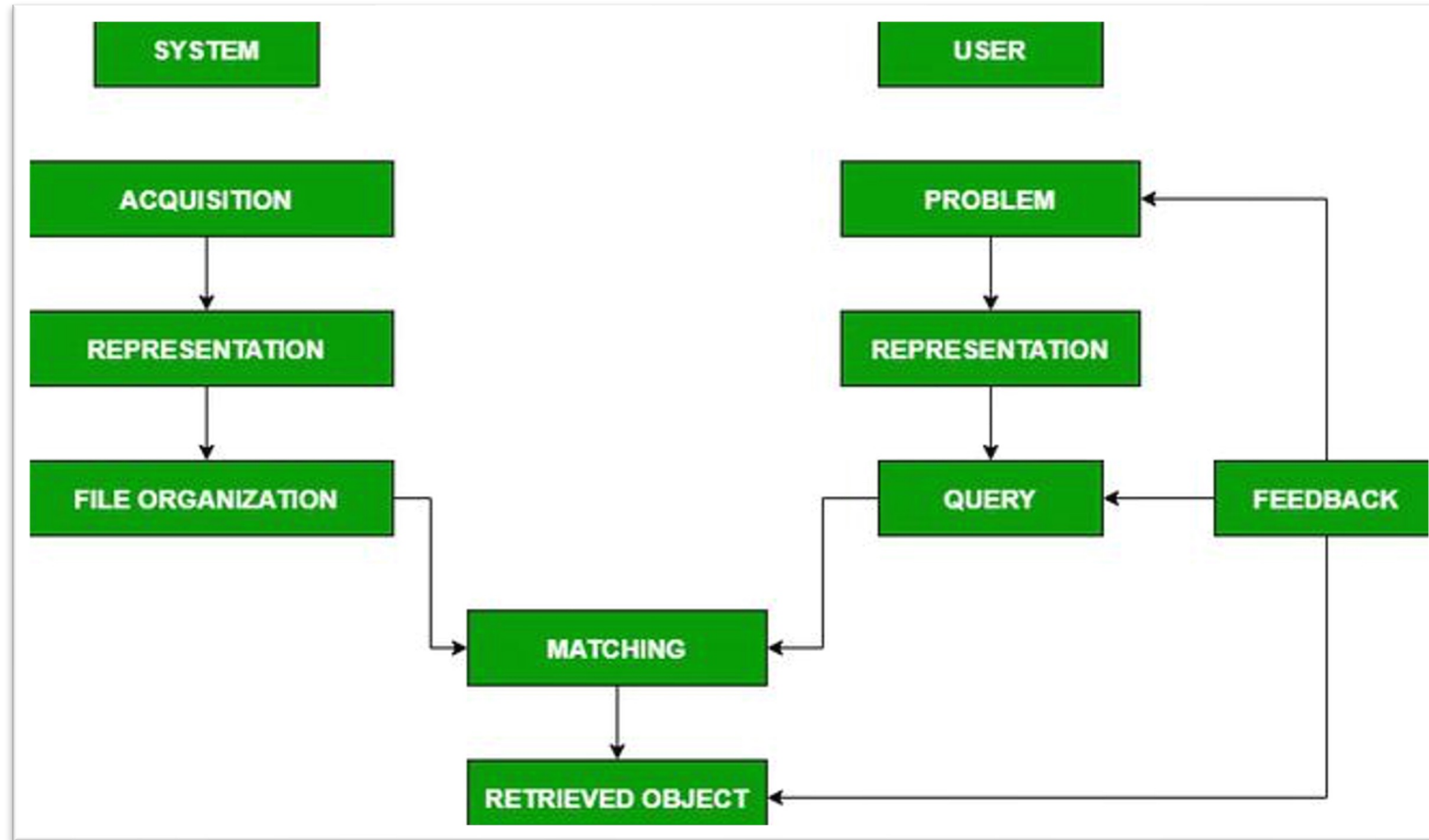


Definición

La Búsqueda y Recuperación de Información es la ciencia de la búsqueda de información en documentos electrónicos y cualquier tipo de colección documental digital, encargada de la búsqueda dentro de estos mismos, búsqueda de metadatos que describan documentos, o también la búsqueda en bases de datos relacionales, ya sea a través de internet, una intranet, y como objetivo realiza la recuperación en textos, imágenes, sonido o datos de otras características, de manera pertinente y relevante.



Componentes de recuperación de información



Adquisición:

- en este paso, se lleva a cabo la selección de documentos y otros objetos de varios recursos web que consisten en documentos basados en texto. Los rastreadores web recopilan los datos requeridos y los almacenan en la base de datos.

Representación:

- Consiste en una indexación que contiene términos de texto libre, vocabulario controlado, técnicas manuales y automáticas. ejemplo: Resumen contiene resumen y descripción bibliográfica que contiene autor, título, fuentes, datos y metadatos.

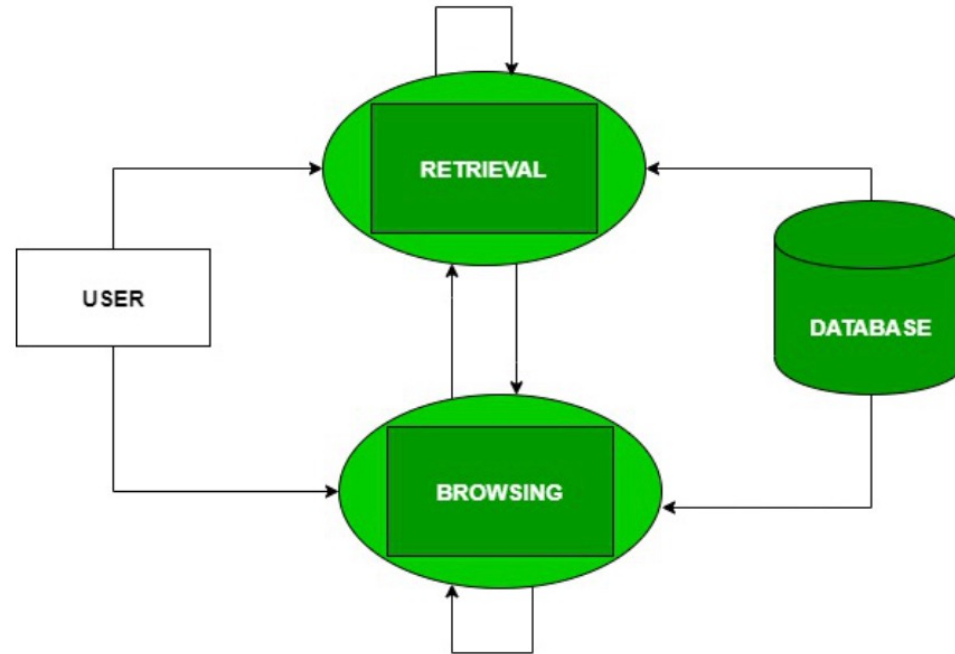
Organización de archivos:

- Hay dos tipos de métodos de organización de archivos. es decir, Secuencial: Contiene documentos por datos de documento. Invertida: Contiene término por término, lista de registros bajo cada término. Combinación de ambos.

Consulta:

- un proceso de IR comienza cuando un usuario ingresa una consulta en el sistema. Las consultas son declaraciones formales de necesidades de información, por ejemplo, cadenas de búsqueda en motores de búsqueda web. En la recuperación de información, una consulta no identifica de manera única un solo objeto en la colección. En su lugar, varios objetos pueden coincidir con la consulta, quizás con diferentes grados de relevancia.

Interacción del usuario IR



La tarea del usuario: se supone que la información primero debe traducirse en una consulta por parte del usuario. En el sistema de recuperación de información, hay un conjunto de palabras que transmiten la **semántica de la información que se requiere**, mientras que, en un sistema de recuperación de datos, se utiliza una expresión de consulta para transmitir las restricciones que satisfacen los objetos. **Ejemplo: Un usuario quiere buscar algo pero termina buscando con otra cosa. Esto significa que el usuario está navegando y no buscando.**

Dato:

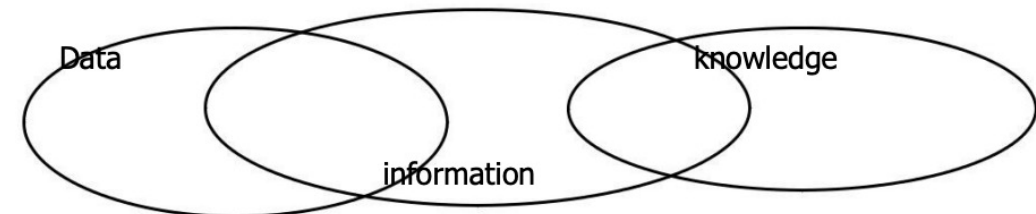
- Cadena de símbolos asociados con objetos, personas y eventos
- Valores de un atributo
 - Los datos no necesitan tener significado para todos
 - Los datos deben interpretarse

Información

- El significado de los datos interpretados por una persona o un sistema.
- Datos que modifican el estado de una persona o sistema que los percibe.
- Datos que reducen la incertidumbre. si los datos no contienen incertidumbre, no hay información con los datos.
- Ejemplos: Nieva en invierno. No nieva este invierno.

Conocimiento

- Información estructurada: a través de la estructuración, la información se vuelve comprensible
- Información procesada a través del procesamiento, la información se vuelve significativa y útil
- información compartida y acordada dentro de una comunidad



$$\text{Precision} = \frac{\text{Number_Retrieved_Relevant}}{\text{Number_Total_Retrieved}}$$

$$\text{Recall} = \frac{\text{Number_Retrieved_Relevant}}{\text{Number_Possible_Relevant}}$$

Precision: la capacidad de recuperar documentos mejor clasificados que son en su mayoría relevantes.

Recall: La capacidad de la búsqueda para encontrar todos los elementos relevantes en el corpus.

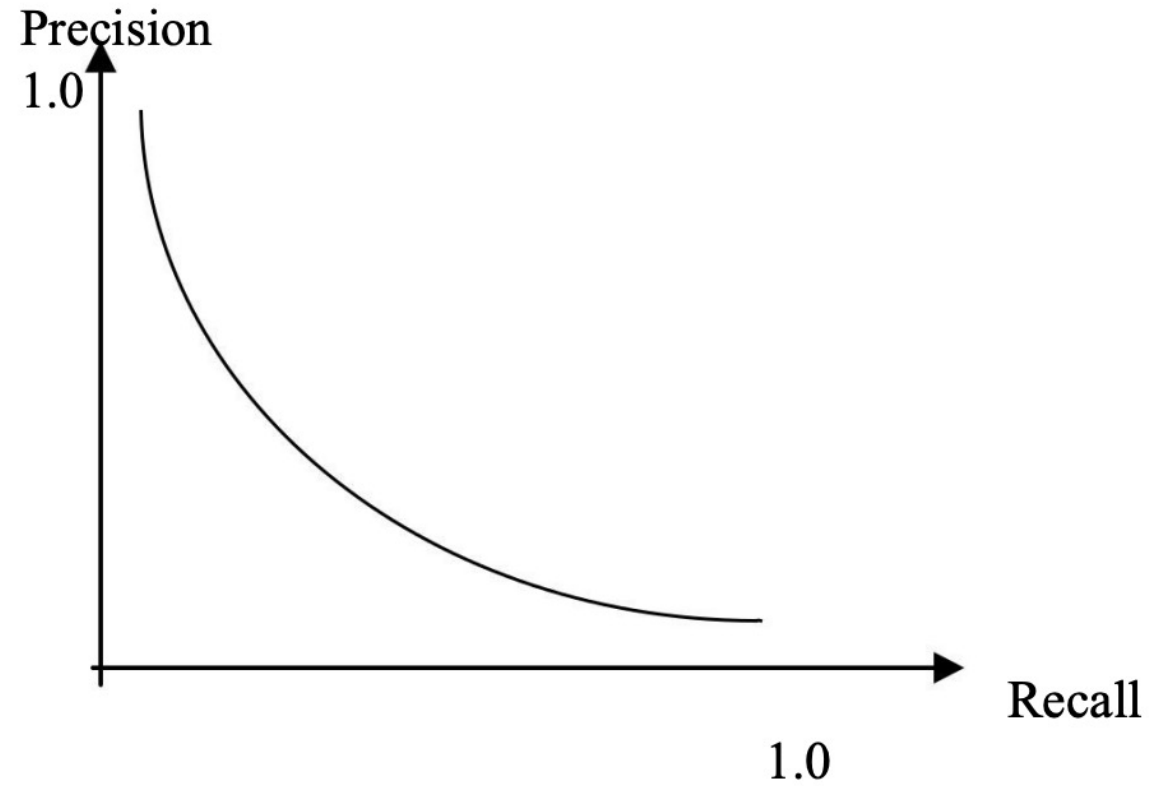
Donde **Number_Possible_Relevant** son el no. de elementos relevantes en la base de datos.

Number_Total_Retrieved: es el número total de elementos recuperados de la consulta.

Number_Retrieved_Relevant: es el no. de elementos recuperados que son relevantes para la necesidad de búsqueda del usuario.

– La precisión mide un aspecto de la sobrecarga de información recuperada para un usuario asociado con una búsqueda en particular.

– Si una búsqueda tiene un 85 %, entonces el 15 % del esfuerzo del usuario se dedica a revisar elementos no relevantes.



- Los sistemas no pueden compararse en un punto de precisión/recuperación
- Precisión media (sobre 11 puntos de recuerdo: 0,0, 0,1, ..., 1,0)

Information Retrieval - Web

- Sin colección de documentos estable (spider, crawler)
- Documento no válido, duplicado, etc.
- Gran cantidad de documentos (colección parcial)
- Documentos multimedia
- Gran variación de la calidad del documento
- Problema multilingüe



	Data retrieval	Information retrieval
Content Data object Matching Items wanted Query language Query specification Model	Data Table Exact match Matching SQL(artificial) Complete Deterministic Highly structure	Information Document Partial match, best match Relevant Natural Incomplete Probabilistic Less structure