

## Evaluation des modèles de RI :

### *Rappel et Précision :*

Soit  $D_q$  : la liste des documents jugés pertinents dans la collection (Exemple : Qrel.txt dans la collection CACM) ; et  $L_q$  : la liste de documents retournés à l'issue du processus de recherche (Le résultats de la recherche, i.e, tous les documents qui ont une similarité  $> 0$ ).

Notons que  $D_q$  est un sous ensemble de la collection  $D$ , ainsi que  $L_q$  qui est aussi un sous ensemble de la collection  $D$ . Idéalement,  $D_q$  est inclus dans  $L_q$  entièrement ou partiellement mais seules les métriques peuvent le confirmer.

Le rappel est défini par : 
$$\frac{|L_q \cap D_q|}{|D_q|}$$

La précision est défini par : 
$$\frac{|L_q \cap D_q|}{|L_q|}$$

Idéalement, on voudrait qu'un système donne de bons taux de précision et de rappel en même temps. Un système qui aurait 100% pour la précision et pour le rappel signifie qu'il trouve tous les documents pertinents, et rien que les documents pertinents. Cela veut dire que les réponses du système à chaque requête sont constituées de tous et seulement les documents idéaux que l'utilisateur a identifiés. En pratique, cette situation n'arrive que très rarement. Plus souvent, on peut obtenir un taux de précision et de rappel aux alentours de 30%.

Les deux métriques ne sont pas indépendantes. Il y a une forte relation entre elles. Quand l'une augmente, l'autre diminue. Il n'est pas opportun de juger la qualité d'un système en utilisant seulement une des deux métriques. En effet, il est facile d'avoir 100% de rappel: il suffirait de donner toute la base comme réponse à chaque requête. Cependant, la précision dans ce cas-ci serait très basse. De même, on peut augmenter la précision en donnant très peu de documents en réponse, mais ça sera fait au détriment du rappel. Il faut donc utiliser les deux métriques ensemble.

Donc l'intérêt de la phase d'évaluation est de déterminer sous quelles conditions de travail (Taille de la réponse et seuil de similarité) il est possible d'avoir les meilleurs taux de rappel et de précision en même temps.

La taille de réponse signifie la quantité de document qu'il est suffisant de retourner à l'utilisateur afin de satisfaire son besoin en information en tenant compte du jugement de pertinence.

Le seuil de similarité signifie le taux de similarité du document par rapport à la requête pour qu'il puisse espérer faire partie de la réponse finale.

### **Concrètement,**

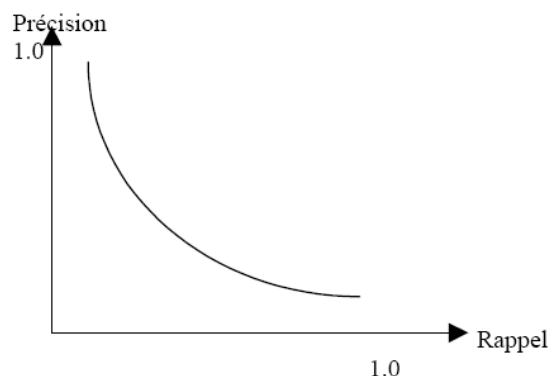
Dans votre projet de tp il est demandé d'implémenter le modèle vectoriel avec quatre mesure de similarité (produit interne, Dice, Jaccard, Cosius). Dans un premier lieu vous allez constater que la réponse fournit par une requête donnée avec les quatre mesure est la même en terme de contenu mais diffère uniquement par rapport à l'ordre des documents dans certains. C'est la raison pour laquelle vous allez retenir le Cosius pour les expérimentations restantes.

A l'issu de la recherche et pour chaque requête un rappel et une précision de référence devront être calculée pour en tenir compte lors de la phase d'évaluation. Pour les besoin de cette phase, vous devrez également, conservée l'ensemble de réponse totale (c'est-à-dire tous les documents qui ont une similarité  $>0$ ). Cet ensemble devra être trié par ordre décroissant de la similarité pour préparer la phase d'évaluation.

La phase d'évaluation consiste à déterminer concrètement le seuil de similarité et la taille de la réponse qui permettent d'avoir les meilleurs taux de rappel et de précision en tenant compte des résultats des requêtes de la collection et des jugements de pertinence. Il s'agit de faire un paramétrage comme pour les métaheuristiques en vue de fixer les deux valeurs en question.

La courbe de rappel-précision : (En option dans le projet)

Les mesures de précision-rappel ne sont pas statiques non plus (c'est-à-dire qu'un système n'a pas qu'une mesure de précision et de rappel). Le comportement d'un système peut varier en faveur de précision ou en faveur de rappel (au détriment de l'autre métrique). Ainsi, pour un système, il existe une courbe de précision-rappel qui a en général la forme suivante:



Pour établir la courbe il est donc utile de faire varier la précision en fonction de rappel ; ce qui n'est possible qu'en faisant varier la taille de la réponse et de noter l'évolution des deux

métriques. En d'autres termes calculer le rappel et la précision en considérant un seul document dans la réponse, puis deux, puis trois et ainsi de suite jusqu'à inclure toute la réponse, et ainsi faire pour toutes les requêtes de la collection. Ensuite procéder à un calcul de moyenne sur les tailles de réponse pour ne garder qu'une seule composante à la fois pour chaque taille de réponse (La moyenne des rappels pour un seul document de réponse, de même pour la précision, ensuite à deux documents, ... etc.)

## Organisation du Rapport de TP

Le rapport de tp est un document technique du projet il ne doit donc en aucun cas contenir d'algorithmes ou de définition des modèles ou autres textes introductifs ou explicatifs des modèles implémenter.

- ✚ Pour la phase d'indexation vous devez présenter dans le rapport les informations suivantes : taille des index en nombre d'entrée et en octet, durée d'indexation (temps) , types de structures utilisées, si vous avez conservé vos index une fois établi sur des fichiers le type de fichier et la structuration. Sans oublier vos critiques de la phase d'indexation (c'est-à-dire les points positifs et négatifs que vous avez constatés).
- ✚ Pour le modèle booléen vous devez présenter la manière avec laquelle vous avez évalué vos expressions booléennes, avec ou sans parenthèses et le temps nécessaire
- ✚ Pour le modèle vectoriel toute la phase d'évaluation depuis les quatre mesures de similarité jusqu'aux expérimentations pour fixer la taille et le seuil de similarité.
- ✚ Tous les résultats doivent être brièvement discuté et commenté.

Le rapport doit être envoyé par email au plus tard le **dimanche 05 Janvier 2020**.