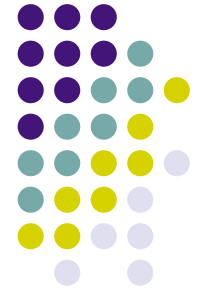


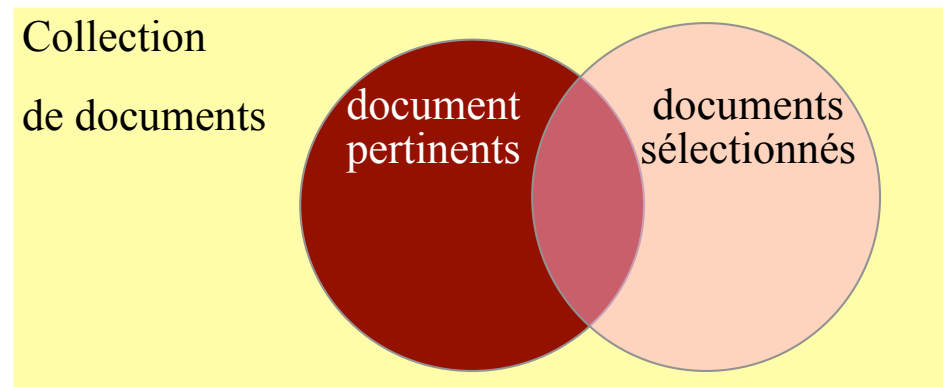
Deux facteurs principaux de pertinence



- Rappel
 - La capacité d'un système à sélectionner **tous** les documents pertinents de la collection
- Précision
 - La capacité d'un système à ne sélectionner **que** des documents pertinents



Précision et Rappel



$$rappel = \frac{\text{Nombre de documents pertinents sélectionnés}}{\text{Nombre total de documents pertinents}}$$

$$précision = \frac{\text{Nombre de documents pertinents sélectionnés}}{\text{Nombre total de documents sélectionnés}}$$



Pourquoi deux facteurs ?



- FACILE de faire du rappel il suffit de sélectionner toute la collection
- MAIS, la précision sera très faible

Calcul du rappel et de la précision



- On suppose qu'on dispose d'une collection de tests
 - Lancer chaque requête sur la collection de tests.
 - Marquer les documents pertinents par rapport à la liste de test.
 - Calculer le rappel et la précision pour chaque document pertinent de la liste.

Calcul du rappel et de la précision

Exemple



n	doc #	relevant		
1	588	x		
2	589	x		
3	576			
4	590	x		
5	986			
6	592	x		
7	984			
8	988			
9	578			
10	985			
11	103			
12	591			
13	772	x		
14	990			

Le nombre total de documents pertinents est = 6

$R=1/6=0.167$; $P=1/1=1$

$R=2/6=0.333$; $P=2/2=1$

$R=3/6=0.5$; $P=3/4=0.75$

$R=4/6=0.667$; $P=4/6=0.667$

$R=5/6=0.833$; $P=5/13=0.38$

Il manque un document pertinent.
On n'atteindra pas le 100% de rappel



Précision moyenne

- On souhaite souvent avoir une valeur unique
 - Par exemple pour les algorithmes d'apprentissage pour contrôler l'amélioration
- La précision moyenne est souvent utilisée en RI
- Plusieurs moyennes
 - Précision moyenne non interpolée (PrecAvg - AP) :
 - Calculer la moyenne des précisions à chaque apparition d'un document pertinent
 - Précision à trois points de rappel (0.2, 0.5, 0.7)

Précision moyenne non interpolée

Exemple



n	doc #	relevant		
1	588	x		
2	589	x		
3	576			
4	590	x		
5	986			
6	592	x		
7	984			
8	988			
9	578			
10	985			
11	103			
12	591			
13	772	x		
14	990			

Le nombre total de document pertinent est = 6

$R=1/6=0.167$; $P=1/1=1$

$R=2/6=0.333$; $P=2/2=1$

$R=3/6=0.5$; $P=3/4=0.75$

$R=4/6=0.667$; $P=4/6=0.667$

$AvgPrec=(1+1+0,75+0,667+0,38)/6$

$R=5/6=0.833$; $p=5/13=0.38$

MAP (Mean Average Precision)



- = Moyenne des AP sur toutes les requêtes



Précision à X documents

- Précision à différents niveaux de documents
 - Précision calculée à 5 docs, 10 docs, 15docs, ...

n	doc #	relevant
1	588	x
2	589	x
3	576	
4	590	x
5	986	
6	592	x
7	984	
8	988	
9	578	
10	985	
11	103	
12	591	
13	772	x
14	990	

Prec. à 5 docs = 3/5

Prec. à 10 docs = 4/10