Published in final edited form as:
Pac Symp Biocomput. 2006 ; : 28–39.

EVALUATION OF LEXICAL METHODS FOR DETECTING RELATIONSHIPS BETWEEN CONCEPTS FROM MULTIPLE ONTOLOGIES*

HELEN L. JOHNSON†, **K. BRETONNEL COHEN**, **WILLIAM A. BAUMGARTNER JR.**, **ZHIYONG LU**, **MICHAEL BADA**, **TODD KESTER**, **HYUNMIN KIM**, and **LAWRENCE HUNTER**
Center for Computational Pharmacology, University of Colorado School of Medicine

Abstract

We used exact term matching, stemming, and inclusion of synonyms, implemented via the Lucene information retrieval library, to discover relationships between the Gene Ontology and three other OBO ontologies: ChEBI, Cell Type, and BRENDA Tissue. Proposed relationships were evaluated by domain experts. We discovered 91, 385 relationships between the ontologies. Various methods had a wide range of correctness. Based on these results, we recommend careful evaluation of all matching strategies before use, including exact string matching. The full set of relationships is available at compbio.uchsc.edu/dependencies.

1. Introduction

Lexical analysis of an ontology is a powerful tool for suggesting relationships between concepts within the ontology [25,28,29,30,32] or among multiple ontologies [7,6]. However, there are many possible text types to match between (e.g. term names, synonyms, and definitions) and variations on matching techniques (e.g. stemming, case normalization, etc.), and there is no reason to expect similar, and equally valid, results for all of them. Most importantly, the mere existence of a match does not prove a valid relationship between concepts. In this paper, we systematically evaluate three text matching techniques in two text types, and use domain experts to evaluate the correctness of the resulting matches.

Recently, [7] demonstrated the utility of an external, publicly-available resource for finding within-ontology relationships. They hypothesized that two GO terms that share a relationship to a single ChEBI term are related to each other. They detected 771, 302 within-GO relationships by finding sets of GO terms and synonyms that lexically matched a ChEBI term or its synonyms. They noted that 55% of all GO terms contained 26% of all ChEBI terms, totalling 20, 497 GO-ChEBI relationships. Implicit in their work is the assumption that relationships found between GO and ChEBI terms are valid and meaningful. More recently [6], they have proposed extending this technique to detect relationships between all OBO ontologies.

Various ontology working groups have become interested in integrating external ontologies into their own, and have pointed out some of the obstacles to doing this [15]. In this paper, we show that a variety of publicly available resources can be exploited for large-scale, automated suggestion of between-ontology relationships. We define a *relationship* as any direct or indirect association between two ontological concepts.

*This work is supported by NLM grant R01-LM00811 to Lawrence Hunter.

†Authors Johnson and Cohen contributed equally to the work reported here.

In this paper, we evaluate the following hypotheses:

1. Valid relationships exist between concepts from GO and from other OBO ontologies.
2. Gene Ontology definitions are a fruitful resource for discovering relationships between concepts in ontologies.
3. Language processing techniques for discovering relationships have quantifiable and variable rates of correctness.

A novel aspect of this paper as compared to [6] and [7] is that we make use of the text of GO definitions. There is some history for using definitions in language processing applications, particularly in the word sense disambiguation task [21]. More recently, [23] shows the value of Gene Ontology definitions for predicting GeneRIFs. We also evaluate simple linguistic processing techniques for term detection and normalization. The findings may be useful for semi-automatically linking ontologies, whether to support reasoning tasks or annotation, and also for detecting terms from ontologies in natural language texts.

1.1. Context and motivation

This research falls into the general category of semantic integration (SI). Semantic integration is a currently active topic of research in the general computer science, artificial intelligence, Internet, and data mining communities [26]. It has crucial roles to play in areas as diverse as interoperability in Semantic Web Services [8], coreference resolution in free text [22], schema and data matching in databases [12], and communication between intelligent agents and resources [13]. There is much related work in the ontology community, e.g. [27] and [24] among many others. Within the biomedical ontology literature, closely related work includes the description-logic-based GONG project [33], in which GO metabolism terms were linked to biological-substance terms from MeSH using lexical tools and term synonyms of UMLS. [2], [19], [5] used various non-lexical techniques to find relations within GO.

Mapping vs. alignment of ontologies—Integration of multiple, independently produced ontologies is an important task in molecular biology. One well-studied aspect of this task is *mapping*, the identification of equivalent concepts in multiple ontologies [15,20,30]. This work has shown some of the difficulties of textual analysis of biological ontologies. For example, [30] points out that biological terminologies pose difficulties for standard normalization procedures, since they often contain alphanumeric modifiers. Other problems include synonymy and morphological variation [20].

Ontology alignment is the task of making overlapping concepts among multiple ontologies compatible. Although mapping may be a part of alignment, the alignment task requires finding meaningful relationships between non-identical concepts. The identification of such relationships may also be valuable within an ontology, e.g. in order to improve compositionality [25,32,28,29] or in defining and populating novel relationships. The work reported here is relevant both to the mapping and to the alignment task.

Natural language processing—The relevance of locating concepts from an ontology in free text is clear from the inclusion of this task in recent “bake-off” competitions in the NLP community. The overall low performance on these tasks [4,16,9] demonstrates their difficulty. The work in this paper can be thought of as a step towards recognizing OBO concepts in free text: GO terms and definitions are themselves a type of semi-structured natural language, fitting the sublanguage model but having enough complexity to be a challenge, while not being as unstructured as the language of scientific abstracts.

2. Methods

Materials

We retrieved the current versions of the GO [1,15], ChEBI [11], Cell Type [3], and BRENDA Tissue [31] ontologies from SourceForge. (In the remainder of this paper, when we say "(other) OBO ontologies," we mean the ontologies other than GO.) We chose these three other ontologies because we expected high degrees of subject-matter overlap between them and GO, and because they are in relatively advanced stages of development.

Finding relationships

We used Lucene [14] to search for the OBO concepts in GO. Lucene is a Java information retrieval library[a]. We modeled the GO concepts as documents to be retrieved, and the other OBO concepts as search engine queries. We indexed the GO concepts, placing the terms and definitions in distinct fields, which allowed us to search them separately. We constructed Lucene *phrasal queries* from the other OBO concepts. This meant that for searches on multi-word OBO concepts, word order could not vary and no words could intrude. Synonym queries were done by constructing phrasal queries for each synonym, and then grouping the phrasal queries with Boolean *OR*. Both indexing and searching require a Lucene class called an *analyzer*. We used the WhiteSpace and Porter-Stemmer analyzers. Lucene gave us an efficient and robust framework for carrying out searches and for manipulating their results.

Evaluation

We drew a random sample from the relationships proposed by each technique for each ontology, for a total of 2,389 relationships. The sample is unevenly distributed across various categories of ontologies, linguistic manipulations, and GO terms vs. GO definitions, but covers all combinations of those categories. These 2,389 relationships were manually examined by domain experts. One domain expert (DE1) has considerable experience in ontologies, biology, and structural chemistry. The other domain expert (DE2) is a bioinformatics doctoral candidate with experience with GO and with protein function and subcellular localization. The experts were presented with (1) the ID and name of a concept from an OBO ontology, and (2) the ID, name, and definition of some concept from the GO. In addition, the experts had access to the definitions of the OBO concept, as well as any other helpful information found in the ontologies themselves. They were instructed to evaluate the output with the following question in mind: *Is this OBO term the concept that is being referred to in this GO term/definition ?* They were permitted to classify all relationships as either true positive or false positive. We calculated *correctness* as the number of true positive relationships divided by the number of proposed relationships (similar to precision or specificity). All relationships are available for public inspection at compbio.uchsc.edu/dependencies.

Inter-annotator agreement (IAA)

DE1 evaluated the majority of the output. A sample of 400 proposed relationships was also evaluated by DE2. Initial IAA between the two was 93.5% (374/400). After dispute resolution, the consensus IAA was 98.2% (393/400). For the remaining seven cases, DE1 had the deciding vote.

Linguistic manipulations

We queried by exact match to the OBO concept name. We also queried using synonyms of OBO concepts. Since all work in this area has observed moderate differences in concept name realization, such as pluralization, we also implemented the standard linguistic manipulations

[a]Previous applications of Lucene to text processing in the biomedical domain are reported in [10] and [18].

of stemming and stop word removal [17]. We evaluated the correctness of the resulting searches individually.

What we counted

For each ontology, we give data on the following:

- Relationships found by matches between the OBO ontology and GO terms (T)
- Relationships found by matches between the OBO ontology and GO definitions (D)
- The union of T and D (T ∪ D)[b]
- The intersection of T and D (T ∩ D)[c]
- The relative complement of T and D (T-D)[d]
- The relative complement of D and T (D-T)[e]

Gain, the magnitude of the increase in the number of relationships detected by examining definitions, rather than just terms, is the relative complement of D and T divided by the union of T and D ((D-T)/(T ∪ D)).

In addition, for each ontology, we calculated the analogous set relations for the various language processing techniques. This allows us to quantify the yield and the correctness of the various techniques with respect to the three ontologies.

For the Cell Type ontology, we filtered out all matches to the terms *cell* (CL:0000000, 3215 matches), *cell by organism* (CL:0000004, 96 matches), and *cell by function* (CL:0000144, 10 matches), since we realized early on that they were either content-free or incorrect.

3. Results

Finding relationships between ontologies

Our initial hypothesis was that there are relationships between GO and the various OBO ontologies. Table 2 summarizes the number of matches between GO and the three other ontologies and the average correctness calculated by manually examining a subset of the matches. Searching GO terms and definitions for terms from the other ontologies resulted in a total of 91,385 proposed relationships. The majority of these links (73,002) are between GO and ChEBI. The average correctness across the three ontologies is 80.62%. This is generally consistent with the precision reported for the mapping task by [30] (range from .36 for BLAST to .94 for exact match) and [20] (range from .25 for Chimaera to 1.0 for PROMPT). These data are consistent with the initial hypothesis, validating the goal expressed in [6], and gives an idea of the size of the set of potential relationships.

Correctness and Error Analysis

To assess the correctness of the matches, a random set of 2,389 was manually examined by domain experts. All results are given in Table 3. Note that although correctness is generally high, *some combinations of ontology and linguistic technique had quite low correctness.* This has important consequences for more ambitious efforts to detect relationships across all OBO

[b] T ∪ D gives the number of relationships that are found in terms or definitions. Some of its relationships are revealed by both. It equals (T ∩ D) + D-T + T-D.
[c] T ∩ D is the number of relationships that are found in both terms and definitions.
[d] T-D is the number of relationships that can be found in terms, but cannot be found by examining definitions. It equals T − (T ∩ D).
[e] D-T is the number of relationships that can be found in definitions, but cannot be found by examining terms. It equals D − (T ∩ D).

ontologies, such as proposed by [6]: we cannot use any technique, including exact matching, without assessing its correctness for a particular pair of data sources.

Exact matching was the most accurate type of search, ranging from 76% to 100% correct. This is consistent with results reported for the mapping task. One source of false positives for exact matching was polysemy, or words with multiple meanings. For example, the word *group* (CHEBI:24433) also has a General English meaning, and often appeared with that sense in GO definitions. Similarly, the BRENDA term *joint* (BTO:0001686), which refers to an anatomical joint, appears as an adjective meaning *combined* in GO concepts. We found examples of false positives related to non-General-English, domain-specific terms as well, e.g. *reticulum* (BTO: 0000347) incorrectly matching the definition of GO:0006614. Incorporating OBO term synonymy resulted in slightly lower correctness, ranging from 42% to 94%, with an average of 67.4% (397/589). Finally, the stemming/stop-word-removal searches show the lowest correctness, ranging from 7% to 92%.

GO terms versus GO definitions

A novel hypothesis of this paper is that GO definitions are a fruitful resource for discovering relationships between GO and other ontologies. Table 4 addresses this hypothesis for BRENDA, and the corresponding data for the other ontologies are given on the website (compbio.uchsc.edu/dependencies). Searching for relationships in the GO definitions in addition to the GO terms had a large impact on the quantity of relationships found between ontologies. The table presents the number of links found in GO terms and in GO definitions, as well as the union, intersection and relative complements of these sets. The number of links found only in GO terms is given by the relative complement of terms and definitions (T-D), listed in the fifth column of the table. The number of links found only in GO definitions is given by the relative complement of definitions and terms (D-T), the sixth column. The final column in Table 4 describes the gain from searching in GO definitions for relationships. It is calculated by dividing the D-T by the union of D and T. For instance, in the first row of Table 4, which displays the number of links found in the GO using an exact BRENDA term search, a gain of 49.59% means that just under half of these between-ontology links could be found only by searching the GO definition. Note that for all ontologies and for all search strategies, the number of relationships is higher when definitions are considered. The gain is never lower than 24.3% (270 additional matches for exact matching of Cell Type concepts), and it is generally higher than 50% (43,146 additional matches just for the case of allowing stemming matches for ChEBI concepts). The correctness (see Table 3) of relationships detected by matches to definitions is comparable to the correctness of relationships detected by matches to terms.

3.1. Linguistic techniques in relationship searches

Using synonyms—Results for including the synonyms associated with BRENDA terms in the search string are given in Table 5; corresponding data for the other ontologies is on the website. *E* is *exact match*, *Syn* adds synonyms for the OBO concept, and the other columns are the union, intersection, and relative complements. Adding synonyms increased the yield of relationships by an average of 36% (23,300/64,987) over using only the exact OBO term query. The set E should be a proper subset of Syn, and the relative complement of E and Syn is the empty set. The yield of using OBO synonyms ranged from 9% (85/925) to 40.69% (8669/21300), and was generally quite similar for GO terms and for GO definitions. Synonyms allowed us to detect some relationships that could not have been found by any other technique, e.g. relating *adipose* (BTO:0000441) to *larval fat body development* (GO :0007504). Correctness for synonymy-based matches was sometimes low, ranging from 42 to 94% (see Table 3)—not surprising in the face of the history of query expansion attempts in information retrieval.

Stemming GO concepts and OBO terms—Results for stemming and stop word removal (labelled *Stem)* for BRENDA are given in Table 6; data for the other ontologies is on the website. Stemming garnered the greatest increase of proposed relationships, with an average gain of 78% (146,777/188,464). However, this increase comes at a price, with a lower average correctness rate of 51% (257/505). Note that the correctness of matching BRENDA to GO terms or definitions by stemming is extremely low (7–15%). Again, searching for stemmed OBO terms also returned the subset of relationships that the exact term searches returned, and E-Stem is the empty set.

Stemming allowed pluralized forms of the same term to be matched. It also picked up other morphological variation in terms, e.g. matching *neuron* (CL:0000540) to *neuronal* in the definition of *syntrophin* (GO:0016013). In a random sample of 97 relationships matched by stemming across the three ontologies, 57% was due to pluralization, 11% to adjectival derivation, and 32% to other morphological variation.

4. Discussion and conclusions

Our results are consistent with the hypotheses that there are many valid relationships between GO and other OBO ontologies, and that in addition to GO terms, GO definitions are an important source for detecting them.

Implications for ontology mapping

One implication of this study comes from the observation that correctness is almost never 100%: even exact string matches do not guarantee a valid match. Ontologists attempting to carry out the goal stated in [6] should not ignore these findings. The results on BRENDA are especially cautionary.

In contrast to work on the mapping task done by the ontology community, the evaluation of work such as ours and Burgun and Bodenreider's has been hampered by the lack of a curated gold standard. One important product of the work reported in this paper is a data set of 2,389 GO/other-OBO concept pairs that has been examined by at least one domain expert. This data set includes 1,926 true positive relationships and 463 known unrelated (i.e., the false positive) pairs. It is publicly available at compbio.uchsc.edu/dependencies, and will allow future researchers in this area to do principled automatic evaluations.

Furthermore, the set of known unrelated pairs can be used in future efforts to filter out terms that are known to produce high numbers of irrelevant, incorrect, or simply unrevealing matches. We suspect that a relatively small set of OBO terms contributed many of the errors, and that correctness can be improved by filtering them. This analysis continues.

Implications for ontology enrichment

One limitation for the application of these relationships to ontology enrichment (the addition of relationships among existing terms) is the fact that most of the relationships that we detect are indirect. For example, our techniques relate *T cell* (CL:0000084) to both *T cell proliferation* (GO:0042098) and *regulation of T cell proliferation* (GO:0042129), but an ontologist would likely prefer to find only the direct relationship from *T cell* to *T cell proliferation.* Another limitation for ontology enrichment is that our methods do not automatically differentiate between relation types (see e.g. [28]). Future work should attempt to differentiate between direct and indirect relationships, and to characterize the nature of the relations between concepts.

Implications for language processing

This study provides cautionary data on the limits of various techniques, even exact string matches. The data provides a list of terms that are likely to produce false-positive matches under conditions of exact match and specific linguistic manipulations; these lists can be used to filter results from any language processing system that seeks to recognize concepts from the ChEBI, Cell Type, and BRENDA ontologies. It also points us towards techniques that might allow us to predict which terms are likely to produce high rates of false positive matches, such as ones at high positions in an ontology (e.g. *cell* (CL:0000000)) and ones that are isomorphic with General English words (e.g. *groups* (CHEBI:24433)). Additionally, it highlights the importance of building biomedical-domain-specific preprocessing tools, such as stemmers.

References

1. Ashburner M, et al. Gene Ontology: tool for the unification of biology. Nature Genetics 2000;25:25–29. [PubMed: 10802651]
2. Bada M, Turi D, McEntire R, Stevens R. Using reasoning to guide annotation with Gene Ontology terms in GOAT. SIGMOD Record 2004;33(2):27–32.
3. Bard J, Rhee SY, Ashburner M. An ontology for cell types. Genome Biology 2005;6:R21. [PubMed: 15693950]
4. Blaschke C, Leon EA, Krallinger M, Valencia A. Evaluation of BioCreative assessment of task 2. BMC Bioinformatics 2005;6(Suppl 1):S16. [PubMed: 15960828]
5. Bodenreider O, Aubry M, Burgun A. Non-lexical approaches to identifying associative relations in the Gene Ontology. PBS 2005 2005:104–115.
6. Bodenreider O, Burgun A. Linking the Gene Ontology to other biological ontologies. ISMB Bio-ontologies SIG meeting. 2005
7. Burgun A, Bodenreider O. An ontology of chemical entities helps identify dependence relations among Gene Ontology terms. Semantic mining in biomedicine. 2005
8. Burstein MH, McDermott DV. Ontology translation for inter-operability among Semantic Web Services. AI Mag 2005;26(1):71–82.
9. Camon EB, Barrell DG, Dimmer EC, Lee V, Magrane M, Maslen J, Binns D, Apweiler R. An evaluation of GO annotation retrieval for BioCreative and GOA. BMC Bioinformatics 2005;6(Suppl 1):S17. [PubMed: 15960829]
10. Carpenter, B. Phrasal queries with LingPipe and Lucene: ad hoc genomics text retrieval; NIST Special Publication: SP 500–261 The Thirteenth Text Retrieval Conference; TREC. 2004; 2004.
11. Degtyarenko, K. Chemical vocabularies and ontologies for bioinformatics; Proc 2003 Int Chemical Info Conf; Nimes, France. 2003.
12. Doan A, Halevy AY. Semantic-integration research in the database community: a brief survey. AI Mag 2005;26(1):83–94.
13. Gruninger M, Kopena JB. Semantic integration through in-variants. AI Mag 2005;26(1):11–20.
14. Gospodnetiæ O, Hatcher E. Lucene in action. Manning. 2005
15. Hill DP, Blake JA, Richardson JE, Ringwald M. Extension and integration of the Gene Ontology (GO): combining GO vocabularies with external vocabularies. Genome Research 2002;12(12):1982–1991. [PubMed: 12466303]
16. Hirschman L, Yeh A, Blaschke C, Valencia A. Overview of BioCreative: critical assessment of information extraction for biology. BMC Bioinformatics 2005;6(Suppl 1):S1. [PubMed: 15960821]
17. Jackson P, Moulinier I. Natural language processing for online applications: text retrieval, extraction, and categorization. John Benjamins. 2002
18. Konrad K, Steinbach R, Stenzhorn H. Competitive intelligence with Lucene in XtraMind's XM-InformationMinder. Gospodnetiæ and Hatcher 2005 2005:344–350.
19. Kumar A, Smith B, Borgelt C. Dependence relationships between Gene Ontology terms based on TIGR gene product annotations. Proc CompuTerm 2004:31–38.

20. Lambrix P, Edberg A. Evaluation of ontology merging tools in bioinformatics. PSB 2003 2003:589–600.
21. Lesk M. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. SIGDOC 1986:24–26.
22. Li X, Morie P, Roth D. Semantic integration in text: from ambiguous names to identifiable entities. AI Mag 2005;26(1):45–58.
23. Lu Z, Cohen KB, Hunter L. Finding GeneRIFs via Gene Ontology annotations. PSB. 2006this volume
24. McGuinness DL, Fikes R, Rice J, Wilder S. The Chimaera ontology environment. AAAI 2000 2000:1123–1124.
25. Mungall CJ. Obol: integrating language and meaning in bio-ontologies. Comparative and Functional Genomics 2004;5:509–520.
26. Noy NF, Doan A, Halevy AY. Semantic integration. AI Mag 2005;26(1):7–9.
27. Noy NF, Musen MA. PROMPT: Algorithm and tool for automated ontology merging and alignment. AAAI 2000 2000:450–455.
28. Ogren PV, Cohen KB, Acquaah-Mensah GK, Eberlein J, Hunter L. The compositional structure of Gene Ontology terms. PSB 2004 2004:PP-214–225.
29. Ogren PV, Cohen KB, Hunter L. Implications of compositionality in the Gene Ontology for its curation and usage. PSB 2005 2005:174–185.
30. Sarkar IN, Cantor MN, Gelman R, Hartel F, Lussier YA. Linking biomedical language information and knowledge resources: GO and UMLS. PSB 2003 2003:427–450.
31. Schomburg L, et al. BRENDA, the enzyme database: updates and major new developments. NRA 2004;32:D431–D433.
32. Verspoor, CM.; Joslyn, C.; Papcun, GJ. Participant notebook of the ACM SIGIR '03 workshop on text analysis and search for bioinformatics. 2003. The Gene Ontology as a source of lexical semantic knowledge for a biological natural language processing application; p. 51-56.
33. Wroe CJ, Stevens R, Goble CA, Ashburner M. A methodology to migrate the Gene Ontology to a description logic environment using DAML+OIL. PSB 2003 2003:624–635.

Table 1

Materials: ontologies, data files, and revisions.

Ontology	terms	synonyms	avg. syn./term	data file	revision_date
Gene Ontology	19,508	8,202	.42	gene_ontology.obo	09:06:2005 17:10
ChEBI	11,549	19,295	1.67	chebi.obo	25:05:2005 10:54
Cell Type	748	215	.29	cell.obo	24:05:2005 17:10
BRENDA	2,222	1,208	.54	BrendaTissue.txt	10:5:2005 13:49:02

Table 2

Counts and correctness of proposed relationships between ontologies. Numbers in parentheses are the correct and total manually evaluated pairs.

Ontology	Relationships to GO	Avg. Correctness
ChEBI	73002	84.2% (977/1161)
Cell Type	1961	92.99% (584/628)
BRENDA	16469	60.83% (365/600)
TOTAL	91385	80.62% (1926/2389)

Table 3

Correctness Rates (correct/evaluated)

Ontology	Exact	Synonyms	Stemming
ChEBI			
GO Term	99.5% (199/200)	42.0% (42/100)	73.0% (73/100)
GO Def	97.8% (451/461)	69.0% (138/200)	74.0% (74/100)
Cell Type			
GO Term	100% (200/200)	94% (44/47)	76% (41/54)
GO Def	98.7% (231/234)	50% (21/42)	92% (47/51)
BRENDA			
GO Term	76.0% (76/100)	83.0% (83/100)	7.0% (7/100)
GO Def	93.0% (93/100)	69.0% (69/100)	15.0% (15/100)

Table 4

Relationships in GO terms vs. GO defs for BRENDA

Ontology	T	D	T∪D	T∩D	T-D	D-T	Gain
Exact	1465	2447	2906	1006	459	1441	49.59%
Exact + synonyms	1875	3093	3686	1282	593	1811	49.13%
Stemmed	3892	15409	15722	3579	313	11830	75.24%

Table 5

Using BRENDA synonyms

GO	E	Syn	EuSyn	EnSyn I	5-Syn	Syn-E	Gain
T	1465	1875	1875	1465	0	410	21.9%
D	2447	3093	3093	2447	0	646	20.9%
TUD	2906	3686	3686	2906	0	780	21.2%

Table 6

Stemming and stop word removal: BRENDA

GO	E	Stem	E∪Stem	E∩Stem	E-Stem	Stem-E	Gain
T	1465	3892	3892	1465	0	2427	62.36%
D	2447	15409	15409	2447	0	12962	84.12%
T∪D	2906	15722	15722	2906	0	12816	81.52%