Annotation of Computer Science Papers for Semantic Relation Extraction

Yuka Tateisi†, Yo Shidahara‡, Yusuke Miyao†, Akiko Aizawa†

†Research Center for Knowledge Media and Content Science, National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo101-8430,Japan
‡Freelance Annotator
E-mail: yucca@nii.ac.jp, yo.shidahara@gmail.com, yusuke@nii.ac.jp, aizawa@nii.ac.jp

Abstract

We designed a new annotation scheme for formalising relation structures in research papers, through the investigation of computer science papers. The annotation scheme is based on the hypothesis that identifying the role of entities and events that are described in a paper is useful for intelligent information retrieval in academic literature, and the role can be determined by the relationship between the author and the described entities or events, and relationships among them. Using the scheme, we have annotated research abstracts from the IPSJ Journal published in Japanese by the Information Processing Society of Japan. On the basis of the annotated corpus, we have developed a prototype information extraction system which has the facility to classify sentences according to the relationship between entities mentioned, to help find the role of the entity in which the searcher is interested.

Keywords: Relation Annotation, Information Retrieval, Document Classification

1. Introduction

Intelligent search engines are in high demand for digital archives of academic research papers. Usually a researcher has a specific search context in mind, such as what is a conditional random field (CRF), what are CRFs used for, and how to improve the performance of CRFs. Currently, almost all search engines are based on matching sets of key words or phrases such as *conditional random field*, and all documents including the phrase or its variants (such as *CRF*) are subsequently displayed without considering the search context. Thus, the researcher has to search for what he/she wants from the results. Classifying the results and grouping them to represent the search context would make the process easier.

Such a facility requires identifying the role of entities and events that are expressed by the key word/phrase. We aim to construct a framework of this identification in the computer science research paper domain, on the assumption that the role can be determined according to the relationships between various entities and events.

We visualise that a computer science paper describes a world with a two-level structure. At the first level, a system, technique, application domain, and the relationship among them are described. At the second level, the author(s) of the paper is involved. In computer science, and arguably in applied science and engineering in general, the author(s) usually creates something (a system, an algorithm, a set of data etc.), uses it for some purpose, and evaluates its performance or investigates the nature of it.

The extraction of such relationships between the author(s) and the entities or events described in the second level of the paper according to the role they play directly suits our purpose of classifying the search results. However, relationships between described entities and events are also necessary for a search system to be useful, as it is often the case that a searcher has questions in mind such as *what can this technology be used for?* and *how good is the system for such-and-such type of data?* In addition, the two levels often interact, in the sense that a relationship at one level is indicative of one at the other.

It is usually the case that information expressed in the relationship at the first (entity-entity) level correlates with the relationship at the second (author-entity) level; therefore, identification of the relationship at both levels may not be separated clearly. In some cases, the relation between the entity and the author(s) may be expressed in the form of relations between entities. For example, in *a specific AL algorithm we have developed is particularly effective in reducing annotation costs*, the fact that the authors have developed the algorithm is explicitly described; however, the facts that the authors used the algorithm for reducing annotation costs and that they regard the algorithm effective for that purpose are not. They can only be deduced from the relation between *algorithm* and *reducing*, and *algorithm* and *effective*. In other words, the relations between the entities and events can be clues to the type of the author involvements.

However, the type of the author involvement can also predict the relations that hold for the described entities and events. For example, creation of a machine learning-based system accompanies the algorithm it is based on and the data for its training; the use of such a system accompanies its purpose and, possibly, characteristics of its users. Evaluation accompanies the results obtained and the evaluation setting.

From these observations, we try to capture the relationships at both levels in a flat structure. This paper describes a corpus annotated with semantic relations according to the view described above. Our corpus is based on computer science research abstracts in Japanese wherein semantic relations among concepts are identified. Unlike traditional relation corpora, our annotation exhaustively covers the relations described in the abstracts. Using the annotated corpus, we expect to achieve a relation-based, refined search system by 1) learning the mapping from the annotated text to relations; 2) mapping a new text to a set of relations based on the concept described in the text; 3) grouping the texts, based on their concepts and relations; 4) showing the group of texts that match the user's query. We also introduce a prototype of such a search system.

2. Related works

Identification of the role of entities described in scientific and technical literature is pursued in technical trend analysis. Gupta and Manning (2011) analyse research articles in terms of Focus (main contribution), Domain (application domain) and Technique (a method or tool used in the article). They have created a corpus on 462 abstracts from the ACL Anthology, in which mentions of Focus, Domain and Technique are annotated, but not relations among them. They use syntactic relations between the entities obtained by a dependency parser as clues for entity classification.

Similarly, in Japanese, Fukuda et al. (2012) extract the Technology and Effect (the Attribute-Value pairs, indicating the results of using the Technology) from patent documents and research papers. The corpus used in the research (Nanba et al., 2010) has entities annotated with Technology, Effect, Attribute and Value. Relationship is not annotated, except in the matching of Attribute and Value.

As the aim of technical trend analysis is to summarise technology trends over time from the literature, it is sufficient to relate the entity mentions to a paper and to determine how the authors of the paper view the entity. As such, they focus on entity classification according to the author's view, but not on identification of relationships between entities.

Extraction of domain-specific events and relations is actively pursued in the biomedical field (Nédellec et al., 2013). A particular feature in biomedical domain is that there are several established ontologies such as Gene Ontology (The Gene Ontology Consortium, 2000), and many studies seek to map the representation in a natural language text to concepts and relations in such ontologies. Another feature of those studies is that they focus on more specialised sub-domains such as gene regulation and protein–protein interaction.

Outside the domain of academic literature, domain-specific roles of entities and relations are annotated in a coprus used for generating flow graphs from culinary recipes (Mori et al., 2014). Entities are tagged with classes based on roles in recipes such as F (food), T (tool), and Ac (action by chef). Using these classified instances of entities as vertices, and connecting them with arcs labelled with relations such as d-obj (direct object of an action), i-obj (indirect object of an action), F-comp (food complement) and F-eq (food equality), they annotate a recipe text so that one flow graph representing the cooking procedure written in the recipe is formed.

Our purpose in classification of the statements is partially achieved through the technique of zoning (Liakata et al., 2010; Guo et al., 2011; Varga et al., 2012), which aims to classify the sentences or paragraphs in the literature to help identify components such as the proposed method or the results obtained in the study. Zoning can be regarded as a means to identify the relation between the author and the propositions stated in the paper, but zoning analysis is not concerned with the relation between the entities involved in events described by the proposition.

In another type of research that identifies the relationship between author(s) and the described entities and events, Thompson et al. (2011) annotate the meta-knowledge such as polarity and the certainty level of the author, as well as whether the event is attributed to the current or previous studies, on top of the annotations of the GENIA event corpus (Kim et al., 2008). For relations within and between the events, they rely on the underlying GENIA annotation.

Search refinement using relations is implemented in the ACL Anthology Searchbench (Schäfer et al., 2011), where an HPSG parser is used to analyse the sentences in computational linguistic research papers to automatically display the extracted topic of a paper, and to search with phrases that have predicate-argument structures to refine the search. For example, the search key *p:improve r: parsing accuracy* returns sentences with *improve* as predicate and *parsing accuracy* as its object, and *s:SVM r:POS tagger* returns sentences with *SVM* as a subject of a predicate and *POS tagger* as the object of the same predicate. Thus, the searcher can relate entities and predicates, but interpretation of the predicate–argument relations is left to the searcher.

3. The annotation scheme

3.1. Basic principles

As stated in Section 1, we annotate relations related to the view of the author of a paper and the entities described therein, and the relationship between the entities, in a flat structure. We attempt to capture this semantically motivated relation structure directly, in contrast to traditional efforts that capture the structure from syntax. This is because semantic relations can be expressed by various linguistic mechanisms such as lexical (e.g. *based* in *CRF-based POS taggers*), syntax (e.g. compare *applying something to CRF* and *applying CRF to something*), and discourse (e.g. *We tried CRF, but the method was not effective in improving accuracy*).

In designing the scheme, we decided not to commit to predefined classifications and took a rather loose approach with entity and relation types: we set minimal classifications of entities, and decided to simultaneously define the relations while annotating the sample texts.

We consider that, unlike biosciences, use of predefined ontologies in the computer science domain is not feasible. One reason for this is simply that no standard, widely-used ontology is available, but we also doubt that a domain ontology in computer science is as helpful as in BioNLP. The applied domain of BioNLP is genomics, a branch of basic natural science which aims to discover the fact or to explain the phenomena involving proteins, genes and other biochemical entities. The intrinsic characteristics of biochemical entities are the target of interest, and domain ontologies that describe the knowledge intrinsic to the entities play an important role in tasks such as constructing databases and extraction of knowledge from literature. In computer science and other applied science, however, the interest is more towards the application of discovered facts and phenomena to some practical purpose, and development of technology to this end. In searching information, the searchers are more interested in the practical use or performance of the entities (such as computer systems) than their intrinsic nature. Therefore, the nature of an entity becomes more relative, determined by the purpose, method and other attributes. In this context, a domain ontology is not as useful as in the context of pure science. Rather, the relation between entities that determines the relative nature of entities is of more interest. In addition, as computer systems are applied to almost all aspects of human activities, a complete ontology of computer science must cover virtually everything in the world, an unfeasible aim.

In contrast, the relations defined in linguistically motivated domain-general frameworks such as WordNet[1] and FrameNet[2] do not sufficiently cover the relations in which we are interested. WordNet has only hyponym–hypernym relationships, and although FrameNet expresses the relations between entities as the frames associated with entities and events, interesting technical entities such as algorithm and corpus are not yet covered.

In addition, FrameNet imposes a semantic type of frame arguments, which we consider to be too restrictive for the current purpose. For example, the frame Education_teaching, to which verbs *learn* and *train* are assigned, has Student and Teacher as its core arguments. The two arguments have type restriction, that is, they must be sentient. This is not the case in a machine learning setting where verbs *learn* and *train* are also used. There a program is *trained* by another algorithm using a set of data, and *learns* a model, and none of these participants in the machine-learning events is sentient themselves. They are just represented by metaphorically using the words for sentient participants. At the same time, a program described as if it is a sentient entity has its attributes as a data file, such as size and format, and other attributes as created artefact such as creator and date of creation, which can also be discussed in a paper. Thus, a program is described as both sentient and non-sentient depending on the context. In other words, the type of a program can be determined by context, in relation to other concepts that appear together. Other upper ontologies have the same concerns, i.e. coverage and type restriction, as Framenet. From these observations we did not rely on a base ontology.

3.2. Entities and relations

We investigated 71 computer science abstracts (498 sentences), and defined an annotation scheme comprising three types of entities and 16 types of semantic relations. Entities

Type	Definition	Example
OBJECT	the name of concrete entities such as a program, a person, and a company	Origin 2400, IPSJ Journal
MEASURE	value, measurement, necessity, obligation, expectation, and possibility	novel, 136.72
TERM	any other	

Table 1: Entity Tags

[1] http://wordnet.princeton.edu/

[2] https://framenet.icsi.berkeley.edu/fndrupal/

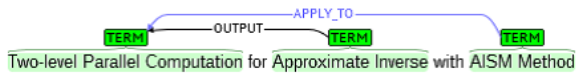

Figure 1: Annotation example

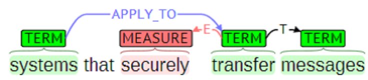

Figure 2: Another annotation example. Note that some arc labels are abbreviated because of space restriction.(E:EVALUATE, T:TARGET)

are items involved in relations, including concepts denoted by verbal and adjectival expressions, classified as shown in 1. An entity is connected to another with directed binary relations summarised in Table 2.

Our relation annotation scheme is based on the concepts related to a computer system, which typically has an input, an output, a method for achieving a function, and its application. For example, in *two-level parallel computation for approximate inverse with AISM method*, *approximate inverse* is the output (result) of *two-level parallel computation*, and *AISM method* is the method for the computation. With our scheme the APPLY_TO relation is annotated from *AISM Method* to *two-level parallel computation*, and the OUTPUT relation is annotated from *approximate inverse* to *two-level parallel computation* (Figure 1).

In addition to the relations that describe the function of a system, we have relations related to its evaluation of such as EVALUATE denoting evaluation results, COMPARE denoting the results of other systems compared and RESULT relating the experiment and the result. Our scheme also includes more general relationships such as ATTRIBUTE, CONDITION, and SUBCONCEPT, and relations between textual representations such as EQUIVALENCE and SPLIT.

We further enhance the use of these relation labels to relate actions and their participants. This is based on the observation that computer programs are often expressed by their function, denoted by verbal expressions. For example, *parser output* and *result of parsing* can denote the same concept. Thus, in the latter case, *result* is related to *parsing* with an OUTPUT relation. Similarly, in *a protocol that combines biometrics and zero-knowledge proof*, *protocol* is the OUTPUT of *combines* as it is the product of combining, and *biometrics* and *zero-knowledge proof* are INPUT of combines as they are the materials for combining.

Note that, by *entity*, we mean all the concepts that can be involved in relations. Our notion of entities includes actions denoted by verbs and verbal phrases, and evaluation denoted by adjectives, adverbs and modals. This is because important concepts such as the function of a system are often denoted by constructions other than simple noun phrases. For example, in *systems that securely transfer messages*, the function of the system is to transfer messages securely. In our annotation, *transfer* and *messages* are annotated as TERMs which are related by a TARGET relation, and *safely* as a MEASURE which is related to *transfer* with EVALUATE. Finally, *transfer* is related to *systems* by

Type	Definition	Example
APPLY_TO(A, B)	A method A is applied to achieve the purpose B or used for conducting B	CRF_A-based $tagger_B$
RESULT(A, B)	A results in B in the sense that B is either an experimental result, a logical conlusion, or a side effect of A	$experiment_A$ shows the $increase_B$ in F-score compared to the baseline
PERFORM(A, B)	A is the agent of an intentional action B	a frustrated $player_A$ of a $game_B$
INPUT(A, B)	A is the input of a system or a process B, A is something obtained for B	$corpus_A$ for $training_B$
OUTPUT(A, B)	A is the output of a system or a process$B$, A is something generated from B	an $image_a$ $displayed_B$ on a palm
TARGET(A, B)	$A$is the target of an action B, which does not suffer alteration	to $drive_B$ a bus_A
ORIGIN(A, B)	A is the starting point of action B	to $drive_B$ from $Shinjuku_A$
DESTINATION(A, B)	A is the ending point of action B	an image $displayed_B$ on a $palm_A$
CONDITION(A, B)	The condition A holds in situation B, e.g, time, loacation, experimental condition	a $survey_B$ conducted in $India_a$
ATTRIBUTE(A, B)	A is an attribute or a characteristics of B	$accuracy_A$ of the $tagger_B$
STATE(A, B)	A is the sentiment of a person B other than the author, e.g. a user of a computer system or a player of a game	a $frustrated_A$ $player_B$ of a game
EVALUATE(A, B) COMPARE(C, B)	A is evaluated as B in comparison to C	experiment shows an $increase_B$ in $F-score_A$ compared to the $baseline_C$
SUBCONCEPT(A, B)	A *is-a*, or is a *part-of* B	a $corpus_A$ such as PTB_a
EQUIVALENCE(A, B)	terms A and B refer to the same entity: definition, abbreviation, or coreference	DoS_B ($denial-of-service_A$) attack
SPLIT(A, B)	a term is split by parenthesical expressions into A and B	DoS_B (denial-of-service) $attack_A$

Table 2: Relation Tags

an APPLY_TO relation to represent that the function of the system is to transfer (Figure 2).

3.3. Preliminary experiment

As a preliminary experiment, two of the current authors annotated 30 abstracts from the IPSJ Journal (a monthly peer-reviewed journal published by the Information Processing Society of Japan) (Tateisi et al., 2013). The annotation was performed using the brat annotation tool (Stenetorp et al., 2012), and no automatic preprocessing was performed.

From 197 sentences in the 30 abstracts 1895 entities and 2269 relations were identified by at least one annotator. Agreement in F-score was calculated in the same manner as previously reported (Brants, 2000), which was 93% for entities and 59.8% for relations. Further analysis revealed several patterns of problematic cases such as evaluation of a feature or an aspect of an event (Figure 3) and causal relationships not marked overtly (Figure 4).

In Figure 3, one annotator interprets *real time*-ness as an attribute of *its processing*, and the other interprets that *real time*-ness is required in the context of *its processing*. Note that, in cases like the one shown in Figure 3, the structure difference counts as two instances of disagreed annotations.

In Figure 4, one annotator recognised the relation but the other interpreted the construction as a simple coordination. There are no overt clues for a causative relation, but it is not considered a particularly bad writing style in Japanese, and not uncommon in academic writing.

If the instances where only one annotator marked up (including the ATTRIBUTE and CONDITION relations in Figure 3 and the RESULT relation in Figure 4) are ignored, the F-scores increase to 96.1% for entities and 76.1% for relations. In instances where both annotators recognised relations, the pair of labels with the lowest agreement

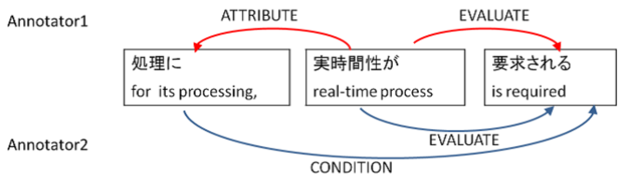

Figure 3: Disagreement in evaluating an asect of an event

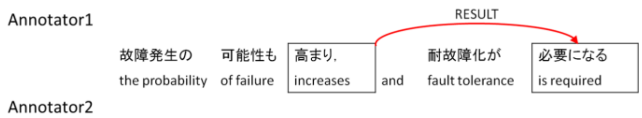

Figure 4: Implicit causal relationship

is OUTPUT-TARGET (85.1% observed agreement and 62.2% in Cohen's kappa), followed by INPUT-TARGET (86.8%/70.2%), INPUT-OUTPUT (87.5%/74.2%) and ATTRIBUTE-CONDITION (88.5%/74.8%). On all other pairs the two annotation results agreed with more than 80% in kappa score. These results indicated that if the relation structure can be identified the relation labels are reliably determined, but distinguishing between the input-output of a system and the target of an action of a system is not so clear. Further detail of the experiment is discussed in our previous paper (Tateisi et al., 2013).

The disagreements in the 30 papers were resolved to produce a gold standard dataset. We also revised the annotation manual to include guidelines regarding the problematic cases found in the preliminary experiment.

Label Pair	Annotator-Gold	Preliminary
ATT-PER	97.1/82.1	96.1/82.3
IN-TAR	93.9/87.4	86.8/70.2
OUT-TAR	96.0/90.6	85.1/62.2
IN-OUT	95.6/90.9	87.5/74.8
APP-IN	96.3/92.5	91.3/82.2

Table 3: Label pairs of low agreement. The numbers before each slash show observed agreement (%) and the numbers after the slash show Cohen's kappa(%). ATT = ATTRIBUTE, APP = APPLY_TO, IN = INPUT, OUT =OUTPUT, PER = PERFORM, TAR = TARGET

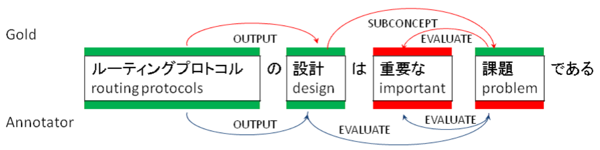

Figure 5: Disagreement involving 課題 (*problem*, *issue*)

4. Annotation experiment

The agreement between an independent annotator and the gold standard was studied to determine whether consistent annotation is possible for a general researcher using the scheme. The annotator, a Japanese researcher in the BioNLP field, was presented with the guidelines and with 100 abstracts not used in the first experiment, but which were also annotated by the authors according to the guidelines for reference. Then, we asked her to annotate the same 30 abstracts used in the first experiment without further instructions or automatic pre-processing.

The agreement between the result and the gold standard in F-score was 96.4% for entities and 79.1% for relations. If the instances found only in either the gold standard or the annotator's result are ignored, the score was 98.0% for entities and 90.3% for relations. These figures suggest that entities and relations can be reliably annotated with example-based training, although the scale of the experiment is small.

For relation instances annotated both in the gold standard and the annotator's results, the pair of labels with the lowest agreement is ATTRIBUTE-PERFORM (97.1%/82.1%). OUTPUT-TARGET, the pair with the lowest agreement in the preliminary experiment, achieved 96.0%/90.6% agreement. The lowest five pairs are shown in Table 3, together with the scores in the preliminary experiment.

By manually comparing the first ten pairs of abstracts (621 entities and 622 relations in the gold standard), we determined that most entity disagreements (16 out of 20) are type mismatches between TERM and MEASURE resulting from in consistency in how to annotate the constructions concerning the EVALUATE relation.

In particular, words like 問題 and 課題, which mean *problem*, *issue*, *task* or *difficulty*, depending on context, are difficult to judge. For example, Figure 5 shows different annotations on a phrase meaning *the design of routing protocols is an important problem*. The words judged as TERMs are lined in green and the words judged as MEASUREs are lined in red. The illustration shows that the interpretation of word

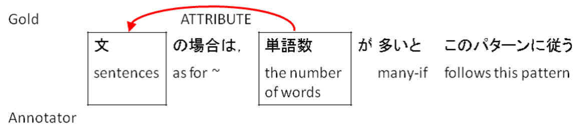

Figure 6: Implicit ATTRIBUTE relationship

課題 (*problem*) resulted in a disagreement; Gold considers it as value-neutral such as technical question to answer and tags it as a TERM, but Annotator considers it as a negative judgement such as difficulty to overcome and tags it as a MEASURE.

Another type of TERM-MEASURE confusion occurs in numeric expressions, in attributive constructions such as 二つの問題 (*two problems*), more often in the gold standard annotation. We had determined numbers to be MEASUREs in the guidelines; therefore, this indicates that there are still errors to be corrected in the gold standard.

The causes of disagreement in relation annotation were varied, but we noted several characteristic cases. Most common is the confusion between EVALUATE and other relation types. This usually involves the TERM-MEASURE confusion discussed above, and an example is shown in Figure 5.

The disagreement also revealed two essentially difficult patterns. One is the cases of implicit relationship. These include the problem of causal relationship shown in Figure 4 in the previous section, and other cases when the relation is between entities that are far apart. A typical example is exemplified by the sentence 文の場合は，単語数が多いとこのパターンに従う (*A sentence follows this pattern when it has a large number of words*), shown in Figure 6. In English, the number of words is shown to be an attribute of the sentence via the pronoun *it*, but in Japanese it is not necessary to use a pronoun; therefore, this relation is not expressed overtly.

Annotation of implicit relations is desirable, but sometimes it requires domain knowledge to decide if a relation must be annotated. For example, in 移動コンピュータ間の無線マルチホップ配送を用いるアドホックネットワーク (*ad-hoc networks where wireless multihop message transmission among multiple mobile computers*), it requires knowledge of an ad-hoc network that it contains mobile computers as its component, and therefore 移動コンピュータ (*mobile computers*) cannot be related to アドホックネットワーク (*ad-hoc networks*) without domain knowledge. After the experiment, we made a new decision that annotators must exclude the use of domain knowledge as much as possible.

The other difficulty concerns the attachment ambiguity of modifiers. For example, in 製造不確実時代における回路のディペンダブル動作を保障する電源配線最適化手法 (*an efficient power grid optimisation algorithm to secure dependable operation of circuits in a manufacture-uncertainty age*), the noun-modifying phrase 製造不確実時代における (*in a manufacture-uncertainty age*) can modify either 回路 (*circuits*), ディペンダブル動作 (*dependable operation*) , or 電源配線最適化手法 (*power grid optimisation algorithm*). This resulted in a disagreement concerning what is related with 製造不確実時代 via the CONDITION

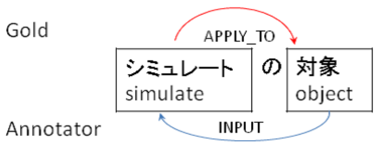

Figure 7: Disagreement involving *simulate*

relation.

As for the label disagreements shown in Table 3, ATTRIBUTE-PERFORM confusion was not found in this subset, but APPLY_TO-INPUT had eight instances. These cases are further divided into two categories. One is when an entity is used as a model or a criterion. The other is the relation between the verb *simulate* and what is simulated. In this case, the direction of the relation is also reversed (Figure 7). These indicate that the distinction of use as a method (APPLY_TO) and use as data (INPUT) is not clear-cut. The model/criterion cases suggest a class of entities that are somewhere in between data or a method. The *simulate* case can be explained by the nature of what is simulated. The objects of simulation are usually events or actions of a system or a machine, but they are presented to a simulation program as input data. Gold takes the former view, interpreting the simulation event as 'applying a process of simulation to an event or action to produce a simulated event'. Annotator takes the latter view, to interpreting the simulation event as 'invocation of a simulation process which has (some parameters of representing) the event or action as input and produce a simulated event as output'.

For those involving INPUT, OUTPUT, and TARGET, there were two instances each of INPUT-TARGET and OUTPUT-TAGET, and five instances of INPUT-OUTPUT. The instances are all the relations between verbs and their direct objects, and a group of verbs related to findings such as 検出 (detect), 指摘 (point out), 評価 (evaluate) and 検討 (examine) comprised five out of nine cases. These results also indicate that, for particular semantic classes, the current scheme cannot effectively handle them.

5. Relations as content classifiers

Using the 130 abstracts used in the two experiments, we constructed a prototype of an SVM-based relation extractor (report in preparation) and a document search system with content classification according to recognised relations. Currently, we are in the process of testing and improving the system in both performance and user interface. The system enables users to search from 3300 abstracts from the IPSJ journal on a web browser interface. The screenshots are shown in Figures 8 and 9. Users can input a keyword or a phrase, including regular expressions, to the upper left window shown in Figure 8. The sentences included in the abstracts are shown in the lower window. They also have the link to the original abstract and the result of automatic annotation shown in brat.

The resulting sentences can be shown as one list, or grouped according to relations automatically annotated by the system, as shown in Figure 8. The original key is shown in yellow, and the related term is shown in green. In the result shown in 8, part of the output is classified into four

Figure 8: Search with content classification

Figure 9: Search with the key phrase and term in relation

groups. The first group consists of sentences in which the key (*SVM*) is related to other terms by APPLY_TO, which indicates that they are sentences stating that SVM (support vector machine) is applied to or used for some purpose. The second group consists of the sentence with OUTPUT relations in which the SVM is the output (result) of some process. In the third group SVM outputs some results, which is represented by the OUTPUT relation in the other direction. In the last group, SVM is related to other terms by the APPLY_TO relation in the opposite direction of the one in the first group, which indicates that some method is applied to the SVM.

Terms in relation to the key are shown in the right window, with frequency. The user can narrow down the results by selecting the term. Figure 9 shows the results of selecting 構築 (*construction*) shown in pink. The first sentence in the lower left window of Figure 9 states that the purpose of the paper which includes the sentence is to construct multiple hypotheses with a high cumulative recognition rate with an SVM. In this sentence, SVM is a method of construction, represented by the APPLY_TO relation. Conversely, the second sentence states that the authors construct a parametric model and an SVM. In this sentence, SVM is the object of construction, represented by the `OUTPUT` relation meaning that the SVM is the output of the constructing action.

Thus, our annotation can be used for classifying sentences according to the role of the entity represented by the user's keyword in the sentence using the relationship that holds between the entity and others stated in the sentence. The current implementation as a prototype system classifies the sentence using the relation annotated on the corpus straightforwardly. We are investigating whether this classification fits the users' needs well, and if not, how to map the annotated relation into an easy-to-use classification.

6. Discussions

The results of the annotation experiments indicate that consistent annotation is achieved with example-based training, although the experiment is on a small scale. However, the disagreement instances found in the experiment also suggest a further direction for refining the scheme.

Our annotation scheme is not based on an existing theory of ontological systems, but the results suggest that it reflects a systematic interpretation of texts, and a human annotator can learn the interpretation system through examples. The interpretation system can be regarded as intuitive, implicit ontology, although it is far more loosely defined compared to the common notion of ontology. Mapping the classes of entities based on the relations such as 'entities that can be an `INPUT` of a particular type of system' to formal ontology may be an interesting theme of research. At the same time, applicability of our annotation to papers outside computer domain is worth investigating.

We are also aware that our current classification based on relations can be made more accurate by combining with the results of zoning research, where the roles of sentences or paragraphs in the paper are identified as stating background, method, results and others. The role of an entity in relation to the author of the paper in which it is described is affected by the role of the statement in which it appears. For example, a problem in a system described in a paper is quite differently judged by the author when it is described in a background section or in an evaluation section.

7. Conclusion

We presented an annotation scheme for semantic relations and showed that semantic relation extraction forms a basis for a search system with content a classification facility. The annotated corpus of 130 abstracts and the guidelines in Japanese are available on request.

8. Acknowledgements

This study was partially supported by the Japan Ministry of Education, Culture, Sports, Science and Technology Grant-in-Aid for Scientific Research (B) No. 22300031.

9. References

Brants, T. (2000). Inter-annotator agreement for a German newspaper corpus. In *Proceedings of LREC 2000*.

Fukuda, S., Nanba, H., and Takezawa, T. (2012). Extraction and visualization of technical trend information from research papers and patents. In *Proceedings of the 1st International Workshop on Mining Scientific Publications*.

Guo, Y., Korhonen, A., and Poibeau, T. (2011). A weakly-supervised approach to argumentative zoning of scientific documents. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 273–283.

Gupta, S. and Manning, C. (2011). Analyzing the dynamics of research by extracting key aspects of scientific papers. In *Proceedings of 5th IJCNLP*.

Kim, J-D., Ohta, T., and Tsujii, J. (2008). Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9.

Liakata, M., Teufel, S., Siddharthan, A., and Batchelor, C. (2010). Corpora for conceptualisation and zoning of scientific papers. In *Proceedings of LREC 2010*.

Mori, S., Maeta, H., Yamakata, Y., and Sasada, T. (2014). Flow graph corpus from recipe texts. In *Proceedings of LREC 2014*.

Nanba, H., Fujii, A., Iwayama, M., and Hashimoto, T. (2010). Overview of the patent mining task at the NTCIR-8 workshop. In *Proceedins of the 8th NTCIR Workshop Meeting*.

Nédellec, C., Bossy, R., Kim, J-D., Kim, J-J., Ohta, T., Pyysalo, S., and Zweigenbaum, P. (2013). Overview of bionlp shared task 2013. In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 1–7, Sofia, Bulgaria, August. Association for Computational Linguistics.

Schäfer, U., Kiefer, B., Spurk, C., Steffen, J., and Wang, R. (2011). The ACL anthology searchbench. In *Proceedings of the ACL-HLT 2011 System Demonstrations*, pages 7–13.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations Session at EACL 2012*.

Tateisi, Y., Shidahara, Y., Miyao, Y., and Aizawa, A. (2013). Relation annotation for understanding research papers. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 140–148, Sofia, Bulgaria, August. Association for Computational Linguistics.

The Gene Ontology Consortium. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29.

Thompson, P., Nawaz, R., McNaught, J., and Ananiadou, S. (2011). Enriching a biomedical event corpus with meta-knowledge annotation. *BMC Bioinformatics*, 12.

Varga, A., Preotiuc-Pietro, D., and Ciravegna, F. (2012). Unsupervised document zone identification using probabilistic graphical models. In *Proceedings of LREC 2012*, pages 1610–1617.