

Introduction

Formal Run Results of Web Retrieval Task at the Third NTCIR Workshop (NTCIR-3)

Executives Committee of NTCIR-3 Web Retrieval Task
ntcadm-web@nii.ac.jp

September 24, 2002

Abstract

This article gives additional information on formal run results of Web Retrieval Task at the Third NTCIR Workshop (NTCIR-3). The run results and this article are available on the CD-ROM and the following URL: ([http://research.nii.ac.jp/ntcir/workshop/web/ Participant/cdrom/](http://research.nii.ac.jp/ntcir/workshop/web/Participant/cdrom/)). Results of additional assessment will be uploaded on the same site.

Please pay attention to the following points in reading ‘Evaluation Result Sheets’ and ‘Summary of Evaluation Results —Graphs—’.

‘Evaluation Result Sheets’

We applied different methods for averaging over the topics between (i) DCG and MRR with (ii) the others: *i.e.*, precision-recall related measures. For the DCG and MRR, the values are always averaged over the topics, each of which causes at least one relevant document on the basis of the relaxed relevance level (level 2). For the precision-recall related measures, the values are averaged over the topics, each of which causes at least one relevant document on the basis of respective relevance levels: rigid level (level 1) and relaxed level (level 2).

‘Summary of Evaluation Results —Graphs—’

The averaging methods are the same as mentioned above.

All the run results are performed ‘automatically’. ‘Interactive’ run results has been submitted, however, the number of them is too small.

In each graph of recall-precision and DCG curves, the explanatory notes report the runs-IDs, which are ranked in order of ‘mean average precision over all topics’ in the case of Survey Retrieval Task or ‘document-level precision after 10 documents retrieved’ in the case of Target Retrieval Task. Here, each of the runs-IDs identifies the best run selected for the individual participating group.