



U N I V E R S I D A D  
**COMPLUTENSE**  
M A D R I D

# **Evaluación de recuperación y expansión de consultas**

Sistemas de Gestión de Datos y de la Información

Enrique Martín - [emartinm@ucm.es](mailto:emartinm@ucm.es)

Máster en Ingeniería Informática

Fac. Informática

# **Evaluación de recuperación**

# Evaluación

- Las funciones de *ranking* producen resultados ordenados para las consultas.
- Si tenemos disponibles varias, ¿cómo sabemos cuál funciona mejor? Existen dos alternativas:
  - Calcular métricas a partir de **colecciones de referencia**.
  - Usar evaluaciones basadas en el **usuario** (resultados visitados).

# Colecciones de referencia

- Están compuestas por:
  - Un conjunto de documentos  $d_j$
  - Un conjunto de consultas  $q_i$
  - Para cada pareja  $(d_j, q_i)$ , un valor de **relevancia**: 0 si no es relevante y 1 si es relevante
- Las colecciones de referencia son producidas por **especialistas**, por tanto no incluyen grandes cantidades de documentos.

# Métricas

- Dada una consulta  $q$  de la colección de referencia, conocemos el conjunto de documentos relevantes **R**.
- Probando esa misma consulta en el sistema a evaluar, obtendremos un conjunto de documentos **A**.
- Considerando estos dos conjuntos obtenemos 2 métrica fundamentales: precisión y exhaustividad.

# Precisión y exhaustividad

- La **precisión** es la proporción de documentos realmente relevantes de todos los producidos por el sistema.
- La **exhaustividad** (*recall*) es la proporción de documentos relevantes que el sistema ha producido.

$$\text{Precisión} = p = \frac{|R \cap A|}{|A|}$$

$$\text{Recall} = r = \frac{|R \cap A|}{|R|}$$

# Precisión y exhaustividad

- La precisión y la exhaustividad son números que indican la calidad de los resultados *como si todos estos se hubiesen examinado*.
- No tienen en cuenta el **orden** en el que aparecen los resultados.
  - A un usuario no le importan estos valores globales sino cómo varían según avanza en el listado de resultados.
- Para ello se suele mostrar una gráfica precisión/exhaustividad.

# Precisión y exhaustividad

- Imaginemos una consulta  $q1$ , cuyo conjunto de documentos relevantes es  $R1 = \{d3, d5, d9, d10, d11, d23, d44, d50, d51, d53\}$ ,  $|R1| = 10$
- Al lanzar la consulta en el sistema, nos devuelve los siguientes resultados:
  - 1: **d9**, 2: d8, 3: **d23**, 4: d67, 5: d68, 6: **d3**, 7: **d44**, 8: d98

$$p = \frac{4}{8} = 0.5$$

$$r = \frac{4}{10} = 0.4$$



# Precisión y exhaustividad

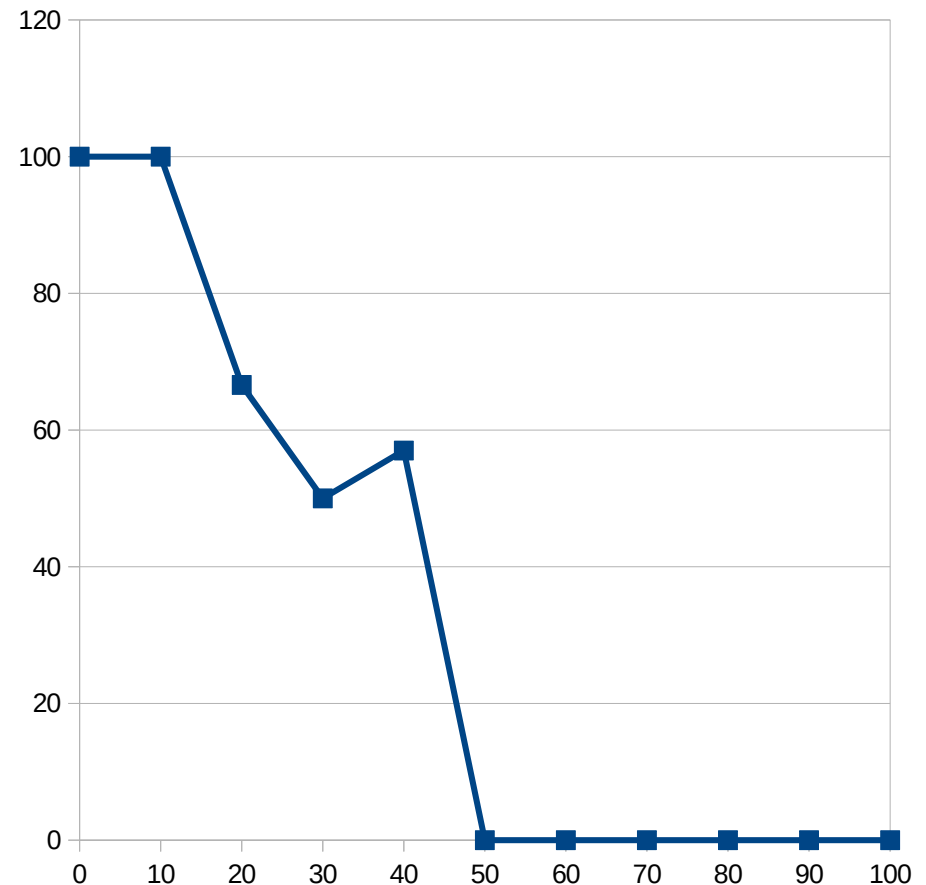
- Veamos cómo evoluciona la precisión según aumenta la exhaustividad:
  - **d9** es el 1<sup>er</sup> resultado relevante. Es el 10% del conjunto relevante, por lo que con exhaustividad 10% tenemos una precisión de 100% (1 de 1).
  - **d23** es el 3<sup>er</sup> resultado relevante, y hasta ese punto hemos obtenido el 20% del conjunto relevante. Por tanto con exhaustividad 20% tenemos precisión 66,6% (2 de 3).
  - **d3** es el 6<sup>o</sup> resultado relevante, y hemos visto el 30% del conjunto relevante → con exhaustividad 30% tenemos precisión 50% (3 de 6).
  - **d44** es el 7<sup>o</sup> resultado relevante, habiendo visto el 40% del conjunto relevante → con exhaustividad 40% tenemos precisión 57% (4 de 7)

Resultados del sistema (relevantes en negrita)

1: **d9**, 2: d8, 3: **d23**, 4: d67, 5: d68, 6: **d3**, 7: **d44**, 8: d98

# Precisión y exhaustividad

- Estos valores se representan en una gráfica.
- A partir de 40% *recall* la precisión es 0 pues no muestra más resultados relevantes



# Precisión y exhaustividad

- En lugar de una sola consulta, se suele calcular la **precisión media** de un conjunto de consultas por cada valor de exhaustividad.
- Midiendo el **área bajo la curva** se obtiene un valor numérico de la calidad de la recuperación.
- Las gráficas precisión/exhaustividad también sirven para **comparar visualmente** varias técnicas de recuperación.

# Otras métricas

- **Precisión media en 'n' ( $P@n$ ):** precisión considerando los  $n$  primeros documentos:  $P@5$ ,  $P@10$ . Útil en buscadores web.
- **R-precisión:** precisión considerando los  $R$  primeros documentos, donde  $R$  es el número de documentos relevantes.

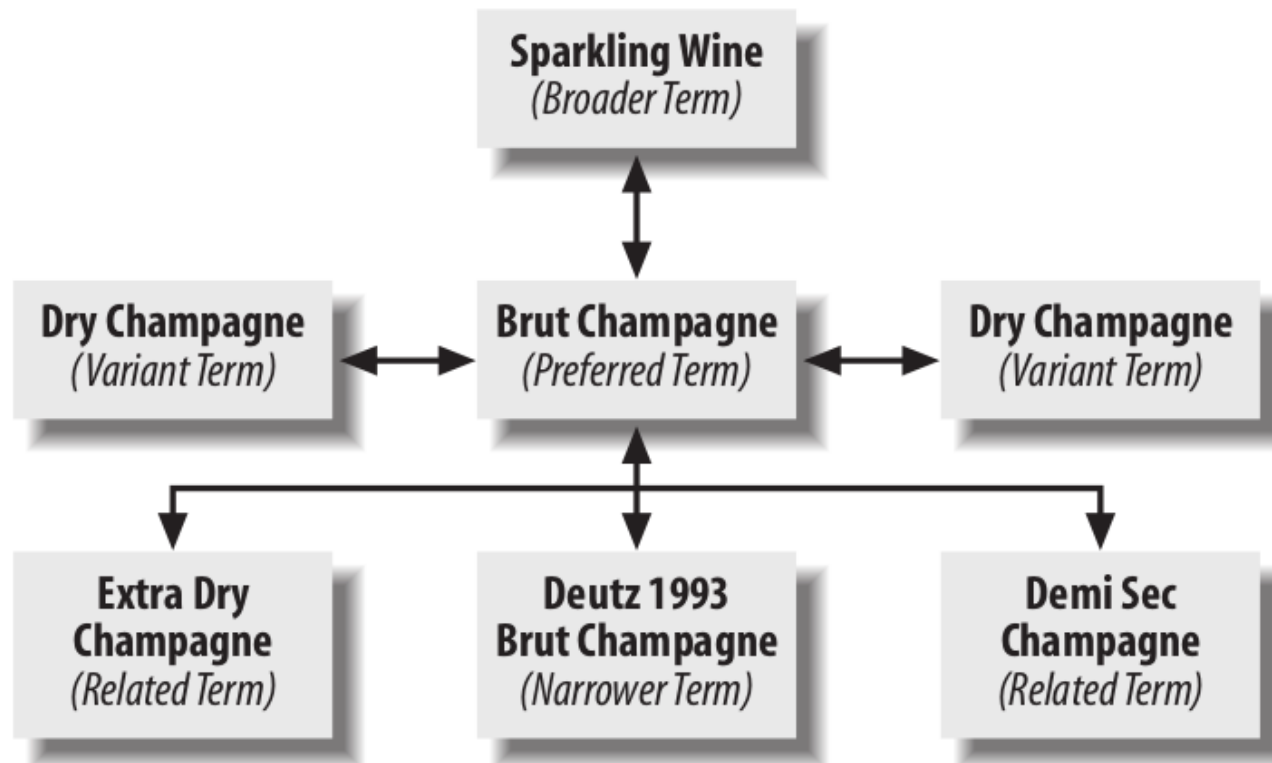
# **Expansión de consultas**

# Expansión de consultas

- Los sistemas de recuperación de información pueden **ampliar las consultas** del usuario automáticamente o proponer **alternativas** → **mejorar la relevancia** de los documentos obtenidos.
- La expansión de consultas se basa en **tesauros**, que son redes semánticas que relacionan términos.

# Expansión de consultas

- Los tesauros enlazan términos con sus **sinónimos**, conceptos **generales** y **particulares**.



# Expansión de consultas

- Usando un tesauro, podríamos sustituir la consulta “French dry champagne” por “French brut champagne”.
- También podríamos proponer consultas más generales como “French sparkling wine” o más particulares “French demi sec champagne”



# Tesauros

- La construcción de tesauros suele ser una tarea manual realizada por expertos.
- Existen para campos concretos como medicina, así que no funcionan para consultas en general.
- Existen técnicas para la generación automática de tesauros relacionando palabras que aparecen juntas (en párrafos o documentos) → su uso en recuperación de información no produce buenos resultados.

# **Bibliografía**

# Bibliografía

- **Modern Information Retrieval, the concepts and technology behind search**, second edition. *Ricardo Baeza-Yates y Berthier Ribeiro-Neto*. Pearson Education Limited (2011).