

Recuperació de la Informació (REIN)

Grau en Enginyeria Informàtica

Departament de Ciències de la Computació (CS)



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

**Escola Politècnica Superior d'Enginyeria
de Vilanova i la Geltrú**

- 1 4. Avaluació en Recuperació de la Informació
 - Avaluació de l'ús de la RI
 - Mesures d'avaluació de l'ús de la RI
 - Altres mesures d'efectivitat
 - Relevance Feedback

- 1 4. Avaluació en Recuperació de la Informació
 - Avaluació de l'ús de la RI
 - Mesures d'avaluació de l'ús de la RI
 - Altres mesures d'efectivitat
 - Relevance Feedback

Avaluació de l'ús de la RI, I

Què farem?

En el model booleà l'especificació no té ambigüitat:

Sabem què farem:

Recuperar i tornar a l'usuari
tots aquells documents
que **satisfacin** la consulta.

Però, això és realment el que vol l'usuari?

Normalment. . . **no**.

Avaluació de l'ús de la RI, II

Què volem optimitzar?

Notació:

- \mathcal{D} : conjunt de **tots els nostres documents** sobre els que l'usuari fa la consulta;
- \mathcal{A} : **conjunt de resposta** amb els documents que el sistema recupera com a resposta;
- \mathcal{R} : **documents rellevants**, aquells que l'usuari realment vol que estiguin en la resposta.
(Però **ningú** coneix aquest conjunt, ni l'usuari!)

Objectiu inassolible: $\mathcal{A} = \mathcal{R}$

és a dir:

$$d \in \mathcal{R} \Rightarrow d \in \mathcal{A} \text{ o també } Pr(d \in \mathcal{A} | d \in \mathcal{R}) = 1 \text{ i}$$

$$d \in \mathcal{A} \Rightarrow d \in \mathcal{R} \text{ o també } Pr(d \in \mathcal{R} | d \in \mathcal{A}) = 1$$

Les mesures de *recall* i *precision*, I

Per evaluar lo bona que es la resposta

Ens conformem amb:

- ***recall*** elevat, $\frac{|\mathcal{R} \cap \mathcal{A}|}{|\mathcal{R}|}$: de tots els documents de la resposta, quants son rellevants
retorna la intercepcio entre A i R i es fa el percentatge
 $Pr(d \in \mathcal{A} | d \in \mathcal{R})$ no massa per sota de 1,
- ***precision*** elevada, $\frac{|\mathcal{R} \cap \mathcal{A}|}{|\mathcal{A}|}$: Sobre el que ha recuperat, quants son rellevants
 $Pr(d \in \mathcal{R} | d \in \mathcal{A})$ no massa per sota de 1.

Més endavant discutirem aquest balanç.

Les mesures de *recall* i *precision*, II

Exemple: test per la tuberculosi (TB)

- 1000 persones, de les quals 50 tenen TB
- el test dóna positiu en 40 persones, de les quals 35 *sí* tenen TB

Recall

% dels que *sí* tenen TB que donen positiu al test = $35/50 = 70\%$

Precision

% dels que donen positiu al test que *sí* tenen TB = $35/40 = 87.5\%$

- ***Recall elevat***: hi haurà poques persones malaltes que no hagin estat detectades.
- ***Precision elevada***: hi haurà poques persones amb qui s'hagin equivocat fent el diagnòstic.

Les mesures de *recall* i *precision*, III

Una definició equivalent

Matriu de confusió

tp -> true positive

fn-> false negative

fp -> false positive

tn-> true negative

		Recuperat	
		rellevant	no rellevant
Realitat	rellevant	tp	fn
	no rellevant	fp	tn

- $|\mathcal{R}| = tp + fn$

- $|\mathcal{A}| = tp + fp$

- $|\mathcal{R} \cap \mathcal{A}| = tp$

- $Recall = \frac{|\mathcal{R} \cap \mathcal{A}|}{|\mathcal{R}|} = \frac{tp}{tp+fn}$

- $Precision = \frac{|\mathcal{R} \cap \mathcal{A}|}{|\mathcal{A}|} = \frac{tp}{tp+fp}$

- 1 4. Avaluació en Recuperació de la Informació
 - Avaluació de l'ús de la RI
 - Mesures d'avaluació de l'ús de la RI
 - Altres mesures d'efectivitat
 - Relevance Feedback

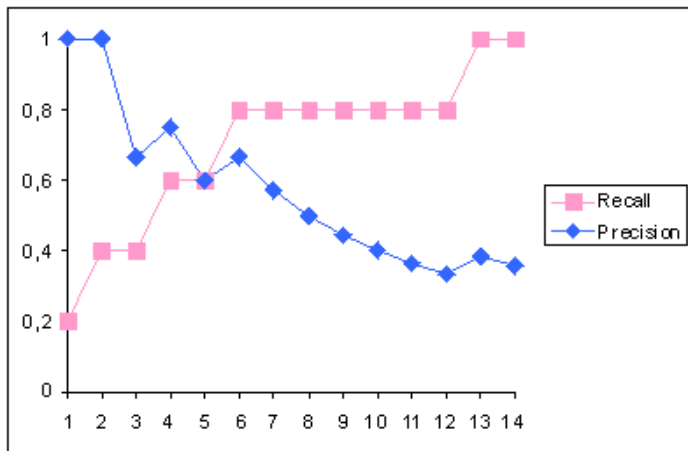
Quants documents cal retornar?

Ordenem tots els documents d'acord amb una mesura. Quants documents hauríem de mostrar?

- Els usuaris no volen llegir respostes massa llargues.
- Les respostes llargues tenen tendència a mostrar *baixa precision*.
- Les respostes curtes tenen tendència a mostrar *baix recall*.

Analitzem la *precision* i el *recall* com a funcions del nombre k de documents tornats com a resultat.

Corbes de *recall* i *precision*, I



(Font: Prof. J. J. Pajmans, Tilburg)

Corbes de *recall* i *precision*, II

Exercici

Suposeu que un sistema de recuperació de la informació torna la llista de *docid* següent: 5, 9, 4, 8, 3, 1, 7, 6, 10, 2.

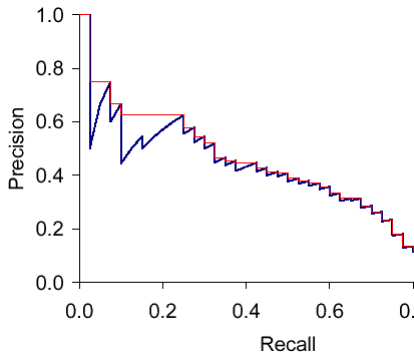
Només els identificadors senars són realment rellevants per la consulta de l'usuari i el sistema té indexats 10 documents.

Construïu les corbes de *precision* i *recall*.

Una única corba “*recall i precision*”

L'eix x pel *recall* i l'eix y per la *precision*.

(Similar a les corbes ROC dels models predictius.)



(Font: Stanford NLP group)

Sovint, es calculen 11 punts de *precision* interpolada al 0%, 10%, 20%,..., 100% de *recall*.

- 1 4. Avaluació en Recuperació de la Informació
 - Avaluació de l'ús de la RI
 - Mesures d'avaluació de l'ús de la RI
 - **Altres mesures d'efectivitat**
 - Relevance Feedback

Altres mesures d'efectivitat, I

- AUC: *Area Under the Curve* de les corbes de *recall* i *precision*, relatives al millor sistema

- F-measure:
$$\frac{2}{\frac{1}{recall} + \frac{1}{precision}} \quad \text{F-measure} = 2 * (recall * precision) / (recall + precision)$$

Mitjana harmònica. Més propera al mínim de les dues que a la seva mitjana aritmètica.

alpha dona més per al recall o a la precisió

- α -F-measure:
$$\frac{2}{\frac{\alpha}{recall} + \frac{1-\alpha}{precision}}$$

Tenir en compte els documents prèviament coneguts per l'usuari:

- **Coverage** (cobertura): valores positivament coses que l'usuari ja el coneix (rellevant i conegut per l'usuari)
 $| \text{rellevant \& conegut \& recuperat} | / | \text{rellevant \& conegut} |$
- **Novelty** (novetat): Es al revés (del que es rellevant i ha recuperat, quantes d'aquestes l'usuari no coneixia)
 $| \text{rellevant \& recuperat \& DESconegut} | / | \text{rellevant \& recuperat} |$

- 1 4. Avaluació en Recuperació de la Informació
 - Avaluació de l'ús de la RI
 - Mesures d'avaluació de l'ús de la RI
 - Altres mesures d'efectivitat
 - Relevance Feedback

Retroalimentació per rellevància (*Relevance Feedback*), I

En alguns casos, podem respondre millor a l'usuari si fem dues rondes d'interacció.

- 1 Obtenir la consulta q .
- 2 Recuperar els documents rellevants per q .
- 3 Mostrem a l'usuari els k més rellevants.
- 4 Demanar a l'usuari que els marqui com a rellevant/irrellevant.
- 5 Usar les respostes per **refinar** q .
- 6 Si es vol, tornar a 2.

Relevance Feedback, II

Com crear una nova consulta?

Model vectorial: consultes i documents són vectors.

Donada la consulta q i un conjunt de documents, **separar-los** entre rellevants R i no rellevants NR , i generar una nova consulta q' :

Relevance Feedback, II

Com crear una nova consulta?

Model vectorial: consultes i documents són vectors.

Donada la consulta q i un conjunt de documents, **separar-los** entre rellevants R i no rellevants NR , i generar una nova consulta q' :

Regla de Rocchio:

$$q' = \alpha \cdot q + \beta \cdot \frac{1}{|R|} \cdot \sum_{d \in R} d - \gamma \cdot \frac{1}{|NR|} \cdot \sum_{d \in NR} d$$

- Tots els vectors q i d han d'estar **normalitzats**.
- Els pesos α , β , γ , són escalars $\alpha > \beta > \gamma \geq 0$; sovint $\gamma = 0$.
 - α : grau de confiança en la consulta original,
 - β : pes de la informació positiva (termes que no apareixen a la consulta però que apareixen en documents rellevants),
 - γ : pes de la informació negativa.

Es pot repetir el procés

però normalment no cal. Sovint, s'obté:

- millora del *recall* en la primera ronda,
- marginal per la segona ronda,
- gairebé res, més enllà.

En la cerca a internet, la *precision* importa més que el *recall*, per tant, el cost computacional extra pot no ser productiu.

Relevance Feedback, IV

Cas particular de *query expansion*

És una forma de Query Expansion:

La nova consulta té pesos diferents de zero per paraules que no estaven a la consulta original.

Pseudorelevance feedback, I

No preguntar res a l'usuari!

No preguntar res a l'usuari!

- Assumim que el resultat inicial ja era prou bo.
- Actuem com si els k primers documents de resposta fossin tots rellevants.
- No interactuem amb l'usuari.
- Però no oblidem que la cerca serà més lenta.

Pseudorelevance feedback, I

No preguntar res a l'usuari!

- Assumim que el resultat inicial ja era prou bo.
- Actuem com si els k primers documents de resposta fossin tots rellevants.
- No interactuem amb l'usuari.
- Però no oblidem que la cerca serà més lenta.
- Parar, com a molt tard, quan els k primers documents no canviïn.

Pseudorelevance feedback, II

Altres fonts per la retroalimentació i el refinament de la consulta:

Altres fonts per la retroalimentació i el refinament de la consulta:

- Enllaços clicats / no clicats.
- Temps per decidir / temps dedicat a mirar una resposta.
- Històric de consultes de l'usuari.
- Altres preferències de l'usuari.
- Paraules cooccurrents: afegir paraules que sovint apareixen amb les paraules de la consulta (per *query expansion*).