	Computational Complexity	Unification	Scalability	Subjectivity
Techniques	Co-occurrence	×	×	×	×	✓	✓	×	×	✓	✓	×	×
	Self Supervised	×	↑	×	×	↑	↑	↑	×	↑	×	×	↑
	Joint Models	×	↑	↑	↑	↑	×	✓	↑	×	×	↑	↑
	Knowledge Based	×	×	×	×	×	✓	×	×	✓	↑	×	×
	Weakly Supervised	×	×	×	×	↑	↑	↑	×	↑	×	×	×
	Feature Based	×	↑	✓	×	×	↑	✓	↑	×	×	↑	↑
	Kernel Based	✓	✓	✓	↑	×	↑	✓	✓	×	×	✓	✓
	Neural Network Models	✓	✓	✓	↑	×	×	✓	↑	×	×	✓	↑

unsolved issue. This (×) mark presets open areas and challenges in biomedical relation extraction. When a method is having (×) mark, it refers to dropping the performance of the method by facing the specific challenge. Cells in Table 1 that have (×) mark show that methods can be developed by considering the challenges. Neural network models are very complex and time-consuming methods, but present great performance and very satisfying results in the biomedical relation extraction field. Generally, methods with (↑) and (×) marks have the high potential and chance to develop and improve to present satisfaction results and resolve the obstacles in the biomedical future works.

5. **Discussion.** Unfortunately, due to the existing challenges, it is not possible to quantitatively compare the biomedical relation extraction techniques. In this section, the advantages, disadvantages and the qualitative comparisons are discussed as shown in Table 1. The major advantage of the co-occurrence method is the extraction of relations from large-scale texts without human interference for class labeling or tagging [17, 34, 42]. The co-occurrence method assumed that the two entities that are frequently mentioned together are related in some way [12]. Thus, co-occurrence methods by limiting computational to the number of word occurrences, and assigning them to words weight ensure simple and fast performance [42]. However, co-occurrence cannot extract complex relations and it is a source of low accuracy and other drawbacks [12]. Knowledge-based methods can detect complex relations in a limited and specific dataset, since this method relies on patterns or the rule of exact matching [43]. When there is need to extract a small set of special relations (subject-object) in a specific context, these techniques are very useful and efficient [43]. Despite being self-supervised and weakly-supervised like two earlier techniques performing on a large scale, performance drops with the slightest change in the word morphology in high dimension biomedical datasets [12, 44]. The low precision obtained by many studies is an evidence of this phenomenon [12]. Accuracy in the self and weakly supervised methods is enhanced by bootstrapping and co-learning [23]. Besides, the self-supervised method is more flexible in pattern matching than weakly-supervised method's ability to manage standardization and subjectivity result [23]. Simplicity and the providing of candidate identifiers are the major characteristics of self and weakly supervised methods [44]. They are helpful when the type of relation is limited (for example, just interaction extraction); however, the detection of coincidental mentions instead of desired relations remains an issue [44]. It is very important to define the relation between

entities for feature-based and kernel-based, joint model, and neural network model methods. Feature engineering is a precise and costly task [6]. Consequently, feature-based methods, joint and neural network models in addition to the complexity and high cost faced with locality and bias to features in relation extraction task [6, 46, 47], decrease the accuracy and generality. Feature-based methods failed when the data-scale and dimension increased and also when the information is heterogeneous and structured [1]. Thus, joint and neural network models use features as weight, and possess more scalability than feature-based methods [1, 45]. Joint models prevent sparseness and affected language process problems [16, 24]; instead, entity recognition, coreference resolution, and relation extraction synchronization are major existing challenges [47]. Kernel-based methods work well in dealing with challenges by implicit data transfer to the vector space, but do not support semantic information [1, 48]. The major advantage of the kernel-based and neural network models is accurate extract of actual relations by using additional information.

6. **Conclusion.** Researchers need to use up-to-date information to conduct biomedical research. Biomedical libraries and repositories are rich sources for obtaining up-to-date biomedical information. Since the complex nature and increasing volume of biomedical information, manual search is costly and practically impossible. Relation extraction is a very beneficial way to discover new biomedical information and manage them. Since this is a challenging task, this article proposed a classification of biomedical relation extraction challenges. The proposed classification divides the challenges into three main categories: data, implementation, and evaluation challenges. The proposed classification covers almost all existing challenges. This paper, by comparing and analyzing the techniques to address these challenges, assists researchers to eliminate the challenges and enhance techniques in future.

REFERENCES

[1] J. Li, Z. Zhang, X. Li and H. Chen, Kernel-based learning for biomedical relation extraction, *Journal of the American Society for Information Science and Technology*, vol.59, no.5, pp.756-769, 2008.

[2] J. Singh and V. Gupta, A systematic review of text stemming techniques, *Artificial Intelligence Review*, vol.48, no.2, pp.157-217, 2017.

[3] M. Moradi and N. Ghadiri, Different approaches for identifying important concepts in probabilistic biomedical text summarization, *Artificial Intelligence in Medicine*, vol.84, pp.101-116, 2018.

[4] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd Edition, New York, Prentice Hall, 2008.

[5] G. Miner, J. Elder, T. Hill and R. Nisb, *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, Academic Press, Oxford, 2012.

[6] C. C. Aggarwal and C. Zhai (eds.), *Mining Text Data*, Springer Science & Business Media, 2012.

[7] Y. Luo, Ö. Uzuner and P. Szolovits, Bridging semantics and syntax with graph algorithms – State-of-the-art of extracting biomedical relations, *Briefings in Bioinformatics*, vol.18, no.1, pp.160-178, 2017.

[8] M. Chowdhury and F. Mahbub, *Improving the Effectiveness of Information Extraction from Biomedical Text*, Ph.D. Thesis, University of Trento, 2013.

[9] P. Buitelaar and P. Cimiano, Ontology learning and population: Bridging the gap between text and knowledge, in *Frontiers in Artificial Intelligence and Applications Series (Book 167)*, IOS Press, 2008.

[10] I. S. Bedmar, *Application of Information Extraction Techniques to Pharmacological Domain: Extracting Drug-Drug Interactions*, Ph.D. Thesis, University of Carlos III de Madrid, 2010.

[11] J. R. Katukuri, Y. Xie and V. V. Raghavan, Biomedical relationship extraction from literature based on bio-semantic token subsequences, *International Journal of Functional Informatics and Personalised Medicine*, vol.3, no.1, pp.16-28, 2010.

[12] Q. C. Bui, *Relation Extraction Methods for Biomedical Literature*, Ph.D. Thesis, University of Amsterdam, 2012.

[13] Z. Huang, *Biomedical Information Extraction: Mining Disease Associated Genes from Literature*, Ph.D. Thesis, Drexel University, 2014.

[14] M. Song, W. C. Kim, D. Lee, G. E. Heo and K. Y. Kang, PKDE4J: Entity and relation extraction for public knowledge discovery, *Journal of Biomedical Informatics*, vol.57, pp.320-332, 2015.
[15] S. Pyysalo, R. Sætre, J. Tsujii and T. Salakoski, Why biomedical relation extraction results are incomparable and what to do about it, *Proc. of the 3rd International Symposium on Semantic Mining in Biomedicine (SMBM 2008)*, pp.149-152, 2008.
[16] M.-F. Moens, *Information Extraction: Algorithms and Prospects in a Retrieval Context*, Springer Science & Business Media, Dordrecht, 2006.
[17] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead and O. E. T. Center, Open information extraction from the web, *Communications of the ACM*, vol.51, no.12, pp.68-74, 2008.
[18] S. Haghani and M. R. Keyvanpour, A systemic analysis of link prediction in social network, *Artificial Intelligence Review*, 2017.
[19] C.-H. Wei, Y. Peng, R. Leaman, A. P. Davis, C. J. Mattingly, J. Li, T. C. Wiegers and Z. Lu, Assessing the state of the art in biomedical relation extraction: Overview of the BioCreative V chemical-disease relation (CDR) task, *Database – The Journal of Biological Databases & Curation*, vol.2016, 2016.
[20] Batista-Navarro and R. T. Bautista, *Information Extraction from Pharmaceutical Literature*, Ph.D. Thesis, The University of Manchester, 2014.
[21] C. Giuliano, A. Lavelli and L. Romano, Exploiting shallow linguistic information for relation extraction from biomedical literature, *The 11th Conference of the European Chapter of the Association for Computational Linguistics*, 2006.
[22] N. Konstantinova, Review of relation extraction methods: What is new out there?, *Analysis of Images, Social Networks and Texts*, 2014.
[23] N. Bach and S. Badaskar, A review of relation extraction, *Literature Review for Language and Statistics II*, vol.2, 2007.
[24] S. Singh, S. Riedel, B. Martin, J. Zheng and A. McCallum, Joint inference of entities, relations, and coreference, *Proc. of the 2013 Workshop on Automated Knowledge Base Construction*, San Francisco, CA, USA, 2013.
[25] F. Li, M. Zhang, G. Fu and D. Ji, A neural joint model for entity and relation extraction from biomedical text, *BMC Bioinformatics*, vol.18, no.1, 2017.
[26] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.
[27] A. M. Cohen and W. R. Hersh, A survey of current work in biomedical text mining, *Briefings in Bioinformatics*, vol.6, no.1, pp.57-71, 2005.
[28] J. Ding, D. Berleant, D. Nettleton and E. Wurtele, Mining MEDLINE: Abstracts, sentences, or phrases?, *Proc. of the Pacific Symposium on Biocomputing*, Hawaii, USA, 2002.
[29] H.-J. Dai, Y.-C. Chang, R. T.-H. Tsai and W.-L. Hsu, New challenges for biological text-mining in the next decade, *Journal of Computer Science and Technology*, vol.25, no.1, pp.169-179, 2010.
[30] N. P. C. Díaz and M. M. M. López, An analysis of biomedical tokenization: Problems and strategies, *Proc. of the 6th International Workshop on Health Text Mining and Information Analysis (LOUHI)*, Lisbon, Portugal, 2015.
[31] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
[32] T. Joachims, *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*, Kluwer Academic Publishers, Norwell, 2002.
[33] R. C. Bunescu and R. J. Mooney, A shortest path dependency kernel for relation extraction, *Proc. of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics*, Vancouver, Canada, 2005.
[34] E. Gonzalez and J. Turmo, Unsupervised relation extraction by massive clustering, *The 9th IEEE International Conference on Data Mining*, Miami, FL, USA, 2009.
[35] D. Zhou, D. Zhong and Y. He, Biomedical relation extraction: From binary to complex, *Computational and Mathematical Methods in Medicine*, vol.2014, p.18, 2014.
[36] B. Min, S. Shi, R. Grishman and C.-Y. Lin, Ensemble semantics for large-scale unsupervised relation extraction, *Proc. of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics*, Jeju Island, Korea, 2012.
[37] R. Baeza-Yates and A. Tiberi, Extracting semantic relations from query logs, *Proc. of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Jose, CA, USA, pp.76-85, 2007.

[38] H. Liu, Y. A. Lussier and C. Friedman, Disambiguating ambiguous biomedical terms in biomedical narrative text: An unsupervised method, *Journal of Biomedical Informatics*, vol.34, no.4, pp.249-261, 2001.

[39] Y. Xu, L. Mou, G. Li, Y. Chen, H. Peng and Z. Jin, Classifying relations via long short term memory networks along shortest dependency paths, *Proc. of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, pp.1785-1794, 2015.

[40] À. Bravo, J. Piñero, N. Queralt-Rosinach, M. Rautschka and L. I. Furlong, Extraction of relations between genes and diseases from text and large-scale data analysis: Implications for translational research, *BMC Bioinformatics*, vol.16, no.1, p.55, 2015.

[41] W. A. Baumgartner, K. B. Cohen and L. Hunter, An open-source framework for large-scale, flexible evaluation of biomedical text mining systems, *Journal of Biomedical Discovery and Collaboration*, vol.3, no.1, p.1, 2008.

[42] T. Hasegawa, S. Sekine and R. Grishman, Discovering relations among named entities from large corpora, *Proc. of the 42nd Annual Meeting on Association for Computational Linguistics*, Barcelona, Spain, p.415, 2004.

[43] S. Sirmakessis, *Text Mining and Its Applications: Results of the NEMIS Launch Conference*, Springer, 2012.

[44] J. Hakenberg, *Mining Relations from the Biomedical Literature*, Ph.D. Thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2010.

[45] A. McCallum and D. Jensen, A note on the unification of information extraction and data mining using conditional probability, relational models, *Computer Science Department Faculty Publication Series*, p.42, 2003.

[46] W. Yin, K. Kann, M. Yu and H. Schütze, Comparative study of CNN and RNN for natural language processing, *ArXiv e-prints arXiv:1702.01923*, 2017.

[47] Q. Li and H. Ji, Incremental joint extraction of entity mentions and relations, *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, USA, vol.1, pp.402-412, 2014.

[48] G. Murugesan, S. Abdulkadhar and J. Natarajan, Distributed smoothed tree kernel for protein-protein interaction extraction from the biomedical literature, *PloS One*, vol.12, no.11, 2017.