Extension of the Interaction Network Ontology for literature mining of gene-gene interaction networks from sentences with multiple interaction keywords

Arzucan Özgür[1*], Junguk Hur[2*], Yongqun He[3ξ]

[1]Department of Computer Engineering, Bogazici University, 34342 Istanbul, Turkey. arzucan.ozgur@boun.edu.tr

[2]Department of Basic Sciences, University of North Dakota, School of Medicine and Health Sciences, Grand Forks, North Dakota 58202, USA. junguk.hur@med.und.edu

[3]University of Michigan Medical School, Ann Arbor, MI 48109, USA. yongqunh@med.umich.edu

* Equal contribution, ξ Corresponding author

Abstract. The Interaction Network Ontology (INO) has been demonstrated to be valuable in providing a structured ontological vocabulary for literature mining of gene-gene interactions from biomedical literature. Our analysis of the Learning Logic in Language (LLL) challenge and vaccine datasets showed that many interactions are signaled with 2 or more interaction keywords used in combination. In this paper, we extend the INO by adding combinatory patterns of two or more literature mining keywords to related INO interaction classes. An INO-based literature mining pipeline was further developed based on SPARQL queries and SciMiner, an in-house literature mining program. The majority of the gene interaction sentences from the LLL and vaccine datasets were found to use multiple keywords to represent interaction types. A comprehensive analysis of the LLL dataset identified 27 gene regulation interaction types each associated with multiple keywords. Special patterns were discovered from the hierarchical structure of these 27 INO types.

Keywords: Interaction Network Ontology, Literature mining, Gene-gene interaction, SciMiner

1 Introduction

Literature mining methods for extracting interactions among biomedical entities including genes and proteins typically formulate the problem as a binary classification task, where the goal is to identify the pairs of entities that are stated to interact with each other in text [1, 2]. Several different methods have been proposed to tackle this problem ranging from relatively simpler co-occurrence based methods [3] to more complex methods that make use of the syntactic analysis of the sentences [4-6], mostly in conjunction with machine learning methods [7-9].

Besides, extracting the existence of interactions among biomolecules, identifying the types of these interactions are vital for a better understanding of the underlying biological processes and for the creation of more detailed and structured models of interactions such as biological pathways. In order to improve the performance of extracting biomolecular events and entities with varying roles (e.g. theme, causes, and etc.), the literature mining community has established collaborative but competitive challenges such as the BioNLP Shared Tasks on Event Extraction [10, 11].

The types of interactions (or events) among biomolecules are in general signaled with specific interaction keywords (trigger words). For example, the interaction keyword “up-regulates” signals an interaction of type positive regulation, whereas the keyword “inhibits” signals an interaction of type negative regulation. We have previously collected over 800 interaction keywords, which we used with support vector machines (SVM) [12] to classify pairs of genes or proteins as interacting or not [13]. We have also shown that the usage of ontologies, such as the Vaccine Ontology (VO), can enhance the mining of gene-gene interactions under a specific domain, for example, the vaccine domain [13] or vaccine induced fever domain [14]. The over 800 interaction-associated keywords provide us tags for mining interactive relation between two genes/proteins.

However, this is basically a binary result of an interaction between two molecules or entities. To extend from the binary yes/no results, we further hypothesized that the ontological classification of these and more keywords would allow us to further identify and classify the types of interactions (e.g., *regulation of transcription*). Based on this hypothesis, we ontologically classified these interaction-related keywords in the Interaction Network Ontology (INO), a community-driven ontology of biological interactions, pathways, and networks [13, 15]. INO classifies and represents different levels of interaction keywords used for literature mining of genetic interaction networks. Its development follows the Open Biological/Biomedical Ontology (OBO) Foundry ontology development principles (*e.g.*, openness and collaboration) [16]. We also showed the utility of using INO and a modified Fisher's exact test to analyze significantly over- and under-represented enriched gene-gene interaction types among the vaccine-associated gene-gene interactions extracted using all PubMed abstracts [15]. Our study showed that INO would provide a new platform for efficient mining and analysis of topic-specific gene interaction networks.

Nevertheless, there still exist two more challenges in regards to the INO-based classification method. The first is that the INO-based data standardization is not easy for tool developers to deploy. The second is that current INO-based classification focuses on the classification of interaction types signaled with one keyword in a sentence. However, it is quite frequent that two or more interaction-related keywords collectively signal an interaction type in a sentence. Such combinations of keywords were discussed in the Discussion section of our previous paper without further exploration [9]. In this article, we report our effort to address these two challenges, including the further development and standardization of INO-based classification method and INO-based classification of multiple interaction keywords representing interaction types in sentences. We have also applied these in two use case studies.

2 Methods

2.1 INO ontology modeling and editing

INO was formatted using the Description Logic (DL) version of the Web Ontology Language (OWL2) [17]. The Protégé OWL Editor [18] was used to add and edit INO specific terms. To identify INO interaction types containing two or more keywords used for literature mining of gene-gene interactions, we manually annotated sentences from selected PubMed abstracts as described later and ontologically modeled each interaction types in INO.

2.2 SPARQL query of the INO subset of interaction keywords used for literature mining of gene-gene interactions

The Ontobee SPARQL endpoint (http://www.ontobee.org/sparql) was used to obtain the literature mining keywords by querying the INO ontology content stored in the He Group RDF triple store [19]. This triple store was developed based on the Virtuoso system. The data in the triple store can be queried using the standard Virtuoso SPARQL queries.

2.3 OntoFox extraction of an INO subset of interaction terms that can be classified by two or more keywords in one sentence

All the INO terms containing literature mining keywords composed of multiple words were identified, and a subset of INO containing these terms and related terms was extracted using the OntoFox tool [20].

2.4 Gold standard LLL data analysis

In order to analyze the characteristics of interactions which are signaled with more than one keywords, we manually annotated the gene/protein interaction data set from the Learning Logic in Language (LLL) Challenge [21] for the interaction types and the keywords that signal them. Two experts reviewed the output of the single-word interaction keywords identified by SciMiner, then carefully examined for multi-keyword interactions. Discrepancy was resolved by agreement between two experts.

2.5 Vaccine gene-gene interaction literature mining use case

In our previous paper, we used ontology-based SciMiner [22] to extract and analyze gene-gene interactions in the vaccine domain using all PubMed abstracts [15]. In this paper, we further annotated those sentences including two or more interaction-related

keywords for annotating gene-gene interactions. The results were then systematically analyzed.

3 Results

3.1 INO representation of interaction terms and literature mining keywords

As defined previously, INO is aligned with the upper level Basic Formal Ontology (BFO) [16]. In INO, a biological interaction is defined as a processual entity that has two or more participants (*i.e.*, interactors) that have an effect upon one another. To support ontology reuse and data integration, INO imports many terms from existing ontologies [15], such as the Gene Ontology (GO) [23], and PSI Molecular Interactions (PSI-MI) [24]. As of August 12th, 2015 INO has 571 terms including 153 terms with INO prefix and 418 terms imported from 10 other ontologies (http://www.ontobee.org/ontostat.php?ontology=INO).

In the present study, we focused on the branch of gene-gene regulation, particularly gene expression regulation. For the INO term 'gene expression regulation', the input interactor is a gene, the output interactor is a gene product including a RNA or protein, and the regulator is typically a protein. There exist different subtypes of 'gene expression regulation', for example, positive or negative regulation of gene expression, and regulation of transcription or translation.

Fig. 1 shows an example of how INO defines the term '*regulation of transcription'*. In addition to its text definition, INO also generates many logic axioms. An equivalent class definition of the term is defined: *regulates some 'gene transcription'*, where 'regulates' is an object property (or called relation) and 'gene transcription' is a gene expression process that transcribes a gene to RNA. In addition to asserted axioms, many axioms are also inherited for the parents of the term '*regulation of transcription'* (Fig. 1).

Various subtypes of 'regulation of transcription' exist. For example, there are different subtypes of positive or negative regulation of transcription. One commonly seen subtype of regulation of transcription is via a promoter. A promoter is a region of DNA located near the transcription start site of a gene, and the binding between a promoter sequence and a transcription factor is required to initiate a transcription. The phrase of a sentence "sigmaB- and sigmaF-dependent promoters of katX" [25] indicates that sigmaB and sigmaF regulate katX through the katX promoters.

Some interactions are characterized with a single interaction keyword. For example, in the sentence "Dephosphorylation of SpoIIAA-P by SpoIIE is strictly dependent on the presence of the bivalent metal ions Mn^{2+} or Mg^{2+}" [26], the type of interaction between SpoIIAA-P and SpoIIE is dephosphorylation reaction, which is characterized with the interaction keyword "dephosphorylation".

On the other hand, there are also more complex interactions that are characterized with two or more interaction keywords. Consider the sentence "In the mother cell compartment of sporulating cells, expression of the sigE gene, encoding the earlier-acting sigma factor, sigmaE, is negatively regulated by the later-acting sigma factor, sigmaK" [27]. The relation between the SigE and SigmaK genes is characterized with

the interaction keywords “expression” and “negatively regulated”. The type of relation is negative regulation of gene expression. SigmaK negatively regulates the expression of SigE. Such relations are represented as complex events in the Genia event corpus [28] used in the BioNLP Shared Tasks, where the expression of SigE is considered as the first event and the negative regulation of this event by the SigmaK gene is considered as the second event. In contrast, INO represents such complex events using a different strategy as described below.

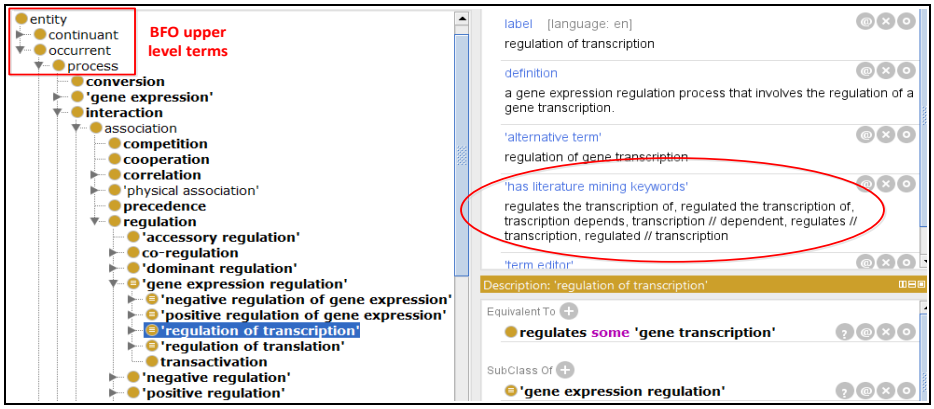


Fig. 1. INO representation of ‘regulation of transcription’. Axioms are defined for this class or inherited from its parent terms including its direct parent term ‘gene expression regulation’. As shown in the figure, INO is aligned with BFO as its upper level ontology. The annotated literature mining keywords for the INO class are highlighted with oval circle.

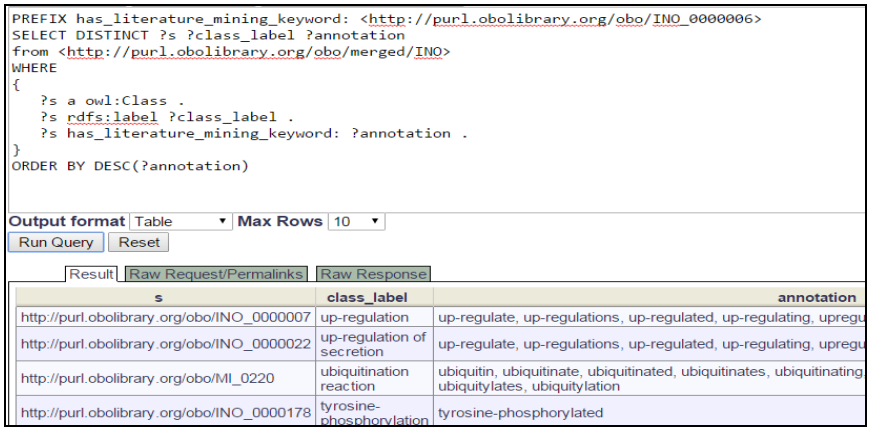
3.2 INO-based standardization of literature mining of gene-gene interactions

As shown in Fig. 1, the literature mining keywords for an INO term are defined as an annotation using the annotation property ‘has literature mining keywords’. To provide a reproducible strategy of representing the literature mining keywords, we used the sign “//” to separate two keywords, which indicates that these two keywords do not have to be next to each other in a sentence (Fig. 1). For example, many keywords are added for the INO term ‘regulation of transcription’ (INO_0000032), including *“transcription // dependent, regulated // transcription, requires // transcription”*. These terms mean that the two keywords such as “requires” and “transcription” can be separate in one sentence, for example, “sspG transcription also requires the DNA binding protein GerE” [29].

Different ways can be used to get the information of keywords. One way is to query INO using SPARQL. To show how we can quickly obtain the INO literature mining keywords, we have shown the usage of a SPARQL query to automatically generate the INO subset for literature mining (Fig. 2).

Before the SPARQL can be executed, the INO ontology content should be first deposited in RDF triple store. Indeed, the INO is included in the Hegroup RDF Triple

Store [19], which is the default RDF triple store for the ontologies in the Open Biological and Biomedical Ontologies (OBO) library (http://www.obofoundry.org/).



```
PREFIX has_literature_mining_keyword: <http://purl.obolibrary.org/obo/INO_0000006>
SELECT DISTINCT ?s ?class_label ?annotation
from <http://purl.obolibrary.org/obo/merged/INO>
WHERE
{
   ?s a owl:Class .
   ?s rdfs:label ?class_label .
   ?s has_literature_mining_keyword: ?annotation .
}
ORDER BY DESC(?annotation)
```

Output format: Table Max Rows: 10

Run Query Reset

Result | Raw Request/Permalinks | Raw Response

s	class_label	annotation
http://purl.obolibrary.org/obo/INO_0000007	up-regulation	up-regulate, up-regulations, up-regulated, up-regulating, upregu
http://purl.obolibrary.org/obo/INO_0000022	up-regulation of secretion	up-regulate, up-regulations, up-regulated, up-regulating, upregu
http://purl.obolibrary.org/obo/MI_0220	ubiquitination reaction	ubiquitin, ubiquitinate, ubiquitinated, ubiquitinates, ubiquitinating, ubiquitylates, ubiquitylation
http://purl.obolibrary.org/obo/INO_0000178	tyrosine-phosphorylation	tyrosine-phosphorylated

Fig. 2. SPARQL query of interaction keywords for INO interaction class terms. This query was performed using the Ontobee SPARQL query website (http://www.ontobee.org/sparql/). This figure is a screenshot of the SPARQL code and a portion of the results.

3.3 Incorporation of INO literature mining system to a software program

SciMiner [22] is our in-house literature mining software program for identifying interactions among genes/proteins/vaccines and analyzing their biological significance. We recently incorporated INO into SciMiner and demonstrated its successful application to the identification of specific interaction types significantly associated with gene-gene interactions in the context of vaccine [15]. SciMiner can also be utilized in identifying and modeling two interaction keywords, which will be eventually used to improve the final literature-mined interaction network.

Fig. 3 illustrates the overall workflow of INO modeling and its application in literature mining for gene-interaction analysis. Briefly, the INO modeling procedure aims at identifying and classifying the interaction patterns of two INO keywords. Sentences with potential multiple interaction keywords (from gold standard sets) are first scanned to identify individual single-word INO keywords and biological entities. For any sentences with two or more interaction keywords identified, combinations of two keywords are queried against the dictionary of keywords associated with existing INO interaction classes. For any two keyword patterns that are not included in the current dictionary, INO experts manually examine the sentences and two-keyword patterns to confirm their valid interactions, update the INO annotations accordingly with new entries, and upload the updated INO to a RDF triple store. Then, SPARQL can be used to create new INO keyword dictionary for literature mining.

Once INO-interaction keyword dictionary is established, it can be applied to constructing interaction networks of biological entities from any set of biomedical literature using SciMiner (as shown in the right part of Fig. 3). Briefly, SciMiner accepts PubMed abstracts or sentences as input. After internal preprocessing of the

abstracts/sentences, SciMiner identifies biological entities such as gene/protein or any ontologies (e.g. vaccine ontology) as well as single-word level INO terms. From the sentences with at least two identified entities and one or more INO terms are used in the interaction modeling. Sentences with two interaction keywords will further go through multi-keyword interaction modeling, and a final interaction network will be generated and subjected to down-stream functional analysis. A standalone command-line based SciMiner, rather than the web version, was used in the current study and the complete standalone pipeline will be available upon completion of the development.

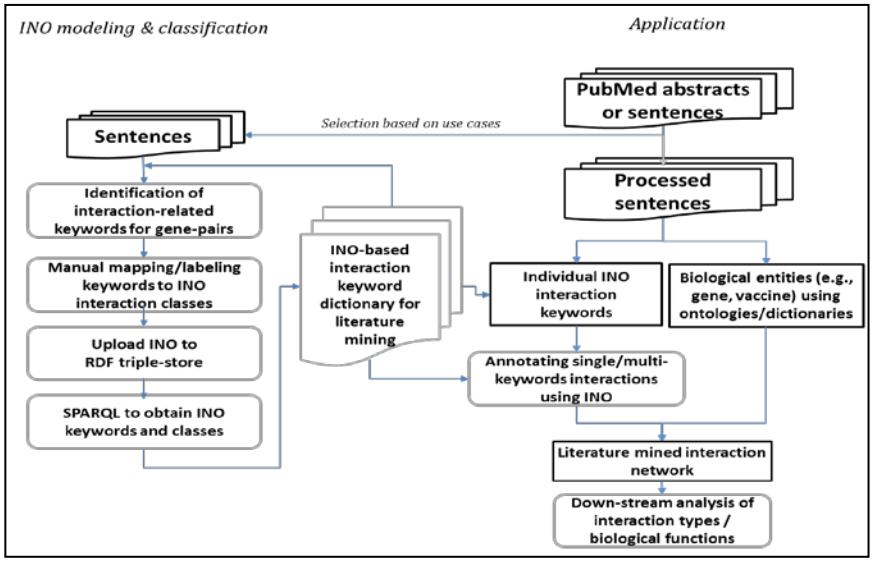


Fig 3. INO modeling and application workflow.

3.4 Annotation of the LLL data set for interaction types

The LLL data set contains gene/protein interactions in *Bacillus subtilis*, which is a model bacterium [6]. The data set contains 77 sentences and 164 pairs of genes/proteins that are described as interacting in these sentences. As an example, consider the sample sentence "Transcriptional studies showed that nadE is strongly induced in response to heat, ethanol and salt stress or after starvation for glucose in a sigma B-dependent manner." [30] from the LLL data set. The interacting protein/gene pairs (i.e., nadE and sigma B) have already been annotated in the data set. Given the sentence and the interacting pair of proteins/genes, we annotated the type of relation between them and the interaction keywords signaling this relation. The type of interaction between nadE and Sigma B is "positive regulation of gene transcription", in other words Sigma B positively regulates the transcription of nadE. The relevant interaction keywords are "transcriptional", "induced", and "dependent". Our

interaction type and keyword annotation of the data set will be made publicly available for future studies.

Our annotation of the LLL data set for interaction types showed that many regulatory relations between gene/protein pairs are represented with multiple keywords. While the interactions among 43 pairs of genes/proteins were represented with a single keyword, the interactions among 116 pairs were signaled using multiple keywords. These interactions correspond to 27 different classes of *regulation* in INO. Fig. 4 shows the hierarchical structure of these 27 classes, their related classes, and the number of gene/protein pairs in the sentences identified for each class.

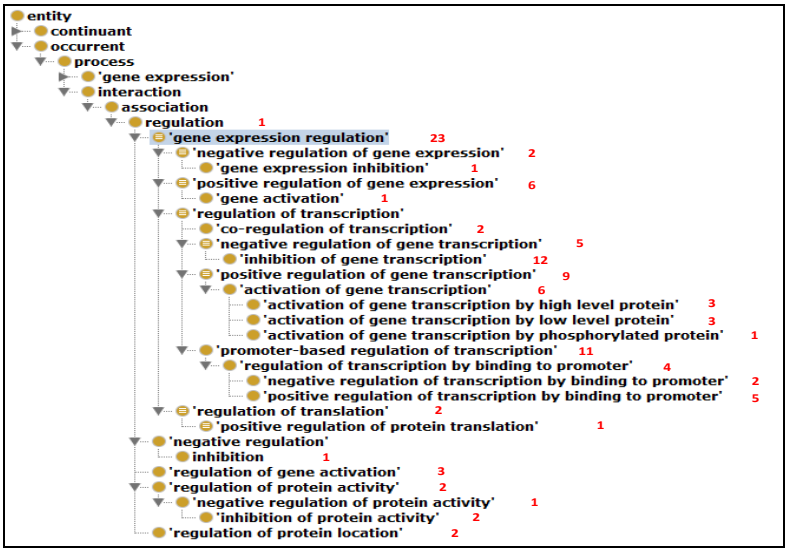


Fig. 4. The hierarchical display of 27 interaction classes and the numbers of sentences associated with these classes in the LLL Data Set. OntoFox was used to generate the INO subset, and the Protégé OWL editor was used to visualize the hierarchical structure.

Our study of the LLL dataset indicated that the majority of the sentences are related to the gene expression regulation, especially in the area of transcriptional regulation. More sentences describe positive regulation rather than negative regulation. An interesting observation is the presence of many sentences focusing on the domain of promoter-based regulation of transcription (Fig. 1). In addition to gene expression regulation, this data set also includes other types of gene regulation, for example, regulation of protein location, regulation of gene activation, and regulation of protein activity. It is noted that protein activity is different from gene expression. Protein activity depends on many factors other than expression, such as correct folding of the protein and the presence of any required cofactors.

Our analysis showed that most multi-keyword interactions are represented with two keywords. Consider the interaction between KinC and Spo0A~P in the sentence “KinC and KinD were responsible for Spo0A~P production during the exponential phase of growth in the absence of KinA and KinB” [31]. This sentence states that

KinC is responsible for Spo0A~P production. The interaction type between these genes is classified as “regulation of translation” in INO. The two keywords signaling this interaction are “responsible” and “production”. The keyword “responsible” signals that this is an interaction of type “regulation”, whereas the keyword “production” signals that this is a specific type of regulation, namely “regulation of translation”. We can consider “responsible” as the main type signaling keyword and “production” as the secondary (sub) type signaling keyword.

There are also more complex interactions, which are represented with more than two keywords. For example, in the sentence “A low concentration of GerE activated cotB transcription by final sigma(K) RNA polymerase, whereas a higher concentration was needed to activate transcription of cotX or cotC.” [32], the interaction between GerE and cotB is signaled with the three keywords “low concentration”, “activated”, and “transcription”. The type of interaction corresponds to the INO class “activation of gene transcription by low level protein”. In another sentence “sigmaH-dependent promoter is responsible for yvyD transcription” [33], four keywords are used: “dependent”, “promoter”, “responsible”, and “transcription”. Such a complex interaction is labeled as “promoter-based regulation of transcription” in INO.

3.5 Analysis of vaccine-based gene-gene interaction literature mining results

Our previous INO-based literature mining study used an INO-based SciMiner program to identify many gene-gene interactions in the vaccine domain using all PubMed abstracts [15]. A statistical method based on the results was also developed to classify significantly over- and under-represented interaction types. Our manual examination of randomly selected 50 sentences identified by SciMiner, a small portion of the whole vaccine corpus, suggested that similar to the LLL data set, over 50% of sentences use two or more keywords to represent specific gene-gene interaction types.

4 Discussion

In this paper, we investigated the interaction types that are characterized with multiple keywords used in combination. The main contributions are: (1) Extending INO by modeling interaction types (classes) each signaled with multiple keywords in literature sentences and adding many new terms by analyzing the LLL and vaccine data sets, (2) Standardizing INO-based literature mining for easy use and testing by future studies. (3) Characterizing and demonstrating multi-keyword interaction type ontology modeling of literature sentences by analyzing the LLL and vaccine-gene interaction data sets.

Multi-keyword interactions have been represented as complex events in the Genia corpus [28], which has also been used in the BioNLP Shared Tasks on Event Extraction. In this representation, in order to identify the complex events, first the simple events (e.g. gene expression, regulation) signaled with individual keywords need to be identified. Next, the simple events are combined to form a complex event.

For instance, given a sentence that states that gene A regulates the expression of gene B, the expression of gene B is represented as Event 1 (i.e., expression of gene B), and Event 2 is a complex event where gene A regulates Event 1. Therefore, we could infer a possible relation between gene A and gene B, by the association of Event 1 – gene B – Event 2 – gene A. Such recognition of the gene A-B interaction is indirect, and may become even more complex when multiple events (with multiple keywords) are applied. Compared to the Genie approach, INO provides a more fine-grained and direct classification of interaction types and can directly model the relation between two biomolecules (e.g., genes or proteins). For instance, the interaction between gene A and gene B in the above example is directly modeled as the interaction type "regulation of gene expression" in INO.

The Gene Regulation Ontology (GRO) [34] models complex gene regulatory events similarly to INO. GRO has recently been used in the Corpus Annotation with Gene Regulation Ontology Task in the 2013 edition of BioNLP Shared Task [35]. The domains of GRO and INO differ. GRO focuses on only gene regulations. However, INO targets the broader scope of interactions and interaction networks. Similar to INO, GRO is also aligned with the Basic Formal Ontology (BFO) and many other ontologies such as the Gene Ontology (GO). However, for the ontology alignments, GRO uses its own identifiers and references back to the original ontologies; in contrast, INO directly imports related terms from other ontologies. Technical representations of entities in INO and GRO also differ in many aspects. Compared to GRO, one of the main advantages of INO is that the interaction types and sub-types are associated with manually compiled comprehensive lists of literature mining keywords. These keywords can be incorporated in dictionary-based or statistical taggers for tagging the interaction keywords in text, which can then be used to map the interactions to their corresponding types in INO.

Future work includes automatic identification and modeling of novel two keyword interactions by SciMiner, and a new notation of multi-keyword interactions using regular expressions to be more systematic rather than the current '//'-based strategy. In this paper we demonstrated our strategy of integrating INO with the SciMiner tagger for ontology-based literature mining. Currently, the integrated INO-SciMiner works as a standalone package, and it can be easily incorporated into other literature mining pipelines, if desired. The current SciMiner system can identify gene/protein and vaccine, but is being upgraded to be able to identify other entities such as drug, tissue, and *etc*., thus, the future version of INO-integrated SciMiner can be applied to not only the typical gene-gene interaction, but also other interactions such as gene-drug interaction, drug-chemical, drug-tissue and various types of interaction.

Acknowledgments. This research was supported by grant R01AI081062 from the US NIH National Institute of Allergy and Infectious Diseases (to YH) and Marie Curie FP7-Reintegration-Grants within the 7th European Community Framework Programme (to AO).

References

1. Arighi, C.N., Lu, Z., Krallinger, M., Cohen, K.B., Wilbur, W.J., Valencia, A., Hirschman, L., Wu, C.H.: Overview of the BioCreative III Workshop. BMC Bioinformatics. 12 Suppl 8, S1 (2011)
2. Krallinger, M., Leitner, F., Rodriguez-Penagos, C., Valencia, A.: Overview of the protein-protein interaction annotation extraction task of BioCreative II. Genome Biol. 9 Suppl 2, S4 (2008)
3. Jelier, R., Jenster, G., Dorssers, L.C., van der Eijk, C.C., van Mulligen, E.M., Mons, B., Kors, J.A.: Co-occurrence based meta-analysis of scientific texts: retrieving biological relationships between genes. Bioinformatics. 21, 2049-2058 (2005)
4. Fundel, K., Kuffner, R., Zimmer, R.: RelEx--relation extraction using dependency parse trees. Bioinformatics. 23, 365-371 (2007)
5. Daraselia, N., Yuryev, A., Egorov, S., Novichkova, S., Nikitin, A., Mazo, I.: Extracting human protein interactions from MEDLINE using a full-sentence parser. Bioinformatics. 20, 604-611 (2004)
6. Temkin, J.M., Gilder, M.R.: Extraction of protein interaction information from unstructured text using a context-free grammar. Bioinformatics. 19, 2046-2053 (2003)
7. Airola, A., Pyysalo, S., Bjorne, J., Pahikkala, T., Ginter, F., Salakoski, T.: All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning. BMC Bioinformatics. 9 Suppl 11, S2 (2008)
8. Tikk, D., Thomas, P., Palaga, P., Hakenberg, J., Leser, U.: A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature. PLoS Comput. Biol. 6, e1000837 (2010)
9. Erkan, G., Özgür, A., Radev, D.R.: Semi-supervised classification for extracting protein interaction sentences using dependency parsing. In: EMNLP-CoNLL, pp. 228-237. (2007)
10. Kim, J.D., Ohta, T., Pyysalo, S., Kano, Y., Tsujii, J.i.: Overview of BioNLP'09 shared task on event extraction. In: Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task, pp. 1-9. Association for Computational Linguistics, (2009)
11. Kim, J.D., Nguyen, N., Wang, Y., Tsujii, J., Takagi, T., Yonezawa, A.: The Genia Event and Protein Coreference tasks of the BioNLP Shared Task 2011. BMC Bioinformatics. 13 Suppl 11, S1 (2012)
12. Joachims, T.: Making large-scale support vector machine learning practical. In: B. Schölkopf, C.J.B., and A. J. Smola, Eds. (ed.) Advances in Kernel Methods: Support Vector Learning, pp. 169-184. MIT Press, Cambridge, MA. (1999)
13. Ozgur, A., Xiang, Z., Radev, D.R., He, Y.: Mining of vaccine-associated IFN-gamma gene interaction networks using the Vaccine Ontology. Journal of Biomedical Semantics. 2 Suppl 2, S8 (2011)
14. Hur, J., Ozgur, A., Xiang, Z., He, Y.: Identification of fever and vaccine-associated gene interaction networks using ontology-based literature mining. Journal of Biomedical Semantics. 3, 18 (2012)
15. Hur, J., Ozgur, A., Xiang, Z., He, Y.: Development and application of an interaction network ontology for literature mining of vaccine-associated gene-gene interactions. Journal of Biomedical Semantics. 6, 2 (2015)
16. Grenon, P., Smith, B.: SNAP and SPAN: Towards Dynamic Spatial Ontology. Spatial Cognition and Computation. 4, 69-103 (2004)
17. http://www.w3.org/TR/2009/REC-owl2-overview-20091027/
18. http://protege.stanford.edu/

19. Xiang, Z., Mungall, C., Ruttenberg, A., He, Y.: Ontobee: A linked data server and browser for ontology terms. In: The 2nd International Conference on Biomedical Ontologies (ICBO), pp. 279-281. CEUR Workshop Proceedings, (2011)
20. Xiang, Z., Courtot, M., Brinkman, R.R., Ruttenberg, A., He, Y.: OntoFox: web-based support for ontology reuse. BMC Research Notes. 3:175, 1-12 (2010)
21. Nedellec, C.: Learning language in logic-genic interaction extraction challenge. In: Proceedings of the 4th Learning Language in Logic Workshop (LLL05). (2005)
22. Hur, J., Schuyler, A.D., States, D.J., Feldman, E.L.: SciMiner: web-based literature mining tool for target identification and functional enrichment analysis. Bioinformatics. 25, 838-840 (2009)
23. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics. 25, 25-29 (2000)
24. Isserlin, R., El-Badrawi, R.A., Bader, G.D.: The Biomolecular Interaction Network Database in PSI-MI 2.5. Database (Oxford). 2011, baq037 (2011)
25. Petersohn, A., Engelmann, S., Setlow, P., Hecker, M.: The katX gene of Bacillus subtilis is under dual control of sigmaB and sigmaF. Molecular Genetics and Genomics. 262, 173-179 (1999)
26. Schroeter, R., Schlisio, S., Lucet, I., Yudkin, M., Borriss, R.: The Bacillus subtilis regulator protein SpoIIE shares functional and structural similarities with eukaryotic protein phosphatases 2C. FEMS Microbiology Letters. 174, 117-123 (1999)
27. Zhang, B., Struffi, P., Kroos, L.: sigmaK can negatively regulate sigE expression by two different mechanisms during sporulation of Bacillus subtilis. Journal of Bacteriology. 181, 4081-4088 (1999)
28. Kim, J.D., Ohta, T., Tsujii, J.: Corpus annotation for mining biomedical events from literature. BMC Bioinformatics. 9, 10 (2008)
29. Bagyan, I., Setlow, B., Setlow, P.: New small, acid-soluble proteins unique to spores of Bacillus subtilis: identification of the coding genes and regulation and function of two of these genes. Journal of Bacteriology. 180, 6704-6712 (1998)
30. Antelmann, H., Schmid, R., Hecker, M.: The NAD synthetase NadE (OutB) of Bacillus subtilis is a sigma B-dependent general stress protein. FEMS Microbiology Letters. 153, 405-409 (1997)
31. Jiang, M., Shao, W., Perego, M., Hoch, J.A.: Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis. Molecular Microbiology. 38, 535-542 (2000)
32. Ichikawa, H., Kroos, L.: Combined action of two transcription factors regulates genes encoding spore coat proteins of Bacillus subtilis. Journal of Biological Chemistry. 275, 13849-13855 (2000)
33. Drzewiecki, K., Eymann, C., Mittenhuber, G., Hecker, M.: The yvyD gene of Bacillus subtilis is under dual control of sigmaB and sigmaH. Journal of Bacteriology. 180, 6674-6680 (1998)
34. Beisswanger, E., Lee, V., Kim, J.J., Rebholz-Schuhmann, D., Splendiani, A., Dameron, O., Schulz, S., Hahn, U.: Gene Regulation Ontology (GRO): design principles and use cases. Stud Health Technol Inform. 136, 9-14 (2008)
35. Kim, J.D., Kim, J.J., Han, X., Rebholz-Schuhmann, D.: Extending the evaluation of Genia Event task toward knowledge base construction and comparison to Gene Regulation Ontology task. BMC Bioinformatics. 16, S3 (2015)