

Retroalimentación por Relevancia (RR)



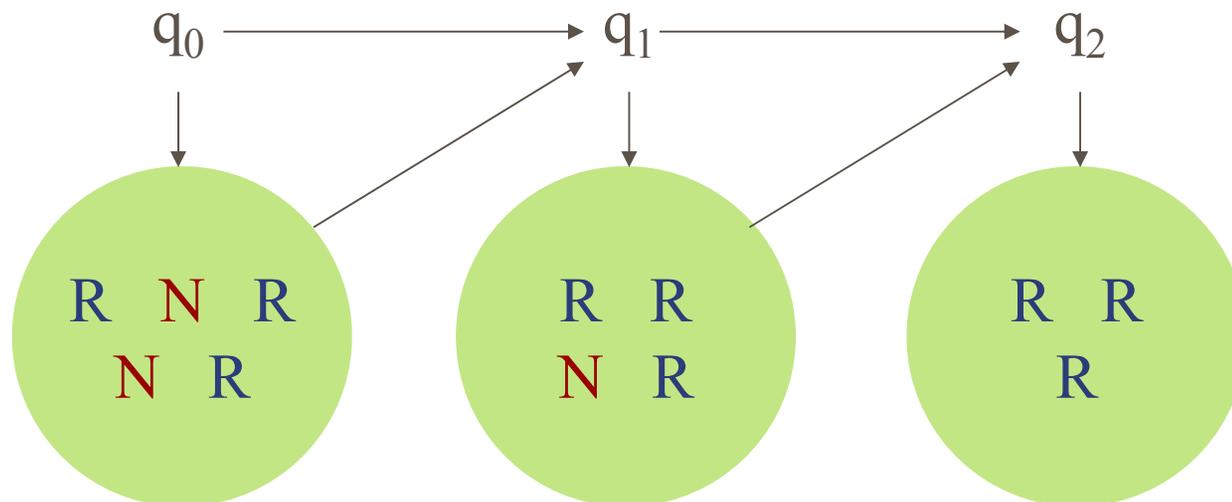
UCR – ECCI

CI-2414 Recuperación de Información

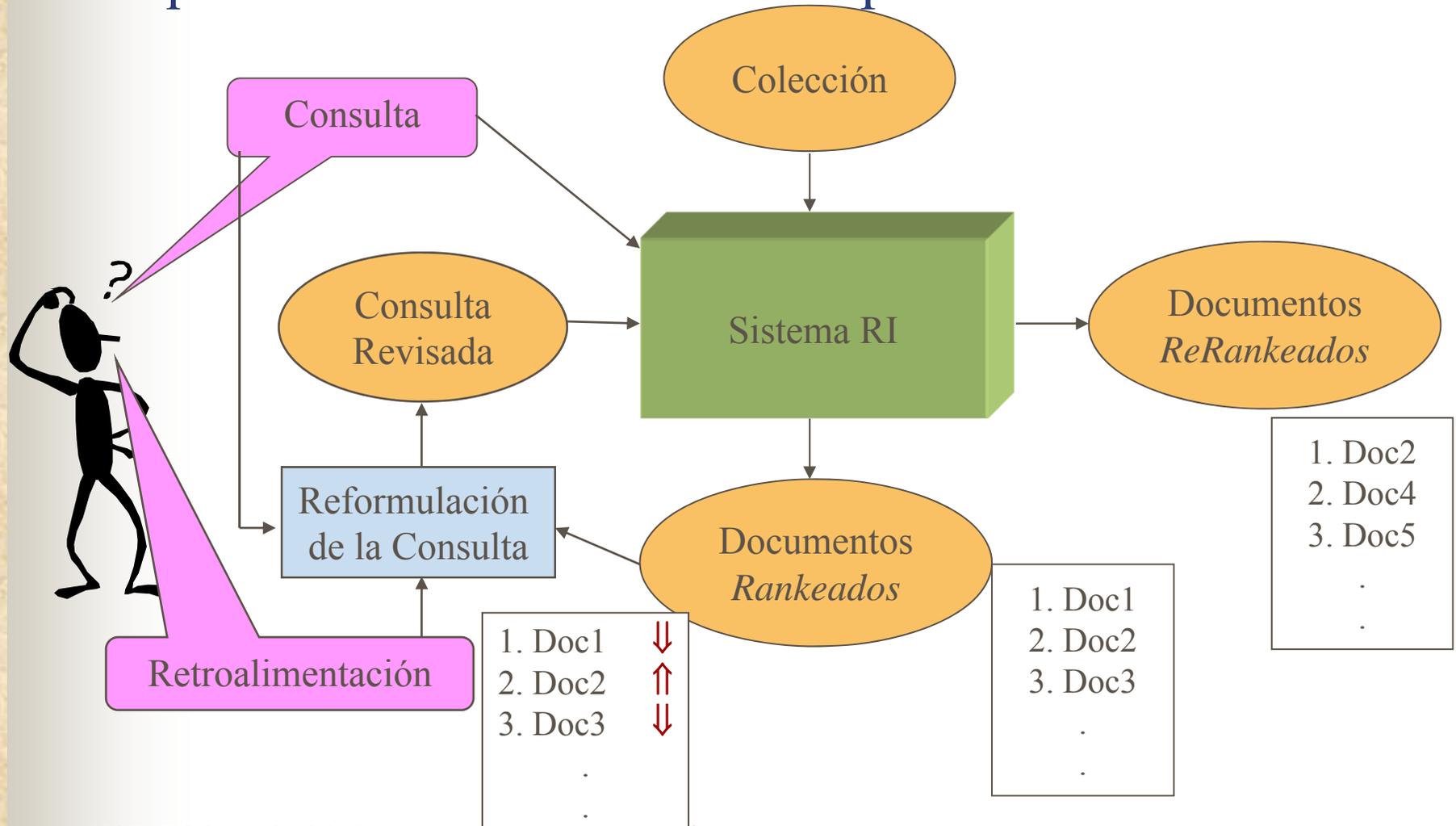
Prof. Kryscia Daviana Ramírez Benavides

Retroalimentación por Relevancia (RR)

- Difícil redactar la consulta que satisfaga la necesidad del usuario.
- RI puede ayudar a modificar automáticamente la consulta del usuario.
- RR ayuda a modificar la consulta de acuerdo a la evaluación del usuario de los documentos recuperados.



Arquitectura Retroalimentación por Relevancia

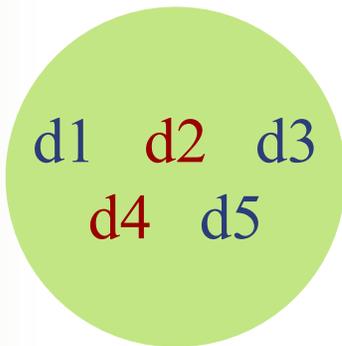




Técnicas RR

- Expansión de la consulta.
- Re-ponderar los términos de la consulta.

¿Cómo Reformular la Consulta?



$$\vec{q}_{opt} = \frac{1}{3}(\vec{d}_1 + \vec{d}_3 + \vec{d}_5) - \frac{1}{2}(\vec{d}_2 + \vec{d}_4)$$

$$\vec{q}_{opt} = \frac{1}{|C_r|} \sum_{\forall \vec{d}_j \in C_r} \vec{d}_j - \frac{1}{|C_n|} \sum_{\forall \vec{d}_j \in C_n} \vec{d}_j$$

■ Donde:

- q_{opt} : Vector consulta óptimo.
- C_r : Número de documentos de la colección relevantes para la consulta.
- C_n : Número de documentos de la colección no relevantes para la consulta.
- d_j : Vector documento.

Standard Rochio

$$\vec{q}_m = \vec{q} + \frac{1}{|D_r|} \sum_{\forall \vec{d}_j \in D_r} \vec{d}_j - \frac{1}{|D_n|} \sum_{\forall \vec{d}_j \in D_n} \vec{d}_j$$

■ Donde:

- q_m : Vector consulta con retroalimentación.
- D_r : Número de documentos relevantes devueltos para la consulta.
- D_n : Número de documentos no relevantes devueltos para la consulta.
- q : Vector consulta original (o anterior).
- d_j : Vector documento.

Ejemplo del *Standard Rochio*

d_1	d_2	d_3
perros	perros	aves
gatos	gatos	
aves		

Vector = (p, g, a)

$$\vec{q} = (0.18, 0.18, 0.09)$$

$$\vec{d}_1 = (0.18, 0.18, 0.18)$$

$$\vec{d}_2 = (0.18, 0.18, 0)$$

$$\vec{d}_3 = (0, 0, 0.18)$$

Similitud:

$$\text{sim}(d_1, q) = 0.96$$

$$\text{sim}(d_2, q) = 0.96$$

$$\text{sim}(d_3, q) = 0.33$$

Suponemos que d_1 y d_2 son relevantes.

Ejemplo del *Standard Rochio* (cont.)

$$\vec{q}_m = \vec{q} + \frac{1}{2}(\vec{d}_1 + \vec{d}_2) - \frac{1}{1}(\vec{d}_3) = (0.36, 0.36, 0)$$

Similitud:

$$\text{sim}(d_1, q) = 0.82$$

$$\text{sim}(d_2, q) = 0.97$$

$$\text{sim}(d_3, q) = 0$$

Variación del Algoritmo de *Rocchio*

$$\vec{q}_m = \alpha \vec{q} + \frac{\beta}{|D_r|} \sum_{\forall \vec{d}_j \in D_r} \vec{d}_j - \frac{\gamma}{|D_n|} \sum_{\forall \vec{d}_j \in D_n} \vec{d}_j$$

- $\beta > \gamma \Rightarrow$ Le da más peso a los documentos relevantes.
- $\gamma > \beta \Rightarrow$ Leda más peso a los documentos no relevantes.
- $\alpha \Rightarrow$ Es la importancia de la consulta anterior, es recomendable un valor ≥ 0.5 .
- $\beta + \gamma = 1$.
- Es recomendable que $\alpha = 1$, $\beta = 0.75$ y $\gamma = 0.25$.



Ventajas y Desventajas de RR

- Ventajas:
 - Sencillez.
 - Buenos resultados.
 - Exime al usuario de los detalles del proceso de reformulación de la consulta.
 - Permite dar énfasis a algunos términos.
- Desventajas:
 - Adopta criterios no óptimos.
 - La variación del algoritmo utiliza constantes con valores tomados al criterio del desarrollador.



Referencias Bibliográficas

- La información fue tomada de:
 - Libro de texto del curso.
 - Presentaciones del curso RI del estudiante Donald Vargas Zúñiga. Universidad de Costa Rica, 2003.