Acquisition of Know-How Information from Web

Shunsuke Kozawa[1], Kiyotaka Uchimoto[2], and Shigeki Matsubara[1]

[1]Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
[2]National Institute of Information and Communications Technology
3-5 Hikari-dai, Seika-cho, Soraku-gun, Kyoto, 619-0289, Japan
kozawa@el.itc.nagoya-u.ac.jp, uchimoto@nict.go.jp, matubara@nagoya-u.jp

Abstract. A variety of know-how such as recipes and solutions for troubles have been stored on the Web. However, it is not so easy to appropriately find certain know-how information. If know-how could be appropriately detected, it would be much easier for us to know how to tackle unforeseen situations such as accidents and disasters. This paper proposes a promising method for acquiring know-how information from the Web. First, we extract passages containing at least one target object and then extract candidates for know-how from them. Then, passages containing the know-how are discriminated from non-know-how information considering each object and its typical usage.

Keywords: know-how, how-to type question answering, object, usage information, procedural question

1 Introduction

A variety of know-how such as recipes and solutions for troubles have been stored on the Web and they are often referred to by using web search. It has been reported that about 40% of all non-factoid questions on Q&A sites are how-to questions [8], and the know-how information has potential to give us the answers to the questions. However, it is not so easy to appropriately find know-how information related to a particular how-to question. If know-how could be appropriately detected, it would become easy to find answers to how-to questions. Also, it would become possible to let us know how to tackle unforeseen situations if know-how information could be stored beforehand and classified into several classes according to their topics.

In this research, we assume that know-how is a procedure or an advice. Figure 1 shows two examples of know-how. In previous researches, know-how information was acquired by using typical words such as "how to" and clue words such as "only if" and "first" [10]. However, it is difficult to efficiently acquire know-how information based on the conventional methods because a variety of clue expressions are required to appropriately detect know-how information which is often described without using clue words or the typical words "procedure", "how to" and "first" as you can see in Figure 1.

Well then, what is a clue to efficiently obtaining know-how information? Our observation for this question is as follows. Firstly, know-how information often

ID	sentence
9	A very easy way to remove the stickers without damaging them is to heat them up with a hair dryer.
10	It sounds kind of odd, but it works well, and doesn't discolor, or melt them.
11	Just heat them up to soften up the adhesive and carefully peel them off.
12	I usually use a razor blade to start peeling it off because you lessen the chance of wrinkling the sticker. (exacto knives work well becaue they have a pointed end)
13	This works well on those unwanted bumber stickers also

ID	sentence
1	How do you treat a heat stroke victim?
2	Victims of heat stroke must receive immediate treatment to avoid permanent organ damage.
3	First and foremost, cool the victim.
4	Get the victim to a shady area, remove clothing, apply cool or tepid water to the skin (for example you may spray the victim with cool water from a garden hose), fan the victim to promote sweating and evaporation, and place ice packs under armpits and groins.
5	Monitor body temperature with a thermometer and continue cooling efforts until the body temperature drops to 101-102°F (38.3-38.8°C).
6	Always notify emergency services (911) immediately.
7	If their arrival is delayed, they can give you further instructions for treatment of the victim.

Fig. 1. Examples of know-how

includes at least one object that plays an important role; for instance, "hair dryer" in the first example and "thermometer" in the second one in Figure 1. Secondly, the typical usage of the object is often described in each know-how; for instance, "to heat something" in the first example, and "to monitor a temperature" in the second one. As seen in the examples, know-how is often characterized by an object name and the description of its usage. We examined 100 lists of know-how information randomly sampled from a web site[1] and found that 75 out of the 100 lists included at least one object and the description of its usage. This supports our intuition.

In this paper, we propose a method for acquiring know-how information by focusing on each object and how it is used. First, we extract passages containing at least one target object and then extract candidates for know-how from them. Then, lists of know-how information are acquired based on the description of the object and its typical usage.

2 Related Works

In previous works, a few studies have addressed the acquisition of procedural texts. Takechi et al. proposed a method to categorize HTML texts tagged with ⟨OL⟩ or ⟨UL⟩ as either procedural or non-procedural by using word N-grams [7]. Aouladomar proposed a method to estimate *questionability* of web texts by using tag information (title, advice, warning, etc.) annotated by rule-base methods [2,

[1] http://know-how.fc2.com/ (in Japanese)

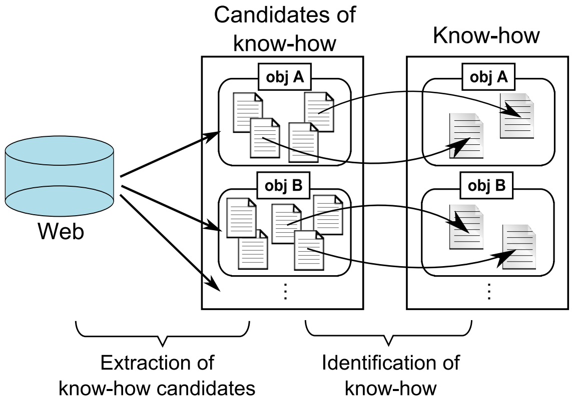

Fig. 2. Flow for acquiring know-how information

3] and clue expressions [1]. Yin et al. proposed a method for measuring the degree of procedurality of texts containing the phrase "how to" by using syntactic tags, morphological tags and cue phrases [10].

In this study, we acquire procedures and advices as know-how information. Our proposed method uses not only clue expressions which were also used in previous works but also an object and how it is used.

3 Acquisition of Know-How

Aouladomar classified procedural texts into three categories [1]; procedures (e.g. recipes, maintenance and construction manuals, etc.), injunctions (e.g. orders, regulations, etc.), and advices (e.g. beauty advices, health management methods, etc.). In this research, we acquire procedures and advices as know-how information since they are frequently asked on Q&A sites, namely there are many demands to obtain information on procedures and advices.

We acquire know-how information by focusing on an object and how it is used. In this research, we assume that objects are noun phrases which are hyponym of "physical entity" in Japanese WordNet 1.1[2]. We use expressions representing utilization of an object (utilization roles) to capture how the object is used. For example, the utilization role of "hair dryer" is to "heat" and that of "thermometer" is to "monitor." Figure 2 illustrates the flow for acquiring know-how information. We assume that a unit of know-how is a passage since know-how is very often composed of multiple sentences, but it is not always composed of all sentences in a document. We also assume that lists of know-how information are acquired for each object. First, given a target object, the candidate passages for know-how are extracted by using several methods since most passages would not contain know-how information. Then, the candidates

[2] http://nlpwww.nict.go.jp/wn-ja

are classified as either they contain know-how information or not by using clue patterns, an object and its utilization roles.

In the following sections, we assume that the target language is Japanese because know-how information is not well organized yet, although the method can be expanded to any languages. We used the Web corpus consisting of 500M Japanese parsed sentences extracted automatically from the Web [5]. The sentences in the corpus have been automatically annotated with morphological and syntactic information. The syntactic information in a sentence is represented as a dependency structure between Japanese phrasal units, *bunsetsu*.

3.1 Extraction of Know-How Candidates

First, passages are extracted in the following way. Every segment that contains at least one target object and is enclosed by a pair of the following HTML tags is extracted:

body, div, table, span, p, blockquote, h1, h2, h3, h4, h5, h6

If the number of sentences contained in a segment is less than or equal to a threshold α, the segment is extracted as a passage. Otherwise, we split the segment into one or more passages with the TextTiling algorithm [4] using α as the window size. This is because there exist Web sites where HTML tags are used incorrectly.

Next, the candidate passages that might contain know-how information are extracted. They are extracted by using the following four conventional methods (A,B,C,D) and our proposed method (E). Multiple methods are used since the acquired know-how information might be limited when only one method is used. We take the union of the candidate passages as the candidates for know-how.

(A) Extraction of the passages containing the term *houhou* (how to) [10] [3].
(B) Extraction of the passages tagged with ⟨OL⟩ or ⟨UL⟩ tags [7].
(C) Extraction of the passages containing any of 47 expressions (e.g. "in order not to", "prefer", "as long as", etc.) which have been manually generated by referring to the policy of generating patterns in the previous works [3, 10].
(D) Extraction of the passages containing any of 638 expressions such as "get well" and "feel good" found in the semantic lexicon constructed by Kobayashi et al. and tagged with *keiken* (experience) tags in the lexicon [6]. This is because we assume that know-how information contains expressions representing experience.
(E) Extraction of the passages containing any of 3-tuples which are composed of an object and its utilization roles. Utilization roles of a given object o are defined as paraphrases of such expressions as "using o" or "enjoying o" and expressed by a pair of a postposition p and a verb v. For example, the utilization role of a hair dryer is $\langle$*de*(by), *atatameru*(heat)$\rangle$. The 3-tuples

[3] The term was originally proposed in English. We manually translated them into Japanese and used them.

$\langle o, p, v \rangle$ (ex. $\langle$ *doraiya*(hair dryer), *de*(by), *atatameru*(heat) $\rangle$) are assumed to appear in the dependent *bunsetsus*; o and p (*doraiya*+*de* (by using hair dryer)) appear in a *bunsetsu* which depends on another *bunsetsu* containing v (*atatameru* (heat)). The method for acquiring utilization roles is described in detail in Section 3.2.

3.2 Identification of Know-How

The extracted candidates are classified as either they contain know-how information or not by using a machine learning model. Clue patterns, part-of-speech information, target objects and their utilization roles (3-tuple) are used as features [4].

- clue patterns
  Clue patterns are manually generated by referring to know-how and non-know-how information in a development data. The patterns are applied to a sentence or *bunsetsu* (Japanese phrasal unit) sequences in a sentence in a target passage. The frequency of the sentences matched with each clue pattern, the total frequency of the sentences matched with the patterns and the number of types of the matched patterns are used as features.
- part-of-speech information
  The appearance frequency of each part-of-speech normalized by the number of the sentences is used as a feature.
- 3-tuples
  The total number of the frequency of the 3-tuples $\langle o, p, v \rangle$ is used as a feature.

The target objects and their utilization roles are our newly added features. Henceforth, we call a set of clue patterns and part-of-speech information **PT** features. The following subsections describe the method for acquiring utilization.

Acquisition of Utilization Roles Torisawa made the following three assumptions about the characteristics of utilization role $\langle p, v \rangle$ for a given noun n: 1) An n marked by p often appears with v. 2) First-person pronouns such as "watashi (I)" often occupy the agent role of v. 3) The postposition "*de*" is a good candidate of a postposition in a utilization role. Utilization roles were acquired by using the following formula reflecting these assumptions [9].

$$U(n) = argmax_{(v',p') \in V \times A}\{Uscore(n, p', v')\}$$

V is a set of verbs, which can be a verb in possible utilization roles. As verbs in V, we used 6,485 verbs which appeared in the Web corpus with the verbal suffix "*tai*," which can be translated to "want." We also manually removed 20 verbs that can never take utilization roles such as "*naru* (become)" or mean

[4] Word n-grams were not used because they did not work well in our preliminary experiments.

Table 1. Sizes of training and testing data

	# of passages containing know-how						# of passages
objects	A	B	C	D	E	total	
air conditioner	14	16	33	5	6	70	492
cell-phone	11	5	11	0	0	27	495
digital camera	8	14	19	5	9	53	487
electric fan	17	29	37	5	11	92	476
iron	34	43	31	9	37	144	468
microwave	24	61	27	13	69	183	482
oven	28	79	32	7	70	208	476
refrigerator	28	56	22	1	60	165	494
vacuum cleaner	14	34	31	5	45	117	478
washer	13	26	14	3	27	78	479
total	201	363	258	53	334	1137	4827

literally "using" and "enjoying" such as "*tsukau* (use)" from V. A is a set of postpositions.

$$Uscore(n, p', v') = \frac{P(n, p', v')P(S|AP, v')Bias(p')}{P(n)}$$

$P(n, p', v')$ is the co-occurrence probability between the verb v' and n marked by the postposition p'. As for $P(S|AP, v')$, S denotes a set of first-person pronouns, and 17 pronouns were used. AP is a set of postpositions which can mark agent roles, and "*ga*" and "*ha*" were used. $P(S|AP, v')$ is the probability that the first-person pronouns occupy the agent role of v'. $Bias(p')$ denotes the bias concerning the postposition "*de*." If p' is "*de*," the bias is 25, otherwise the bias is 1 by referring to Torisawa's method [9].

4 Evaluation

4.1 Preparation

Construction of Training and Testing Data For our experiment, ten objects were selected from electric products that had appeared 10,000 times or more in the web corpus [5] and also appear in Japanese WordNet 1.1. We constructed training and testing data from the web corpus using the methods in Section 3.1. Passages were extracted using 20[5] as the window size α. We chose two sets of 100 passages at random from each group of passages extracted by using the methods A and B, and three sets of 100 passages in decreasing order of frequency from the groups of passages extracted based on the C, D and E methods, and judged manually whether they contained know-how information

[5] The same window size was used throughout all of the experiments mentioned in this paper, and it was not well tuned to particular data.

Table 2. Size of development data

# of passages containing know-how						# of passages
A	B	C	D	E	total	
43	44	41	16	34	148	444

or not. As utilization roles, we used top 25 pairs of a proposition and a verb automatically produced by the method described in Section 3.2. In this research, we assume that the passage is judged to be correct if a given passage contains know-how information. Note that if a given passage contained only a fragment of know-how information, the passage was judged as incorrect. The sizes of the data are shown in Table 1.

Generation of Clue Patterns We constructed the development data in the same way as training and testing data by extracting candidates for know-how from the web corpus with the methods described in Section 3.1 and judging them manually whether they contained know-how information or not. The breakdown of the development data is shown in Table 2.

We manually generated 79 types of patterns by referring to the development data. Some examples of the patterns are shown in Table 3. The symbols "|", "+" and ".∗" in column 2 represent a disjunction, a word boundary and any word sequences, respectively. Column 3 represents the targets that the patterns are applied to. S, L, R1, R2 and R3 represent a sentence, the leftmost *bunsetsu* in a sentence, the rightmost *bunsetsu* in a sentence, the rightmost two *bunsetsus* in a sentence and the rightmost three *bunsetsus* in a sentence, respectively.

In our experiments, the patterns were generated based on the development data constructed focusing on the object "hair dryer." However, we found that the patterns tended to depend not on an object but on the types of know-how information such as procedures and advices according to our preliminary investigation.

4.2 Settings

In the experiment, the 3-tuples and PT features (clue patterns and part-of-speech information) were used to train machine learning models. We used the following two types of 3-tuples:

3T$_{auto}$ As utilization roles, the top 25 pairs of a postposition and a verb produced by the method described in Section 3.2, which were the same pairs used for constructing the training and testing data, were used.

3T$_{man}$ As utilization roles, pairs of a postposition and a verb were manually selected from among top 100 pairs produced by the method described in Section 3.2, and were used. The average number of the manually selected pairs was 25.

Table 3. Clue patterns

<table>
<tr><th>No.</th><th>patterns</th><th>target</th></tr>
<tr><td>1</td><td>mazu|hajimeni|hajime+ha|saisho+ha (first, primarily, to begin with)</td><td>L</td></tr>
<tr><td>2</td><td>sore+kara|tsugini|konoato|sonoato|soshite (then, next, secondly)</td><td>L</td></tr>
<tr><td>3</td><td>dekiagari|kansei|shuryo|kanryo (finish, end, complete, accomplish)</td><td>R1</td></tr>
<tr><td>4</td><td>verb+hou+ga.*yoi|verb+no+ga.*anshinda (prefer, preferable)</td><td>S</td></tr>
<tr><td>5</td><td>verb+yasui (easier to + verb)</td><td>R2</td></tr>
<tr><td>6</td><td>kinmotsu|dameda|genkin|kiken (danger, caution, prohibition)</td><td>S</td></tr>
<tr><td>7</td><td>verb+nai+youda.*verb|verb+nu+ni.*verb (never, avoid + verb)</td><td>S</td></tr>
<tr><td>8</td><td>ki+wo+tsukeru|te+wo+nuku+nai|tyuui (see to, warning)</td><td>R3</td></tr>
<tr><td>9</td><td>wo+taishou|ni+gentei|ni+kagiru (limit, target, restrict)</td><td>R3</td></tr>
<tr><td>10</td><td>hitsuyouda|youi|kakaseru+nai|hissuda (necessary, need, essential, vital)</td><td>R3</td></tr>
<tr><td>11</td><td>yakudatsu|katsuyaku|benrida|kouritsu (useful, helpful, efficient)</td><td>R2</td></tr>
<tr><td>12</td><td>teineida|shintyouda|kinnitsu|shikkari (carefully, advisedly, fastly)</td><td>S</td></tr>
</table>

Support Vector Machines were used as machine learning models and they were trained using LibSVM[6]. For the experiments, we prepared the following baseline method:

baseline (conventional method) Identify whether a given passage contains know-how information or not by using the model based only on **PT** features.

In the next section, the effectiveness of using both an object and its utilization roles is shown by comparing models with and without the 3-tuples.

4.3 Experimental Results

We split the data for each object in five and carried out 5-fold cross validation by using data composed of ten objects. The experimental results are shown in Table 4. In comparison with the model using only PT features (baseline), the models using both PT and 3T features significantly improved in both the precision and the recall. This indicates that lists of know-how can be efficiently acquired by focusing on an object and its utilization roles. In addition, the model using both PT and $3T_{auto}$ features achieved good results although the rates were a little lower than ones using both PT and $3T_{man}$ features.

In the above experiment, both the training and testing sets contained the same 10 objects. Thus, the results do not show that the proposed models would achieve good results for the data containing objects which are not contained in the training set since the patterns and the 3-tuples appearing in know-how information might depend on an object. Therefore, we carried out 10-fold cross validation by training the data containing nine objects out of ten and testing the data containing the rest (one object). The experimental results are shown in Table 5. The results show that both the precision and the recall were significantly

[6] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Table 4. Experimental results by 5-fold cross validation

Feature	Precision	Recall	F-measure
PT (baseline)	73.0% (588/805)	51.7% (588/1137)	60.6
PT + $\mathbf{3T}_{auto}$	74.6% (660/885)	58.1% (660/1137)	65.3
PT + $\mathbf{3T}_{man}$	75.6% (666/881)	58.6% (666/1137)	66.0

Table 5. Experimental results by 10-fold cross validation

Feature	Precision	Recall	F-measure
PT (baseline)	71.9% (550/765)	48.4% (550/1137)	57.8
PT + $\mathbf{3T}_{auto}$	72.8% (600/824)	52.8% (600/1137)	61.2
PT + $\mathbf{3T}_{man}$	73.4% (614/836)	54.0% (614/1137)	62.2

improved by using the 3-tuples. This indicates that using an object and its utilization roles is useful even though the training set does not contain data for the target object. The differences between the proposed methods and the baseline method are statistically significant according to Mcnemar's test ($p < 0.01$).

The 3-tuples feature of our proposed method is a template or a type of features while the patterns used as baseline features are surface strings or tokens. Therefore, the training data do not have to contain all of the 3-tuples appearing in the testing data. This indicates that the proposed method works without creating training data for all objects.

4.4 Discussion

Error Analysis In order to detect the causes of errors, we investigated the results by 10-fold cross validation with the model that uses both **PT** and $\mathbf{3T}_{man}$.

We investigated 222 passages acquired incorrectly and found following three main causes:

- Passages containing only a fragment of know-how information
  There were 93 (41.9%) passages. This is because passages were extracted inaccurately from the documents. If these passages had been extracted accurately, the passages would have been judged as correct. There were 190 passages containing only a fragment of know-how information in the training and testing data. It would be important to develop a method for accurately extracting passages if the acquired passages are separated from documents and then used. In the case that people refer to the acquired know-how information, it is a matter of no importance that the boundaries of passages are incorrect since they can understand know-how by referring to the know-how information in conjunction with surrounding contexts.
- Passages containing utilization roles
  There were 69 (31.1%) passages. They were either diaries or commercial articles which contained typical usage of the target object and did not contain

Table 6. Experimental results using arbitrary pairs and results combining models

Feature	Prec	Rec	F1
PT (baseline)	71.9%	48.4%	57.8
PT+3T$_{all}$	71.5%	49.3%	58.3
***PT+3T**$_{auto}$	72.8%	52.8%	61.2
***PT+3T**$_{man}$	73.4%	54.0%	62.2
PT (without the method E)	71.5%	46.0%	56.0

know-how information. This type of errors would be reduced by taking into account co-occurrences and sequences of the patterns, the target object and its utilization roles.

- Passages containing injunctions
  There were 21 (9.5%) passages. We did not target them in this research. However, it is a type of procedural texts, and it could be a type of know-how.

We investigated 523 passages that are know-how but could not be acquired. Objects besides the target object were used in 450 (86.0%) passages. That is to say, these know-how information might be acquired by targeting other objects. Furthermore, we expect that the recall will be more improved by simultaneously considering two or more objects and their utilization roles.

Effect of an Object and Its Utilization Roles In order to show the contribution of our 3-tuples compared with that of arbitrary 3-tuples to acquisition of know-how information, we carried out 10-fold cross validation using the following feature:

3T$_{all}$ The total number of the frequency of the 3-tuples $\langle o, p, v\rangle$, where o is a target object and $\langle p, v\rangle$ is the pairs of any postposition and any verb, was used.

In addition, we carried out experiments by using only conventional methods in both extraction of know-how candidates and identification of know-how. That is, know-how candidates were extracted by using the method A through D in Section 3.1 (without the method E considering 3-tuples) and know-how information was identified by using the model based on PT features. Experimental results are shown in Table 6. The models attached with * in Table 6 are statistically significant than the baseline method according to Mcnemar's test (p < 0.01). The model based on PT and 3T$_{all}$ features is not statistically significant than the baseline method. Moreover, the models attached with * in Table 6 are statistically significant than the model based on PT and 3T$_{all}$ according to Mcnemar's test (p < 0.01). These results indicate that the 3-tuples should be selected for efficiently acquiring know-how information and one of the promising selection method is based on an object and its utilization roles. The last row in

Table 7. Experimental results of open-domain experiments

Feature	Precision
PT (baseline)	68.4% (171/250)
PT + $\mathbf{3T}_{auto}$	70.5% (172/244)
PT + $\mathbf{3T}_{man}$	69.0% (176/255)

Table 6 shows the results obtained by using only conventional methods. The results show that the 3-tuples play important roles in both extraction of know-how candidates and identification of know-how although we can not simply compare them since the size of the data are different.

To our knowledge, this study is the first trial to show that an object and its usage play an important role in acquiring know-how information.

4.5 Open-Domain Tests

We performed experiments on data for different domains. For the experiments, we selected eight objects (curtain, hunger, ladder, lighter, mirror perfume, scissors, stove) which are different types of objects from electric products. They were selected almost randomly except that frequent nouns were preferred to others. The passages were automatically extracted in the same way as the construction of the training and testing data and classified whether they contained know-how information by the models trained using the training and testing data.

Table 7 shows the precisions for acquiring know-how information. The model using automatically acquired 3-tuples improved the precision and the recall (the number of lists of successfully acquired know-how information).

Thus, we expect that know-how information can be acquired by focusing on an object and how it is used without depending on domain.

5 Conclusion

This paper presented a method for acquiring know-how information by focusing on an object and how it is used. First, we extracted know-how candidates for each object. Then, by using both an object and its utilization roles, passages containing know-how information were efficiently acquired from them. Various lists of know-how information could be acquired by expanding the number of target objects although they were restricted in our experiments as a first trial.

In this paper, we used manually extracted patterns to compare our method with those in the previous works. We would like to acquire know-how information using automatically extracted patterns. In the future, we would like to consider (grammatical) objects of verbs for utilization roles since we expect that they are more useful information for identifying know-how information. For example, to "monitor a temperature with thermometer" is more important than to "monitor something with thermometer" in the second example in Figure 1. In addition, we would like to focus on multiple objects and how they are used since it is often the case that multiple objects are used in know-how. Know-how information would be more efficiently acquired by simultaneously considering two or more objects and their utilization roles.

In this study, we assume that know-how information could be mostly covered by the physical entities of WordNet. However, the coverage should be improved to get statistically valid data for obtaining utilization roles. We are planning to generalize nouns not found in WordNet.

References

1. Aouladomar, F.: Towards answering procedural questions. In: Proceedings of the IJCAI Workshop on Knowledge and Reasoning for Answering Questions. pp. 21–31 (2005)
2. Delpech, E., Saint-Dizier, P.: Investigating the structure of procedural texts for answering how-to questions. In: Proceedings of the 6th International Conference on Language Resources and Evaluation. pp. 46–51 (2008)
3. Fontan, L., Saint-Dizier, P.: Analyzing the explanation structure of procedural texts: dealing with advice and warnings. In: Proceedings of the 2008 Conference on Semantics in Text Processing. pp. 115–127 (2008)
4. Hearst, M.A.: TextTiling: Segmenting text into multi-paragraph subtopic passages. Computational linguistics 23(1), 33–64 (1997)
5. Kawahara, D., Kurohashi, S.: A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis. In: Proceedings of the 7th Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. pp. 176–183 (2006)
6. Kobayashi, N., Inui, K., Matsumoto, Y., Tateishi, K., Fukushima, T.: Collecting evaluative expressions for opinion extraction. In: Proceedings of the 2nd International Joint Conference on Natural Language Processing. pp. 584–589 (2004)
7. Takechi, M., Tokunaga, T., Matsumoto, Y., Tanaka, H.: Feature selection in categorizing procedural expressions. In: Proceedings of the 6th International Workshop on Information Retrieval with Asian Languages. pp. 49–56 (2003)
8. Tamura, A., Takamura, H., Okumura, M.: Classification of multiple-sentence questions. In: Proceedings of the 2nd International Joint Conference on Natural Language Processing. pp. 426–437 (2005)
9. Torisawa, K.: Automatic acquisition of expressions representing preparation and utilization of an object. In: Proceedings of the 5th Recent Advances in Natural Language Processing. pp. 556–560 (2005)
10. Yin, L., Power, R.: Adapting the naive bayes classifier to rank procedural texts. Lecture Notes in Computer Science 3936, 179–190 (2006)