

Une chaîne globale de fouille de textes

Evaluation Terminologie

Mathieu Roche

Cours ECDA

2016/2017

Evaluation des méthodes de RI

BioTex

BIOmedical Term EXtraction

[Extraction](#) [About and Citation](#) [JAR and Documentation](#) [Contact Us](#)

 **Alert:** This application **does not work with Firefox**, it only works with Google Chrome, Safari and Internet Explorer

Patterns Information

Number of linguistic patterns 
Patterns extracted from [UMLS](#) for English and Spanish, and from [MeSH](#) for French
Ex: Noun Noun; Noun Prep:det Noun; ... [more examples](#)

Type of terms to extract

All Terms single-word + multi-word term Multi Terms multi-word term

Measures selection and data

Select ranking measure [read more](#)

Type of documents Single Document Set of Documents

File source Aucun fichier sélectionné.
Only ".txt" accepted as file extension

Language of your text

Institutions



Laboratoire
Informatique
Robotique
Microélectronique
Montpellier



Sponsors

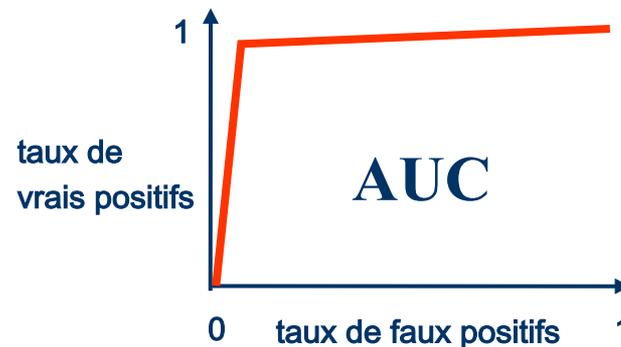


Evaluation des méthodes de RI

- **Front de Pareto (courbe Rappel/Précision) : plusieurs critères sont abordés**
- Le front de Pareto est défini par l'ensemble des approches telles qu'il n'existe pas une solution qui soit la meilleure pour tous les critères (ici précision et rappel).
- Les approches qui ne sont pas sur le front de Pareto sont dites “dominées”.
- Classement possible par front de Pareto

Evaluation des méthodes de RI

- Evaluation des fonctions de rang
- Utilisation des **courbes ROC** (Receiver Operating Characteristic) :
courbe dont le taux de vrais positifs est représenté en ordonnées et
le taux de faux positifs est représenté par l'axe des abscisses
- *Avantage* : **pas de sensibilité** dans le cas d'un **déséquilibre** entre les
classes.



Evaluation des méthodes de RI

- Evaluation des fonctions de rang

Exercice : Soit les deux listes ci-dessous :

liste 1 : - - + + -

liste 2 : + - - - +

Quelle est la liste la plus pertinente selon le “critère AUC” ?