

Overview of the TMREC Tasks

NTCIR Workshop TMREC Group:

Kyo Kageura, Masaharu Yoshioka, Koichi Takeuchi & Teruo Koyama

Department of Research and Development, NACSIS

{kyo,yoshioka,koichi,koyama}@rd.nacsis.ac.jp

Advisors & Task Forces: Keita Tsuji & Fuyuki Yoshikane

Graduate School of Education, University of Tokyo

i34188@m-unix.cc.u-tokyo.ac.jp, fuyuki@p.u-tokyo.ac.jp

Secretary: Maho Okada, NACSIS, okadama@rd.nacsis.ac.jp

1 Tasks

The TMREC tasks consist of three different sub-tasks, i.e.

- (1) **Automatic Term Recognition Task:** To extract technical terms which are considered to characterise the given set of document (or domain) as extensively or consistently as possible from the corpus.
- (2) **Keyword Extraction Task:** As an optional task to the automatic term recognition task, the participants may distinguish “technical terms” that characterise the domain from “keywords” that represent individual documents. The task of extracting keywords which represent individual documents are optionally and separately defined as keyword extraction task.
- (3) **Role Analysis Task:** To extract the descriptive structure of academic documents in the form of the triplets of the following constituents, i.e. (a) what a document treat (main subject), (b) by what means (means or method), and (c) what is applied (procedure or action applied to the main subject).

The main among these three is the automatic term recognition task, and the number of participants to this task is the largest among the three tasks.

2 Participants

The number of groups that applied to the TMREC tasks was eight, plus one advisor & task force.

- (1) **Automatic Term Recognition Task:** Eight groups with one advisor submitted the results. Among the eight groups, one group submitted the result later than the deadline. In total, 17 result files are submitted.
- (2) **Keyword Extraction Task:** Three groups submitted the results, of which one group submitted two files. So four result files are submitted.

- (3) **Role Analysis Task:** Two teams submitted the results, one for each.

For details of these, please refer to the evaluation reports for each of the three subtasks.

3 Evaluation Reports

There are three evaluation reports, each corresponding to each of the three tasks listed above. The names of the teams are kept anonymous. Evaluation and comparison of the results were carried out by the four main member of the TMREC group, i.e. the first four author of this overview. Two authors listed as “advisors & task forces” contributed to making the manually selected term candidates, and the secretariat contributed to the final polishing up of the tagged-corpus. As the main task is the automatic term recognition task, we devote much of the volume to the evaluation of the automatic term recognition task.

As will be clarified in the evaluation reports, they are not intended to determine which methods are better, etc. Rather, as was clearly mentioned in our original task descriptions, our intention is to stimulate the constructive discussion of both technical and conceptual aspects of the automatic processing of terms and terminology. As such, the evaluation reports are for triggering constructive discussions concerning various methods of the TMREC tasks, but it is also the case that the evaluation reports are the target of the discussion and evaluation.

Acknowledgement

Although we cannot present here the names of the participants in order to keep them anonymous, we would like to thank all the participants who submitted the excellent and really intriguing results for the TMREC tasks.