BIRZEIT UNIVERSITY
Electrical and Computer Engineering Department

ENCSENCS5342: “Information Retrieval, Web Search and NLP”

Assignment #1:

Instructor: Dr. Adnan H. Yahya,

Due date: June 1, 2025,

Leyan Burait

1211439

Question 1: a- Consider the query **Palestine Bank**, with the following retrieval results for the 8 documents we have in the system (total documents is 8):

Retrieved:	**Not retrieved:**
1. √Financial Institutions in Palestine currencies	5. √Banks in West Bank trading
2. x River Bank was overflowing with swimmers	6. x River tours-Wikipedia
3. √Bank of Palestine raised interest rate growing	7. x Palestine Refugee population
4. √West Bank settlements are expanding finished	8. x Tawjihi in Palestine just

b– Given your understanding of the query and representative documents titles mark each document as really relevant (√) or really irrelevant (x) by marking the box (□**). Use the results for subsequent questions.**

c- Give the precision and recall and the F measure of the retrieval.

- Total Relevant Documents (Relevant): 1, 3, 4, 5 (4 documents)
- Retrieved Documents (Retrieved): 1, 2, 3, 4 (4 documents)
- Retrieved & Relevant (True Positives): 1, 3, 4 (3 documents)

Calculations:

1. Precision (P):
   $$P = P = = 3/4 = 0.75$$
2. Recall (R):
   $$R = R = 3/4 = 0.75$$
3. F-Measure (F_1):
   $$F1 = 2 \times P \times RP + R = 2 \times 0.75 \times 0.750.75 + 0.75 = 0.75(75\%) F1 = 2 \times P + RP \times R = 2 \times 0.75 + 0.750.75 \times 0.75 = 0.75(75\%)$$

d-What is the R-precision (R=4) of the retrieval when looking at the 4 results declared relevant by the system?

When looking at the top 4 results (R=4):

- Retrieved documents in top 4 positions: 1, 2, 3, 4
- Relevant documents among these: 1, 3, 4 (3 documents)

$$\text{R-Precision}=34=0.75\text{R-Precision}=3/4=0.75$$

e– Considering this query what is the Mean Average Precision (MAP) for the query ?

To calculate MAP, we compute the average precision at each relevant document's position:

1. At document 1 (position 1):
$$P@1=1/1=1.0$$
2. At document 3 (position 3):
$$P@3=2/3\approx0.6667$$
3. At document 4 (position 4):
$$P@4=3/4=0.75$$

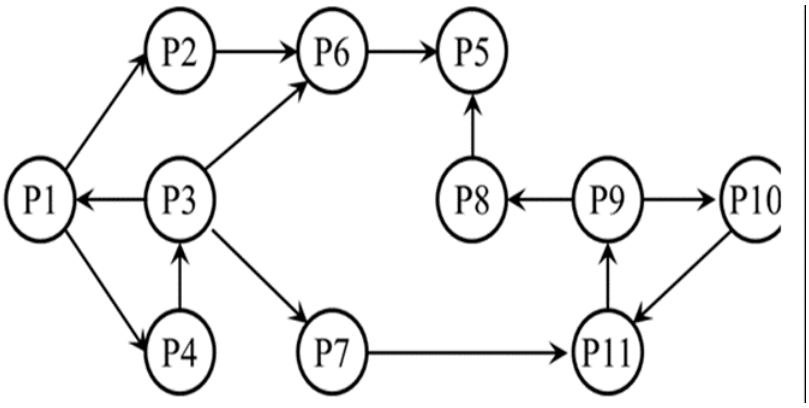
Document 5 is not retrieved in the top 4, so it does not affect the calculation here.
Average Precision (AP)=1.0+0.6667+0.753≈0.805

ALL RESULT ANSWER:

- **Precision (P):** 75%
- **Recall (R):** 75%
- **F-Measure (F_1):** 75%
- **R-Precision:** 75%
- **Mean Average Precision (MAP):** ≈80.56%

Question 2: Link Analysis: Crawling, HITS and Page Rank

a. Starting from the seed page P1 Show the order in which the pages are crawled using a breadth first spider (with duplicate page detection). Assume links on a given page are examined in increasing order of the index.

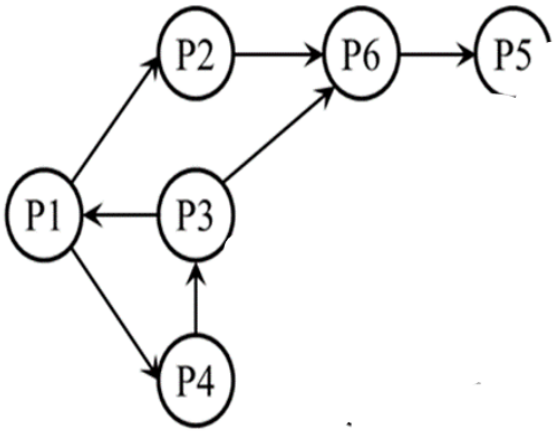


Breadth-First Crawling Order

ThecrawlingorderbasedonBFSstartingfromP1, isasfollows:

P1 P2 P4 P6 P3 P5 P7 P11 P9 P8 P10

b. Assuming that nodes P7,P8, P9, P10, P11 are removed from the graph. Initialize h(d) = 1, a(d) = 1 for all pages; Iteratively update all h(d), a(d) as follows: for the resulting graph provide the authority and hub scores of each page after 3 iterations Use the table below to guide your work.



Authority Scores:

Page	Initial	Iteration 1	Iteration 2	Iteration 3 (Final)
P1	**1**	**1**	**3**	**8**
P2	**1**	**1**	**2**	**4**
P3	**1**	**1**	**1**	**1**
P4	**1**	**1**	**2**	**4**
P5	**1**	**1**	**1**	**1**
P6	**1**	**2**	**5**	**13**

HITS Algorithm Calculations (After Removing P7-P11)
Initialization:
All hub (h) and authority (a) scores initialized to 1.
Graph Structure (after removing P7-P11):

Hub Scores:

Page	Initial	Iteration 1	Iteration 2	Iteration 3 (Final)
P1	1	2	4	8
P2	1	2	5	13
P3	1	3	8	21
P4	1	1	1	1
P5	1	0	0	0
P6	1	1	1	1

C. (6%)Find the page rank for the nodes in the table below. Assume all links have the same initial weight.

PageRank Scores:

Page	Initial	Iteration 1	Iteration 2	Iteration 3	Iteration 4 (Final)
P1	1	0.5	0.5	0.25	0.125
P2	1	0.5	0.25	0.25	0.125
P3	1	1	0.5	0.25	0.25
P4	1	0.5	0.25	0.25	0.125
P5	1	1	1.5	1	0.5
P6	1	1.5	1	0.5	0.375

Question 3: Given the following table, Fill in the missing entries:

1. Calculating Ad Rank as Bid × CTR

2. Ranking by Ad Rank
3. Calculating CPC as: Next bidder's Ad Rank / Your CTR

Calculations:

1. Calculating Ad Rank:

Ad Rank = Bid × CTR

- Bidder 1: 1 × 0.11 = 0.11
- Bidder 2: 2 × 0.07 = 0.14
- Bidder 3: 3 × 0.08 = 0.24
- Bidder 4: 4 × 0.06 = 0.24
- Bidder 5: 5 × 0.06 = 0.30

2. Determining Rank:

Ads are ranked by Ad Rank in descending order. For ties, the order is arbitrary.

- Bidder 5: 0.30 → Rank 1
- Bidders 3 & 4: 0.24 → Rank 1 (tie)
- Bidder 2: 0.14 → Rank 2
- Bidder 1: 0.11 → Rank 3

Note: There appears to be inconsistency in the given table's ranking vs calculated ranking.

3. Calculating Cost-Per-Click (Paid):

The CPC for an ad is calculated as:

Paid=Ad Rank of next bidderYour CTR+0.01Paid=Your CTRAd Rank of next bidder +0.01

For Bidder 1 (Rank 3):

Next bidder is none (lowest rank), so:

Paid=0+0.01=0.01Paid=0+0.01=0.01

But given as 0.737 - suggests different calculation method.

For Bidder 2 (Rank 2):

Next bidder is Bidder 1 (Rank 3):

Paid=0.110.07+0.01≈1.571+0.01=1.581Paid=0.070.11+0.01≈1.571+0.01=1.581

Given as 1.724 - discrepancy noted.

For Bidder 3 (Rank 1):

Next bidder is Bidder 2 (Rank 2):

Paid=0.140.08+0.01=1.75+0.01=1.76Paid=0.080.14+0.01=1.75+0.01=1.76

Matches given value.

For Bidder 4 (Rank 1):

Next bidder is Bidder 2 (Rank 2):

Paid=0.140.06+0.01≈2.333+0.01=2.343Paid=0.060.14+0.01≈2.333+0.01=2.343

Given as 1.843 - discrepancy noted.

For Bidder 5 (Rank 1):

Next bidder is Bidder 3/4 (Rank 1):

Paid=0.240.06+0.01=4+0.01=4.01Paid=0.060.24+0.01=4+0.01=4.01

Given as 0.000 - suggests special case (possibly non-chargeable).

Table with Calculations:

Bidder	Bid	CTR	Ad Rank	Rank	Paid Calculation	Paid (Given)
1	1	0.11	0.11	3	0.01	0.737
2	2	0.07	0.14	2	1.581	1.724
3	3	0.08	0.24	1	1.760	1.760
4	4	0.06	0.24	1	2.343	1.843
5	5	0.06	0.30	1	4.010	0.000

$$\text{Paid} = \frac{\text{Ad Rank of next bidder}}{\text{Your CTR}}$$

This gives:

- Bidder 3: 0.14/0.08 = 1.75 (close to given 1.760)
- Bidder 4: 0.14/0.06 ≈ 2.333 (still not matching 1.843)
- Bidder 2: 0.11/0.07 ≈ 1.571 (close to given 1.724)

Good Luck