Exploiting Wikipedia-based Information-rich Taxonomy for Extracting Location, Creator and Membership Related Information for ConceptNet Expansion

Marek Krawczyk[1], Rafal Rzepka[2], and Kenji Araki[2]

[1] Future Processing, ul. Bojkowska 37A 44-100 Gliwice, Poland,
mkrawczyk2@future-processing.com

[2] Hokkaido University, Kita-ku, Kita 14, Nishi 9, Sapporo, Japan
{rzepka, araki}@ist.hokudai.ac.jp

Abstract. In this paper we present a method for extracting IsA assertions (hyponymy relations), AtLocation assertions (informing of the location of an object or place), LocatedNear assertions (informing of neighboring locations), CreatedBy assertions (informing of the creator of an object) and MemberOf assertions (informing of group membership) automatically from Japanese Wikipedia XML dump files. We use the Hyponymy extraction tool v1.0, which analyses definition, category and hierarchy structures of Wikipedia articles to extract IsA assertions and produce information-rich taxonomy. From this taxonomy we extract additional information, in this case AtLocation, LocatedNear, CreatedBy and MemberOf types of assertions, using our original method. The presented experiments prove that both methods produce satisfactory results: we were able to acquire 5,866,680 IsA assertions with 96.0% reliability, 131,760 AtLocation assertion pairs with 93.5% reliability, 6,217 LocatedNear assertion pairs with 98.5% reliability, 270,230 CreatedBy assertion pairs with 78.5% reliability and 21,053 MemberOf assertions with 87.0% reliability. Our method surpassed the baseline system in terms of both precision and the number of acquired assertions.

Keywords: common sense knowledge, knowledge extraction, ConceptNet

1 Introduction

The effectiveness of systems dealing with textual-reasoning tasks depends on the scope of large-scale general knowledge bases they utilize. Just to enumerate few examples of such bases we could mention Cyc [1], YAGO [2] and ConceptNet [3]. In this paper we will focus on the last of the three - ConceptNet, a knowledge representation project that provides a large semantic graph describing general human knowledge. ConceptNet was designed to contain knowledge collected by Open Mind Common Sense project's website [4]. Further releases incorporated knowledge from similar websites and online word games which automatically collect general knowledge in several languages. Current goal of ConceptNet is to expand the knowledge base with data mined from Wiktionary[3], a multilingual, web-based free content dictionary, and Wikipedia[4], a free-access, free content Internet encyclopedia. This open-source knowledge base is used for many applications such as topic-gisting [5], affect-sensing [6], dialog systems [7] and so on. Manual expansion of the knowledge base would be a long and labor-intensive process, as seen in nadya.jp[5], an online project aiming at gathering knowledge by using a game with a purpose [8]. Since its launch in 2010 it was able to introduce little over 43,500 entries to the ConceptNet. It is therefore evident that we need to employ automatic methods to gather new data.

Projects such as NELL [9] or KNEXT [10] aim at extracting semantic assertions from unstructured text data found on the Internet. Alternatively we could transfer information from the existing semi-structured sources into a knowledge base. As a considerable amount of human validation has already been involved in the process of creating such sources, the reliability of information gathered this way would be considerably higher. Wikipedia is probably the best example of open-source, large-scale information pools. Apart from previously mentioned YAGO, DBpedia project also aims at transferring knowledge gathered in Wikipedia into more formalized, digitally processable form [11]. English part of DBpedia has already been merged to ConceptNet, however the Japanese part has not been transferred yet, leaving this part of the knowledge base at the size of roughly 1/10 of the English language domain. The problem with using DBpedia repository is that the information gathering algorithms used to prepare the knowledge base were designed for multilingual input processing and therefore introduce a considerable amount of noise. As the knowledge gathered in ConceptNet is in considerable proportion language-specific, it is vital to widen the scope of Japanese part independently.
The current paper elaborates on efforts of [12]. We extended the scope of acquired assertions as well as explored possibilities of deriving commonsense knowledge from instance related information triplets.

2 Hyponymy relation as IsA relation

In our approach we use the Hyponymy extraction tool v1.0[6], an open-source program for extracting hyponymy relation pairs from Wikipedia's XML dump files. The tool has been developed specifically to process Japanese language entries. It consists of four modules, three of which deal with extraction of hyponymy pairs from different parts of Wikipedia content: definition, category and hierarchy

[3] http://www.wiktionary.org/
[4] http://www.wikipedia.org/
[5] http://nadya.jp/
[6] http://alaginrc.nict.go.jp/hyponymy/

structures [13]. The program utilizes Pecco library[7] (SVM-like machine learning tool) to assess the plausibility level of the extracted hyponymy relation pairs and boost the precision and recall of the system [14]. The extracted hyponymy pairs may be transferred to ConceptNet as two concepts related to each other by IsA relationship (Table 1 lists examples of the extracted pairs). According to [15] these pairs are not informative enough to be useful for NLP tasks such as Question Answering, however they do fall into the scope of ConceptNet, a domain representing commonsense and general knowledge. They are simple enough not to interfere with the ConceptNet's usage flexibility, yet informative enough to introduce new and valuable input to the knowledge base.

Table 1. Examples of extracted 'IsA' relationship pairs.

Hypernym	Hyponym
kouen [8] (park)	*Motomiya-kouen* (Motomiya Park)
kougu (tool)	*baisu* (vice)
Werudaa Bureemen-no senshu (Werder Bremen player)	Klaus Allofs
Nihon-no SF shousetsu (Japanese SF novel)	*Maikai Suikoden* (Hell's Water Margin)

3 Extracting other relations

The fourth module of the Hyponymy extraction tool v1.0 generates intermediate concepts of hyponymy relations using the output of the first three modules [15]. The tool executes the following procedure: first it acquires basic hyponymy relations from Wikipedia using the method proposed by [14]. Next, it augments each acquired hypernym with the title of the Wikipedia article from which the basic hyponymy relation was extracted and consolidates the basic hypernym with the newly generated augmented hypernym (so called 'T-INTER'). Finally it generates additional intermediate concept ('G-INTER') by generalizing the enriched hypernym. As a result, it acquires four-level, information-rich hyponymy relations.

Examples of augmented hyponymy relations include: *tojo-jinbutsu* (character) – *SF eiga no tojo-jinbutsu* (character of SF movie) – *WALL-E no tojo-*

[7] http://www.tkl.iis.u-tokyo.ac.jp/∼ynaga/pecco/

[8] All Japanese language phrases are transliterated and written in italics.

jinbutsu (character of WALL-E) – M.O; *seihin* (product) – *kigyo no seihin* (product of a company) – *Silicon Graphics no seihin* (product of Silicon Graphics, Inc.) – IRIS Crimson; *sakuhin* (work) – *America no shosestu-ka no sakuhin* (work of American novelist) – *J.D. Salinger no sakuhin* (work of J.D. Salinger) – A boy in France; *machi* (town) – *England no shu no machi* (town in a county in England) – *East Sussex no machi* (town in East Sussex) – Uckfield. As we can see from the examples, the generated augmented hypernyms are too specific to be incorporated into ConceptNet directly. However some additional information about their corresponding hyponyms may be extracted from them, such as information concerning location, neighboring locations, creator, membership and so on. Knowledge about location, creator and membership may be directly transferred into ConceptNet through already built-in AtLocation, LocatedNear, CreatedBy and MemberOf relations. It should be noted that according to the ConceptNet documentation[9] CreatedBy relation relates to processes, however inspection of the existing CreatedBy assertions show that they include creations and their authors as well. The remaining part of the acquired information related to the hyponyms may be represented by a more general RelatedTo relation.

The procedure of acquiring additional information is presented in Figure 1 and exemplified in Figure 2. First (Step 1), we scan the G-INTER using our handcrafted primary rule base in search of tags referring to locations, creators or members, for example {city}[10], {district}, {cartoonist}, {writer}, {member} and so on. In the case of acquiring LocatedNear pairs, we confirm that the basic hypernym contains a marker indicating physical proximity (such as the Chinese character meaning 'neighboring'). Next (Step 2), we filter the basic hypernym through a secondary rule base to exclude items that would introduce noise. For example, we can extract information about the birthplaces of famous people; however this does not mean that we can build an AtLocation kind of relationship between the person and his or her birthplace. If so, hypernyms indicating people are excluded from the analysis of location. When analysing LocatedNear pairs we filter out ambiguous items. If the basic hypernym is positively assessed by the secondary rule base, then (Step 3) we assume that the phrase acquired by deleting the basic hypernym from the G-INTER is a valid location, creator or member tag. Using the example from Figure 2, we check that 'adjacent municipality' is a valid tag to describe a nearby location. In the next stage (Step 4) we compare the validated location, creator or member tag with the content of the T-INTER. This way, using the previous example, we can extract the knowledge that the municipality we refer to is *Tomi-shi*. Finally (Step 5), we join the newly acquired information to the base hyponym with a proper relationship tag to extract a new relation, for example *Komoro-shi*-LocatedNear-*Tomi-shi*.

The effectiveness of the method mainly depends on the number and nature of introduced rules to both primary and secondary rules base. Our method is still work in progress and at this stage we used 58 primary rules and 16 secondary

[9] https://github.com/commonsense/conceptnet5/wiki/Relations

[10] Curly brackets were used to mark the tags' representations.

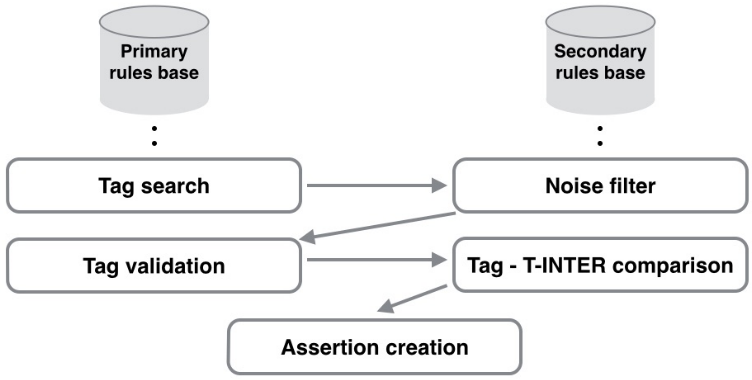

Fig. 1. Flowchart of our proposed method.

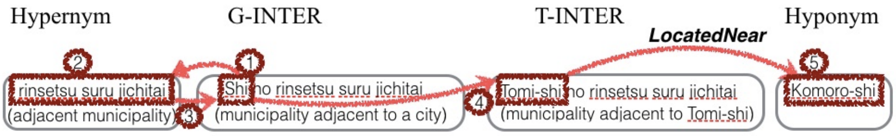

Fig. 2. Procedure of our proposed method exemplified on the extracted relation.

rules, which allowed us to extract assertions concerning location, neighboring locations and creators. The manually crafted rules have been created using heuristics after the analysis of the input data. The reason why we chose this kind of approach is because the information units contain Chinese characters indicating a type of location, a city, province, school or a creator. We use the rules to detect these characters, and this way we are able to get the named entities referring to locations and creators. Because of the qualities of Japanese language writing system these rules are often very simple, containing a single character, but still effective for detecting language units we want to extract. For example secondary rules used for detecting people include suffix '∼sha', which describes different professions. For English such shortcut would be harder to apply, and therefore person detection would require a much larger rules base covering a long list of names of professions and appropriate suffixes (like '∼er', '∼or' or '∼ist').

As our experiments revealed, extracting creator information is more complex and creates some challenges. While extracting location and member-related information, the introduced rules may be simple and straightforward. In the case of creators, the rules not only have to cover the qualities of the writing system, but also take into consideration the importance of particular roles while creating a given piece of work. For example our annotators indicated that a number of professionals taking part in the creation of films may not be considered as the creators of these films. Actors, actresses and voice actors, even if they make a great contribution to the work, should not be labeled as its creators. Further experiments have shown that similarly animators, animation directors, sound directors, and storyboard creators, according to the annotators, do not qualify to be included in the common sense CreatedBy assertions. The question whether all these roles should be indeed excluded from the creator category is open to discussion. If we changed our perspective and considered that not only one person or role is to be credited as the creator of a given piece of work, then we could assess some of these roles as correct in the CreatedBy assertions. The problem of different opinions on this matter would however remain. As the algorithm bases on keywords, it is unable to distinguish, for example, between director and sound director. Such distinction would be possible if we employed an additional, concept-based knowledge base.

In future we would like to investigate the possibility of combining heuristics with automated rule discovery methods in order to achieve higher precision and recall. The number and reliability level of the data acquired with our method is presented in the Evaluation section.

4 Evaluation

To verify the reliability level declared by Sumida [14] and evaluate our proposed method for obtaining additional relations we used the 2014-11-04 version of the Japanese Wikipedia dump data. We ran the definition, category and hierarchy modules of the Hyponymy extraction tool v1.0 at 93% precision rate using the biggest available training set, and obtained 6,014,194 hypernym-hyponym pairs. The number of unique hyponymy pairs was 5,866,680, which indicates that 147,514 pairs have been extracted by more than one module. The 93% reliability level declared by the authors of the method has been verified by three human annotators, whose task was to evaluate a sample of the data and decide whether the extracted pairs a) represent a correct hyponymy relation, b) represent related concepts, but not in a hyponymy relation, or c) represent unrelated concepts. The annotators assigned 1, 0.5 and 0 points respectively to 300 randomly selected assertions. We decided to assign 0.5 points to related concepts as they may be used to create correct assertions (see Future Work section). If two or more annotators assessed an item as belonging to one category, their decision was regarded as the evaluation output. In cases where their decisions varied (which happened 10 times), the first author decided the score. The procedure follows a modified Sumida *et al.* [14] evaluation method.

Table 2 presents the evaluation results. 283 pairs were assessed as representing a correct hyponymy relation, 10 pairs as related concepts, but not in a hyponymy relation and 7 as unrelated concepts.This results in 96.0% precision value of the tested sample, which surpasses the 93% declared by Sumida *et al.* The level of overall agreement between annotators was 86.9%, and the

Kappa value[11] was 0.80, which indicates that the annotation judgement was in substantial agreement [16].

Table 2. Evaluation results for IsA relations.

Correct hyponymy	Related concepts	Unrelated concepts	Precision	Total number of pairs
0.943 (283/300)	0.033 (10/300)	0.023 (7/300)	0.960	5,866,680

Running the fourth 'extended' module of the Hyponymy extraction tool v1.0 on the same Wikipedia dump data resulted in obtaining 2,738,211 basic hypernym–G-INTER–T-INTER–basic hyponym sets. By applying our method for extracting additional information, we were able to produce 131,760 pairs representing AtLocation relation, 6,217 pairs representing LocatedNear relation, 270,230 pairs representing CreatedBy relation and 21,053 pairs representing MemberOf relation. For comparison, nadya.jp, the baseline system, has provided only 8,706 AtLocation relations and no LocatedNear, CreatedBy or MemberOf relations in four years of its operation. In the case of AtLocation pairs, we evaluated 100 pairs[12] randomly selected from our method's output and 100 pairs randomly selected from nadya.jp's AtLocation assertions [8]. While evaluating LocatedNear, CreatedBy and MemberOf relations, a comparison with the baseline was not possible, as ConceptNet 5.3 does not yet contain any LocatedNear, CreatedBy or MemberOf pairs in its Japanese language section. These assertions were therefore evaluated independently. The evaluation procedure follows the previously applied one: 1 point being applied to correct AtLocation, LocatedNear, CreatedBy or MemberOf assertions, 0.5 point to related concepts, but not in the evaluated relation, and 0 points to unrelated concepts. In 15 cases the annotators' evaluation was inconsistent, and therefore the first author decided the score.

Table 3 shows the evaluation results of our AtLocation pairs generation method in comparison with the baseline system. 88 pairs generated by our method were evaluated as representing a correct AtLocation relation, 11 pairs as related concepts, but not in an AtLocation relation, and 1 as unrelated concepts. This results in a 93.5% precision value. In the case of the baseline system, 64 pairs were evaluated as correct AtLocation assertions, 20 as related concepts, but not in an AtLocation relation, and 16 as unrelated concepts. The precision value for the baseline system is 74.0%. The level of overall agreement between

[11] To measure the agreement level between judges, we used Randolph's free marginal multirater kappa instead of Fleiss' fixed-marginal multirater kappa, due to high agreement low kappa paradox.

[12] We adjusted the number of evaluated pairs to balance the proportion between the total number of pairs and the test sample.

annotators was 73.6% and the Kappa value was 0.60, which indicates that the annotation judgment was in moderate agreement. Examples of the extracted AtLocation assertions are presented in Table 4.

Table 3. Evaluation results for AtLocation relations in comparison with the nadya.jp baseline.

	Correct AtLocation	Related concepts	Unrelated concepts	Precision	Total number of pairs
Proposed	0.880 (88/100)	0.110 (11/100)	0.010 (1/100)	0.935	131,760
Baseline	0.640 (64/100)	0.200 (20/100)	0.160 (16/100)	0.740	8,706

$p < 0.001$, t-score = 4.6291

Table 4. Examples of generated AtLocation assertions.

Tomato Ginkou (Tomato Bank)	AtLocation	*Okayama-shi* (Okayama city)
Outao hoikuen (Outao nursery)	AtLocation	*Sakai-shi* (Sakai city)
Sandifukku (Sandy Hook)	AtLocation	*Eriotto-gun* (Elliott County)
Hoteru Kadoya (Kadoya Hotel)	AtLocation	*Tochigi-shi* Tochigi city)

Table 5 contains the evaluation result of the generated LocatedNear relations. 97 pairs were evaluated as correct LocatedNear pairs, 3 as related concepts and none as unrelated concepts, which results in 98.5% precision. The level of overall agreement between annotators was 86.6% and the Kappa value was 0.80, which indicates that the annotation judgment was in substantial agreement. Examples of the extracted LocatedNear assertions are presented in Table 6.

Table 7 contains the evaluation result of the generated CreatedBy relations. 60 pairs were evaluated as correct CreatedBy pairs, 37 as related concepts and 3 as unrelated concepts, which results in 78.5% precision. The level of overall agreement between annotators was 71.6% and the Kappa value was 0.57, which indicates that the annotation judgment was in moderate agreement. Examples of the extracted CreatedBy assertions are presented in Table 8.

Table 5. Evaluation results for LocatedNear relations

Correct LocatedNear	Related concepts	Unrelated concepts	Precision	Total number of pairs
0.970 (97/100)	0.030 (3/100)	0.000 (0/100)	0.985	6,217

Table 6. Examples of generated LocatedNear assertions.

Ougoe-machi (Ougoe city)	LocatedNear	*Ono-machi* (Ono city)
Iseri-gawa (Iseri river)	LocatedNear	*Konoha-gawa* Konoha river
Daiting	LocatedNear	Monheim
Kumotori-yama (Mount Kumotori)	LocatedNear	*Karamatsuo-yama* (Mount Karamatsuo)

Table 7. Evaluation results for CreatedBy relations.

Correct CreatedBy	Related concepts	Unrelated concepts	Precision	Total number of pairs
0.600 (60/100)	0.370 (37/100)	0.030 (3/100)	0.785	270,230

Table 8. Examples of generated CreatedBy assertions.

Dark Horse	CreatedBy	George Harrison
Kaze (Wind)	CreatedBy	*Kubota Koutarou*
Manuke-na Oukami (Sheep Wrecked)	CreatedBy	Michael Lah
The Point of View	CreatedBy	Alan Crosland

The analysis of the relatively low precision score of the assessed CreatedBy assertions revealed the following: in 24 cases it was the annotators' opinion that actors, voice actors, animators, storyboard creators or sound directors cannot be considered as creators of works they contribute to. Although it would be valid to include such persons in the RelatedTo kind of relationship with the work they helped to create, defining them as creators would go against common sense. This is a valid observation and it will be taken into consideration when re-designing and expanding the rule base for the next version of the algorithm. There were also cases of assertions assessed as invalid due to errors passed from the output of the Hyponymy extraction tool to the proposed method. Table 9 contains examples of assertions that were assessed as erroneous by the annotators.

Table 9. Examples of erroneous CreatedBy assertions.

Road 88	CreatedBy	*Tomita Yasuko* (actress)
Kaiketsu Zorori (Incredible Zorori)	CreatedBy	*Yamada Etsuji* (sound director)
Kishin Douji Zenki (Zenki)	CreatedBy	*Hayashi Akemi* (animator)
Human (incomplete name error)	CreatedBy	Nicholson Baker

Table 10 contains the evaluation result of the generated MemberOf relations. 76 pairs were evaluated as correct MemberOf pairs, 22 as related concepts and 2 as unrelated concepts, which results in 87.0% precision. The level of overall agreement between annotators was 80.6% and the Kappa value was 0.71, which indicates that the annotation judgment was in substantial agreement. Examples of the extracted MemberOf assertions are presented in Table 11.

Table 10. Evaluation results for MemberOf relations.

Correct MemberOf	Related concepts	Unrelated concepts	Precision	Total number of pairs
0.760 (76/100)	0.220 (22/100)	0.020 (2/100)	0.870	21,053

In the 13 cases the annotators decided that the generated MemberOf assertion refer to the former member of relative group, and therefore assigned it as the related concepts. The question whether these pairs should be considered as representing concepts in MemberOf relation is currently under discussion. If we would consider that the status of a member, once granted, is not temporary, then the precision rate of the tested sample would be higher, reaching 93.5%.

The results show that IsA relation pairs generated by the definition, category and hierarchy of the Hyponymy extraction tool v1.0, as well as AtLocation, LocatedNear and MemberOf relation pairs extracted by our proposed method may

Table 11. Examples of generated MemberOf assertions.

Henning Schmitz	MemberOf	*Kurafutowaaku* (Kraftwerk)
Dir.F	MemberOf	*Suiyoubi no Kanpanera* (Wednesday Canpanella)
Oono Satoshi	MemberOf	*Arashi*
Nishimura Akihiro	MemberOf	*Nikkan Giin Renmei* (Japan-Korea Parliamentarians' Union)

be incorporated into ConceptNet. Considering the number of the newly acquired assertions as well as reliability of the data in comparison with the resources already present in the knowledge base, such operation would be beneficial for ConceptNet. CreatedBy relation pairs could also be added after the revision of introduced rules and a substantial increase of the precision rate.

5 Generalizing over assertions

Wikipedia contains a lot of information about instances of certain concepts, such as Salvador Dali as an instance of a painter. Filling up ConceptNet with instances is a valid task, as it is very hard to establish the boundaries of commonsense knowledge - facts obvious for one group of people in large proportion overlap with knowledge of another group, but there is always a discrepancy. This issue raises a question: would it be possible to come to more general conclusions on the basis of the numerous instances? In order to solve this problem we created and performed an initial test of the following method: we took each of the additional information lists (representing LocatedAt, LocatedNear and CreatedBy relations) and analyzed each assertion one by one. For both concepts in the assertion we found their hypernyms in the generated IsA relations list. Next we generated assertions representing all possible combinations between concept's A hypernyms and concept's B hypernyms. We have repeated the process for all assertions in the additional information list and calculated the generated hypernym assertions' occurrence frequency. As predicted the assertions with the highest occurrence frequency represent general, commonsense observations. This is true for AtLocation and CreatedBy lists, but it is not the case when processing the LocatedNear list because of the relatively low number of LocatedNear assertions. It became apparent that the higher number of initial assertions increases the probability of generating meaningful general assertions. See Table 12 for the examples of generated general assertions. The procedure requires further development in terms of the method of frequency calculations and automatic filtering of non-general assertions.

Table 12. Examples of generated general assertions.

toshi oyobi machi (city and town)	AtLocation	*gun* (province)
shougakkou (elementary school)	AtLocation	*machi* (city)
douro (road)	AtLocation	*machi* (city)
sakuhin (work)	CreatedBy	*zonmei jinbutsu* (living person)
anime sakuhin (anime)	CreatedBy	*anime kankeisha* (people involved in making anime)
shutsuen sakuhin (performance art)	CreatedBy	*bunkajin* (cultural figure)

6 Conclusion

In this paper we presented a method for automatic acquisition of common sense knowledge triplets from the Japanese Wikipedia. It allowed us to mine IsA, AtLocation, LocatedNear, CreatedBy and MemberOf assertions with precision estimated at the levels of 96.0%, 93.5%, 98.5%, 78.5% and 87.0% respectively. We also demonstrated the possibility of formulating common sense assertions on the basis of generated instances data. As the Japanese part of the current ConceptNet 5.3 consists of 1,071,046 assertions, a contribution of 6,295,940 new assertions would be significant. It would mean an almost sixfold increase and could potentially make ConceptNet applicable to many Japanese language analysis problems. Moreover, as Wikipedia is a constantly expanding source, we could acquire more assertions simply by applying our method to the updated Wikipedia XML dump files.

7 Future work

In order to extend the functionality of our proposed method, we intend to update the primary and secondary rules, which would allow the system to increase its precision and the scope of extracted information. We would also like to explore the possibility of using a machine learning algorithm for automatic rule generation combined with the already present heuristics. Such a combination could potentially be more effective in increasing precision and recall, as well as finding new rules to extract even more relations.

We also plan to create an interface for the evaluation of the method's output by Japanese native speakers, which would allow us to utilize the pairs representing related concepts.

References

[1] Lenat, D. B.: CYC: A large-scale investment in knowledge infrastructure. Communications of the ACM 38, no. 11, 33–38 (1995)

[2] Suchanek, F. M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. Proceedings of the 16th international conference on World Wide Web, 697–706 (2007)

[3] Liu, H., Singh, P.: ConceptNet ? a practical commonsense reasoning tool-kit. BT technology journal 22, no. 4, 211–226 (2004)

[4] Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., Zhu, W. L.: Open Mind Common Sense: Knowledge acquisition from the general public. On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE, 1223–1237 (2002)

[5] Speer, R. H., Havasi, C., Treadway, K. N., Lieberman, H.: Finding your way in a multi-dimensional semantic space with Luminoso. Proceedings of the 15th international conference on Intelligent user interfaces, 385–388 (2010)

[6] Cambria, E., Hussain, A., Havasi, C., and Eckl, C.: SenticSpace: visualizing opinions and sentiments in a multi-dimensional vector space. Knowledge-Based and Intelligent Information and Engineering Systems, 385–393 (2010)

[7] Korner, S. J., Brumm, T.: RESI - A Natural Language Specification Improver. IEEE International Conference on Semantic Computing, 1–8 (2009)

[8] Nakahara, K., Yamada, S.: Development and Evaluation of a Web-Based Game for Common-Sense Knowledge Acquisition in Japan. Unisys Technology Review no. 107, 295–305 (2011)

[9] Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr, E. R., Mitchell, T. M.: Toward an Architecture for Never-Ending Language Learning. AAAI 5, 3 (2010)

[10] Schubert, L.: Can we derive general world knowledge from texts? Proceedings of the second international conference on Human Language Technology Research, 94–97 (2002)

[11] Mendes, P. N., Jakob, M., García-Silva, A., Bizer, C.: DBpedia spotlight: shedding light on the web of documents. Proceedings of the 7th International Conference on Semantic Systems, 1?8 (2011)

[12] Krawczyk, M., Rzepka, R., Araki, K.: Extracting ConceptNet Knowledge Triplets from Japanese Wikipedia. Proceedings of the 21st Annual Meeting of The Association for Natural Language Processing, 1052–1055 (2015)

[13] Sumida, A., Torisawa, K.: Hacking Wikipedia for Hyponymy Relation Acquisition. IJCNLP 8, 883–888 (2008)

[14] Sumida, A., Yoshinaga, N., Torisawa, K.: Boosting Precision and Recall of Hyponymy Relation Acquisition from Hierarchical Layouts in Wikipedia. LREC, (2008)

[15] Yamada, I., Hashimoto, C., Oh, J., Torisawa, K., Kuroda, K., De Saeger, S., Tsuchida, M., Kazama, J.: Generating information-rich taxonomy from Wikipedia. 4th International Universal Communication Symposium (IUCS), 97–104 (2010)

[16] Randolph, J. J.: Free-Marginal Multirater Kappa (multirater K [free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. Online Submission, (2005)