Semantic Relations: From Thesaurus to Ontology

Abstract

Thesaurus is frequently used in building ontology: the hierarchy of the thesaurus contains *is-a* relations and *part-whole* relations between broader terms (BT) and narrower terms (NT), and other relations between BT and related terms (RT). In this paper we propose how to elicit *is-a* and non-taxonomic semantic relations for ontology from BT-NT/RT relations in thesaurus. First, we propose a set of rules based on lexical information and definition of concepts to identify *is-a* relations and *not is-a* relations from BT-NT/RT relations. Second, a similarity-based approach is proposed to identify non-taxonomic semantic relations from *not is-a* relations, which semantic relations are defined for domain ontology considering criteria of generality, popularity and uniqueness. A rule-based system to identify *is-a* relations is implemented. Its performance (accuracy of 85%) on IT domain ontology is comparable to agreement among human experts (86%). A similarity-based system to identify non-taxonomic semantic relations is implemented and its performance is quite promising.

1 Introduction

Ontology consists of information including concepts, semantic relations between concepts, instances, and axioms on the domain. Existing lexical knowledge bases such as thesaurus contain part of semantic information including terms/concepts and taxonomic relations. It is the reason why thesaurus is frequently used for ontology building. But thesaurus gives only roughly defined relations without clear identification. The hierarchy of the thesaurus contains *is-a* relations and *part-whole* relations between BT and NT, and other related relations between BT and RT [17]. We attempt to transform BT-NT/RT relations in thesaurus to *is-a* and non-taxonomic semantic relations for ontology**.**

For an example, for given relations of *btnt*(active antennas, antennas) and *btnt*(antenna theory, antennas)[1], the former one is *is-a* relation *isa*(active antennas, antennas); and the latter one, as a relation between NT "antenna theory" and BT "antennas", should be identified to specific semantic relations, for example, "*p_isTheoryAbout*", for ontology.

In this paper, we propose a set of rules based on lexical information and definition of concepts to identify *is-a* relations from BT-NT/RT relations. The rules are adopted to compare the identities of the concepts. The concepts in BT-NT/RT relations are considered to hold *is-a* relation if they share the same identity with each other, and other relations are identified as *not is-a* relations.

A similarity and pattern-based approach is proposed to identify non-taxonomic semantic relations from *not is-a* relations of thesaurus. BT and NT/RT are classified to semantic categories first with similarity-based approach, then the given *not is-a* relation is identified to a semantic relation based on the patterns constructed semi-automatically.

IT domain ontology is adopted as a case study. Inspec thesaurus [10] is taken as source thesaurus. Experiment shows that its performance of *is-a* relation identification is comparable to the consistency of experts. A similarity-based system to identify non-taxonomic semantic relations is implemented, and its performance is quite promising.

2 Previous Work

There are many researches performed on extracting semantic relations from text. [4, 6, 8, 9, 16] However, in spite of its necessity, most of these works are more tend to used for the extension of existing ontology. Because the extracted relations are normally excursive ones, how to use these excursive relations to build a basis hierarchy of ontology is a problem. It's the reason why using existing knowledge base such as thesaurus to build

---

[1] Hereafter, *btnt*(NT, BT) denotes BT-NT and BT-RT relations, *isa*(NT, BT) *is-a* relations, and *n-isa*(NT, BT) *not is-a* relations.

ontology gets attentions in these years. Existing thesaurus or thesaurus-like knowledge bases normally provide generally used terms in certain domains as their vocabulary, and present relations between BT and NT with a hierarchy tree structure. Some of these knowledge bases including also constraint information can be used for prediction or reasoning [15]. But they often provide only part of the information required for ontology building.

Some researches focus on converting existing knowledge base to ontological expression without creating ontology [2]. The original thesaurus formats are converted into ontological expressions like RDF or OWL. Case study on the format of individual thesaurus is required, then a pattern or rule-based conversion is performed. Some of other researches address problem of transforming the source knowledge into ontological one by extracting useful information from existing knowledge bases [1, 15, 20]. Constraints of ontology are derived from the source code of original logic programs, and then transformed to ontological knowledge.

There are also some of the researches remodeling thesaurus to ontology with relation level enrichment [17, 20, 23]. Some of them [20] extends thesaurus by inserting case relations and semantic relations into taxonomic hierarchy of thesaurus, where the case relations are from existing machine translation systems and dictionaries, and the semantic relations acquired from correlation information extracted from corpus. In the other researches [17, 23], the BT-NT/RT relations are identified to more specific semantic relations of ontology by rules or patterns defined by human developers.

From the point of task to solve, what we proposed in this paper is similar to these researches [23, 17]. The differences of our task including, first, IT domain and Inspec thesaurus cover rather broader domain, and so there is limitation to define patterns manually. Adopt what kind of semantic relations for the domain ontology would be also a problem. To solve these problems, a domain taxonomy with semantic categories for IT domain are defined semi-automatically first, then semantic relations are induced from these semantic categories, with the domain and range specified automatically with the semantic categories. The semantic relations with domain and range constraints are adopted as patterns in further identification of semantic relations. Another difference of our task is, the BT-NT relations in Inspec thesaurus actually are a mixture of both BT-NT and BT-RT relations[2], while the relations in [23, 17] already labeled with distinguished BT/NT, USE/UF, and RT relations. To solve this problem, Inspec relations are identified to *is-a* and *not is-a* relations first, and then *not is-a* relations are identified to semantic relations latterly, with approach of similarity-based term classification and pattern-based semantic relations identification. A series of experiments are performed, and the performances are reported very promising in the paper.

3 Thesaurus and Target Ontology

Introducing Inspec thesaurus, semantic relations of ontology to target are described. The details on how to define and identify these semantic relations are presented in Section 5.

3.1 Inspec Thesaurus

Inspec [10], made by the Institution of Electrical Engineers (IEE), is bibliographic information service in English. Inspec covers more than 14 areas including computing, control engineering, electrical and electronic engineering, information technology and physics, and etc. Inspec also provides thesaurus, which contains more than 8,300 terms and 15,901 of BT-NT relation[3]. The relations are a mixture of BT-NT and BT-RT relations, and there is no labeling (types of relations) assigned to the relations (Figure.1).

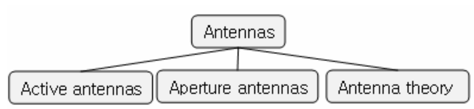

Figure. 1. The BT-NT relations of Inspec

3.2 Target Ontology

The BT-NT relations in Figure. 1 should be identified to semantic relations for ontology in Figure. 2.

For IT domain ontology, 108 semantic relations are defined, where the domains and ranges are defined in terms of 77 semantic categories of IT

[2] Hereafter BT-NT refers the mixture of BT-NT and BT-RT relations in Inspec, and NT refers either NT or RT indeed.

[3] Korean lexical expressions and definitions to the Inspec terms are complemented for our work.

domain. Some of the semantic relations are shown in Table 1: as an example, the domain of relation "*p_isTheoryAbout*" is an instance of category "cTheory", and its range can be an instance of either "cStructure" or "cEquipment".

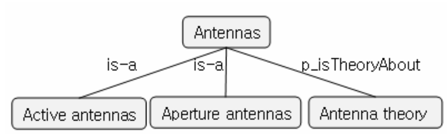

Figure. 2. Identified relations for domain ontology

Relation	Domain	Range
p_isFunctionFor	cFunction	cAnalysis
p_isFunctionIn	cFunction	cLogic
p_isFunctionOf	cFunction	cPlan
p_isTheoryAbout	**cTheory**	**cStructure**
p_isTheoryAbout	**cTheory**	**cEquipment**
p_isTheoryOf	cTheory	cInformation

Table 1. Semantic relations defined for IT domain

4 Is-a Relations

Our approach to identify *is-a* relation is relied on two assumptions: **the first assumption** is that two concepts in BT-NT relation hold *is-a* relation if they have the same identity; **the second assumption** is that the headword of a lexical expression reflects the identity of its concept. **Identity** of a concept is the essential property that distinguishes the concept from other concepts, and a NT which has the same identity with its BT and so has *is-a* relation with BT inherits the properties of the BT. In Figure.2, "Active antenna" has the same identity with its BT's, "Antennas", and so they are in a *is-a* relation. But "Antenna testing" has different identity with its BT's, "Antennas", and so they are in a *not is-a* relation.

Under the two assumptions, we propose a set of rules based on the headwords to identify *is-a* relations: the rules of the same headword, transitivity of the relation of headwords, tolerance, and abbreviation.

4.1 Headword Recognition

To a concept in IT domain, headword is normally the last word in the compound noun which expresses the concept. If the concept is expressed by only one word, then the word itself is the headword. If it is not in above cases, we adopt pattern-based approach to recognize headword. Let *head*(concept) be a function of headword recognition, we have patters as following:

- <headword><preposition><otherword>, where
  – <preposition>={by, in, on, of, from, for, with, about}
  – Ex) *head*(learning by example) = learning
- <headword>_<domain>, where
  – <domain> indicates domain information of the concept
  – Ex) *head*(network_circuits) = circuits
- <otherword>-<headword>
  – Ex) *head*(unsolicited_e-mail) = mail
- <otherword&headword>
  – This pattern indicates that there is no space between headword and other words.
  – Ex) *head*(radiotelephony) = telephony

4.2 The Rule of the Same Headword

According to the second assumption introduced at the start of the Section 4, if two concepts under BT-NT relation[4] have same headword, then we know they share same identity, and so their relations is is-a according to the second assumption: *isa*(elastic waves, waves).

Let h_n and h_t be headwords of the concepts in a BT-NT relations, and m, n be word strings, the rule of the same headword could be expressed with: if $btnt(mh_n, nh_t)$ and $h_n=h_t$, then we have $isa(mh_n, nh_t)$.

4.3 Rule of Transitivity

If the headwords of two BT-NT concepts have *is-a* relation, then according to above two assumptions and transitivity of *is-a* relations, these two concepts should have *is-a* relation also. Suppose we have *isa*(listings, programs) and *isa*(methods, theory) in IT domain, according to transitivity rule, we have *isa*(JAVA listings, complete computer programs) and *isa*(smoothing methods, filtering theory) as results. The headwords that have *is-a* relations between each other could be extracted by either experts, or from existing thesaurus like WordNet, or extracted automatically as it described in Section 5.1 on domain taxonomy building.

Let h_n, h_b be headwords of two terms, and m, n be word strings, the rule of transitivity can be expressed with: if $isa(h_n, h_b)$, then we have $isa(mh_n, nh_b)$.

[4] In following descriptions, all of the target concept pairs are all under given BT-NT relations (which are actually a mixture of BT-NT and BT-RT relations in our task) if there are no special notations.

4.4 Rule of Tolerance

There are some of the concepts have tolerance in lexical expressions. For example, given BT "equipment", it could have many terms like "receivers", "antennas", "cameras", "tubes", and "transmitters" as its NT. They are all equipments in identity, so the relations can be considered as *is-a* relations. The example of concepts in IT domain which have tolerance in lexical expressions including "equipment", "accessories", and "applications"[5]. If BT has one of these words as its headword, then very likely it has *is-a* relations with its NTs even these NTs have different headwords in their lexical expressions. Applying this rule in IT domain, we have: *isa*(radio receivers, radio equipments), *isa*(radio tracking, radio applications).

4.5 Rule of Abbreviation

Some of the BT headwords have many NT in abbreviation form. "Languages", "standards", "networks" are the examples of such BT headwords. Similar with the rule of tolerance, if BT has one of these words as its headword, and the NT is an abbreviation, then most likely they have *is-a* relation. For example, we have: *isa*(BASIC, high level languages), *isa*(ISDN, telecommunication networks).

4.6 Identification Using other Information

There are also some of the *is-a* relations can not be established with above headword based rules. For these cases, definitions of concepts are also useful information to identify *is-a* relations, because definitions usually provide hypernyms and identity information of the concepts. We describe how definition information is explored with some examples as follows.

Using definition of NT: Given *btnt*(antenna array, antenna), the definition of "antenna array" from web [11] which derived by Google Glossary is "a group of identical antennas arranged and interconnected…", and its definition in Korean [22] is "a kind of antenna that consists of more than two arranged antennas…". It shows although "antenna array" is "a group of antennas" by its definition in English, so that it might have *part-whole* relation with "antenna", but considering that it inherits the main function properties of its BT "antenna" according to its definition in Korean, its identity is more of the same with "antenna". According to the first assumption, their relation is identified to *is-a* relation. The rule adopted in this example can be represented with: if *isa*(B, A) and *n-isa*(C,B), then we have *n-isa*(C,A).

Using definition of both BT and NT: Given relation *btnt*(course ware, computer aided instruction), the definition of NT "course ware" is "**Software** that is designed for an educational program", and the definition of BT is "Refers to a **system** of educational instruction performed almost entirely by computer". The definition of BT shows that the BT is a kind of "software system", which has the same identity with "software", and "software" is the hypernym of NT "course ware" according to its definition. So the given relation can be identified to *is-a* relation.

Using synonym and sibling information of BT: Let A, B, C be three different concepts, *syn*(B, C) be synonym or sibling relations between concepts B and C, we have rule of: if *isa*(B,A) and *syn*(C,B), then *isa*(C,A). As an example, to a given relation of *btnt*(digital television, digital communication), we know "digital television" can be considered as a sibling of "digital radio" when it is put to the relation with the common BT "digital communication". We already have *isa*(digital radio, digital communication), and so, the given relation can be identified to *is-a* relation.

Ambiguity of BT and NT: When BT or NT had more than two senses, then if they could have *is-a* relation with one of the senses of either BT or NT, then we identified their relation as *is-a* relation. To avoid abuse this rule, there was a constraint that, there should be clear statement on each sense. This rule can be expressed with: if *isa*(B, C), *isa*(B, D), and *isa*(D,A), then *isa*(B, A).

Given relation *btnt*(digital radio, digital communication). "digital radio" has senses of hardware system, technology, and broadcast. Considering "broadcast" is a kind of "data transmission", while "digital communication" is a kind of "electronic transmission of information", it is identified to *is-a* relation.

5 Non-taxonomic Semantic Relations

In Section 4, the *is-a* and *not is-a* relations are identified from BT-NT relations in thesaurus. In

[5] These tolerance related terms, and the abbreviation related terms in Section 4.5 are come from our observation in practice.

this section, non-taxonomic semantic relations are defined for IT domain ontology in Section 5.1, and then, the *not is-a* relations are identified to target semantic relations of ontology. Semantic relation is identified by two phases: first, classify both terms in a BT-NT relation to semantic categories in Section 5.2; then, identify the BT-NT relation to certain semantic relation of which domain contains the category of NT and of which range contains the category of BT.

5.1 Defining Semantic Categories and Relations

As top-level classes for domain ontology, **semantic categories** should be popular enough in certain domain. The popularity of a semantic category can be measured by the number of terms belong to the category.

220K of bilingual terms in both English and Korean expressions are collected from several IT domain dictionaries, and bout 26K of headwords with frequency information are obtained from these terms. Top 180 of the headwords with the highest frequencies are found to cover more than 23% of all terms, these terms are mapped to a general domain thesaurus CoreNet [3], by performing term classification on their headwords - as a kind of word sense disambiguation, term classification in this phase used a simple frequency based approach. The CoreNet category popularities in IT domain are measured by the number of terms belong to the category after classification. As the result, 77 of the CoreNet categories are finally selected as IT domain categories. These semantic categories, with their hierarchy borrowed from CoreNet, and the headwords classified in each category, form a small-size IT domain taxonomy.

Semantic relations are defined in terms of the semantic categories. Terms in the Inspec thesaurus are all either domains or ranges of BT-NT relations. Classification of Inspec terms into semantic categories results in a domain-range list of BT-NT relations, where domains and ranges are represented in semantic categories. Semantic relations are decided by experts. In Figure.3, given relation of *btnt*(calculus, mathematical analysis), classifying the terms, we have *btnt*(cComputation, cAnalysis). A semantic relation of "*p_isComputationFor*(cComputation, cAnalysis) can be induced by expert.

These relations with domain and range constraints are kind of relation patterns. As the result, 108 semantic relations with 258 relation patterns are defined semi-automatically, using 527 BT-NT relations in Inspec thesaurus.

5.2 Classification on BT and NT

A similarity-based approach is adopted to classify concepts in BT-NT relations to semantic categories. The similarities between a concept and semantic categories are calculated, the category showing the highest similarity with the headword of a concept is decided as the category of the concept.

Let t be the lexical expression of a concept in Inspec and h_t its headword. Let $C=\{C_1, C_2,\ldots C_i, \ldots C_n\}$ be the corresponding nodes of the semantic categories, where n=77 as it described in Section 5.1. The category of t, $C(t)$, is classified by Eq. (1):

$$C(t) = C(h_t) = \arg\max_{C} \operatorname*{Sim}_{i=1}^{n}(h_t, C_i) \qquad \textbf{(1)}$$

Assume that h_t has multi senses of m. Let s_j be one of its senses, and $depth(s_j)$ be its depth of the node in domain taxonomy. The similarity between h_t and category C_i is the maximum reciprocal of the distances between C_i and the senses of h_t in the taxonomy. The similarity is zero if a h_t is not a hyponym of C_i (Eq. (2)-(3)).

$$Sim(h_t, C_i) = \begin{cases} 0, \text{ if } h_t \text{ is not hyponym of } C_i; \\ \max_{j=1}^{m} (1/\, path(s_j, C_i)), \text{ else.} \end{cases} \qquad \textbf{(2)}$$

$$path(s_j, C_i) = 1/(depth(s_j) - depth(C_i) + 1) \qquad \textbf{(3)}$$

5.3 Identifying Semantic Relations

The semantic relations with the semantic constraint of domain and range (Table 1) can be considered as relation patterns. As it described in Section 3.2, there were 108 semantic relations defined for IT domain, with domain and range in 77 semantic categories. With the BT and NT already classified in Section 5.1, given BT-NT relations can be identified to semantic relations by adopting relation patterns.

For example, to given BT-NT relation of *btnt*(bubble chambers, particle track visualisation), NT "bubble chambers" can be classified to category "cEquipment", BT "particle track visualisation" can be classified to category "cProcessing". Considering that the specific semantic relation of *btnt*(cEquipment, cProcessing) can be only "*p_isEquipmentFor*" in the defined relations (Figure.4), the given relation can be identified as *p_isEquipmentFor*(bubble chambers, particle track visualization).

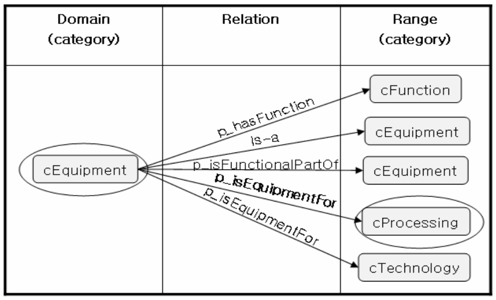

Figure**.** 3. Semantic relations used as patterns

6 Experiments

6.1 Evaluation on Is-a Relation

Based on the rules proposed in Section 4.2-4.5, a rule-based identification system of *is-a* relations is implemented. The BT-NT relations of Inspec thesaurus are used as test data. The system identifies as either *is-a* or *not is-a* relations.

Let R1 be identified relations by the system, and R2 be the relations decided by experts. The accuracy of the system is calculated by Eq. (4):

$$Accu = \frac{|R1 \cap R2|}{|R2|} \quad (4)$$

Identifying all of the BT-NT relations of Inspec to *is-a* relations by default is taken as baseline, which as 79.59%. The accuracy of the system tested with 4,999 BT-NT relations showed 85.83%, which is much higher than the baseline.

Furthermore, evaluation by the experts was done. Relation identification can be very confusing even to the experts, and so there could be different opinions among experts on a relation. That motivated us to evaluate the consistency of the identification by experts[6]. Let R3 and R4 be identified relations by different experts, the consistency is evaluated by Eq. (5). The consistency tested with 2,994 relations was 86.44%.

$$Cons = \frac{|R3 \cap R4|}{(|R3|+|R4|)/2} = \frac{|R3 \cap R4|}{|R3|} \quad (5)$$

[6] Experts identified *is-a* relations based on the rules proposed in Section 4.6 with their own knowledge and any information they can get from either web or offline knowledge.

This test result shows that the accuracy of the automatic identification system is compatible to the consistency among experts.

6.2 Evaluation on Semantic Relations

A similarity-based system described in Section 5 was implemented. As we described in Section 6.1, the BT-NT relations of Inspec thesaurus were identified to either *is-a* or *not is-a* relations. The *not is-a* relations were used as test data.

$$Accu = \frac{|R1 \cap R2|}{|R2|} \quad (6)$$

$$\text{Coverage} = \frac{|\text{ Identified NotISA relations }|}{|\text{ NotISA relations }|} \quad (7)$$

Both accuracy and coverage are evaluated with Eq. (6) and (7). Among 12,821 BT-NT relations in Inspec thesaurus, there were 3,307 relations were *not is-a* relations. They were identified to 108 non-taxonomic semantic relations. The coverage was 31.09% with about 90% of accuracy.

7 Discussion and Conclusion

In this paper we proposed how to elicit information (BT-NT/RT relations) represented in the thesaurus to the semantic information for ontology: 1) rules to identify *is-a* relations from BT-NT relations were proposed, and a rule-based system was implemented; 2) a similarity and pattern-based approach was proposed and implemented to identify semantic relations from *not is-a* relations.

Extensive experiments were performed with our proposed approaches. It shows that our approaches are very promising to build domain ontology automatically from thesaurus. The identification accuracy of is-a relations was comparable to the consistency among experts, and the accuracy of identifying non-taxonomic semantic relations from *not is-a* relations was also high enough for practical using, although coverage was relatively low compare to the accuracy.

The reason of low coverage is that many domain concepts could not be successfully classified to semantic categories, because the domain taxonomy which was automatically built does not contain enough many domain concepts in it. To improve the coverage, an ongoing work is that enriching domain taxonomy by utilizing existing lexical resources such as CoreNet, WordNet and domain dictionary.

References

[1] Harith Alani, Position paper: Ontology Construction from Online Ontologies. In *Proceedings of the 5th International Semantic Web Conference*, November 2006

[2] Mark van Assem, Véronique Malaisé, Alistair Miles, and Guus Schreiber: A Method to Convert Thesauri to SKOS. In *Proceedings in the 3rd European Semantic Web Conference*, June 2006, pp. 95-109

[3] Key-Sun Choi, Hee-Sook Bae, Procedures and Problems in Korean-Chinese-Japanese Wordnet with Shared Semantic Hierarchy , In *Proceedings of the Global WordNet Conference*, pp. 320~325, 2004.1, Brno, Czech.

[4] Philipp Cimiano, Aleksander Pivk, Lars Schmidt-Thieme, Steffen Staab. Learning Taxonomic Relations from Heterogeneous Sources of Evidence. In *Proceedings of the ECAI 2004 Ontology Learning and Population Workshop*

[5] Zhendong Dong and Qiang Dong, *HowNet and the Computation of Meaning*. Published by World Scientific Publishing Co. Pte. Ltd. 1995. pp. 36.

[6] Christiane Fellbaum.*WordNet,: An Electronic Lexical Database*. Cambridge, MA: MIT Press. 1998

[7] Roxana Girju, Adriana Badulescu and Dan Moldovan. Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations. In *Proceedings of HLT-NAACL 2003*, Edmonton, May-June 2003. pp. 1-8

[8] Roxana Girju, Adriana Badulescu, Dan Moldovan. Automatic Discovery of Part-Whole Relations. *Computational Linguistic*, 32(1), March 2006

[9] Willem van Hage, Hap Kolb, Guus Schreiber. A Method for Learning Part-Whole Relations. In *Proceedings of the 5th International Semantic Web Conference*, November 2006

[10] Inspec v2.0 Getting Started Guide. http://scientific.thomson.com/media/scpdf/inspec_gettingstarted_en.pdf

[11] Toshinobu Kasai, Haruhisa Yamaguchi, Kazuo Nagano and Riichiro Mizoguchi. Building an ontology of IT education goals. *Int. J. Cont. Engineering Education and Lifelong Learning*, Vol. 16, Nos. 1/2, pp.1-17, 2006

[12] Online Glossary provided by Broadband Cable Association of Pennsylvania: www.bcapa.com/about/glossary.php

[13] Pease, A., Niles, I., and Li, J. 2002. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, Edmonton, Canada, July 28-August 1, 2002.

[14] Alan L. Rector: Modularisation of domain ontologies implemented in description logics and related formalisms including OWL. *K-CAP 2003*: 121-128

[15] D. Sleeman, S. Potter, D. Robertson, and M. Schorlemmer. Ontology Extraction for Distributed Environments. In *Proceedings of Workshop on Knowledge Transformations for the Semantic Web* (affiliated to ECAI-02), July 2002

[16] R Snow, R., Jurafsky, D., and Ng, A. Y. Semantic taxonomy induction from heterogenous evidence. In *Proceedings of COLING / Association for Computational Linguistics*, 2006.

[17] Dagobert Soergel, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer and Stephen Katz. Reengineering Thesauri for New Applications: the AGROVOC Example. *Journal of Digital Information*, 4(4), March 2004.

[18] SUMO (Suggested Upper Merged Ontology), http://ontology.teknowledge.com/#download

[19] Wielinga, B., Schreiber, G., Wielemaker, J., & Sandberg, J.A.C. From thesaurus to ontology. In *Proceedings of International Conference on Knowledge Capture*, Victoria, Canada, October 2001

[20] Golbeck, Jennifer, Gilberto Fragoso, Frank Hartel, Jim Hendler, Jim Oberthaler, Bijan Parsia "The National Cancer Institute's Thesaurus and Ontology," *Journal of Web Semantics*, 1(1), December 2003.

[21] Sin-Jae Kang and Jong-Hyeok Lee, Semi-Automatic Practical Ontology Construction by Using a Thesaurus, Computational Dictionaries, and Large Corpora, In *Proceedings of ACL 2001 Workshop on Human Language Technology and Knowledge Management*, Toulouse, France, July 6-7, 2001

[22] Telecommunications Technology Association, Information and Telecommunication Glossary, http://word.tta.or.kr/faq/faq.jsp

[23] Asanee Kawtrakul, Aurawan Imsombut, Aree Thunkijjanukit, Dagobert Soergel, Anita Liang, Margherita Sini, Gudrun Johannsen, and Johannes Keizer, Automatic Term Relationship Cleaning and Refinement for AGROVOC, *Workshop on The Sixth Agricultural Ontology Service, July 25-28, 2005. Vila Real, Portugal*.