Biological Ontology Enhancement with Fuzzy Relations: A Text-Mining Framework

Muhammad Abulaish
Department of Mathematics
Jamia Millia Islamia(A Central Univ.)
Jamia Nagar, New Delhi – 25, India
mdabulaish@rediffmail.com

Lipika Dey
Department of Mathematics
Indian Institute of Technology
Hauz Khas, New Delhi – 16, India
lipika@maths.iitd.ernet.in

Abstract

Domain ontology can help in information retrieval from documents. But ontology is a pre-defined structure with crisp concept descriptions and inter-concept relations. However, due to the dynamic nature of the document repository, ontology should be upgradeable with information extracted through text mining of documents in the domain. This also necessitates that concepts, their descriptions and inter-concept relations should be associated with a degree of fuzziness that will indicate the support for the extracted knowledge according to the currently available resources. Supports may be revised with more knowledge coming in future. This approach preserves the basic structured knowledge format for storing domain knowledge, but at the same time allows for update of information. In this paper, we have proposed a mechanism which initiates text mining with a set of ontological concepts, and thereafter extracts fuzzy relations through text mining. Membership values of relations are functions of frequency of co-occurrence of concepts and relations. We have worked on the GENIA corpus and shown how fuzzy relations can be further used for guided information extraction from MEDLINE documents.

Keywords: Biological Information extraction, Text mining, Fuzzy ontology, Fuzzy relation.

1. Introduction

The field of Molecular Biology has witnessed a phenomenal growth in research activities in the recent past. Consequently there is an increasing demand for automatic Information Extraction (IE) schemes to extract contents from scientific documents and store them in a database to locate information from this huge unstructured collection intelligently [9]. Existing search engines often fail to do this efficiently, since most of them rely on a precise articulation of query terms to perform well. But, generally these terms are often difficult to coin for a user in an unaided way. Besides, knowledge that is deeply embedded in natural language is difficult to extract using simple pattern matching techniques.

Ontology-guided information extraction and query processing mechanism can help users to access information stored within unstructured or semi-structured text documents effectively. Ontology represents domain knowledge in a structured form and is increasingly being accepted as the key technology wherein key concepts and their inter-relationships are stored to provide a shared and common understanding of a domain across applications [1]. Ontology is a conceptualization of a domain into a human understandable, but machine-readable format consisting of classes, attributes, relationships and axioms and also entities [2]. Ontology also supports taxonomy and other association relations between classes. Since ontology describes a domain of interest in an unambiguous way, ontology-based Information Extraction (IE) schemes can help in alleviating a wide variety of natural language ambiguities present in a given domain and still perform effective semantic analysis of texts. However, ontology based text document processing, though promising, has not been exploited fully.

One of the bottlenecks in developing ontology-based text processing systems stems from the fact that the conceptual formalism supported by typical ontology may not be sufficient to represent uncertainty information that is commonly found in many application domains [8]. Generally in an ontology, inter-concept relations are crisp and there is no provision for defining the strength of a relation. However, as Wallace and Avrithis [14] states, relations among real life entities are always a matter of degree, and are, therefore, best modeled using fuzzy relations. Again, though many relations can be theoretically feasible, it is observed that entities participating in a relation, decides the significance of a relation in a domain. For example, Sekimizu and Tsujii [12] state that some of the most commonly occurring, hence useful relations in the domain of Molecular Biology

are *activate, bind, interact, regulate, encode, signal* and *function*. In addition to these, we further observe that the relational verb "bind" can be associated with different biological entities with different degrees of strength, where strength of a relation can be directly proportional to the frequency of occurrence of the specific associations in a standard collection of documents. For example we have observed that the relation "inhibitor of" occurs more frequently between *protein molecule* and *protein complexes* while the same relation is rare, though not impossible, between two *protein molecules*. Most of the existing biological ontologies like GENIA ontology, GENE ontology etc. represent concepts and taxonomical relationships like "is-a" or "part-of" and do not capture relations with variable degree of associations. Determining these relations and their strengths is a challenging problem required for designing efficient document retrieval schemes, for which deeper content analysis is required than just looking for presence or absence of concepts.

In this paper, we propose an intelligent information extraction system with twin objectives. Firstly, the system uses an existing biological ontology (GENIA in this case) for intelligent information extraction from text documents. Secondly, the extracted information is used to enhance existing crisp ontology structures to fuzzy ontology structures, which can accommodate inter-concept relationships with varying degrees of strength. The ontology-based text mining system employs named entity recognition in conjunction with NLP techniques to mine *fuzzy relations* linking ontology concepts from text documents. The definition of a fuzzy relation as defined in [10] is produced below. A fuzzy relation defined in the Cartesian products of the crisp sets U_1, U_2, , U_n, is a fuzzy set R such that

$$R = \{((U_1, U_2, \ldots, U_n), \mu_R (U_1, U_2, \ldots, U_n)) \mid (U_1, U_2, \ldots, U_n) \in U_1 \times U_2 \times \ldots \times U_n\}$$

Where $\mu_R: U_1 \times U_2 \times \ldots \times U_n \rightarrow [0,1]$ and the crisp Cartesian product $U \times V$ is defined as:

$$U \times V = \{(u, v) \mid u \in U \text{ and } v \in V\}$$

The proposed system alleviates several problems associated with biological entity extraction from text documents like irregular naming conventions used by the biologists, extensive crossover in vocabulary between classes, existence of synonymy, etc. through appropriate entity-concept mapping, which can be identified through natural language processing. The information extracted through deep text mining is not only used for ontology enhancement but is also stored in a structured data base, the design of which is also guided by the underlying ontology structure. The system helps users pose intelligent queries in an ontology-guided manner over the structured database. Queries can be formulated at various levels of specificity. For example, a user may pose a query containing concepts with different degrees of specification like "Which *protein molecules* are **activated** by NF-Kappa B?", in which NF-Kappa B is a specific entity which is of type protein molecule, which is a more generic concept.

1.1. The GENIA Ontology

Our work uses the GENIA[1] ontology designed by Tateisi *et al.* to provide the base collection of molecular biological entity types and relationships among them [4]. We show how the existing GENIA ontology can be enhanced using information extracted from tagged MEDLINE documents. The existing GENIA ontology describes biological substances and sources (substance locations) in which substances are classified according to their chemical characteristics rather than their biological roles, since chemical classification of substances is quite independent of the biological context in which it appears, and therefore more stably defined. The biological role of a substance may vary depending on the biological context. Therefore, in this model substances are classified as proteins, DNA, RNA etc. They are further classified into families, complexes, individual molecules, subunits, domains and regions. The sources are classified into natural and cultured sources that are further classified as an organism (human), a tissue (liver), a cell (leukocyte), a sub-location of a cell (membrane or a cultured cell line (HeLa)). Organisms are further classified into multi-cell organisms, viruse, and mono-cell organisms other than virus.

In this work we have worked with 2000 tagged MEDLINE abstracts, included in the GENIA corpus. The tags in these documents correspond to GENIA ontology classes. We have applied our framework on this corpus to extract and incorporate relevant Fuzzy relational concepts describing interactions between the biological substances, functions of the biological entities etc. into the GENIA ontology.

The rest of the paper is organized as follows. We present a brief overview of some related works on ontology-based text processing systems and fuzzy ontology generation in section 2. Our framework for ontology-based text mining system to extract fuzzy relations text documents is explained in section 3. The applicability of the fuzzy ontology structure is presented in section 4. Finally, we conclude the paper with some future works in section 5.

---

[1] http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/topics/ Corpus/ genia-ontology.html

2. Related Work

The use of ontological models to access and integrate large knowledge repositories in a principled way has an enormous potential to enrich and make accessible unprecedented amounts of knowledge for reasoning. Ontology representation languages like DAML+OIL and OWL are based on Description Logics (DLs) thus enabling the knowledge representation systems to provide reasoning support as well [3]. A number of systems exist to help in the process of construction of domain ontologies, of which the most famous one is the Protégé[2].

Liddle *et al.* [5] have developed an integrated tool that helps domain experts to create domain ontology. The created ontology is used to extract data from web documents that are thereafter stored in structured form. Velardi *et al.* in [13] have developed a text-mining tool to identify, define, and enter concept descriptions into ontology structures. Li and Zhong [6] presents an ontology-based abstract web mining model to gather information from multiple Web resources, based on user profiles. All the above work were targeted at general-purpose ontology building. Muller et al. [7] have developed a system called "Textpresso" which provides facilities for ontology based information retrieval from Biological documents. This system stores the extracted information and facilitates semantic querying rather than simple keyword based search. Though, nearest to our system conceptually, Textpresso does not consider ontology enhancement.

Creation of Fuzzy ontology structures have also received a lot of attention in recent times. Wallace and Avrithis [14] introduce fuzzy membership of concepts in existing semantic relations. They have also utilized such relations to estimate the context of the document, the context of the user and the context of the query, for the purpose of intelligent information retrieval. In their work, a few commonly encountered semantic relations are identified and their combinations are used to generate fuzzy, quasi-taxonomic relations. Traditionally, fuzzy ontology is generated and used in text retrieval [15], in which membership values are used to evaluate the similarities between concepts on a concept hierarchy. In this work, abstracts of papers from several IEEE Transactions have been manually typed and tagged based on their title, authors, publication date, abstract body, and author supplied keywords. In addition, the system extracts some keywords from the abstract body and then a fuzzy ontology is built on the collection of keywords. However manual generation of fuzzy ontology from a predefined concept hierarchy is a difficult and tedious task that often requires expert intervention. Quan *et al.* in [8] have

proposed an automatic fuzzy ontology generation framework on uncertainty data stored in a database. They have incorporated fuzzy logic into formal concept analysis to handle uncertainty information for conceptual clustering and concept hierarchy generation. But, this system is not designed to work on the fuzzy relational concepts present in unstructured or semi-structured text documents.

3. System architecture

The proposed ontology-based text mining system is aimed at extracting fuzzy relational concepts from biological documents, which have been tagged according to a domain ontology. Extracted relations are accompanied by membership values that indicate the degree of co-occurrence of the concepts. The ontology is enhanced by incorporating the fuzzy relations into it. Entities are not added to the ontology. All entities are stored in a structured database along with other relevant information for answering user queries efficiently. Figure 1 demonstrates the architecture of the proposed ontology-guided text mining system to enrich a biological-domain-ontology with Fuzzy relations. Each relation has a strength based on its co-occurrence with a pair of Biological entities. The proposed system architecture has five major components.

- *Document Processor* – This module accepts ontology-based tagged text documents as input and extracts sentences from it. The module uses a Parts-of-Speech (POS) Tagger to assign a grammatical tag to every word in a sentence. Each document is converted into a tree of sentences. This module does not use the ontology tags.
- *Relational Verb Extractor* – This module uses the ontology tags of the input document and also the tree-structure generated by the *Document Processor.* It implements deep text mining principles based on NLP techniques to mine all relational verbs and their morphological variants from the document.
- *Biological Tag Association Extractor* – This module exploits the domain ontology taxonomy relations in conjunction with the tree structure to identify *frequently occurring* associations between different types of biological entity *tags*.
- *Fuzzy Biological Relation Extractor* – This module uses the relational verbs and their morphological variants extracted by *Relational Verb Extractor,* and the *strongly-associated* entity-tag pairs extracted by the *Biological Tag Association Extractor*, to identify *feasible* fuzzy biological relations, where *feasible* relations are those that have *strength* greater than system-specified minimum support. The same relational verb may be associated to multiple entity-tag pairs with differing strengths.
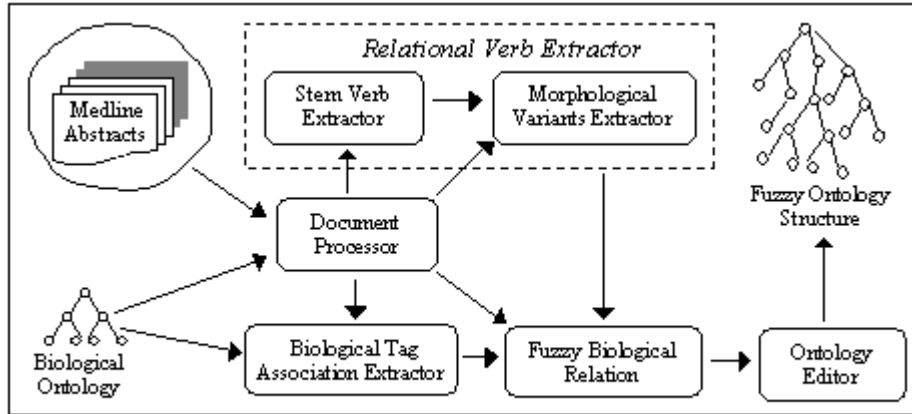
Figure 1. System Architecture

- *Ontology Editor* – This module is used to incorporate the extracted Fuzzy biological relations and their membership values into the existing domain ontology.

Functional details about each module are given in the following sub-sections.

3.1. Document processor

The *Document Processor* parses the text documents into sentences and assigns a document id, a name (Medline number), and a line number to them. It then filters stop words and Meta Language (ML) tags from the tagged documents and passes the documents through a Parts of Speech (POS) tagger to assign parts of speech to different words. A sub-set of stop words used by PubMed database is removed prior to processing. Some of these words are *significantly, further* etc. We however retain the prepositions in order to find the relational variants. Finally, every parsed document is converted into a binary tree structure which is recursively defined as follows:

```
struct sentence{
        string  root_word
        struct sentence * right_segment
        struct sentence * left_segment;
}
```

Each sentence of a document and thereby the whole document is converted into an instance of the tree by distributing the tags in the following way:

Root (R): A node that contains the right most VERB tag.

Right Segment (R_s): A node that contains all tags that are to the right of the tag considered at R

Left Segment (L_s): A node that contains all tags that are to the left of the tag considered at R.

The equivalent context-free grammar for this is given as follows:

Document (D) $\rightarrow$ L_s R R_s D | $\in$

L_s $\rightarrow$ L_s R R_s | (E+N+A+J+R)* | $\in$

R_s $\rightarrow$ (E+N+A+J+R)*

R $\rightarrow$ V

where, N, A, J, R and V are Noun, Adverb, Adjective, Preposition, and Verb tags respectively assigned by the POS tagger. E is the tag assigned by document processor to biological entity tags.

An Example Sentence from GENIA Corpus
We demonstrate, through the deletion of the <cons sem="G#DNA_domain_or_region">human <cons sem="G#protein_molecule"> UDG </cons> promoter sequences</cons>, that expression of <cons sem="G#protein_molecule">E2F-1</cons> activates the <cons sem="G# DNA_domain_or_region"><cons sem="G#protein_molecule">UDG</cons> promoter</cons> through several <cons sem ="G# DNA_domain_or_region"><cons sem="G#protein_family_or_group">E2F</cons> sites</cons>.

- -

Filtered and POS Tagged Sentence
demonstrate, through deletion of DNA-domain-or-region
    V   R   N   R   E
expression of protein-molecule activates
    N   R   E   V
DNA-domain-or-region through DNA-domain-or-region.
   E   R   E

- -

Generated Tree structure

activates (V)

demonstrate (V)   DNA-domain-or-region (E) through (R) DNA-domain-or-region (E)

Null   through (R) deletion (N) of (R) DNA-domain-or-region (E) expression (N) of (R) protein-molecule (E)
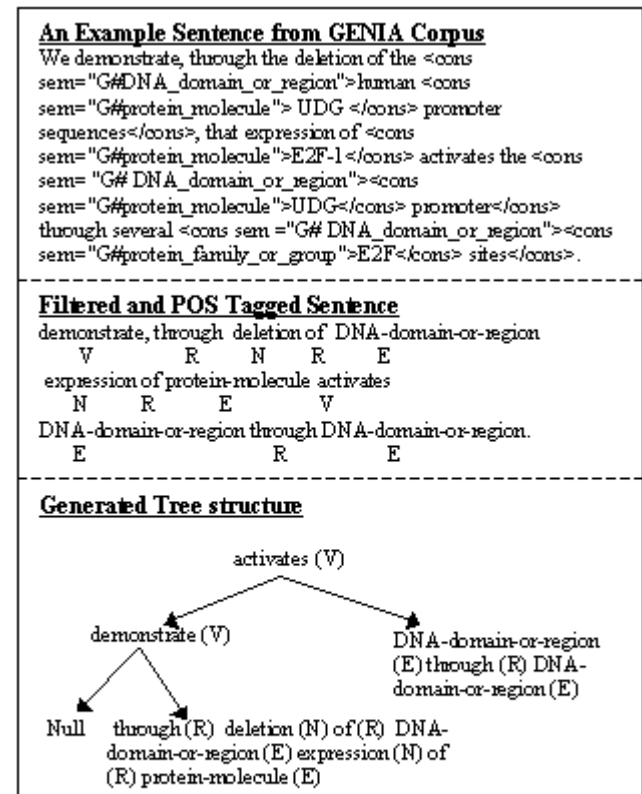
Figure 2. A sentence from MEDLINE: 95197524, its filtered and POS tagged form and corresponding tree structure

An example sentence, taken from GENIA corpus, its filtered form along with the POS tags assigned by POS tagger to different words and the generated tree structure is shown in figure 2. The tree structure encodes all relevant information which can be effectively exploited by the extractor of relational verbs and their morphological variants, whose working principle is elaborated in the next section.

frequencies of the various verbs observed in the collection. These frequencies include all occurrences of the root verb or its morphological variants. It may be noted that the most frequent set of relations identified include those seven that were identified by Sekimizu and Tsujii [12].

After this the *Relational verb extractor* module finds all possible morphological variants of the extracted verbs,
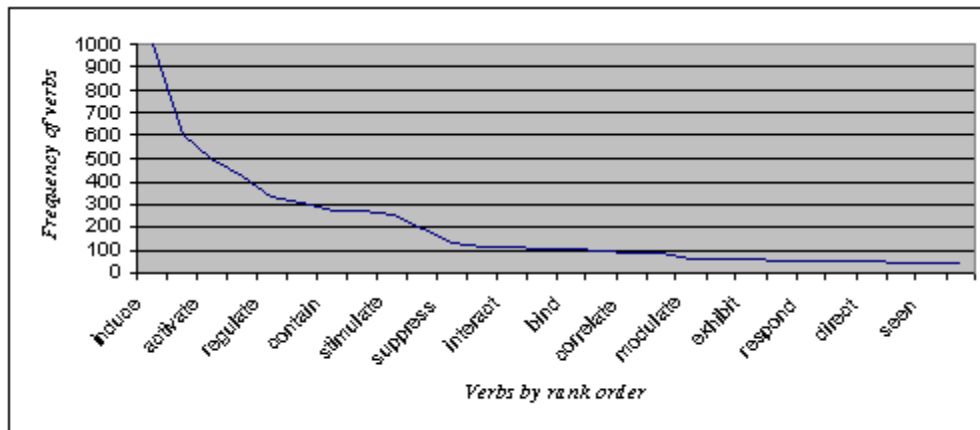

Figure 3. A plot of the frequency of relational verbs occurrence and their rank order

3.2. Biological relational verb extractor

This module has dual purpose. It uses the binary tree representation of the document sentences to extract the relational verbs present in the text documents. To obtain only relevant Biological verbs, we have considered only those verbs that are surrounded by biological entities. This constraint automatically filters the common verbs used in the text documents, even if they have very high TF. To extract biological relational verbs, only those non-leaf nodes of the tree are considered, whose immediate inorder predecessor and successor both contain at least one biological entity tag. Each relational verb is accompanied by its strength computed by its Term Frequency (TF). We employed a lower cut-off of 50 i.e. only those biological relations that occur at least 50 times in the collection are retained for further consideration. By applying these rules on the GENIA Corpus, 24 stem verbs were extracted by the system that can represent a valid relationship among biological entities and their locations. The selected verbs along with their frequency counts are:
Induce (1000), *inhibit* (607), *activate* (507), *express* (426), *regulate* (335), *mediate* (308), *contain* (267), *associate* (265), *stimulate* (252), *enhance* (197), *suppress* (127), *affect* (115), *interact* (110), *produce* (109), *bind* (107), *contribute* (97), *correlate* (90), *encode* (90), *modulate* (60), *promote* (59), *exhibit* (57), *characterize* (52), *respond* (52), *generate* (50). Figure 3 depicts the

from the document collection. It uses the extracted verbs along with the binary tree structure created by the *document processor*. Algorithm *Morphological Variants* presented in Figure 4 states how morphological variants are recognized. Some of the key behavioral features of the morphological variants extraction mechanism are:

Algorithm Morphological_variants (ROOT)
Input: List of stem verbs; Sentence binary tree structure
Output: Morphological variants of the verbs.
Steps :
1. Ptr = ROOT // start from root node
2. If (Ptr ≠ Null)   // If the tree is non-empty
3.   Search_Stem_Verb (Stem_Verb_List, Ptr->root_word) // Search
        // the stem verb list for the partial match of the root word
4.   If the root word has a match with a stem verb
5.   Check the first element of the text segment pointed by Right Segment (Rs)
6.   If it is a "by", "in", " with", or "of" preposition
7.   Store the root word along with the proposition as a morphological variant of the verb that was matched in step 3.
8.   Go to step 13
9.   Else
10.   Store the root word only as a morphological variant of the verb that was matched in step 3.
11.   Go to step 13
11.   End if
12. Else // the root word has not a match with any stem verb
13.   Ptr = Ptr -> Left_segment (Ls) // Consider the next sub-tree
        Go to step 2 // Repeat the above steps for the next sub-tree
14. End if
15. End if
16. End

Figure 4. Morphological variants generation

(i) Prepositions and partial pattern matches play key roles in identification of morphological variants of relational verbs. (ii) Only those prepositions are considered whose immediate inorder predecessor is a biological relational verb (iii) A sentence may have zero or more morphological variant(s).

3.3. Biological tag association extractor

This module uses the underlying domain ontology and the binary tree representation of the documents to extract frequently co-occurring tags in the document. Each association between a pair of tags is assigned a strength computed using term frequency (TF) and inverse document frequency (IDF) [11] as follows. Since an entity occurs in a document in association to a tag, the TF of a tag in a document denotes the total number of times the tag occurs in the document in the context of any entity of that type. Weight W_{ij} of tag E_j in document D_i is computed by the following formula

$$W_{ij} = E_{ij} \times \log(N/n_j) \quad \ldots\ldots\ldots\ldots\ldots\ldots.(i)$$

where E_{ij} is the frequency of the jth entity E_j in document d_i, $\log(N/n_j)$ is the inverse document frequency of entity E_j, N is the total number of documents in the collection, and n_j is the number of documents that contain the entity E_j. The strength of association between a tag pair E_j and E_k is computed as

$$\mu(E_j, E_k) = \cfrac{\sum\limits_{i=1}^{N} W_{ij} \otimes W_{ik}}{\sum\limits_{i=1}^{N} W_{ij}} \quad \ldots\ldots\ldots\ldots\ldots.(ii)$$

where $\otimes$ denotes a fuzzy conjunction operator which is taken as a min operator in our case.

Table 1. Tag pair associations

Entity-1 (E1)	Entity-2 (E2)	Strength μ (E1, E2)
protein_family_or_group	Protein_molecule	0.77
DNA_family_or_group	DNA_domain_or_region	0.68
DNA_family_or_group	cell_line	0.33
cell_type	DNA_family_or_group	0.27
RNA_molecule	protein_complex	0.14
protein_subunit	Virus	0.09
Nucleotide	mono_cell	0.00
Nucleotide	Carbohydrate	0.00

A threshold value may be applied to filter all weakly associated fuzzy tag associations. A partial list of

extracted tag associations along with their strength of co-occurrence is shown in table 1.

3.4. Fuzzy biological relation extractor

This module compiles the *frequently occurring* combinations of the relational verbs and their morphological variants in conjunction with the associated tag pairs extracted by the *Biological Tag Association Extractor*. These relations are called the *frequent fuzzy relations,* where a relation defined between a pair of biological tags (E_i, E_k), is associated with a membership value μ (Ej, Ek). It is found that a number of tag-pair combinations never co-occur in the corpus. This information is utilized during guided query formulation. Thus the system obtains the frequently occurring triplets of the form <E1, V, E2> very efficiently. The membership value associated with a relation triplet is its normalized frequency of occurrence.

Table 2. Instances of Fuzzy Biological Relations extracted from GENIA corpus

Entity-1	Relational Verb	Entity-2	Membership Value
Mono_cell	Induces in	Protein_molecule	1.00
protein_molecule	Inhibit in	Protein_molecule	1.00
protein_molecule	expressed	cell_type	1.00
protein_molecule	Affects	cell_type	0.67
protein_subunit	associates	Protein_subunit	0.50
Cell_type	activated	Protein_molecule	0.40
Cell_type	contained in	cell_type	0.33
Nucleotide	Mediate	Lipid	0.00

Obviously, triplets <E1, V, E2> and <E2, V, E1> may not be associated with same membership value. For example, *E1 activates E2* does not imply that *E2 activates E1*. The tree structure of a sentence determines whether it supports the relation <E1, V, E2> or <E2, V, E1> when all three of them are present. A partial list of generated *Frequent Fuzzy Relations* along with their membership values is shown in table 2. As is obvious from table 2, two of the most frequently occurring relations are "induces-in" and "expressed with" defined between mono-cell and protein molecule, and a protein molecule and cell type respectively.

3.5. Ontology editor

This module is used to incorporate the extracted fuzzy relations along with their membership values into the existing domain ontology. We have used Protégé for the said purpose.

4. Fuzzy ontology guided query processing

The ontology-guided query interface shown in figure 5, helps the user to pose his/her queries intelligently by activating different tags and displaying the extracted fuzzy relations. The user has the flexibility to choose GENIA tag names or provide specific entity names for formulating queries. For a GENIA tag, or a known entity, the interface guides the user to extend the queries by displaying valid relation names and domain tags, where the relations are arranged according to decreasing strengths. Fuzzy relations can also be intelligently used to accept/reject the user's queries. For example, if a user query includes nucleotide and mono_cell along with any relating verb (indicated by *) then the query is rejected
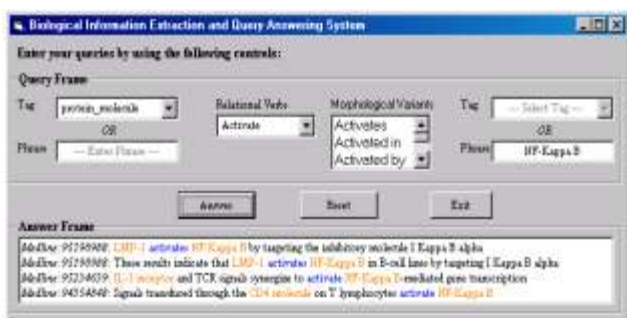

Figure 5. Query Interface

even without searching the huge corpus, since it is known, due to their association strength zero, that no document contains any such combination of tags. A sample query "protein-molecule activates NF-Kappa B" is shown in the figure, and the corresponding MEDLINE sentences that satisfy it are also shown. It may be noted that information extraction can be done from all tagged abstracts thereafter efficiently, provided the distribution of verbs remain similar. However, when distributions change drastically, fuzzy relations may be learnt afresh.

5. Conclusion and future work

In this paper we have presented an ontology-based text mining system that extracts fuzzy relations from biological texts and enhances the traditional biological ontological structures with these. The enhanced ontology structure helps in increasing the effectiveness of the information extraction process and also guides the user to pose queries in a more focused way. Presently, the system is being integrated with a tag predictor system, for tagging new MEDLINE documents. The integrated system is also being extensively experimented for focused query processing.

References

[1]   D. Fensel, F. van Harmelen, I. Horrocks, D.L. McGuinness, and P.F. Patel-Schneider, "OIL: An ontology Infrastructure for the Semantic Web", *IEEE Intelligent Systems*, 16 (2), 2001, pp. 38-45.

[2]   N. Guarino, and P. Giaretta, "Ontologies and Knowledge Bases: Towards a Terminological Clarification", *Toward Very Large Knowledge Bases: Knowledge Building and Knowledge sharing*, N. Mars (ed.) IOS Press, Amsterdam, 1995, pp. 25-32.

[3]   I. Horrocks, and U. Sattler, "Ontology Reasoning in the SHOQ(D) Description Logic", *In Proceedings of IJCAI'01*, Morgan Kaufmann, 2001, pp. 199-204.

[4]   J.-D. Kim, T. Ohta, Y. Tateisi, and J. Tsujii, "GENIA corpus – a semantically annotated corpus for bio-textmining", *Bioinformatics*, Vol. 19 Suppl. 1 2003, pp. i180-i182.

[5]   S. W. Liddle, K. A. Hewett, and D. W. Embley, "An Integrated Ontology Development Environment for Data Extraction", *In Proceedings of ISTA'03*, 2003, pp. 21-33.

[6]   Y. Li, and N. Zhong, "Web Mining Model and its Applications for Information Gathering", *Knowledge-Based Systems 17*, 2004, pp. 207-217.

[7]   H. M. Muller, E. E. Kenny, and P. W. Sternberg, "Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature", *PloS Biology*, Volume 2(11), e309, Nov. 2004. URL: www.plosbiology.org

[8]   T. T. Quan, S. C. Hui, and T. H. Cao, "FOGA: A Fuzzy Ontology Generation Framework for Scholarly Semantic Web", *In Proceedings of the 2004 Knowledge Discovery and Ontologies Workshop (KDO'04)*, Pisa, Italy, Sep. 24, 2004.

[9]   D. Rebholz-Schuhmann, H. Kirsch, and F. Couto, "Facts from Texts-Is Text Mining Ready to deliver?", *PlosBiology*, Vol. 3(2), e65, Feb. 2005. URL: www.plosbiology.org

[10] R. A. Ribeiro, and A. M. Moreira, "Fuzzy Query Interface for a Business Database", *Int'l. Journal of Human-Computer Studies 58*, 2003, pp. 363-391.

[11] G. Salton, *Automatic Text processing*, Addison-Wesley, 1989.

[12] T. Sekimizu, H.S. Park, and J. Tsujii, "Identifying the interaction between genes and genes products based on frequently seen verbs in Medline abstract", *Genome Informatics 9*, 1998, pp. 62–71.

[13] P. Velardi, P. Fabriani, and M. Missikoff, "Using Text Processing Techniques to Automatically Enrich a Domain Ontology", *In Proceedings of ACM Conference on Formal Ontologies and Information Systems (FOIS'01)*, Ogunquit, Maine, 2002, pp. 270-284.

[14] M. Wallace, and Y. Avrithis, "Fuzzy Relational Knowledge Representation and Context in the Service of Semantic Information Retrieval", *In Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Budapest, Hungary, July 2004.

[15] D. H. Widyantoro, and J. Yen, "A Fuzzy Ontology-based Abstract Search Engine and Its User Studies", *In Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, 2001, Melbourne, Australia, 2001, pp. 1291-1294.