Read Me file : folder contents

1 Last Place Ranking

The **last place ranking** is defined as $L = 1 - \frac{Rank-n}{N-n}$, where $Rank$ indicates the rank at which the last relevant object is found, n is the number of relevant items, and N is the size of the whole database ($N = 400$). The performance of a method is as good as L is high; in fact $L \in [0, 1]$, and the best value, occurring when the last relevant object is in the n^{th} position, is $L = 1 - \frac{n-n}{400-n} = 1$.
The figure refers to highly relevant items, and it is an average with respect to the queries.

2 Precision-Recall folders

Recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the database, while precision is the ratio of the number of relevant records retrieved to the size of the return vector. Depending on which kind of relevant items are considered, 2 different folders are available.
Graphs refer both to each query and to the average.

3 Vectors

In order to compute the Normalized Discounted Cumulated Gain (NDCG) vector, the gain vector G has to be defined accordingly to the ground-truth. The gain vector is obtained by the ranked list where the model's identifiers are substituted with their ''relevance scores'' and where the relevance scores depend on the definition of the ground-truth. In particular, highly relevant, marginally relevant and non-relevant models have relevance scores 2, 1 and 0 respectively. The Discounted Cumulated Gain vector is recursively defined as:

$$DCG[i] = \begin{cases} G[i] & \text{if } i = 1 \\ DCG[i-1] + (G[i]/\log i) & \text{otherwise} \end{cases}$$

where $G[i]$ represents the value of the gain vector at the position i. The normalized discounted cumulated gain vector NDCG is obtained by dividing DCG by the ideal cumulated gain vector.