

BLOQUE 2. MODELOS Y EVALUACIÓN DE LA RECUPERACIÓN

- **MC-F-006.** Tema 6. **Modelos de Recuperación de Información** ([PDF](#)).
- **MC-F-007.** Tema 7. **Evaluación de la recuperación de información** ([PDF](#)).

1. Indique cuáles de las siguientes son características de la indización de documentos y consultas:

- a) Si utiliza todo el texto se denomina a *documento completo*
- b) Frecuentemente se eliminan palabras vacías
- c) Se da más importancia a verbos y sustantivos que a sustantivos
- d) En ningún caso se debe normalizar número y género para responder a la consulta

2. Indique cuáles de las siguientes afirmaciones sobre índices inversos son correctas:

- a) Registran los documentos enlazando para cada uno los términos que incluyen
- b) Se denominan inversos porque los términos se ordenan de mayor a menor
- c) El registro de cada término nunca incluye el número de documentos en los que aparece
- d) El registro del documento puede incluir la ubicación del término (párrafo, línea, número de palabra)

3. Indique cuáles de las siguientes afirmaciones sobre modelos clásicos de recuperación es falsa:

- a) utilizan una función que calcula y valora la relevancia de los documentos de una colección para una consulta dada
- b) Se consideran modelos clásicos: Booleano, Vectorial, Exponencial y Probabilístico
- c) Los documentos son descritos por una serie de términos $K=\{k_1...k_n\}$
- d) Los términos tienen asociados pesos que pueden variar en función del documento (W_{ij} = peso del término i en el documento j)

4. Seleccione cuáles de las siguientes son características del modelo de recuperación booleano:

- a) Asigna pesos en función del número de sinónimos de cada término en el documento
- b) Los pesos toman valores enteros aunque suelen normalizarse $[0,1]$
- c) Cuando los pesos son normalizados no existe relevancia parcial
- d) Admite consultas en lenguaje booleano: and, or, not incluyendo paréntesis

5. Indique cuáles de las siguientes afirmaciones sobre la frecuencia de términos es falsa:

- a) Suele utilizarse en el cálculo del peso del término
- b) Mide la relevancia de un documento para un término en base a su representatividad
- c) Si un término aparece mucho en el documento su peso en el documento es alto
- d) Se suele dividir entre el número de documentos para normalizar los valores entre 0 y 1

6. Indique cuáles de las siguientes afirmaciones sobre la frecuencia inversa de documentos es falsa:

- a) Es la frecuencia de un término en la colección calculada a partir del número de documentos en los que aparece el término y el número de documentos de la colección
- b) Es una medida del poder de discriminación de un término para recuperar un documento específico
- c) Si un término aparece en muchos documentos su IDF será alto
- d) Al medir la capacidad de discriminación del término el IDF prima la precisión: a mayor IDF mayor precisión

7. Seleccione cuáles de las siguientes NO ES UNA CARACTERÍSTICA del modelo de recuperación vectorial:

- a) Cada documento y consulta se representa con un vector con tantas dimensiones como elementos tenga el diccionario de términos
- b) La similitud entre una consulta y un documento se mide como la distancia entre ambos vectores
- c) El modelo vectorial ordena los documentos por grado de similitud con la consulta
- d) La función más utilizada para calcular la similitud es el producto escalar entre vectores

8. Indique cuáles de las siguientes afirmaciones sobre la evaluación mediante colecciones estándar son correctas:

- a) Son colecciones de documentos relevantes para una necesidad de información
- b) El sistema de recuperación de información debe recuperar un conjunto de documentos mediante consultas definidas en función de la necesidad de información
- c) Para medir la efectividad del sistema de recuperación debe compararse el conjunto de documentos recuperados con el conjunto de documentos de la colección, cuanto más se aproximen más eficiente será el sistema
- d) Requiere un conjunto controlado de documentos para su aplicación en Internet, dado que presentan una alta exhaustividad en la recuperación

9. Seleccione cuáles de las siguientes afirmaciones sobre medidas de eficacia son correctas:

- a) Precisión: define qué porcentaje de documentos recuperados son relevantes (correcto)
- b) Ruido: define qué porcentaje de documentos irrelevantes son recuperados
- c) Silencio: define qué porcentaje de documentos no relevantes recuperados son recuperados
- d) Exhaustividad: define qué porcentaje de documentos relevantes no han sido recuperados

10. Seleccione cuáles de las siguientes afirmaciones sobre medidas de eficacia es falsa:

- a) Las curvas de Precisión-Exhaustividad representan la precisión para cada punto de corte en el que se recupere un % más de documentos relevantes
- b) La Precisión Promedio (Average Precision) calcula la media de precisiones para cada documento relevante recuperado
- c) La Ganancia Acumulada Descontada (DCG) a diferencia de la Ganancia Acumulada (CG), penaliza la necesidad de recorrer más resultados para obtener información relevante
- d) El Rango Recíproco Esperado (Expected Reciprocal Rank) se basa en la probabilidad denominada persistencia, proporcionando un ratio estimado de utilidad (Expected Utility Rate)