

NTCIR-3特許検索タスクで用いた特許データコレクション

2002年11月28日
 NTCIR-3特許検索タスクオーガナイザー
 ntcadm-patent@nii.ac.jp

このテストコレクションは、覚書に従って使用してください。
 覚書の内容は、以下のURLを参照してください。
<http://research.nii.ac.jp/ntcir/permission/perm-en.html#ntcir3-patent>
<http://research.nii.ac.jp/ntcir/permission/perm-ja.html#ntcir3-patent>

=====
 ファイル構成
 =====

README.txt 本ファイル(英語版)
 README-euc.txt 本ファイル(EUC-Ja版)
 README-sjis.txt 本ファイル(SJIS版)

kkh/ 「公開特許公報全文データ(98,99)」

出願から18ヶ月たって公開される特許全文情報で、基本的には特許庁より公開されているものと同じである。特許査定を経て公開される「特許公報」とは異なり、出願した特許全件を含む。配布するのは98年99年公開分である。ただしテキスト情報のみでそれ以外の図表情情等は含まない。データの詳細は"kkh/READMEJ-kkh.txt"を参照。

jsh/ 「JAPIO出願抄録データ(95,96,97,98,99)」

JAPIO((財)日本特許情報機構)により作成された特許抄録コーパスである。公開公報には、もともと出願人により要約が付与されているが、JAPIO出願抄録は、これらを専門家が(必要であれば)修正したものである。専門家は、以下の基準により修正が必要な要約を選別する。

- + 適切な長さか(400字程度を目安)
- + 該当特許にふさわしいものか
 - 請求項の内容を含んでいるか
 - 使用語句について発明の詳細からの引用がおこなわれているか
 - その他

95年、96年、97年、98年、99年公開分を配布する。
 データの詳細は"jsh/READMEJ-jsh.txt"を参照。

paj/ 「日本国英語特許出願抄録データPAJ(95,96,97,98,99)」

JAPIO出願抄録を英語に翻訳したコーパスである。95年、96年、97年、98年、99年公開分を配布する。なお、全てのJAPIO出願抄録が翻訳されているわけではない点に注意。
 データの詳細は"paj/READMEJ-paj.txt"を参照。

ptc/ 「パトリス標準検索課題」

本データに関するの問い合わせは(株)パトリス("http://www.patolis.co.jp")までお願いします。本アーカイブ(ptc.zip)には日本語ファイル名のファイルが含まれています。

JAPIOにより作成され、(株)パトリスに引き継がれたデータセットであり、検索課題とその正解集合から成る。検索課題は、3分野34課題である。正解集合の作成手順は以下の通りである。まず各課題につき、専門家が論理検索式を作る。この検索式は本データには含まれない。次に全文検索を行い正解候補集合を得る。最後に専門家が判定を行い、最終の正解集合を得る。正解集合の特許数は総計1500程度である。

tools/ フォーマット変換プログラム

kkh2ntc.pr1 公開特許公報 --> NTC標準形
 jsh2ntc.pr1 JAPIO出願抄録 --> NTC標準形
 paj2ntc.pr1 PAJ --> NTC標準形

詳細は各プログラムを参照。

judgments/ 適合性判定
 topics/ 検索トピック