Advanced Information Retrieval Systems Assignment #1

2020-06-01

Takashi Yukawa

Assignment #1
レポート課題（その1）

Assignment #1 (1)

- Assume a document collection
  - the number of documents: 1000
    - $\{d_1, d_2, d_3, d_4, d_5, ..., d_{1000}\}$
  - For ease, we write d_n as n
    - with this notation, the document collection can be expressed as {1, 2, 3, 4, 5, ..., 1000}
- For a given query, the relevant (correct) documents are
  - {26, 45, 135, 186, 192, 208, 239, 254, 346, 468, 511, 517, 595, 624, 627, 653, 752, 785, 987, 1000}

Assignment #1 (2)

▶ An IR system gives the following retrieval result (answer set)
{595, 346, 208, 880, 186, 808, 192, 33, 517, 653, 125, 92, 624, 363, 654, 135, 393, 913, 452, 45, 447, 799, 1000, 987, 468, 924, 870, 766, 540, 752, 872, 461, 232, 904, 126, 26, 287, 785, 97, 627, 204, 511, 176, 681, 719, 455, 953, 254, 6, 258}

Assignment #1 (3)

1. Draw the Recall-Precision graph

2. Compute the following performance indices

   ▶ Average Precision

   ▶ 11-point Average Precision

   ▶ R-Precision

Submission

- Submission
  - Deadline: June 15th 13:00
  - Accept the report with ILIAS
    - PDF file only

  - On-line submission: use ILIAS ("assignment1 submission")
    - File name has to be s<student_id(upper 6 digit)>.pdf
      - Example: s012345.pdf (if the student_id is 01234567)
      - Use only ASCII characters for the file name; do not use Zenkaku characters
    - Submission will be closed at 13:00 with the ILIAS server's clock
      - assume that the clock on the server is not quite accurate