

README.txt

Patent Data Collection for the NTCIR-3 Patent Retrieval Task

November 28, 2002  
NTCIR-3 Patent Retrieval Task Organizers  
ntcadm-patent@nii.ac.jp

Please use this test collection according to the user agreement  
which is referable at URL:  
<http://research.nii.ac.jp/ntcir/permission/perm-en.html#ntcir3-patent>

=====  
Files  
=====

README.txt This file  
README-euc.txt This file (EUC-ja)  
README-sjis.txt This file (SJIS)

kkh/ "Publication of unexamined patent applications" (98,99)

The full texts of the Japanese unexamined patents which were published in 1998 and 1999. These patents had not been examined by Japan Patent Office at their publishing date. Note that each patent has only text, and does not have images nor formulas. See "kkh/README-kkh.txt" for detail.

jsh/ "JAPIO Patent Abstracts" (95,96,97,98,99)

Japanese abstracts of the patents published in 1995, 1996, 1997, 1998 and 1999. These abstracts were created by JAPIO (Japan Patent Information Organization). Although applicants write abstracts of their patents, experts of JAPIO revised the abstracts if necessary. The standards whether to revise the original abstract are as follows.

- + The length of the abstract should be around 400 Japanese characters.
- + The abstract should represent its major claims.
- + Terms (words) used in the abstract are consistent with those used in the other parts of the patent.

See "jsh/README-jsh.txt" for detail.

paj/ "Patent Abstracts of Japan (PAJs)" (95,96,97,98,99)

English abstracts translated from the JAPIO Patent Abstracts published in 1995, 1996, 1997, 1998 and 1999. Note that not all the JAPIO Patent abstracts were translated into English. See "paj/README-paj.txt" for detail.

ptc/ "Patolis Test Collection"

NO WARRANTY. PLEASE ASK QUESTIONS ABOUT THIS TEST COLLECTION TO PATLIS CORPORATION ("<http://www.patolis.co.jp>"). THE ARCHIVE (ptc.zip) CONTAINS FILES WHOSE NAMES ARE JAPANESE.

A Japanese patent test collection containing search queries and the relevant patents (patent numbers) to the queries. This collection was created by JAPIO and the copyright of the collection was transferred to Patolis Co. The number of the search queries is 34 of the three major industries. The process of constructing the test collection is as follows.

- 1) Experts construct boolean search formulas from the search queries. These formulas are NOT included in the collection.
- 2) Patents are retrieved from the formulas using the PATOLIS (Patent On-Line Information System) served by PATOLIS Corporation.
- 3) Experts judge all of the retrieved patents (being relevant or not).

tools/ data converting tools

kkh2ntc.prl for "Publication of unexamined patent applications"  
jsh2ntc.prl for "JAPIO Patent Abstracts"  
paj2ntc.prl for "Patent Abstracts of Japan (PAJs)"

README.txt

All the programs convert input files to NTCIR-form files. See each program for detail.

judgments/	relevance judgments
topics/	search topics