

# Operaciones sobre Consultas



UCR – ECCI

CI-2414 Recuperación de Información

Prof. Kryscia Daviana Ramírez Benavides



# Tipos de Consulta

- Texto:
  - Palabra única.
  - En contexto:
    - Frase.
    - Proximidad.
  - Booleana.
  - Lenguaje natural.



## Tipos de Consulta (cont.)

- Correspondencia de patrones:
  - Lenguaje natural.
  - Palabra.
  - Prefijos.
  - Sufijos.
  - *Substring*.
  - Rango.
  - Permitiendo errores.



## Tipos de Consulta (cont.)

- Correspondencia de patrones (cont.):
  - Expresiones regulares:
    - Unión.
    - Concatenación.
    - Repetición.
  - Patrones extendidos:
    - Clase de caracteres.
    - Expresiones condicionales.
    - Combinaciones.

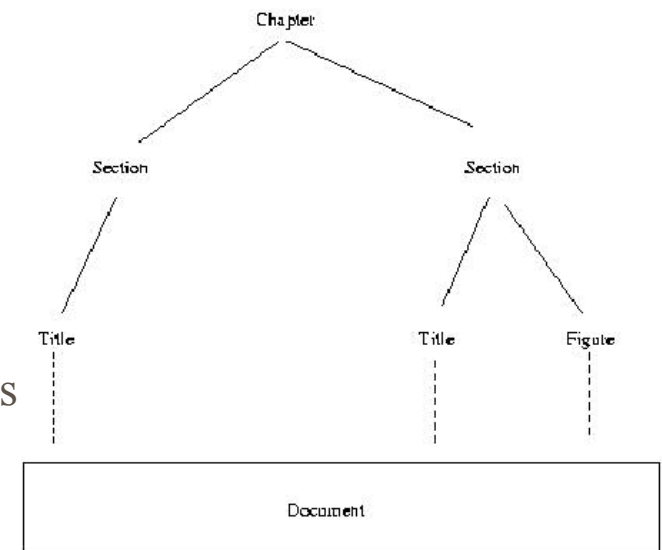
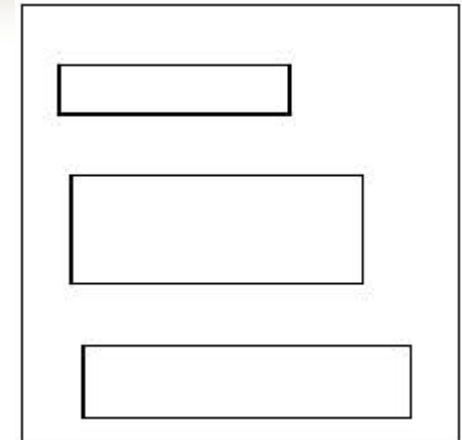
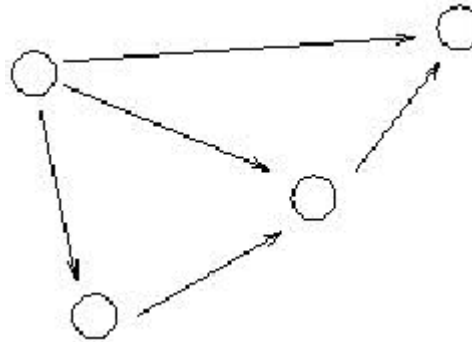
## Tipos de Consulta (cont.)

### ■ Estructural:

- Estructura fija.
- Hipertexto.
- Estructura jerárquica.

### ■ Información espacial:

- Ventana (qué).
- Objetos (dónde).
- Consulta combinada o estructural:
  - Objetos.
  - Relaciones.
  - Problema de satisfacción de restricciones







## Operaciones sobre Consultas

- La idea es mejorar la formulación de la consulta inicial a través de la expansión de la consulta y re-peso de palabras claves.
- Estos enfoques se agrupan en tres categorías:
  - Retroalimentación de parte de usuario.
  - Derivación del grupo de documentos inicialmente derivados (análisis local).
  - Derivados de la información global de los documentos en la colección (análisis global).



## Retroalimentación del Usuario

- Al **usuario** se le presenta una lista de **documentos recuperados** y después de examinarlos, **marca los que son relevantes**.
- A los términos o palabras claves de los documentos seleccionados como relevantes se les da más importancia en la reformulación de la consulta.



## Retroalimentación del Usuario (cont.)

- En el caso del modelo vectorial, la retroalimentación considera que los vectores de peso de los documentos identificados como relevantes son similares entre ellos.
- Más aún, documentos no relevantes tienen vectores que son distintos a los de los relevantes.
- Entonces, la consulta es reformulada de manera que sea más cercana al espacio de vectores de pesos de los documentos relevantes.
- Este proceso es llamado retroalimentación por relevancia.





## Análisis Local

- Realiza la expansión con los **documentos recuperados por una consulta** donde se realizan *clusters* basados en la **correlación de términos**.
- Utiliza esta información para expandir las consultas añadiendo a las mismas los términos correlacionados con los de la consulta.
- Este enfoque trata de obtener un conjunto más grande de objetos relevantes automáticamente.
- Esto usualmente consiste en identificar sinónimos, variaciones terminales, o términos que están cercanos a los términos de la consulta en el texto.



## Análisis Local (cont.)

- En el análisis local, los documentos recuperados para una consulta son examinados para determinar términos de expansión, lo cual es hecho sin el apoyo del usuario.
- Existen dos claros enfoques:
  - Agrupamiento local.
  - Análisis de contexto local.



## Análisis Global

- Realiza la expansión basado en la **construcción de tesauros utilizando la colección completa de documentos** para añadir a la consulta los términos más cercanos a los de la consulta en el tesauro.
- Se puede hacer basadas en la similaridad de conceptos o en las estadísticas de aparición de los términos que representan conceptos.
- La idea de este tipo de expansión de la consulta es considerar todo el conjunto de documentos en la colección.
- Existen dos claros enfoques:
  - Expansión basada en un tesauro de similitud.
  - Expansión basada en un tesauro estadístico.



## Referencias Bibliográficas

- La información fue tomada de:
  - Libro de texto del curso.