

Patent Machine Translation Task at NTCIR-9

December 6-9, 2011, Tokyo, Japan

Call for Participation

- **Chinese to English** subtask newly added
- **Human evaluations** will be carried out
- Parallel corpora provided: **1 million** Chinese-English, and **3 million** Japanese-English sentence pairs

Interested participants are invited to the Patent Machine Translation task (PatentMT) at NTCIR-9. Patents constitute one of the **challenging domains** for machine translation because other than cumbersome sentences, they can contain new technical terms and legalistic language. Moreover, there is a significant **practical need** for machine translation of patent documents. Let us cultivate this challenging and significant practical research field with patent machine translation! Participants can use a large-scale patent parallel corpus for research and will benefit from reliable human evaluation of their MT quality.

PatentMT, while cast in a framework of friendly competition, has the ultimate goal to foster scientific cooperation. In this context, the organizers propose a research task and an open experimental infrastructure for the scientific community working on machine translation research.

Task

- Subtasks:

Subtasks	Parallel corpus
Japanese to English	3 million sentence pairs from patent descriptions
English to Japanese	
Chinese to English	1 million sentence pairs from mostly patent descriptions

Test data: 2,000 patent description sentences

- Participants choose the subtasks in which they would like to participate.
- Resources to be provided
 - Chinese to English subtask: A parallel corpus consisting of 1 million Chinese-English sentence pairs from mostly patent descriptions, a large-scale monolingual patent corpus in English, and a test set of patent descriptions
 - Japanese to English subtask: A parallel corpus consisting of 3 million Japanese-English sentence pairs from patent descriptions, a large-scale monolingual patent corpus in English, and a test set of patent descriptions
 - English to Japanese subtask: A parallel corpus consisting of 3 million Japanese-English sentence pairs from patent descriptions, a large-scale monolingual patent corpus in Japanese, and a test set of patent descriptions
- Use of the data will be governed by NTCIR-9 agreement which is under preparation.
- Participants are requested to machine translate the test sets.
- The submitted translation results will be evaluated by **human evaluation** and automatic evaluation. The primary evaluation is human evaluation whose criteria will be *adequacy* and *acceptability*. In this task, *acceptability* is defined as in Fig. 1. Evaluation of *acceptability* will be undertaken for some selected systems in line with budgetary

limitations. We will select systems based on the following criteria: (i) multiple methods are preferred and (ii) systems with higher adequacy are given priority over others with the same methodology but lower adequacy.

- The task definition is basically the same as that for the NTCIR-8 Patent Translation Task.
- Participants are requested to submit a paper describing the MT system, the utilized resources, and their results using the provided test data, and are requested to present their papers at the workshop.

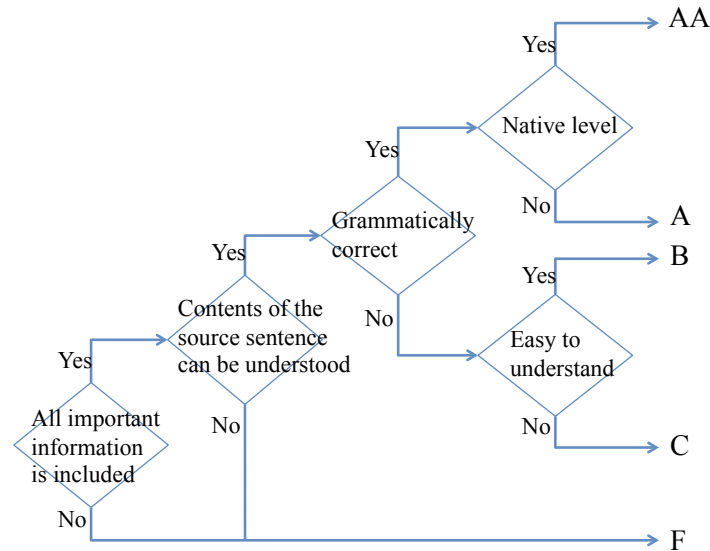


Fig. 1 Acceptability

Schedule

2010.12.20	2011.1.20:	Task registration due (extended)
2011.1.5:		Training data release
2011.5.9:		Test data release
2011.5.22:		Translation results submission due
2011.8.19:		Evaluation results release
2011.9.20:		MT system description due
2011.11.4:		Camera-ready due
2011.12.6-9:		NTCIR-9 workshop

(NII will release the NTCIR-8 PATMT training data to the public for research use before 2011.1.5. Participants are allowed to use the data before the 2011.1.5 release date.)

Organizers

Japanese-English Task:

Isao Goto (NICT)
Eiichiro Sumita (NICT)

Chinese-English Task:

Benjamin K. Tsou (Hong Kong Institute of Education/City University of Hong Kong)
Kapo Chow (Hong Kong Institute of Education)
Bin Lu (City University of Hong Kong/ Hong Kong Institute of Education)

For more information, please visit the NTCIR-9 PatentMT web site:

<http://ntcir.nii.ac.jp/PatentMT/>