

## TD 3 - Evaluation d'un SRI

### Exercice 1 – Métriques d'évaluation

Soit un corpus de documents comprenant 1000 documents au total. On dispose de deux systèmes de RI (S1 et S2) qui ont chacune ordonnancés 20 documents pertinents en réponse à une requête  $Q$ . On associe ainsi pour chaque système une liste de document trié par ordre décroissant du score de pertinence système (de la gauche vers la droite). Ces documents sont annotés par TP pour traduire "Très pertinent", P pour traduire "Pertinent" ou N pour traduire "Non pertinent". On suppose qu'il existe 12 documents pertinents (au sens pertinents (P) ou très pertinents (TP)) au total pour cette requête.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
S1	N	N	N	N	N	N	N	TP	P	N	TP	N	N	N	P	N	N	TP	P	P
S2	TP	TP	TP	TP	N	N	N	P	TP	P	N	N	N	N	N	N	N	N	N	N

**Q 1.1** Calculer les valeurs de rappel et précision aux différents points de rappel, en considérant une pertinence binaire P/NP, donc TP est assimilé à P.

**Q 1.2** Calculer les valeurs de précision interpolées des systèmes S1 et S2 puis tracer sur un seul graphique la courbe de rappel/précision interpolée associée. Commenter la courbe en faisant une analyse comparative des performances des deux systèmes S1 et S2.

**Q 1.3** Calculer les mesures NDCG des systèmes S1 et S2

**Q 1.4** Compléter votre interprétation de la courbe rappel/précision interpolée en faisant le lien avec les mesures NDCG.

### Exercice 2 – Courbes rappel-précision

La figure qui suit présente la courbe de rappel-précision interpolée de deux SRI qui indexent des collections d'articles scientifiques. On suppose qu'il n'existe pas de différences dans les noyaux de ces deux systèmes hormis la fonction de pondération des documents. Supposons que vous êtes un scientifique qui recherche tous les articles qui traitent d'un sujet donné et donc vous ne souhaitez de préférence rater aucun article. Lequel de ces SRI vous paraît le plus pertinent et pourquoi ?

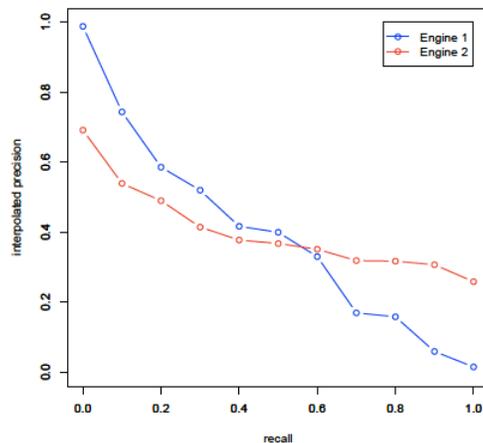


FIGURE 1 – Exercice de Lynda Tamine