

TD1 – RI

Rappel :

	Pertinent	Non pertinent
Ramené (positif)	vrais positifs	faux positifs
Non ramené (négatif)	faux négatifs	vrai négatifs

- **La précision:** La proportion des vrais positifs dans le résultat r .

$$\text{Précision} = \frac{t_p(r)}{t_p(r) + f_p(r)} = \frac{t_p(r)}{|r|}$$

- **Le rappel:** la proportion des documents pertinents qui sont inclus dans r .

$$\text{Rappel} = \frac{t_p(r)}{t_p(r) + f_n(r)}$$

Exercice 1 :

- Évaluez les systèmes de recherche d'information ci-dessous, en fonction de leur précision et rappel.
 - Le SIF à la suite d'une requête avec le mot *pomme*, retourne une liste de 60 documents. Sur 100 documents traitant du fruit « pomme », il en fournit 50, mais les 10 documents restants portent plutôt sur la compagnie Pomme et Fils qui vend des tournevis.
 - Jean cherche un document qui explique comment planter un arbre. Il existe 12 documents de ce type dans la base de données. Suite à la requête « planter arbre », le système lui fournit une liste de 60 documents, dont 6 sont pertinents.
 - Soit une base de données contenant trois documents pertinents sur un total de 10. Un algorithme sélectionne 5 documents, dont deux sont pertinents, quel est le rappel et la précision ?
- Une société informatique prétend pouvoir identifier les artistes qui auront du succès, à l'aide d'un logiciel, avec une précision de 99 %. Est-ce possible ?

Exercice2 : précision et rappel

Voici une matrice donnant les faux positifs et faux négatifs, vrais positifs et vrais négatifs pour une recherche.

Table de contingence		
	Pertinent	Non pertinent
Ramené (positif)	50 (vrais positifs)	10 (faux positifs)
Non ramené (négatif)	30 (faux négatifs)	120 (vrai négatifs)

Quelques questions :

- Donnez la précision et le rappel.
- Quelle méthode stupide donne toujours un rappel maximal?
- Quelle méthode stupide donne une précision infinie ?

Exercice 3 :

Soit un système qui affiche systématiquement 20 documents, ni plus ni moins, pour toutes les recherches. Indiquez quel est la précision et le rappel dans les cas suivants :

- Mon besoin de recherche correspond à un document unique de la collection, il est affiché parmi les 20.
- Ma base contient 100 documents, je veux tous les obtenir, le système m'en renvoie 20.
- Je fais une recherche dont je sais qu'elle devrait me ramener 30 documents. Parmi les 20 que renvoie le système, je n'en retrouve que 10 parmi ceux attendus.