Acquiring Conceptual Relationships from a MRD and Text Corpus

Masaki Kurematsu[1], and Naomi Nakaya[2] and Takahira Yamaguchi[2]

[1] Faculty of Software and Information Science, Iwate Prefectural University
152–52 Takizawasugo Takizawa Iwate 020-0193 JAPAN
kure@soft.iwate-pu.ac.jp
[2] Dept. Computer Science, Shizuoka University
3–5–1 Johoku Hamamatsu Shizuoka 432-8011 JAPAN
{cs7068, yamaguti}@cs.inf.shizuoka.ac.jp

Abstract. How to exploit a machine-readable dictionary (MRD) and text corpus in supporting the construction of domain ontologies that specify taxonomic and non-taxonomic relationships among given domain concepts are discussed here. a) In building taxonomic relationships (hierarchy structure) of domain concepts, some hierarchy structure can be extracted from MRD with marked sub-trees that may be modified by a domain expert, using both matching result analysis and trimmed result analysis. Domain-specific hierarchical structure can also be extracted from text corpus, using pairs of concepts that turn to be located near and have similar context by WordSpace. Thus two different kinds of hierarchical structure change into unified one with additional modification by a domain expert. b) In building non-taxonomic relationships (specification templates) of domain concepts, we construct concept specification templates that come from pairs of concepts that turn to be located near and have similar context by WordSpace. A domain expert does the task based on them later. The case study with some law called CISG shows us that the trade-off between precision and recall is so important in practically building domain ontologies.

1 Introduction

Although ontologies have been very popular in many application areas, we still face the problem of high cost associated with building up them manually. In particular, since domain ontologies have the meaning specific to application domains, human experts have to make huge efforts for constructing them entirely by hand.

In order to reduce the costs, automatic or semi-automatic methods have been proposed using knowledge engineering techniques and natural language

This paper is identical to the one published in IJCAI'01 Workshop on Ontology Learning

processing ones (cf. Ontosaurus [Swartout et. al. 1996]). The authors have also developed a domain ontology refinement support environment called LODE [Kurematsu and Yamaguchi 1997] and a domain ontology rapid development environment called DODDLE [Sekiuchi et. al. 1998], using machine readable dictionaries. However, these environments facilitate the construction of only a hierarchically structured set of domain concepts, in other words, taxonomic conceptual relationships.

As domain ontologies have been applied to widespread areas, such as knowledge sharing, knowledge reuse, software agents and information integration, we need software environments that support a human expert in constructing the domain ontologies with not only taxonomic conceptual relationships but also non-taxonomic ones. In order to develop the environments, it seems better to put together two or more techniques such as knowledge engineering, natural language processing, machine learning and data engineering, as seen in the workshop on ontology learning in ECAI2000 (e.g. [Maedche and Staab 2000]).

Here in this paper, we extend DODDLE into DODDLE II that constructs both taxonomic and non-taxonomic conceptual relationships, exploiting WordNet [Fellbaum 1998] and domain-specific texts with the automatic analysis of lexical co-occurrence statistics, based on WordSpace [Marti and Schutze] that has the idea that a pair of terms with high frequency of co-occurrence statistics can have non-taxonomic conceptual relationships. Furthermore, we evaluate how DODDLE II works in the field of law, the Contracts for the International Sale of Goods (CISG). The empirical results show us that DODDLE II can support a law expert in constructing domain ontologies.

2 DODDLE II: A Domain Ontology Rapid Development Environment

Figure 1 shows an overview of DODDLE II, "a Domain Ontology rapiD Development Environment" that has the following two components:

- Taxonomic relationship acquisition module using WordNet
- Non-taxonomic relationship learning module using domain-specific texts

A domain expert gives a set of domain terms to the system.
A) The taxonomic relationship acquisition module (TRA module) does "spell match" between the input domain terms and WordNet. The "spell match" links these terms to WordNet. Thus the initial model from the "spell match" results is a hierarchically structured set of all the nodes on the path from these terms to the root of WordNet. However the initial model has unnecessary internal terms (nodes). They do not contribute to keeping topological relationships among matched nodes, such as parent-child relationship and sibling relationship. So we can trim the unnecessary internal nodes from the initial model into a trimmed model, as shown in Figure 2. In order to refine the trimmed model, we have the following three strategies that we will describe later in the context of interaction with an user:

- Matched result analysis

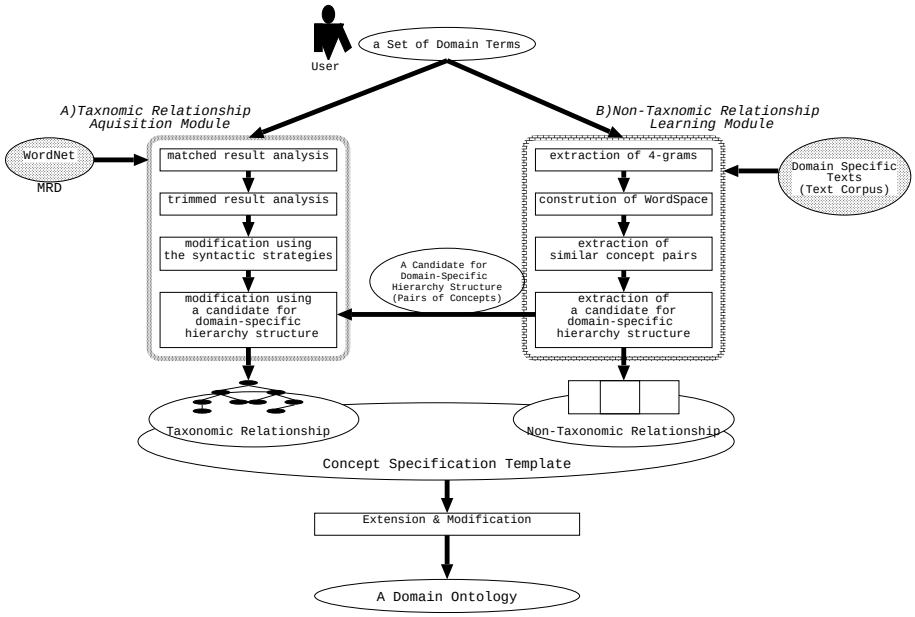

Fig. 1. DODDLE II overview

- Trimmed result analysis
- Using a candidate for domain-specific hierarchy structure extracted from text corpus.

B) The non-taxonomic relationship learning module (NTRL module) extracts the pairs of terms that should be related by some relationship from domain-specific texts, analyzing lexical co-occurrence statistics, based on WordSpace that is a multi-dimensional, real-valued vector space where the cosine of the angle between two vectors is a continuous measure of their semantic relatedness. Thus the pairs of terms extracted from domain-specific texts are the candidates for non-taxonomic relationships. We can built concept specification templates by putting together taxonomic and non-taxonomic relationships for the input domain terms. The relationships should be identified in the interaction with a human expert.

3 Taxonomic Relationship Acquisition

After getting the trimmed model, TRA module is refined by interaction with a domain expert, using the following three strategies: matched result analysis, trimmed result analysis and using domain-specific hierarchy structure extracted from text corpus.

Looking at the trimmed model, it turns out that it is divided into a PAB (a PAth including only Best spell-matched nodes) and a STM (a Sub-Tree that

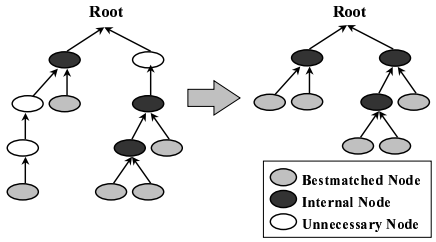

Fig. 2. Trimming Process

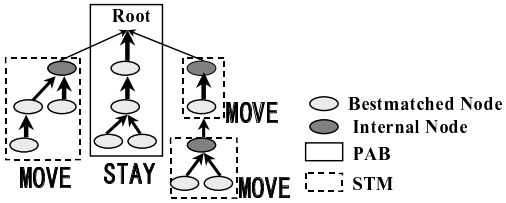

Fig. 3. Matched Result Analysis

includes best spell-matched nodes and other nodes and so can be moved) based on the distribution of best-matched nodes. On one hand, a PAB is a path that includes only best-matched nodes that have the senses good for given domain specificity. Because all nodes have already been adjusted to the domain in PABs, PABs can stay in the trimmed model. On the other hand, a STM is such a sub-tree that an internal node is a root and the subordinates are only best-matched nodes. Because internal nodes have not been confirmed to have the senses good for a given domain, a STM can be moved in the trimmed model. Thus DODDLE II identifies PABs and STMs in the trimmed model automatically and then supports a user in constructing a conceptual hierarchy by moving STMs. Figure 3 illustrates the above-mentioned matched result analysis.

In order to refine the trimmed model, DODDLE II can use trimmed result analysis as well as matched result analysis. Taking some sibling nodes with the same parent node, there may be many differences about the number of trimmed nodes between them and the parent node. When such a big difference comes up on a sub-tree in the trimmed model, it is be better to change the structure of the sub-tree. DODDLE II asks the user if the sub-tree should be reconstructed or not. Based on the empirical analysis, the sub-trees with two or more differences

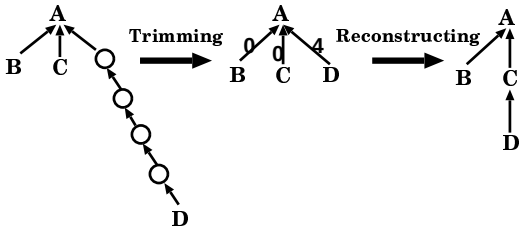

Fig. 4. Trimmed Result Analysis

may be reconstructed. Figure 4 illustrates the above-mentioned trimmed result analysis.

TRA module can not make suggestions about domain-specific hierarchy structure using above-mentioned strategies. Because these strategies don't know domain specific information. So, in addition to do that, TRA module makes suggestions using a candidate for domain-specific hierarchy structure extracted from text corpus by NTRL module. Domain-specific hierarchy structure is a set of pair of concepts. We will describe how to extract the structure later. When there are two different kinds of hierarchy structure between two concepts, DODDLE II asks the user if the hierarchy structure should be changed into unified one or not.

Finally DODDLE II completes taxonomic relationships of the input domain terms manually from the user.

4 Non-Taxonomic Relationship Learning

Non-taxonomic Relationship Learning almost comes from WordSpace, which derives lexical co-occurrence information from a large text corpus and is a multi-dimension vector space (a set of vectors). The inner product between two word vectors works as the measure of their semantic relatedness. When two words' inner product is beyond some upper bound, there are possibilities to have some non-taxonomic relationship between them.

4.1 Construction of WordSpace

WordSpace is constructed as shown in Figure 5.

1. extraction of high-frequency 4-grams Since letter-by-letter co-occurrence information becomes too much and so often irrelevant, we take term-by-term co-occurrence information in four words (4-gram) as the primitive to make up co-occurrence matrix useful to represent context of a text. We take high frequency 4-grams in order to make up WordSpace.

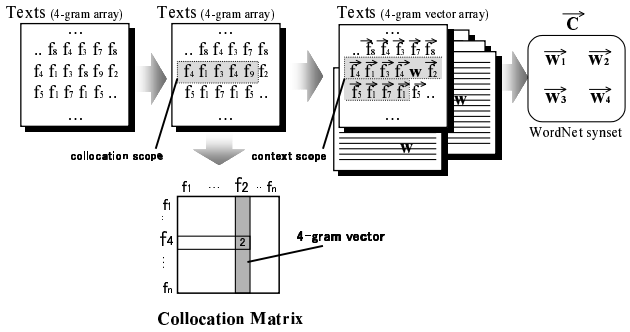

Fig. 5. Construction Flow of WordSpace

2. construction of collocation matrix A *collocation matrix* is constructed in order to compare the context of two 4-grams. Element $a_{i,j}$ in this matrix is the number of 4-gram f_i which comes up just before 4-gram f_j (called *collocation area*). The collocation matrix counts how many other 4-grams come up before the target 4-gram. Each column of this matrix is the *4-gram vector* of the 4-gram f.
3. construction of context vectors A *context vector* represents context of a word or phrase in a text. A sum of 4-gram vectors around appearance place of a word or phrase (called *context area*) is a context vector of a word or phrase in the place.
4. construction of word vectors A word vector is a sum of context vectors at all appearance places of a word or phrase within texts, and can be expressed with the following formula. Here, $\tau(w)$ is a vector representation of a word or phrase w, $C(w)$ is appearance places of a word or phrase w in a text, and $\varphi(f)$ is a 4-gram vector of a 4-gram f. A set of vector $\tau(w)$ is WordSpace.

$$\tau(w) = \sum_{i \in C(w)} \Big(\sum_{f \text{ close to } i} \varphi(f) \Big)$$

5. construction of vector representations of all concepts The best matched "synset" of each input terms in WordNet is already specified, and a sum of the word vector contained in these synsets is set to the vector representation of a concept corresponding to a input term. The concept label is the input term.

4.2 Constructing and Modifying Concept Specification Templates

Vector representations of all concepts are obtained by constructing WordSpace. Similarity between concepts is obtained from inner products in all the combination of these vectors. Then we define certain threshold for this similarity. A

concept pair with similarity beyond the threshold is extracted as a similar concept pair. A set of similar concept pairs becomes concept specification templates. Both of the concept pairs, whose meaning is similar (with taxonomic relation), and has something relevant to each other (with non-taxonomic relation), are extracted as concept pairs with context similarity in a mass. However, by using taxonomic information from TRA module with co-occurrence information, DODDLE II distinguishes the concept pairs which are hierarchically close to each other from the other pairs as TAXONOMY.

A user constructs a domain ontology by considering the relation with each concept pair in the concept specification templates, and deleting an unnecessary concept pair.

4.3 Extracting Domain-Specific Hierarchy Structure

In order to make suggestions about domain-specific hierarchy structure, NTRL module tries to extract pairs of concepts which form part of a candidate for domain-specific hierarchy structure. In order to do that, we pay attention to the distance between two concepts in a document. In this paper, the distance between two concepts means the number of words between them. If the distance between two concepts is small and the similarity between them is close, we suppose that one concept explains the other. If the distance is large and the similarity is close, we suppose that they form part of domain-specific hierarchy structure. According to above-mentioned idea, we calculate the proximally rate between two concepts within a certain scope. It is the number of times both concepts occur within the scope divided by the number of times only one concept occurs within it. We define certain threshold for this proximally rate. Pairs of concepts whose proximally rate is within this threshold and the similarity between them is beyond the threshold for similarity are extracted as part of a candidate for domain-specific hierarchy structure.

5 Case Studies for Taxonomic Relationship Acquisition

In order to evaluate how DODDLE is doing in practical fields, case studies have been done in a particular field of law called Contracts for the International Sale of Goods (CISG). Two lawyers joined the case studies. In the first case study, input terms are 46 legal terms from CISG Part-II. In the second case study, they are 103 terms including general terms in an example case and legal terms from CISG articles related with the case. One lawyer did the first case study and the other lawyer did the second.

Table 1 shows the result of the case studies . Figure 6 shows how much of the intermediate products is included in final domain ontology at each DODDLE activity.

Generally speaking, in constructing legal ontologies, 70 % or more support comes from DODDLE. About half portion of the final legal ontology results in the information extracted form WordNet. Because the two strategies just imply

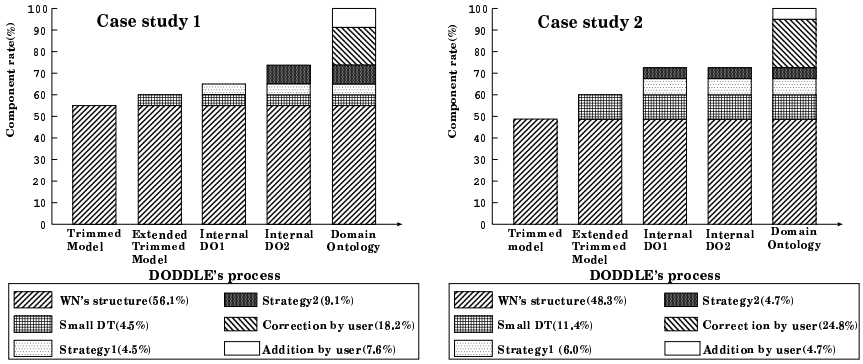

Fig. 6. The Component Rate of the Final Domain Ontology

Table 1. The Case Studies Results

The number of X	The first case study	The second case study
Input terms	46	103
Small DT(Component terms)	2(6)	6(25)
Nodes matched with WordNet(Unmatched)*	42(0)	71(4)
Salient Internal Nodes(Trimmed nodes)	13(58)	27(83)
Small DT integrated into a trimmed model(Unintegrated)	2(0)	5(1)
Modification by the user(Addition)	17(5)	44(7)
Evaluation of strategy1**	4/16(25.0%)	9/29(31.0%)
Evaluation of strategy2**	3/10(30.0%)	4/12(33.3%)

* "Nodes matched with WordNet" is the number of input terms which have be selected proper senses in WordNet and "Unmatched" is not the case.

** The number of suggestions accepted by a user/The number of suggestions generated by DODDLE

the part where concept drift may come up, the part generated by them has low component rates and about 30 % hit rates . So one out of three indications based on the two strategies work well in order to manage concept drift. Because the two strategies use such syntactical feature as matched and trimmed results, the hit rates are not so bad. In order to manage concept drift smartly, we may need to use more semantic information that is not easy to come up in advance in the stratagies.

6 A Case Study for Non-Taxonomic Relationship Learning

DODDLE II, domain ontology rapid development environment, which refer to MRD and domain-specific texts, is being implemented on Perl/Tk now. Figure 7 shows the ontology editor (left window) and the concept graph editor (right window).

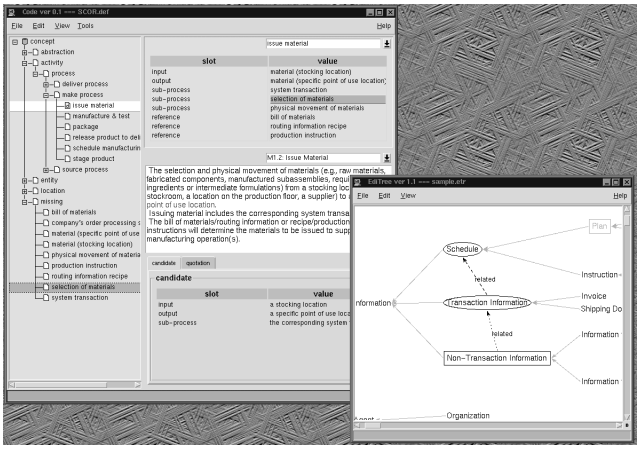

Fig. 7. The Ontology Editor

Table 2. significant 46 concepts in CISG part II

acceptance	delivery	offer	reply
act	discrepancy	offeree	residence
addition	dispatch	offerer	revocation
address	effect	party	silence
assent	envelope	payment	speech act
circumstance	goods	person	telephone
communication system	holiday	place of business	telex
conduct	indication	price	time
contract	intention	proposal	transmission
counteroffer	invitation	quality	withdrawal
day	letter	quantity	
delay	modification	rejection	

Subsequently, as a case study for non-taxonomic relationship acquisition, we constructed the concept definition for significant 46 concepts of having used on the first case study (Table 2) with editing the concept specification template using DODDLE II, and verified usefulness. The concept hierarchy, which the lawyer actually constructed using DODDLE in the first case study was used here (Figure 8).

6.1 Construction of WordSpace

High-frequency 4-grams were extracted from CISG (about 10,000 words) and 526 kinds of 4-grams were obtained. In order to keep density of a collocation matrix high, the extraction frequency of 4-grams must be adjusted according to the scale of text corpus. As CISG is the comparatively small-scale text, the extraction frequency was set as 8 times this case. Then, the collocation matrix was constructed by counting the number of each 526 kinds 4-gram just before a 4-gram for each kind. Since 526 kinds of 4-grams were extracted, the collocation

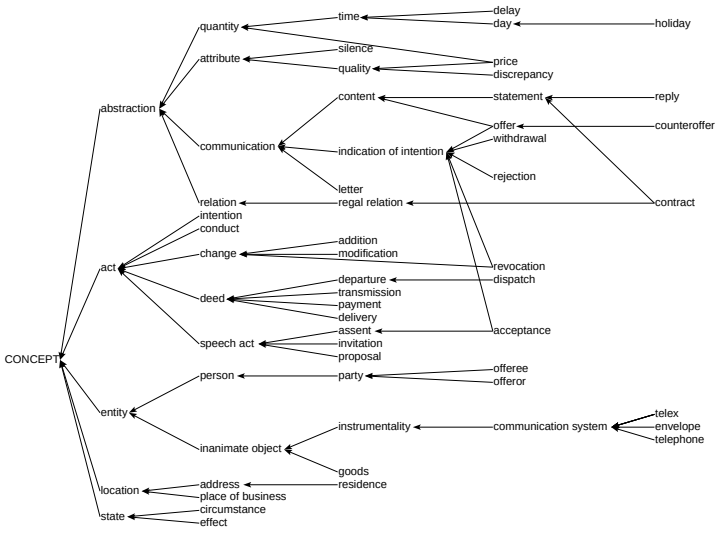

Fig. 8. domain concept hierarchy of CISG part II

matrix also became 526 dimensions. In order to construct a context vector, a sum of 4-gram vectors around appearance place circumference of each of 46 concepts was calculated. One article of CISG consists of about 140 4-grams. The number of 4-gram vectors in context area was set as 60 from experience. For each of 46 concepts, the sum of context vectors in all the appearance places of the concept in CISG was calculated, and the vector representations of the concepts were obtained. The set of these vectors is used as WordSpace to extract concept pairs with context similarity.

6.2 Constructing and Modifying Concept Specification Templates

Having calculated the similarity from the inner product for the 1035 concept pairs which is all the combination of 46 concepts, and having used threshold as 0.9993, 90 concept pairs were extracted, and concept specification templates were constructed. Table 3 is the list of the extracted similar concepts corresponding to each concept. A concept in bold letters is either an ancestor, descendant or a sibling to the left concept in the concept hierarchy constructed using DODDLE in the first case study. In concept specification templates, such a concept is distinguished as TAXONOMY relation. As taxonomic relationships and non-taxonomic relationships may be mixed in the list based on only context similarity, the concept pairs which may be concerned with non-taxonomic relationships are obtained by removing the concept pairs with taxonomic relationships. Figure 9 shows concept specification templates extracted about the concept "assent". The concepts underlined are in taxonomic relation with each other.

Table 3. the concept pairs extracted according to context similarity (threshold 0.9993)

CONCEPT	CONCEPT LIST IN SIMILAR CONTEXT
acceptance	**communication**, **offer**, **indication**, telex
act	offeror, **assent**, effect, payment, person, quantity, time, goods, **delivery**, **dispatch**, price, contract, delay, withdrawal, offeree, place, quality
assent	offeror, **act**, effect, offer, person, offeree, withdrawal, time, **proposal**
communication	**acceptance**, **offer**, **telex**, conduct, **indication**
conduct	party, telex, communication
contract	effect, act, person, delivery, payment, quantity
delay	delivery, offer, act, payment
delivery	**payment**, quantity, goods, place, **act**, delay, time, contract, person, effect, quality
dispatch	goods, price, **act**, person, quantity, offeror
effect	person, assent, act, offeror, contract, proposal, payment, time, withdrawal, party, delivery
goods	dispatch, quantity, delivery, payment, act, person, price, quality
indication	intention, **acceptance**, **communication**
intention	indication
offer	**acceptance**, assent, **communication**, delay
offeree	withdrawal, **offeror**, assent, act, price
offeror	act, assent, withdrawal, **offeree**, **person**, effect, time, price, dispatch
party	conduct, effect, place, **person**
payment	quantity, **delivery**, place, act, goods, quality, delay, effect, person, contract, time
person	effect, **offeror**, act, proposal, goods, assent, withdrawal, contract, dispatch, payment, delivery, **party**, place, price
place	payment, delivery, time, quantity, party, act, person
price	dispatch, act, offeror, goods, withdrawal, offeree, person
proposal	person, effect, withdrawal, **assent**
quality	quantity, payment, goods, act, delivery
quantity	payment, delivery, goods, act, quality, dispatch, place, contract, **time**
telex	conduct, **communication**, acceptance
time	act, offeror, delivery, place, effect, payment, **quantity**, assent
withdrawal	offeree, offeror, person, price, act, assent, effect, proposal

The final concept definition is constructed from consideration of concept pairs in the templates. Figure 10 shows the definition of the concept "assent" constructed from the templates. Although relation AGENT exists also in assent-offeree and assent-offeror, it is represented by definition inheritance and not described.

6.3 Extracting Domain-Specific Hierarchy Structure

We have defined the threshold for the proximally rate as 0.78, the certain scope as the same sentence and tried to extract domain-specific hierarchy structure in the first case study. As a result, DODDLE II extracted 128 pairs of concepts regarded as part of domain-specific hierarchy structure from text corpus. 8 pairs out of them have occurred in the concept hierarchy constructed by the user and have not occurred in the trimmed model. That is, they and modifications by the user were same. It shows that DODDLE II can make useful suggestions about domain-specific hierarchy structure using candidate for them extracted from text corpus. But the rate of same suggestions as modification by the user is about 6%(8/128) and is not good. So, we have to improve extraction of candidate.

assent	*non-TAXONOMY?* :	**offeror**
	TAXONOMY :	**act**
	non-TAXONOMY? :	**effect**
	non-TAXONOMY? :	**offer**
	non-TAXONOMY? :	**person**
	non-TAXONOMY? :	**offeree**
	non-TAXONOMY? :	**withdrawal**
	non-TAXONOMY? :	**time**
	TAXONOMY :	**proposal**

Fig. 9. The concept specification templates for "assent"

assent	AGENT	: **person**
	LEGAL-SEQUENCE	: **offer**
	ANTONYM	: **withdrawal**

Fig. 10. The concept definition for "assent" with editing the templates

6.4 Results and Evaluation

The user with legal knowledge did evaluation about extraction of concept pairs. Note that the concept definition constructed in this case study is only for the 46 concepts as input terms, and is not the whole concept definition which should be constructed from CISG. The detail of the extracted concept pairs in this case study are shown in Table 4.

Taxonomic or non-taxonomic relationships existed in 59% from the top of the list of concept pairs with high context similarity between the concepts. Since a concept pair with high context similarity has a high possibility that it has some kind of relation, concept definitions can be led by considering these pairs.

The problems obtained from this case study are the follows.

Determination of a Threshold Threshold of the context similarity changes in effective value with each domain. It is hard to set up the most effective value in advance. Figure 11 is the relation between the numbers of the extracted concept pairs and recall and precision in this case study.

Specification of a Concept Relation Concept specification templates have only concept pairs based on the context similarity, it requires still high cost to specify relationships between them. It is needed to support specification of concept relationships on this system in the future work.

Ambiguity of Multiple Terminology For example, the term "transmission" is used in two meanings, "transmission (of goods)" and "transmission (of communication)", in the article, but DODDLE II considers these terms as the same and creates WordSpace as it is. Therefore constructed vector expression may not be exact. In order to extract more useful concept pairs, semantic specialization of a multisense word is necessary, and it should be considered that the 4-grams with same appearance and different meaning are different 4-grams.

Table 4. The detail of the extracted concept pairs

Threshold	Extracted concept pair	Advisable	Unknown	Improper
0.9993	90	53	14	23

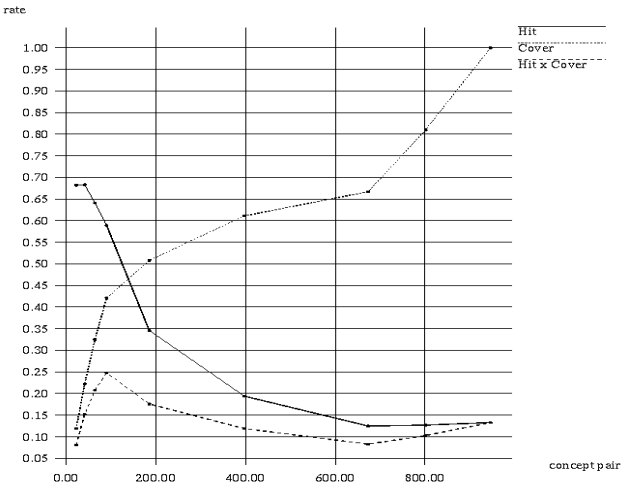

Fig. 11. recall and precision

7 Related Work

In the research using verb-oriented method, the relation of a verb and nouns modified with it is described, and the concept definition is constructed from these information (e.g. [Hahn 1998]). In [Faure and Nédellec 1999], taxonomic relationships and Subcategorization Frame of verbs (SF) are extracted from technical texts using a machine learning method. The nouns in two or more kinds of different SF with a same frame-name and slot-name is gathered as one concept, base class. And ontology with only taxonomic relationships is built by carrying out clustering of the base class further. Moreover, in parallel, Restriction of Selection (RS) which is slot-value in SF is also replaced with the concept with which it is satisfied instantiated SF. However, proper evaluation is not yet done. Since SF represents the syntactic relationships between verb and noun, the step for the conversion to non-taxonomic relationships is necessary.

On the other hand, in ontology learning using data-mining method, discovering non-taxonomic relationships using a association rule algorithm is proposed by [Maedche and Staab 2000]. They extract concept pairs based on the modification information between terms selected with parsing, and made the concept pairs a transaction. By using heuristics with shallow text processing, the generation of a transaction more reflects the syntax of texts. Moreover, RLA, which is their original learning accuracy of non-taxonomic relationships using the existing taxonomic relations, is proposed. The concept pair extraction method in our

paper does not need parsing, and it can also run off context similarity between the terms appeared apart each other in texts or not mediated by the same verb.

8 Conclusion

In this paper, we discussed how to construct a domain ontology using existing MRD and domain-specific texts. In order to acquire taxonomic relationship, two strategies have been proposed: matched result analysis and trimmed result analysis. Furthermore, in order to learn non-taxonomic relationships, concept pairs may be related to concept definition, extracted on the basis of the co-occurrence information in domain-specific texts, and a domain ontology is developed by the modification and specification of concept relations with concept specification templates. It serves as the guideline for narrowing down huge space of concept pairs to construct a domain ontology.

It is almost craft-work to construct a domain ontology, and it is still difficult to obtain the high support rate on system. The DODDLE II mainly supports for construction of a concept hierarchy with taxonomic relationships and extraction of concept pairs with non-taxonomic relationships now. However a support for specification concept relationship is indispensable. The future work follows: improvement in the scalability of the definition support by learning of heuristics, introduction of the useful data-mining method instead of WordSpace, and system integration of taxonomic relationship acquisition module and non-taxonomic relationship learning module (now implementing).

Acknowledgments

We would like to express our thanks to Mr. Takamasa Iwade (a graduate student of shizuoka university) and the members in the Yamaguchi-Lab.

References

[Swartout et. al. 1996] Bill Swartout, Ramesh Patil, Kevin Knight and Tom Russ: "Toward Distributed Use of Large-Scale Ontologies", Proc. of the 10th Knowledge Acquisition Workshop (KAW'96), (1996)

[Kurematsu and Yamaguchi 1997] Masaki Kurematsu and Takahira Yamaguchi: "A Legal Ontology Refinement Support Environment Using a Machine-Readable Dictionary", "*Artificial Intelligence and Law 5*", 119-137, (1997)

[Sekiuchi et. al. 1998] Rieko Sekiuchi, Chizuru Aoki, Masaki Kurematsu and Takahira Yamaguchi: "DODDLE : A Domain Ontology Rapid Development Environment", PRICAI98, (1998)

[Maedche and Staab 2000] Alexander Maedche, Steffen Staab: "Discovering Conceptual Relations from Text", ECAI2000, pp.321-325 (2000)

[Fellbaum 1998] C.Fellbaum ed: "Wordnet", The MIT Press, 1998. see also URL: http://www.cogsci.princeton.edu/~wn/

[Marti and Schutze] Marti A. Hearst, Hinrich Schutze: "Customizing a Lexicon to Better Suit a Computational Task", in Corpus Processing for Lexical Acquisition edited by Branimir Boguraev & James Pustejovsky, pp.77-96

[Hahn 1998] Udo Hahn, Klemens Schnattinger: "*Toward Text Knowledge Engineering*", AAAI98, IAAAI-98 proceedings, pp.524-531 (1998)

[Faure and Nédellec 1999] David Faure, Claire Nédellec, "Knowledge Acquisition of Predicate Argument Structures from Technical Texts Using Machine Learning: The System ASIUM", EKAW'99

[Sono and Yamate 1993] Kazuaki Sono, Masasi Yamate: United Nations convention on Contracts for the International Sale of Goods, Seirin-Shoin(1993)