RECUPERACIÓN DE LA INFORAMACIÓN

INVESTIGACIÓN

Teoría de la información

Marcos Damián Pool Canul  
Universidad del Caribe, México  
31 de agosto del 2023

A black sign with white text

Description automatically generated

Table of Contents

[Recuperación de información 2](#_Toc144382134)

[Definición y contexto 2](#_Toc144382135)

[Tipos básicos de sistemas de RI 3](#_Toc144382136)

[Sistema 3](#_Toc144382137)

[Procesos Intelectuales 3](#_Toc144382138)

[Procesos Automáticos 3](#_Toc144382139)

[Explicación y ejemplos 3](#_Toc144382140)

[Métodos de recuperación de información 4](#_Toc144382141)

[1. Recuperación basada en texto 4](#_Toc144382142)

[2. Recuperación basada en relevancia 4](#_Toc144382143)

[3. Recuperación de información en la web 4](#_Toc144382144)

[4. Recuperación de información multimedia 4](#_Toc144382145)

[5. Recuperación de información en redes sociales 4](#_Toc144382146)

[Referencias bibliográficas 5](#_Toc144382147)

# Recuperación de información

## Definición y contexto

Recuperar significa volver a tener. Recuperar información significa volver a tener una información que alguna vez, hace unos minutos o hace unos años, ha sido producida por alguien, bien por nosotros mismos o bien por terceras personas.

La *Recuperación de Información* (RI, a partir de ahora) es la disciplina que estudia la representación, la organización y el acceso eficiente a la información que se encuentra registrada en documentos.

De las operaciones propias de la RI, sin duda la más característica consiste en la *selección* de documentos, bien a partir de las características de su *contenido*, (los temas tratados), bien a partir de características de su *contexto* (p.e. la fecha de publicación,) bien a partir de alguna combinación de ambas cosas (p.e: "documentos sobre desarrollo humano publicados por UNESCO entre 2003 y 2005").

Ahora bien, para que la RI tenga sentido se presupone un entorno en el cual no es trivial, precisamente, el hecho de acceder a los documentos por su contenido. Este contexto lo genera, típicamente, cualquier fondo documental a partir del momento que contenga unos centenares o unos miles de documentos. Empresas pequeñas, medianas o grandes, con ejecutivos, abogados, químicos o ingenieros que necesitan encontrar una información en fondos internos o externos es un ejemplo. Universitarios e investigadores que necesitan consultar bases de datos bibliográficas para asegurarse de que no reinventan la rueda es otro. Finalmente, la Web, que en realidad es un enorme sistema de información documental con varios miles de millones de documentos es el ejemplo extremo de contexto característico de RI.

# Tipos básicos de sistemas de RI

|  |  |  |  |
| --- | --- | --- | --- |
| Sistema | Procesos Intelectuales | Procesos Automáticos | Explicación y ejemplos |
| **Tipo 1** | Intensivos | Semi intensivos | Bases de datos referenciales con uso de herramientas complejas de indización intelectual y sin indización de texto completo (Ej: *LISA*, *ERIC*) |
| **Tipo 2** | Intensivos | Intensivos | Bases de datos con indización de texto completo y con uso de herramientas complejas de indización intelectual (Ej.: *Sosig*) |
| **Tipo 3** | Semi-intensivos | Semi-intensivos | Bases de datos referenciales con uso de herramientas simples de indización (p.e. listas de descriptores) Como ejemplo, se puede decir que es un caso típico de muchas bases de datos de uso corporativo |
| **Tipo 4** | Ninguno | Intensivos | Motores de búsqueda (*Google*), programas de indización de texto completo (*ZyLab*), bases de datos con indización exclusivamente automática (*FindArticles*) |

# Métodos de recuperación de información

1. Recuperación basada en texto:
   * **Búsqueda Booleana**: Este método utiliza operadores booleanos como AND, OR y NOT para combinar palabras clave en consultas y recuperar documentos que cumplan con esos criterios.
   * **Recuperación vectorial**: Se basa en la representación de documentos y consultas como vectores en un espacio multidimensional, donde se calcula la similitud entre ellos.
2. Recuperación basada en relevancia:
   * **Modelos probabilísticos**: Estos modelos calculan la probabilidad de que un documento sea relevante para una consulta dada. Ejemplos incluyen el modelo de Okapi BM25 y el modelo de lengua.
   * **Aprendizaje automático**: Se utilizan algoritmos de aprendizaje automático para clasificar documentos como relevantes o no relevantes en función de ejemplos previamente etiquetados.
3. Recuperación de información en la web:
   * **Motor de búsqueda web**: Los motores de búsqueda utilizan algoritmos sofisticados para indexar y recuperar páginas web relevantes en función de las consultas de los usuarios.
   * **Minería de datos en la web**: Se emplean técnicas de minería de datos para extraer información útil de la web, como opiniones de usuarios, tendencias o patrones.
4. Recuperación de información multimedia:
   * **Recuperación de imágenes y videos**: Se utilizan características visuales y metadatos para buscar y recuperar contenido multimedia relevante.
5. Recuperación de información en redes sociales:
   * **Análisis de redes sociales**: Se centra en la extracción de información relevante de plataformas de redes sociales como Twitter o Facebook, incluyendo la detección de tendencias y la evaluación de opiniones.

# Referencias bibliográficas

MARTÍNEZ MÉNDEZ, F. J. (2004). *Recuperación de información: modelos, sistemas y evaluación.* JMC Kiosko.

ABADAL, E., & CODINA, L. (2005). *Recuperación de información. Bases de Datos Documentales: Características, funciones y método*, 29-92.

LANCASTER, F. W. (2002). *El control del vocabulario en la recuperación de información* (Vol. 12). Universidad de València.

MÉNDEZ RODRÍGUEZ, E. M. (2001). *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales.*