

# Système simple de recommandation de documentation

## Matrices d'utilité et de cooccurrence

Dans le TP1 nous avons réalisé un programme permettant d'obtenir la matrice Usagers Thèmes (**Mut**) ainsi que la liste des usagers et la liste des thèmes. On vous soumet un nouveau fichier de log (**log-reco.txt**) ayant un format semblable au fichier de log précédent. On vous demande d'en extraire les 5 matrices suivantes :

- **Mut**: Matrice Usagers Thèmes correspondant au nouveau fichier **log-reco.txt**
- **Mut-binaire** : Transformation binaire de Mut (voir plus loin)
- **Mtt** : Matrice des distances entre thèmes
- **Mtt-binaire** : Transformation binaire de Mtt
- **MutR** : Matrice contenant les thèmes recommandés aux usagers

Pour construire **Mtt** on commence par transformer **Mut** en matrice binaire (i.e selon la présence ou l'absence d'une relation usager thème dans Mut, 1 si une relation existe, 0 sinon). On appliquera ensuite le calcul de distance\* entre thèmes sur la matrice **Mut-binaire** pour obtenir **Mtt**.

On convertira ensuite **Mtt** en matrice binaire pour extraire les règles d'associations. Pour cela, on considérera que 2 thèmes peuvent être liés s'ils y a une distance inférieure à 0,5 (**seuil**) entre eux. **Mtt-binaire** contiendra 1 dans ce cas, zéro sinon.

## Recommandations

Pour chaque couple de thèmes liés dans la matrice **Mtt-binaire** (i.e. contenant un 1), on vérifiera la présence des 2 thèmes dans la matrice **Mut-binaire** de manière à construire la nouvelle matrice **MutR** (usagers thèmes recommandés) comme suit.

Au départ **MutR** a le même format que Mut-binaire (i.e usagers-thèmes), excepté qu'elle est initialisée à zero. La construction de **MutR** se fait selon la règle de recommandation suivante :

*Pour tous les couples de thèmes liés (i.e. dans Mtt-binaire), si l'utilisateur a consulté seulement un des deux thèmes (i.e. dans Mut-binaire), on mettra 1 dans la cellule du thème non consulté (i.e. dans MutR).*

La matrice **MutR** sera ensuite utilisée pour lister les utilisateurs et les documentations qui leurs seront recommandées.

## Résultats attendus (voir fichier format\_rendu.txt)

Les résultats à rendre sont les suivants :

- 5 Matrices Mut, Mut-binaire, Mtt, Mtt binaires et MutR
- La liste des recommandations (à qui recommande t'on quoi ?)
- Un compte rendu d'une demi page environ. Vous analyserez en particulier le rôle du seuil. Vous indiquerez ce qui pour vous est important dans ce TP, vos remarques, suggestions, ....
- Les sources de votre programme

\* On utilisera la distance de Jacquard qui correspond au complément du rapport entre la taille de l'intersection sur la taille de l'union entre deux ensembles. 
$$D = 1 - \frac{(U1 \cap U2)}{(U1 \cup U2)}$$