Task 1: Evaluation Metrics and α-NDCG

1.1 Name two aspects that are not considered when using the standard information retrieval evaluation metrics precision and recall to measure the relevance of a ranked document by a search engine.

- Ranking
- Diversity

1.2 Given the equations for α-NDCG@k and for the gain (for explanations see appendix), answer the following questions for each letter in "α-NDCG":

- What aspect does the letter stand for?
- What is that aspect's role in α-NDCG?
- In which part of the equations is that aspect covered?

$$\alpha - NDCG(q,k) = Z_{kq} \sum_{m=1}^{k} \frac{G[m]}{log_2(1+m)}\ ,\ \ G[m] = \sum_{i=1}^{|S|} J(d_m, n_i)(1-\alpha)^{r_{i,m-1}}$$

Letter	Name	Description	Part of the Equation
G	Gain	The information gained from a specific document at position *m*.	$G[m]$
α	Diversification	When computing the gain, redundant results are discouraged.	$(1-\alpha)^{r_{i,k-1}}$
D	Discounted	The gain is reduced logarithmically proportional to the position of the result. That means, relevant documents at the top of the list are valued more.	$log_2(1+m)$
C	Cumulative	The gain of each document at positions 1..k is added to the score.	$\sum_{m=1}^{k}$
N	Normalized	The score is normalized by the best possible ranking's score.	$Z_{k,j}$

Task 2: Computing the Gain Vector for α-NDCG

Given is the following ranking of documents and the information nuggets they contain.

Document			Information Nuggets						Gain
ID	Rank	Title	n_1	n_2	n_3	n_4	n_5	n_6	
a	1	Carnival Re-Enters Norway Bidding		X		X			2
b	2	NORWEGIAN CRUISE LINE SAYS OUTLOOK IS GOOD		X					0.5
c	3	Carnival, Star Increase NCL Stake		X					0.25
d	4	Carnival, Star Solidify Control							0
e	5	HOUSTON CRUISE INDUSTRY GETS BOOST WITH...	X					X	2
f	6	TRAVELERS WIN IN CRUISE TUG-OF-WAR	X						0.5
g	7	ARMCHAIR QUARTERBACKS NEED... THIS CRUISE			X				1
h	8	EUROPE, CHRISTMAS ON SALE	X						0.25
i	9	TRAVEL DEALS AND DISCOUNTS							0
j	10	HAVE IT YOUR WAY ON THIS SHIP							0

2.1 Compute the gain for each document.

$$G[1] = 0 \cdot (1-0.5)^0 + 1 \cdot (1-0.5)^0 + 0 \cdot (1-0.5)^0 + 1 \cdot (1-0.5)^0 + 0 \cdot (1-0.5)^0 + 0 \cdot (1-0.5)^0 = 2$$

$$G[2] = 0 \cdot (1-0.5)^0 + 1 \cdot (1-0.5)^1 + 0 \cdot (1-0.5)^0 + 0 \cdot (1-0.5)^1 + 0 \cdot (1-0.5)^0 + 0 \cdot (1-0.5)^0 = \frac{1}{2} = 0.5$$

$$G[3] = 0 \cdot (1-0.5)^0 + 1 \cdot (1-0.5)^2 + 0 \cdot (1-0.5)^0 + 0 \cdot (1-0.5)^1 + 0 \cdot (1-0.5)^0 + 0 \cdot (1-0.5)^0 = \frac{1}{4} = 0.25$$

$$G[4] = 0$$

$$G[5] = 1 \cdot (1-0.5)^0 + 1 \cdot (1-0.5)^0 = 2$$

$$G[6] = 1 \cdot (1-0.5)^1 = 0.5$$

$$G[7] = 1 \cdot (1-0.5)^0 = 1$$

$$G[8] = 1 \cdot (1-0.5)^2 = 0.25$$

$$G[9] = 0$$

$$G[10] = 0$$

Task 3: Computing α-NDCG

Compute the α-NDCG for the top-10 results ($k = 10$) for a single query q that returns the document ranking from Task 2.

3.1 Create the perfect ranking for that document ranking and compute the corresponding gains and its DCG for $\alpha = 0.5$.

Document	Information Nuggets						Gain
ID	n_1	n_2	n_3	n_4	n_5	n_6	
a		X		X			2
e	X					X	2
g			X				1
b		X					0.5
f	X						0.5
c		X					0.25
h	X						0.25
d							0
i							0
j							0

$$DCG(Q,k) = \frac{2}{log_2(1+1)} + \frac{2}{log_2(1+2)} + \frac{1}{log_2(1+3)} + \frac{0.5}{log_2(1+4)} + \frac{0.5}{log_2(1+5)} + \frac{0.25}{log_2(1+6)} + \frac{0.25}{log_2(1+7)} + \frac{0}{log_2(1+8)} + \frac{0}{log_2(1+9)} + \frac{0}{log_2(1+10)} \approx 4.343$$

3.2 Compute the α-NDCG of the ranking in Task 2.

$$\alpha - NDCG(Q,k) = Z_{10,q} \cdot \sum_{m=1}^{10} \frac{G[m]}{log_2(1+m)}$$

$$= \frac{1}{4.343} \cdot \left(\frac{2}{log_2(1+1)} + \frac{0.5}{log_2(1+2)} + \frac{0.25}{log_2(1+3)} + \frac{0}{log_2(1+4)} + \frac{2}{log_2(1+5)} + \frac{0.5}{log_2(1+6)} + \frac{1}{log_2(1+7)} + \frac{0.25}{log_2(1+8)} + \frac{0}{log_2(1+9)} + \frac{0}{log_2(1+10)}\right)$$

$$\approx \frac{1}{4.343} \cdot 3.804 \approx 0.858$$

4. Diversification algorithm

Given is the following list L[l] of top-k queries ranked by relevance:

Rank	ID	P(Q|K)	Query Interpretation
1	Q_1	0.95	{“consideration”: movie.title, “christopher guest“:director.name}
2	Q_2	0.87	{“christopher guest“:director.name}
3	Q_3	0.78	{“guest”: movie.title, “consideration”: movie.title}
4	Q_4	0.23	{“consideration”: movie.title}

Use the diversification algorithm to return the list R[r] of the relevant and diverse query interpretations.

4.1 Use the Jaccard similarity to compute the similarity between each pair of query interpretations.

	Q_1	Q_2	Q_3	Q_4
Q_1		½	⅓	½
Q_2			0	0
Q_3				½
Q_4				

<u>4.2 Given these queries, interpretation probabilities and similarity values, perform the diversification algorithm to generate a diversified ranking for $r = 3$ and $\alpha = 0.5$.</u>

$R =< Q_1 >$

1. $best\_score = 0$
    a. $Score(Q_2) = \frac{1}{2} \cdot P(Q_2, K) - (1 - \frac{1}{2})\Sigma_{q \in QI} \frac{Sim(Q_2,q)}{|QI|}$
    $= \frac{1}{2} \cdot 0.87 - \frac{1}{2} \cdot \frac{Sim(Q_2,Q_1)}{1} = 0.435 - \frac{1}{2} \cdot \frac{1}{2} = 0.185$
    $\rightarrow best\_score = 0.185$
    b. $Score(Q_3) = \frac{1}{2} \cdot 0.78 - \frac{1}{2} \cdot \frac{1}{3} = 0.22\overline{3}$
    $\rightarrow best\_score = 0.22\overline{3}$
    c. $Score(Q_4) = \frac{1}{2} \cdot 0.23 - \frac{1}{2} \cdot \frac{1}{2} =- 0.135$
2. $R =< Q_1, Q_3 >$
3. $best\_score = 0$
    a. $Score(Q_2) = \frac{1}{2} \cdot P(Q_2, K) - (1 - \frac{1}{2})\Sigma_{q \in QI} \frac{Sim(Q_2,q)}{|QI|}$
    $= \frac{1}{2} \cdot 0.87 - \frac{1}{2} \cdot (\frac{Sim(Q_2,Q_1)}{2} + \frac{Sim(Q_2,Q_3)}{2})$
    $= 0.435 - \frac{1}{2} \cdot (\frac{(\frac{1}{2})}{2} + \frac{0}{2}) = 0.31$
    $\rightarrow best\_score = 0.31$
    b. $Score(Q_4) \leq \frac{1}{2} \cdot 0.23 = 0.115 \leq 0.31 = best\_score$
    $\Rightarrow$ break
    c. $R =< Q_1, Q_3, Q_2 >$
4. Stop, because $|R| = r$.

Appendix

α-NDCG@k for a single query q

$$\alpha - NDCG(q,k) = Z_{kq} \sum_{m=1}^{k} \frac{G[m]}{log_2(1+m)}$$

k – the number of results in the ranking that are considered

Z_{kq} – a normalization factor such that the $\alpha - NDCG$ of the perfect ranking for q at *k* is 1.

$G[m] = \sum_{i=1}^{|S|} J(d_m, n_i)(1-\alpha)^{r_{i,m-1}}$ – gain of a result

$J(d_m, n_i)$ – an assessor judged that document d_m contains nugget n_i.

α – a factor that balances relevance and novelty

$r_{i,m-1}$ – the number of documents ranked up to position $m-1$ that have been judged to contain nugget n_i.

$S = \{n_1, ..., n_s\}$ – the set of information nuggets

Jaccard Similarity between two queries Q_1, Q_2 and their interpretations (I_1, I_2)

$$Sim(Q_1, Q_2) = \frac{I_1 \cap I_2}{I_1 \cup I_2}$$

Query Score combining Relevance and Similarity

$$Score(Q) = \lambda \cdot P(Q|K) - (1-\lambda) \cdot \Sigma_{q \in QI} \frac{Sim(Q,q)}{|QI|}$$

QI – a set of previously selected query interpretations in the ranked list

λ – a factor to balance relevance and novelty in the diversification procedure

Diversification Algorithm

Input: list $L[i]$ of top-k query interpretations ranked by relevance

Output: list $R[r]$ of the relevant and diverse query interpretations

```
Proc Select Diverse Query Interpretations
copy the first element of L in R;
while (less than r elements selected) {
      //select the best candidate for R[i]
      initialize best_score=0;
      while (more candidates for R[i] in L) {
            if (Score(L[j]) > best_score) set new best_score;
      }
      add the candidate from L with best_score to R
}
End Proc;
```