Advanced Information Retrieval Systems Assignment #4

2023-07-03

Takashi Yukawa

Assignment #4
レポート（その4）課題

Assignment #4 (1)

- Assume Web pages

  - $\{p_1, p_2, p_3, p_4, p_5\}$

- Assume links between Web pages

  - p_1: $\{p_2, p_3, p_5\}$

  - For ease, we write a link p1 to p_n as n

    - with this notation, outgoing links of the web page p_1 can be expressed as 2,3,5.

- The file "IRSys23_WebLinks.csv" can be downloaded from ILIAS

  - each line corresponds each of the Web pages

Assignment #4 (2)

▶ assg2WebLinks.csv$\{p_1, p_2, p_3, p_4, p_5\}$

   3,4

   1,4

   …

> p1 has outgoing link to p_3 and p_4

> p2 has outgoing link to p_1 and p_4

▶ Compute the transition probability matrix

   ▶ follow class #11 slide 8

▶ Compute the score for each page based on (simple) PageRank algorithm

   ▶ follow class #11 slide 9

   ▶ **Normalize the score** (i.e. sum of scores = 1)

   ▶ Denoting the score of the page p_i as r_i

Assignment #4 (3)

- Submission format of the transition probability matrix
  - CSV file
  - Each line corresponds each row
  - Each Value has to be rounded to 4 digit after the decimal point
- Submission format of the score for each page
  - One line CSV file
  - r_1, r_2, r_3, r_4, r_5
  - Each Value has to be rounded to 4 digit after the decimal point
- Computing eigenvalues and eigenvectors
  - use online (cloud) computation web site
  - use Maxima (free software)
  - use Python with NumPy (**carefully read the reference manual**)

Assignment #4 Submission

- Accept the report with ILIAS only

- Upload at "Assignments" > "Assignment #4 Submission Transition Probability Matrix" and "Assignments" > "Assignment #4 Submission Page Score List"

  - Upload 2 separate files: transition probability matrix and scores of the pages

    - File name for the transition probability matrix has to be T<student_id_6digits>.csv

    - File name for the score to be P<student_id_6digits>.csv

    - Use only ASCII characters for the file name; do not use Zenkaku characters

- Filename example

  - Assume your student ID 12345678

  - The transition probability matrix: T123456.csv

  - The scores for the pages: P123456.csv

Submission Format

▶ Follow the instruction in each assignment **strictly**

▶ **A submission file with illegal format is not evaluated (i.e. 0 point)**

Deadline

- Deadline: July 17th **14:30**
    - On July 17th, we have no face-to-face class
    - That period is for working on the assignment **#4**

- Submission will be closed at the deadline with the ILIAS server's clock
    - do not assume that the clock on the server is quite accurate