

KYOTO EVALUATION

The ultimate goal of KYOTO is the extraction of facts from text. This is done by the Kybot system at the end of the pipeline of processing. The Kybots depend on the results of all the previous processing modules. Errors in the processing may be stacked. For example, if the parser assigns the wrong part-of-speech to a word, the word-sense-disambiguation will fail to assign the proper concept, and the onto tagger will not insert the correct ontological implications into the KAF file. Consequently, the Kybots may fail because a profile specified for a certain part-of-speech does not match or the wrong profile specified for a concept is matched, thus affecting recall and or precision.

This being said, modules may also introduce errors that have no effect on the extraction of facts. Therefore, it is a wise thing to evaluate both the modules and the end-applications that depend on it to learn about the effects and relevance of errors. Likewise, the results of KYOTO have been evaluated in different ways:

1. System evaluation:
 1. Evaluation of different modules
 2. Evaluation of the facts extracted by the KYOTO system

3. User evaluation:
 1. Use of the knowledge editor
 2. Use of the Semantic search in the extracted facts

Evaluation of modules

Module evaluations have been carried out for Word-Sense-Disambiguation and Named-Entity-Recognition. These are the most important modules that connect text to concepts.

1. Word-Sense-Disambiguation (WSD): the results of this have been described in the SemEval2010 workshop that was organized by KYOTO:

<http://aclweb.org/anthology-new/S/S10/S10-1013.pdf>

2. Named-Entity-Recognition: the results of this have been described in the KYOTO deliverable [D09.2](#).

The domain-based approach to WSD in KYOTO scored on the 4th place of the Knowledge-Based systems for English, with a precision of 0.481 and a recall of 0.481. The system was not optimized for the task. We also applied the KYOTO-WSD to Chinese (0.322 precision, 0.296 recall), Dutch (0.526 precision, 0.526 recall) and Italian (0.529 precision, 0.529 recall).

Named-entity recognition (NER) has been developed for dates and places. Dates and places are necessary to turn event-relations into facts with a time and place. The NER module detects 85.3% of the locations and 92.5% of the dates. 95.5% of the dates are correctly interpreted; of the locations, 89.1% are disambiguated to the correct country and 42.0% to the correct feature type.

Evaluation of facts:

We developed a complete package to evaluate the output of the event/fact mining by the Kybots. We defined a neutral triplet representation for representing the facts, which consists of:

1. a relation
2. a list of word token identifiers that represent the event
3. a list of word token identifiers that represent the participant

If a fact has more than one participant, it is broken down into separate triplets. We created a gold-standard and defined a baseline. We experimented with various settings. The details are described in [D5.4](#). With the best settings, we obtain a precision of 50% and a recall of 40%. We recover 100% of the events with a precision of 29%.

The evaluation data and software can be downloaded from the following URL:

<https://kyoto.let.vu.nl/~kyoto/files/data/kybotevaluation/11767/KybotEvaluationDataMarch2011.zip> (12 MB)

We also carried out an open competition evaluation in combination with the 2nd KYOTO workshop for which we created a different gold-standard. Two other groups participated. The details can be found on the [workshop-event page](#).

Evaluation of semantic search

The facts extracted by the Kybots are indexed and searchable in a cross-lingual semantic index. To evaluate the usefulness of these facts, we carried out a benchmark evaluation of the search indexes and an end-user evaluation. In this evaluation, we compare the search on Kybot facts with a standard text search solution. The benchmark evaluation checks to what extent the kybot facts represent all the information in the database. We compared the recall of the semantic search in the Kybot system with the recall of a standard text search. This showed that 35% of the coverage of a standard system is achieved. The extraction of the kybots is thus a good step towards a full text representation but still there is room for improvement. The details of the benchmark can be found in the KYOTO deliverable [D09.5](#).

For the end-user evaluation, we had 21 users (students and environmentalists) that had to find answers to questions using the standard retrieval system, a mash-up fact retrieval system and the semantic search in the kybot facts. The evaluation showed no significant difference in quality and performance but the users had major difficulties understanding and appreciating the

innovative features of the semantic search in the fact index. Some less-conservative users on the other hand exactly liked these features. The details are discussed in the KYOTO deliverable [D09.6](#).

Evaluation of the Knowledge editor

The Wikyoto was used by environmentalist that have no training in linguistics or knowledge engineering. They created an English domain wordnet with mappings to the central ontology. The details are described in deliverable [D08.4](#).