

1. Purpose

- We annotate TECHNOLOGY tags and VALUE tags for patent map generation.

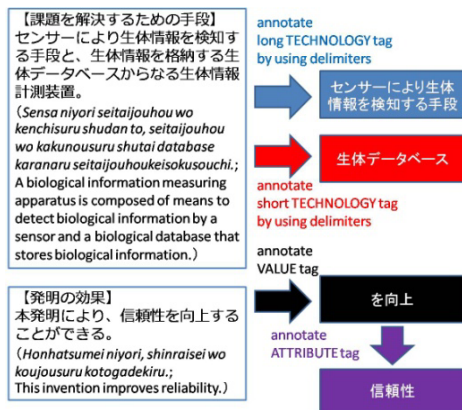


Fig. 1 Overview of our method

2. Our method

Annotation of EFFECT tags

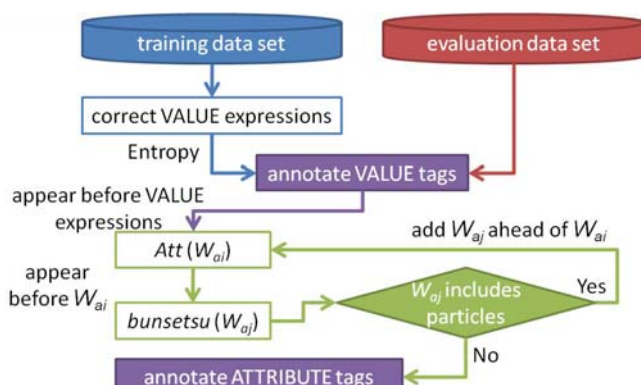


Fig. 2 Flowchart of annotation VALUE tags

Our method for annotation VALUE tags is as follows:

- Step 1: Collect an expression that is enclosed by VALUE tags and these expressions are candidates of VALUE expressions.
- Step 2: Select an appropriate VALUE expression from the candidates by using entropy-based score.

After annotation VALUE tags, our method for annotation ATTRIBUTE tags is as follows:

- Step 1: Extract *bunsetsu* W_{ai} that appears before a VALUE expression and the word is selected as a candidate for ATTRIBUTE expression *Att*.
- Step 2: Extract *bunsetsu* W_{aj} that appears before the candidate for ATTRIBUTE expression *Att*.
- Step 3: Add *bunsetsu* W_{aj} to *Att* when *bunsetsu* includes Japanese particles “の”, “を”, “が”, “や”, “および”, and “及び”, otherwise select *Att* as an appropriate ATTRIBUTE expression.
- Step 4: Repeat Steps 2 and 3

Annotation of TECHNOLOGY tags

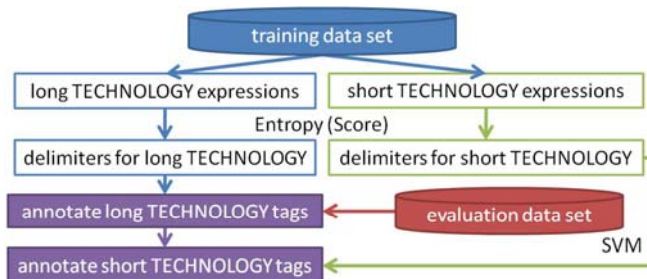


Fig. 3 Flowchart of annotation TECHNOLOGY tags

Table 1. Delimiters of long TECHNOLOGY tags

Group	Delimiters
First	は、(ha,)、において、(nioite,)、ば、(ba,)
Second	と、(to,)、とを(towo)、設け(mouke)、備え(sonae)、有し(yushi)、有する(yusuru)、more

Table 2. Delimiters of short TECHNOLOGY tags

Delimiters
と(to,)、とを(towo)、設け(mouke)、備え(sonae)、有し(yushi)、有する(yusuru)、用い(mochii)、介し(kaishi)、含む(fukumu)

Extraction of delimiters is as follows:

- long TECHNOLOGY tags: Entropy (Score)

$$H(t) = - \sum_{t \in S(t)} P(t) \log_2 P(t), \quad t = \text{delimiter}$$

- short TECHNOLOGY tags: annotated by using SVM

3. Results and Discussions

- We achieved high precision on each task, in particular, task of annotating VALUE tags.
- Recall of our method is a low value.
 - The collection of appropriate delimiters and heuristic rules are non-exhaustive.

Table 3. Comparison of performance

Team	Recall	Precision	F-measure
HCU	0.43	0.55	0.48
HTC	0.22	0.38	0.28
ONT	0.21	0.37	0.27
TRL	0.44	0.51	0.47
Our method	0.27	0.55	0.36

4. Conclusion

- We annotated TECHNOLOGY and EFFECT tags by using delimiters and machine learning.
- In order to choose a delimiter, we used a score based on entropy.
- We achieved 0.55 precision and 0.27 recall.