

# Información no estructurada

## Práctica 2 -Evaluación

Javier Aróstegui Martín

### Contenido

Implementación: .....	2
Evaluación Offline: .....	2
Métricas: .....	2
Precisión:.....	2
Recall:.....	2
Media Armónica: .....	2
Mean Average Precision (MAP):.....	3
Mean Reciprocal Rank (MRR):.....	3
nDCG: .....	3
Expected Reciprocal Rank (ERR):.....	3
Rank-Biased Precision (RBP):.....	4
Resultados:.....	4
Evaluación Online:.....	4
Métricas: .....	4
Clicks por consulta:.....	4
Tasa de abandono: .....	4
Mean RR:.....	4
Max RR: .....	4
Resultados:.....	5

## Implementación:

La implementación se ha realizado en Python con la ayuda de las librerías de pandas para el manejo de los datos y de numpy para operaciones matemáticas. Lo primero que se ha realizado es un fichero `juntarFicheros.py` que se encarga de hacer un merge de `qrels.txt` con cada uno de los resultados de cada buscador para así crear una base de datos para cada buscador que contiene toda la información que se necesita. Luego en el fichero `metricsOffline.py` se lee este fichero y se calculan las métricas de cada uno de los 4 buscadores iterando las consultas de cada uno de ellos y luego realizando la media. El fichero `metricsOnline.py` se encarga de la parte de evaluación online y por tanto calcula las métricas correspondientes a esa parte.

## Evaluación Offline:

Lo primero que ha sido necesario para resolver esta tarea ha sido tener los datos de una manera con la que se pueda trabajar cómodamente. Mediante la librería pandas se han juntado el fichero `qrels.txt` con cada uno de los resultados de cada uno de los buscadores creando así un fichero con las columnas necesarias de consulta, posición y relevancia para cada uno de ellos.

## Métricas:

### Precisión:

$$P = \frac{|\text{Relevant} \cap \text{Returned}|}{|\text{Returned}|}$$

Para cada consulta dentro de un mismo buscador para luego hacer la media de las precisiones de todas las consultas en ese buscador.

### Recall:

$$R = \frac{|\text{Relevant} \cap \text{Returned}|}{|\text{Relevant}|}$$

En este caso para calcular el número total de relevantes se han utilizado todos los documentos devueltos por todos los buscadores sin repetición.

### Media Armónica:

$$\frac{2PR}{P + R}$$

Con los valores de Precisión y Recall, este cálculo es trivial.

Mean Average Precision (MAP):

$$AP = \text{avg}_{d_k \text{ relevant}} P@k$$

Para el cálculo de esta métrica debemos realizar la media de la precisión@k de cada uno de los k documentos relevantes (solo los devueltos) . Tras realizar este cálculo en cada una de las consultas, debemos realizar la media para que se convierta en MAP.

Mean Reciprocal Rank (MRR):

$$RR = \frac{1}{\min\{k | d_k \text{ relevant}\}}$$

Aquí calculamos 1 partido de la posición donde se encuentra el primer documento relevante en cada ranking para, como en el caso anterior, posteriormente realizar la media y obtener MRR.

nDCG:

$$nDCG = \frac{DCG}{IDCG}$$

nDCG se calcula como DCG dividido entre el DCG idóneo (iDCG), por lo tanto, DCG está normalizado.

$$DCG = \sum_{d_k \in R} \frac{g(d_k)}{\log_2(k + 1)}$$

Para calcular DCG calculamos la suma de los grados de relevancia de los documentos devueltos en cada consulta y lo dividimos por el logaritmo en base 2 de su posición más 1.

$$IDCG = \max_{R \in \sigma(\mathcal{D})} DCG(R, q)$$

Para calcular IDCG realizamos el mismo cálculo, pero ordenando los documentos relevantes devueltos en el orden perfecto, es decir, los más relevantes primero.

Expected Reciprocal Rank (ERR):

$$ERR = \sum_{k=1}^{|R|} \frac{1}{k} p(r|d_k) \prod_{j=1}^{k-1} (1 - p(r|d_j))$$

En este caso calculamos la utilidad y lo vamos acumulando, multiplicándolo al mismo tiempo por 1/k (posición del documento en el ranking) y por 1 menos la utilidad pasada. De la misma

manera que todas las demás métricas, esto se calcula para cada consulta y luego se realiza la media para ese buscador.

Rank-Biased Precision (RBP):

$$RBP = (1 - p) \sum_{d_k \in R} p^{k-1} g(d_k)$$

Los resultados que se presentan se han realizado con  $p = 0.8$ . El valor  $p$  indica la probabilidad que tiene el usuario de pasar al siguiente documento y es constante.

Resultados:

	Precisión	Recall	Media Armónica	MAP	MRR	nDCG	ERR	RBP
Google	0.658	0.543	0.595	0.701	0.903	0.591	0.638	0.94
DuckDuckGo	0.557	0.412	0.473	0.621	0.797	0.462	0.582	0.825
Ecosia	0.537	0.396	0.456	0.564	0.751	0.433	0.527	0.756
Bing	0.527	0.383	0.443	0.58	0.746	0.426	0.529	0.774

Podemos observar que Google obtiene los mejores resultados en todas las métricas y además por un margen significativo. El resto de los buscadores se encuentran más cercanos entre sí y obtienen un rendimiento similar en la mayoría de las métricas.

Evaluación Online:

En el caso de la evaluación online los ficheros que se han utilizado han sido click-count.txt y click-log.txt que se han separado para cada buscador. A partir de ahí ya podemos iterar por consulta dentro de cada buscador y calcular las métricas.

Métricas:

Clicks por consulta:

Esta métrica se calcula como la media de clicks que se han realizado en las consultas.

Tasa de abandono:

Porcentaje de consultas en las que no se ha hecho ningún click.

Mean RR:

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{rank_i}$$

Media de 1 entre el ranking de todos los clicks.

Max RR:

Max RR es igual que la métrica anterior, pero utilizando solamente el ranking del click más alto.

## Resultados:

	Clicks por consulta	Tasa de abandono	Max RR	Mean RR
Ask	1.55	0.375	0.816	0.522
Bing	0.041	0.968	0.968	0.875
DuckDuckGo	1.478	0.229	0.741	0.524
Google	1.663	0.261	0.751	0.487

En este caso podemos observar que Google obtiene los mejores resultados en general, aunque DuckDuckGo lo supera en tasa de abandono y Max RR. Bing obtiene unos resultados pésimos porque en la base de datos hay muy pocos patrones con el buscador Bing y además en muchos de ellos no se ha hecho click.