

**BLOQUE 3. TÉCNICAS AVANZADAS DE RECUPERACIÓN**

- **MC-F-008. Tema 8. Técnicas de Procesamiento del Lenguaje Natural** ([PDF](#)).
- **MC-F-009. Tema 9. Extracción de información** ([PDF](#)).
- **MC-F-010. Tema 10. Realimentación de consultas** ([PDF](#)).

1. Indique cuáles de las siguientes afirmaciones sobre indización de documentos son correctas:

- a) La indización manual se realiza mediante asignación intelectual de conceptos, la automática por extracción de palabras**
- b) La indización automática suele ser menos exhaustiva (menos descriptores) pero también da menos ruido
- c) La indización manual: suele proporcionar más descriptores pero menos precisos
- d) La indización manual para ser eficaz debe de ser hecha por los usuarios genéricos y utilizando solamente su sentido común

2. Seleccione cual no es una técnica utilizada en PLN para preparar el documento antes de su indización:

- a) Tokenización
- b) Lematización
- c) Part-of-Speech (POS) tagging
- d) Remasterización**

3. Seleccione cuáles de las siguientes afirmaciones sobre Lematización y Stemming son correctas:

- a) La lematización aplica heurísticas para eliminar flexiones y afijos
- b) El stemming utiliza análisis morfológico para determinar el tipo de palabra y aplica reglas en consecuencia para obtener la base de la palabra
- c) La lematización consume menos tiempo pero es menos preciso que el stemming
- d) En ambos casos se aumenta la exhaustividad a cambio de reducir la precisión en la recuperación**

4. Indique cuáles de las siguientes afirmaciones sobre Extracción de información son correctas:

- a) A diferencia de la Recuperación de información, parte de dicciones sobre el que se formulan consultas
- b) El resultado de la Extracción de información será un conjunto de documentos como resultado de la consulta
- c) Permite obtener entidades como personas, sitios u organismos, a partir de los documentos**
- d) Garantiza que los sucesos, entidades y relaciones obtenidas son resultados correctos para la consulta realizada

5. Indique cuáles de las siguientes son técnicas para el reconocimiento de entidades:

- a) Patrones de captura
- b) Bootstrapping
- c) Landscaping**
- c) Aprendizaje supervisado

6. Indique cuáles de las siguientes afirmaciones sobre Web Scraping es **falsa**:

- a) Permite la extracción de datos de sitios web
- b) Suele implicar un acceso a páginas web mediante navegadores, selección y copia de los datos de la página, almacenamiento y procesamiento
- c) Utiliza el protocolo HTTP o navegadores web
- d) Se realiza manualmente para garantizar la calidad de los datos extraídos**

7. Seleccione cuáles de las siguientes afirmaciones relativas a la Realimentación de consultas **es falsa**:
- a) Se asemejan al “more like this” de motores de recuperación como Google
  - b) Al obtener los resultados de la consulta el usuario indica cuáles son relevantes y cuáles no
  - c) El sistema recalcula la consulta en función de la información proporcionada por el usuario
  - d) El cálculo de la nueva consulta se basa en maximizar la distancia con los documentos no relevantes del corpus**
8. Seleccione cuáles de las siguientes afirmaciones relativas a la Realimentación de consultas son correctas:
- a) Reduce problemas de polisemia al incluir términos relacionados**
  - b) Mejora la precisión pero empeora la exhaustividad
  - c) Reduce el tiempo de procesamiento gracias a la información que facilita el usuario
  - d) No es un método de recuperación válido si no se conoce el vocabulario específico del dominio que utilizan todos los documentos
9. Indique la afirmación que no es falsa en relación a la Reformulación de consultas:
- a) Es resultado de la Realimentación de consultas**
  - b) Permite corregir consultas incorrectas o con faltas ortográficas
  - c) Utiliza información histórica como “preguntas más frecuentes”
  - d) Puede proporcionar un vocabulario controlado para formular consultas más precisas
10. Seleccione cuáles de las siguientes afirmaciones relativas a la Expansión de consultas son correctas:
- a) Como la Realimentación de consultas, requiere intervención del usuario para mejorar la consulta
  - b) Al igual que la Reformulación de consultas, tiene en cuenta los resultados para optimizar la consulta
  - c) Hace uso de recursos léxicos para obtener términos relacionados con los incluidos en la consulta**
  - d) Incluir nuevos términos relacionados mejor la exhaustividad y la precisión de los resultados