MURASAKI: Web-based Word Sense Description System

Fumito Masui[1] Yoshifumi Kawamuraand[1] Jun'ichi Fukumoto[2] and Naoki Isu[1]
[1]Graduate School of Information Engineering, Mie University,
Kurima-machiya-cho 1577, Tsu 514-8507, Japan
[2]Department of Media Technology, Ritsumeikan University,
Noji-Higashi 1-1-1, Kusatsu 525-8577, Japan
E-mail : [1]masui,yosi,isu@ai.info.mie-u.ac.jp, [2]fukumoto@media.ritsumei.ac.jp

Abstract: In this paper, we propose a method to descriptive a Japanese word sense with associative knowledge. The associative knowledge, which are called as $descriptors$, are extracted from world wide web(www) by focusing on collocation with comparative structure. For extracting descriptors, the specific comparative expressions with context which was generated with a pair of a query word and each related descriptor are retrieved from www. Some experiments have been conducted and those experimental results showed that the performace of our method overcome the baseline and we regard our method as a promising approach.

1. Introduction

When we ask about an unkown word, definition sentences or descriptive sentences are responded. For example, we would like to know and imagine about "*Randy Johnso*", the sentences to describe is given as below.

Ex.1 Descriptive sentences in Wikipedia[1]

Randy Johnson

Randall David Johnson (born September 10, 1963), nicknamed "the Big Unit", is a left-handed American starting pitcher who currently plays for Major League Baseball's Arizona Diamondbacks. The 6-foot-10 Johnson has been celebrated for having one of the most dominant fastballs in the game. He regularly approached, and occasionally exceeded, 100 miles per hour during his prime. However, his signature pitch is a hard, biting slider. Johnson has won the Cy Young Award five times, second only to Roger Clemens' seven. . . .

To generate those definition sentences as a response to question automatically, several methods such as question answering, information retrieval, automatic text summarization and so on are needed[1][2][3].

Considering question answering against non-factoid task, some elements such as a noun and a named entity are extracted from question to detect a appropriate passage or a paragraph. Talking about automatic text summarization, sentence extraction method as a common practice for automatic summarization is a simple and useful to make a sentence. But ultimately the elements should be selected to extract sentence. To realize a capable information retrieval system which deals with several specific requests, a query should be expanded flexibly. At any rate it is important how to collect enough elements or components to generate sentences.

Consequently, we emphasize a focus on purveying enough elements to imagine the word sense instead of answering without exaggeration and without omission about a query word. It is assumed that a set of knowledge pieces stimulates use's association to understand the word sense as effectively as set of sticky notes.

Ex.2 Description with knowledge pieces

Randy Johnson

- a starting pitcher
- a left handed pitcher

- Major League Baseball
- Arizona Diamondbacks

- 6-foot-10
- a tall man

- biting slider
- fastballs
- 100 miles per hour

- the Big Unit

⋮

We propose a method which gleans and coordinates knowledge pieces to describe a query word from www and profer a set of them as a word sense description.

First comparative expressions is generated by applying a query word to stylized frame expression. Second, descriptive words relates to the query word are extracted by retrieving the comparative expression from www. Then the descriptive words are ranked and integrated by scores computed based on frequencies. Depending on user's choice, a result is visualized as a describe list, a graph of score or a map of description. Utilizing a set of knowledge pieces in description, the user can imagine the query word and understand sense of the word by association.

Hereafter, methodology of an implemented system called *MURASAKI* will illustrated in section 2. The experiments were performed to evaluate the performance of the system in section 3 and we discuss the experimental results in section 4.

2. Web-based Word Sense Description System

We implemented a prototype of japanese word sense description system *MURASAKI* which extracts a set of descriptors against an input query word. *MURASAKI* extracts relation between two words based on comparative structure.

[1] $http://en.wikipedia.org/wiki/Randy_johnson$

No statistical method could extract such as word relations as *MURASAKI* can. Moreover, Societal attitude and topic transition against a query word will be followed flexibly because the system also extracts such a word relation from www dynamically.

Hereinafter, the construction of our implemented system is described with Figure 1. There are nine steps in the system process. (STEP1)query accession, (STEP2)comparative structure expression generation(A), (STEP3)expression retrieval(A), (STEP4)extraction of descriptor candidate, (STEP5)comparative structure expression generation(B), (STEP6)expression retrieval(B), (STEP7)scoring descriptors, (STEP8)integration of descriptors and (STEP9)visualization of result.

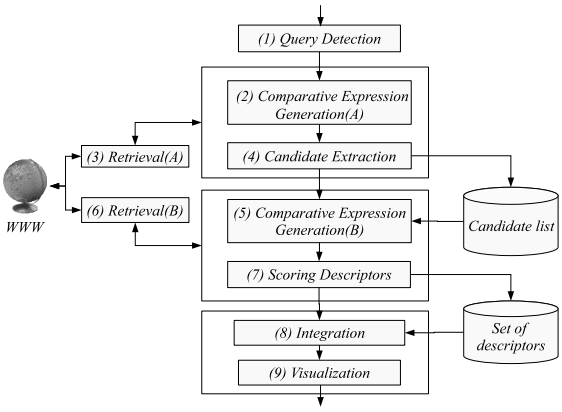

Figure 1. Description Process of *MURASAKI*

The detailed of each step are described as follows

STEP1 an input query word x is recognized.

STEP2 a comparative expression is generated by applying the query word x to the slot X in a context frame "$(X)\_noyouna\_(\gamma)$". The context frame "$(X)\_youna\_(\gamma)$", which generates a comparative expressions such as "$yuki\_noyouna\_hai(ash\ \ like\ \ snow)$", has been registered beforehand. The patten "$\_noyouna\_$" is an indicator for a comparative expression.

STEP3 documents are retrieved from www using the comparative expression generated at STEP2 as a query. Snippets are received as a result of the retrieval.

STEP4 all of the retrieved snippets are analyzed by a morphological analysis. Afterward candidates for descriptors are extracted from each analyzed snippet including "$X\_noyouna\_\gamma$" context.

STEP5 a comparative expression is generated by applying the each descriptor candidate γ to the slot γ in a context frame "$(X)\_noyouna\_(\gamma)$" registered beforehand.

STEP6 documents are retrieved from www using the comparative expression generated at STEP5 as a query. Number of hit cases are received as a result of the retrieval.

STEP7 the frequency of the comparative expression is approximated by the number of hit cases. The frequency is threshold α or more is identified as a proper descriptor.

STEP8 a set of descriptors which describes a query word X are integrated. On basis of frequency gotten in STEP7, a score for each descriptor are computed.

STEP9 descriptors are ranked in order of each score and read out. As a default, the output is graphed out in Figure2. Following each link on a descriptor, a web document includes a co-occurence of a query word and the descriptor are shown. Alternatively the format can be chosen from a graph out, a mapping the relation between descriptors and a list of snippet which highlights the comparative expression including the descriptor.

Scoring descriptors

We represent each word sense by scores of descriptors. Each score of a descriptor corresponds to a relation between a descriptor and a query word in context. In this case, two frame patterns "$\_noyouna\_$" and "$\_noyouni\_$ are the contexts.

The relation between the descriptor and the query word is calculated by formula(1).

$$score(\gamma_i|X) = \quad F_g(\gamma_i|X) \ * \ IWF_g(\gamma_i) \tag{1}$$

Let g be a context and $F_g(\gamma_i|X)$ be the frequency count of a descriptor γ_i occurring with a query word X in context g.

$IWF_g(\gamma_i)$ is defined as inverted frequency of a query word against a descriptor γ_i. $F_g(\gamma_i)$ be the frequency of words occuring with a descriptor γ_i in context g. and N be a total number of words occurs in context g.

$$IWF_g(\gamma_i) = \quad log\frac{N}{F_g(\gamma_i)} \ + \ 1 \tag{2}$$

The results are shown as a graph , a list of snippet include each descriptor, a list of descriptors links to relevant web pages in which each descriptor appears with a query word co-occurrently. One of output example is shown in Figure2.

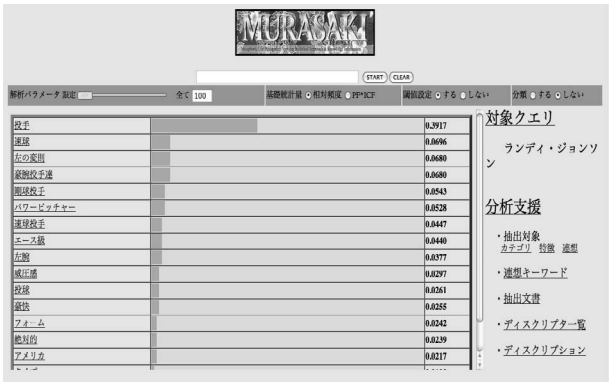

Figure 2. An Example of *MURASAKI*'s Output

3. Evaluation and Discussion

Experiments were performed to evaluate the performance of our implemented method.

We implemented the other method as a baseline for comparison of the performance. Basically, the baseline has the same STEPs as the proposed method. The difference is that the baseline uses n sized-window to extract a candidate of descriptor in place of the context frame such as "$(X)\_youna\_(\gamma)$". if the query word appears in window, the nearest neighbor noun ψ is regarded as a candidate of descriptor. The window size were set $n = 6$ as the optimum value by preliminary experiment results.

We list 173 query words by sampling common nouns and proper nouns from intellectual training books and internet news sites. Each descriptor set on the query word were extracted by our method and the baseline respectively. The descriptors were evaluated by 20 university students within it to judge acceptable or not acceptable.

Our method were compared with the baseline based on two measures: precision and relative recall. And the ranking performance was evaluated by MRR(mean reciprocal rank).

3.1 Evaluation of Extraction Performance

With precision and relative recall, the experiment to evaluate extraction performance was conducted. Precision is defined as a measure of the proportion of selected items that the system got right.

We computed precision for a descriptor set for each the query word. As an evaluated value, the average of all precision for the query words is calculated. The result shows that the proposed method marks 0.53 precision overcome 0.35 of the baseline.

Relative Recall is defined as the proportion of the target items that the system selected[4]. According to Pantel et. al[5], although knowing the total number of correct instances of a particular relation in any nontrivial corpus is impossible, it is possible to compute the recall of a system relative to another system's recall. And the relative recall of sysytem A is defined with given system B, $R_{A|B}$, as below.

$$R_{A|B} = \frac{R_A}{R_B} = \frac{\frac{C_A}{C}}{\frac{C_B}{C}} = \frac{C_A}{C_B} = \frac{P_A * |A|}{P_B * |B|}$$

where R_A is the recall of A, C_A is the number of approprate descriptors extracted by A, C is the total number of appropriate descriptors in the corpus, P_A is precision of A in the our experiments, and $|A|$ is the total number of descriptors estemated by $A(MURASAKI)$.

On basis of the precision, we computed relative recall of the proposed method against the baseline. Thereby the proposed method marks 1.11 on relative recall. It means that our system demonstrates advantage of recall over the baseline.

3.2 Evaluation of Ranking Performance

Utilizing MRR, first experiment to evaluate ranking performance was done. MRR is a statistic for evaluating any process that produces a list of possible responses to a query, ordered by probability of correctness. MRR is calculated as an average of the reciprocal ranks of a sample of queries. The reciprocal rank of a query response is the multiplicative inverse of the rank of the correct answer.

We sampled descriptor sets on 70 query words from 173 at random and computed MRR for top 5 descriptors. After the process was repeated 50 times, we also calculated an average of MRR as an evaluated value.

The result shows that the average MRR marks 0.74.

4. Discussion

Considering the results of extraction performance, our method provides high performance for extraction. The results of MRR means that our method ranks an appropriate descriptor at the top more than 70% possibilities and there will be definitely appropriate descriptor in top 3. Figure3 shows a rank-frequency distribution of top appropriate descriptor ranks. According to the figure, the most common rank is top and almost of all appropriate descriptors appears in top 3. Our consideration is also supported by the figure.

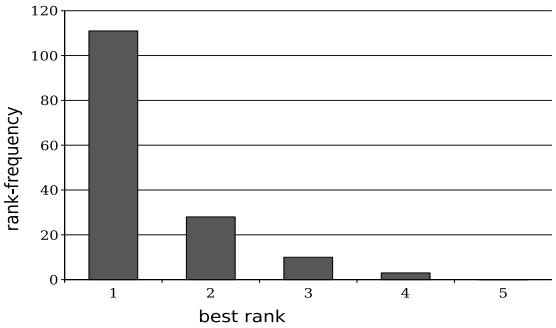

Figure 3. A Rank-frequency Distribution of Top Appropriate Descriptors

Examples of description for two query words are listed in table 1 and 2. Table 1 describes "$Randy\ Johnson$". It can be said that totally similar descriptions, such as "$a\ pitcher$", "$fastballs$", "$unorthodox$" and so forth, were extracted as an example in introduction. Table 2 lists descriptors for "$Ringo(Apple)$". The proposed method is effective in description of not only proper nouns but also common nouns.

We conducted an experiment to evaluate the effectiveness of the description for query expansion of document retrieval. By expanding with pair of the query word and each descriptor ranking in the top 5, possible queries are generated Utilizing those queries, we retrieved documents from www[2].

Each descriptor set on the query word were extracted by our method and the baseline respectively. The snippets were evaluated by 4 university students. Decision criterion are defined as:

(A) the snippet is descriptive
(B) the document is descriptive
(C) not descriptive
(D) unclear

[2]We used the goo search (http://search.goo.ne.jp/).

Table 1. Description of "Randy Johnson"

Randy Johnson		
rank	descriptor	score
1	toushu (a pitcher)	0.392
2	sokkyu (fast ball)	0.070
3	hidari no hensoku (unorthodox left hand)	0.068
4	gowan toshu tachi (rubber arm pitchers)	0.068
5	gokyu toushu (a strong-armed pitcher)	0.054
6	power pitcher (a power pitcher)	0.053
7	sokkyu toushu (a fastball pitcher)	0.045
8	ace kyuu (ace class)	0.044
9	sawan (left hand)	0.038
10	iatsukan (dominating)	0.030

Table 2. Description of "Ringo(apple)"

ringo(apple)		
rank	descriptor	score
1	kaori (aroma)	0.147
2	sawayaka (invigorating)	0.075
3	fruity (fruity)	0.062
4	sanmi (sourness)	0.053
5	kajitsu (furuit)	0.045
6	furuti (furuity)	0.041
7	fresh (fresh)	0.035
8	aji (taste)	0.034
9	sawayaka (invigorating)	0.027
10	hoo (cheeks)	0.026

Our method were compared with the baseline which use no query expansion based on precision.

In the case that (A) is regarded as a effective result, average precision marks 0.18 for the proposed method and 0.35 for the baseline. In the case that (A) and (B) are regarded as effective results, average precision marks 0.62 for the proposed method and 0.81 for the baseline. A result shows that the expansion with a descriptor is not effective.

By scrutinizing the results of both methods to consider the cause of the result, we found there are much difference between both retrieved results. For example of "$pansy$", in top 10, 8 snippets are related to "$pansy$ " but two about a company which is called "$pansy$". in case of using "$pansy + hana(flower)$", all of snippets in top 10 are related to "$pansy\ flower$" There are snippets about definition of "$pansy$", instances of" $pansy$", description of "$pansy\ flower$", how to cultivate "$pansy$" and photo album of "$pansy\ flower$". In the sense that snippets includes context, there is far more chance of the expansion than no expansion.

Furthermore, we also focus on the point that there are some kinds of descriptors such as broader class, attribute, instance and metaphor. In the experiment, it can be considered that mixing those kinds of descriptors caused conflicts of retrieval.

Applying the classified descriptor to expansion a query will be effective to retrieval document for specific purposes.

5. Conclusion

We proposed a method to descriptive a word sense with descriptor set and implemented a prototype system *MURASAKI*. The descriptors are extracted from www with context frame such as "$X\_noyouna\_\gamma$". We have verified performance of the method by comparing with the baseline and investigated advantages and problems. Experimental results shows that our method provides high performance for extraction and ranking descriptors. Moreover, we discussed that classified descriptor will be useful to information retrieval for specific purposes even tough simple query expansion has weak effectiveness.

There are many avenues of future work in improving an our implemented method. We plan to investigate the method to classify descriptors based on other context frames and support query expansion.

References

[1] Ellen M. Voorhees and Donna K. Harman: "TTREC: Experiment and Evaluation in Information Retrieval", *The MIT Press*(2005)

[2] Junichi Fukumoto, Tsuneaki Kato, Fumito Masui and Tatsunori Mori: "An Overview of the 4th Question Answering Challenge (QAC-4) at NTCIR Workshop 6", *In Proceedings of NTCIR-6 Workshop Meeting*, pp.433-440(2007)

[3] Hidetsugu Nanba and Manabu Okumura: " An Automatic Method for Summary Evaluation Using Multiple Evaluation Results by a Manual Method ". *In Proceedings of COLING/ACL 2006*, pp.603–610(2006)

[4] Keen, E. M.: "Evaluation parameters. In G. Salton (Ed.)", The SMART retrieval system-Experiments in automatic document processing. Prentice-Hall(1971)

[5] Patrick Pantel, Deepak Ravichandran and Edward Hovy: "Towards Terascale Knowledge Acquisition", In Proceedings of COLING-04, pp.771–777(2004)