

Faculty of Engineering and Technology

Electrical and Computer Engineering Department

INFORMATION RETRIEVAL WITH APPLICATIONS OF NLP ENCS5342

**Assignment #4**

Prepared by: Mazen Batrawi - 1190102

Instructor: Dr. Adnan Yahya

Section: 1

BIRZEIT

JUNE – 2023

Q.1

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | docID | Graded Relevance | Binary Relevance |
| 1 | 43 | 3 | 1 |
| 2 | 531 | 0 | 0 |
| 3 | 183 | 4 | 1 |
| 4 | 102 | 2 | 1 |
| 5 | 10 | 2 | 1 |
| 6 | 1051 | 0 | 0 |
| 7 | 1031 | 1 | 1 |
| 8 | 332 | 1 | 1 |
| 9 | 573 | 0 | 0 |
| 10 | 128 | 2 | 1 |

1. Precision = Relevant retrieved / documents retrieved

P@6 = 4 / 6 = 0.6667

P@10 = 7 / 10 = 0.7

1. Recall = relevant retrieved / relevant documents

R@6 = 4 / 7 = 0.571

R@10 = 7 / 7 = 1

1. For the maximization needed, we have to put at least 6 relevant documents at the beginning, so that P@6 will become 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | docID | Graded Relevance | Binary Relevance |
| 1 | 43 | 3 | 1 |
| 2 | 128 | 2 | 1 |
| 3 | 183 | 4 | 1 |
| 4 | 102 | 2 | 1 |
| 5 | 10 | 2 | 1 |
| 6 | 1031 | 1 | 1 |
| 7 | 1051 | 0 | 0 |
| 8 | 332 | 1 | 1 |
| 9 | 573 | 0 | 0 |
| 10 | 531 | 0 | 0 |

1. The value of P@10 will always be 7 / 10 in this example, because the number of relevant documents is 7, and for any permutation, when we are at row 10, we must have seen the 7 relevant documents, so the answer will be 7 / 10 = 0.7.
2. From the above table, the recall at 6 is 6 / 7, which is the maximum we can get.
3. The value of R@10 will always be 7 / 7 in this example, because the number of relevant documents is 7, and for any permutation, when we are at row 10, we must have seen the 7 relevant documents, so the answer will be 7 / 7 = 1.
4. The average precision AP is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | docID | Graded Relevance | Binary Relevance | Precision |
| 1 | 43 | 3 | 1 | 1 / 1 = 1 |
| 2 | 531 | 0 | 0 | 1 / 2 = 0.5 |
| 3 | 183 | 4 | 1 | 2 / 3 = 0.6667 |
| 4 | 102 | 2 | 1 | 3 / 4 = 0.75 |
| 5 | 10 | 2 | 1 | 4 / 5 = 0.8 |
| 6 | 1051 | 0 | 0 | 4 / 6 = 0.667 |
| 7 | 1031 | 1 | 1 | 5 / 7 = 0.714 |
| 8 | 332 | 1 | 1 | 6 / 8 = 0.75 |
| 9 | 573 | 0 | 0 | 6 / 9 = 0.6667 |
| 10 | 128 | 2 | 1 | 7 / 10 = 0.7 |

AP = (1/1 + 2/3 + 3/4 + 4/5 + 5/7 + 6/8 + 7/10) / 7 = 0.769

1. To maximize the AP, we have to put all relevant documents at the beginning. The idea is to increase the numerator of APs equation, and to increase it, we have to guarantee the maximum precision achievable at each entry. The following table will demonstrate the solution:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | docID | Graded Relevance | Binary Relevance | Precision |
| 1 | 43 | 3 | 1 | 1 / 1 = 1 |
| 2 | 128 | 2 | 1 | 2 / 2 = 1 |
| 3 | 183 | 4 | 1 | 3 / 3 = 1 |
| 4 | 102 | 2 | 1 | 4 / 4 = 1 |
| 5 | 10 | 2 | 1 | 5 / 5 = 1 |
| 6 | 1031 | 1 | 1 | 6 / 6 = 1 |
| 7 | 332 | 1 | 1 | 7 / 7 = 1 |
| 8 | 1051 | 0 | 0 | 7 / 8 = 0.875 |
| 9 | 573 | 0 | 0 | 7 / 9 = 0.7778 |
| 10 | 531 | 0 | 0 | 7 / 10 = 0.7 |

AP = (1+1+1+1+1+1+1) / 7 = 1.

1. See the below table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | docID | Graded Relevance | Log n | Rel / Log n | DCG n |
| 1 | 43 | 3 | - | - | 3 |
| 2 | 531 | 0 | 1 | 0 | 3 |
| 3 | 183 | 4 | 1.58 | 2.53 | 5.53 |
| 4 | 102 | 2 | 2 | 1 | 6.53 |
| 5 | 10 | 2 | 2.32 | 0.86 | 7.39 |

1. 1. The ideal top documents are found using the graded relevance, the higher graded relevance the higher rank.

The documents are, in order: 183, 43, 102, 10, 128.

2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | docID | Graded Relevance | Log n | Rel / Log n | GCG n |
| 1 | 183 | 4 | - | - | 4 |
| 2 | 43 | 3 | 1 | 3 | 7 |
| 3 | 102 | 2 | 1.58 | 1.27 | 8.27 |
| 4 | 10 | 2 | 2 | 1 | 9.27 |
| 5 | 128 | 2 | 2.32 | 0.86 | 10.13 |

3. NDCG5 = DCG5 / IDCG5 = 7.39 / 10.13 = 0.729

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | docID | Graded Relevance | Binary Relevance | Recall | Precision |
| 1 | 43 | 3 | 1 | 0.143 | 1 |
| 2 | 531 | 0 | 0 | 0.143 | 0.5 |
| 3 | 183 | 4 | 1 | 0.286 | 0.6667 |
| 4 | 102 | 2 | 1 | 0.429 | 0.75 |
| 5 | 10 | 2 | 1 | 0.571 | 0.8 |
| 6 | 1051 | 0 | 0 | 0.571 | 0.6667 |
| 7 | 1031 | 1 | 1 | 0.714 | 0.714 |
| 8 | 332 | 1 | 1 | 0.857 | 0.75 |
| 9 | 573 | 0 | 0 | 0.857 | 0.6667 |
| 10 | 128 | 2 | 1 | 1 | 0.7 |

|  |  |
| --- | --- |
| Recall | Precision |
| 0.143 | 1 |
| 0.143 | 0.5 |
| 0.286 | 0.6667 |
| 0.429 | 0.75 |
| 0.571 | 0.8 |
| 0.571 | 0.6667 |
| 0.714 | 0.714 |
| 0.857 | 0.75 |
| 0.857 | 0.6667 |
| 1 | 0.7 |

Q2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Company | Bid | CTR | Bid \* CTR | Rank | Paid |
| A | 1 | 0.08 | 0.08 | 5 | minimum |
| B | 2 | 0.05 | 0.10 | 4 | 1.61 |
| C | 3 | 0.07 | 0.21 | 1 | 2.8671 |
| D | 4 | 0.04 | 0.16 | 3 | 2.51 |
| E | 5 | 0.04 | 0.20 | 2 | 4.01 |

* Piad (Rank 1) = Bid2 \* CTR2 / CTR1 = 0.20 / 0.07 = 2.8571

The second price auction rule says that the advertiser pays the minimum amount necessary to maintain their position in the auction (plus 1 cent). So, the paid value for the first rank will be 2.8671.

* Piad (Rank 2) = Bid3 \* CTR3 / CTR2 = 0.16 / 0.04 = 4, plus 1 cent = 4.01
* Piad (Rank 3) = Bid4 \* CTR4 / CTR3 = 0.10 / 0.04 = 2.5, plus 1 cent = 2.51
* Piad (Rank 4) = Bid5 \* CTR5 / CTR4 = 0.08 / 0.05 = 1.6, plus 1 cent = 1.61

Q.3

A(x) = x mod 5

B(x) = (3x + 1) mod 5

C(x) = (2x + 1) mod 5

|  |  |  |  |
| --- | --- | --- | --- |
|  | Doc1 | Doc2 | Doc3 |
| A (x)  B (x)  C (x) | **∞**  **∞**  **∞** | **∞**  **∞**  **∞** | **∞**  **∞**  **∞** |
| A (1) = 1  B (1) = 4  C (1) = 3 | 1 1  4 4  3 3 | 1 1  4 4  3 3 | - **∞**  - **∞**  - **∞** |
| A (2) = 2  B (2) = 2  C (2) = 0 | 2 1  2 2  0 0 | 2 1  2 2  0 0 | 2 2  2 2  0 0 |
| A (3) = 3  B (3) = 0  C (3) = 2 | 3 1  0 0  2 0 | - 1  - 2  - 0 | 3 2  0 0  2 0 |
| A (4) = 4  B (4) = 3  C (4) = 4 | - 1  - 0  - 0 | 4 1  3 2  4 0 | 4 2  3 0  4 0 |
| A (5) = 0  B (5) = 1  C (5) = 1 | - 1  - 0  - 0 | - 1  - 2  - 0 | 0 0  1 0  1 0 |

J (D1, D2) = 2 / 3 = 0.667

J (D1, D3) = 2 / 3 = 0.677

J (D2, D3) = 1 / 3 = 0.333