

Prénom : \_\_\_\_\_ Nom de famille : \_\_\_\_\_

Matricule : \_\_\_\_\_ Signature: \_\_\_\_\_

**École Polytechnique de Montréal  
Génie informatique**

**INF6403 (Interfaces intelligentes)**

Professeur : Michel Desmarais

**Final (partie théorique), automne 2006**

**4 décembre 2006**

**Durée : 150 minutes**

**Pondération : 30%**

**(pour la partie théorique seulement)**

**La documentation et la calculatrice programmable sont permises**

L'examen comporte **5** pages et **9** questions. Le chiffre entre crochet au début de chaque question indique le nombre de points assigné à la question.

Répondre à toutes les questions.

**Écrivez vos réponses sur le questionnaire**

Réservé au correcteur :

	Résultat
1 (3)	
2 (2)	
3 (2)	
4 (2)	
5 (2)	
6 (1)	
7 (2)	
8 (4)	
9 (2)	
<b>Total (20)</b>	

1. Dans le cadre des modèles d'espaces vectoriels, Il existe plusieurs mesures de similarité dont celle basée sur le cosinus :

$$CS(Q, D_i) = \frac{\sum_{j=1}^t w_{qj} d_{ij}}{\sqrt{\sum_{j=1}^t (d_{ij})^2 \sum_{j=1}^t (w_{qj})^2}}$$

et celle basée sur le produit scalaire :

$$CS(Q, D_i) = \sum_{j=1}^t w_{qj} \times d_{ij}$$

où  $Q$  représente la requête et  $D_i$  le document  $i$  et où  $w$  et  $d$  sont basés sur le produit de  $tf$  et  $idf$ .

- (a) [2 pts.] En supposant que l'on effectue une recherche par mot-clés pour trouver un cours dans la liste des paragraphes de descriptions de cours de l'École, indiquez sous quelles conditions on peut s'attendre à ce que les deux mesures retournent le même résultat et, si ces conditions ne sont pas rencontrées, quel résultat serait le plus approprié.

**Solution:** Le cosinus est normalisé par la longueur de la requête ( $\sqrt{\sum_{j=1}^t (w_{qj})^2}$ ) et la longueur du document ( $\sqrt{\sum_{j=1}^t (d_{ij})^2}$ ). La requête étant toujours la même, ce facteur ne jouera donc pas. Dans la mesure où la longueur des descriptions est elle aussi similaire, les résultats seront les mêmes. Autrement, les descriptions plus longues plus longs auront plus de chances d'être retournée.

- (b) [1 pt.] Supposons maintenant que l'on cherche plutôt quels cours sont les plus similaires en comparant leurs descriptions. Laquelle des mesure est la plus adéquate et pourquoi?

**Solution:** C'est le même problème mais, cette fois, la longueur des deux documents peut varier lorsqu'on mesure la similarité (alors que dans le premier cas la longueur de la requête est constante). Si les paragraphe ont tous la même longueur, les deux donneront le même résultat, sinon les paragraphes plus longs seront favorisés. Comme la longueur n'est pas un indice de pertinence pour la similarité, le cosinus est plus adéquat.

2. [2 pts.] Les mesures de fréquence de termes ( $tf$ , *term frequency*) et de fréquence inversée de documents ( $idf$ ) peuvent-elles servir dans des applications de filtres collaboratifs? Si oui, comment?

**Solution:** Voir Hanani et al. (2001) ou Breese et al. (1998). Essentiellement, comme les filtres collaboratifs utilisent des modèles d'espaces vectoriels, les mêmes techniques peuvent s'appliquer. Il s'agit donc, par exemple, de diminuer le poids de caractéristiques (votes) très communes chez l'ensemble des utilisateurs au bénéfice des caractéristiques moins fréquentes.

3. [2 pts.] Expliquez la différence entre l'approche item-item et l'approche utilisateur-utilisateur pour les applications de filtres collaboratifs et indiquez dans quelles circonstances une est préférable à l'autre.

**Solution:** À élaborer.

4. [2 pts.] Expliquez pourquoi l'algorithme PageRank surpasse la simple recherche booléenne par mot clé.

**Solution:** L'algorithme exploite la notion de popularité (vote) des pages, qui s'apparente aux principes de filtres collaboratifs, pour fournir une mesure de pertinence. La popularité d'une page, mesurée par le nombre pondéré de liens entrants, s'avère un excellent indicateur de la pertinence d'une page. L'algorithme permet par exemple de trouver des pages pertinentes même si elles ne contiennent pas aucun des mots de la requête.

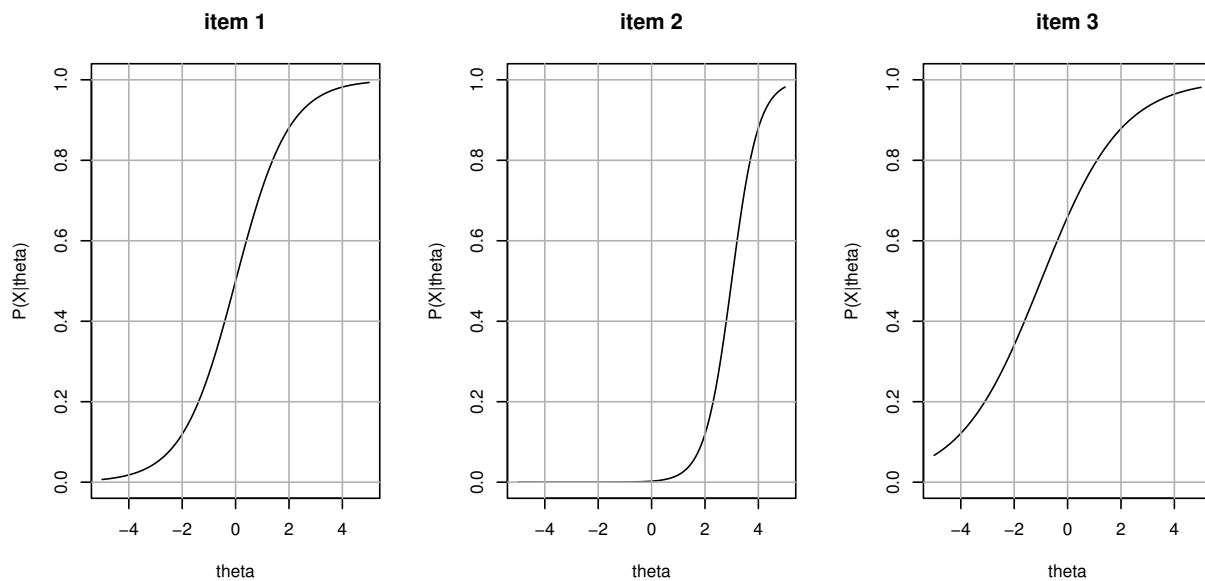
5. [2 pts.] Comment, selon-vous, un modèle utilisateur pourrait-il être utilisé dans le cadre de la recherche d'information?

**Solution:** Un modèle utilisateur peut représenter des intérêts qui peuvent par la suite être utilisés pour déterminer la pertinence de documents. Par exemple, les intérêts peuvent être représentés sous la forme d'un vecteur de mots dans un espace réduit, qui peut servir à filtrer les documents les plus pertinents.

6. [1 pt.] Expliquez le rôle de l'hypothèse d'indépendance faite dans la Théorie des réponses aux items.

**Solution:** La formule de base de la TRI,  $P(X_1, X_2, \dots, X_n | \theta) = \prod_i P(X_i | \theta)$ , repose sur l'hypothèse d'indépendance des termes  $P(X_i | \theta)$ , c.-à-d. la probabilité de chaque item étant donné  $\theta$ . Sans elle on ne peut calibrer le modèle avec une quantité raisonnable de données.

7. [2 pts.] Les graphiques suivants indiquent la relation entre l'habileté d'un individu et ses chances de succès pour trois items. Indiquez les caractéristiques respectives de chacun des 3 items en termes de difficulté et de discrimination relatives (un item par rapport à l'autre).



**Solution:** La difficulté (point d'inflexion en  $x$ ) va croissante dans cet ordre : 3,1,2. La discrimination (pente) va croissante dans le même ordre : 3,1,2.

8. Le fameux trombone animé de l'application MS Office utilisait certaines techniques vues au cours mais il n'a jamais eu le succès attendu.

(a) [2 pts.] Donnez deux exemples des techniques que "M. Trombone" pouvait utiliser pour offrir une fonctionnalité d'assistance. Décrivez la fonctionnalité correspondante.

**Solution:** (1) Il peut utiliser une technique de reconnaissance de plan identifier une tâche réalisée de manière inefficace et pour suggérer une méthode alternative mais plus efficace de faire un tâche. (2) Il peut utiliser un schème bayésien pour prédire le répertoire associé à une tâche donnée en fonction d'une série d'indice (voir Stumpf et al. (2005).

(b) [2 pts.] Expliquez pourquoi il n'a peut-être pas eu le succès escompté en vous référant aux principes des interfaces à initiatives mixtes énoncés par Horvitz.

**Solution:** Plusieurs recommandations et interventions n'étaient pas pertinentes. Le "trombone" ne faisait pas d'analyse de l'utilité escomptée de l'intervention et avait ainsi tendance à être plus souvent encombrant et dérangerant qu'utile. Il aurait dû avoir un mécanisme pour inhiber une intervention jugée non utile par rapport au bénéfice escompté.

9. [2 pts.] Quel sont, selon vous, les applications les plus prometteuses des techniques d'interface intelligentes? Justifiez.

**Solution:** Question de synthèse jugée selon le niveau d'assimilation des concepts du cours.

**Fin de l'examen**  
**Nombre de pages : 5**  
**Total des points : 20**