

Polytech' Grenoble – Info 4
Examen Accès et Recherche d'Information – 8 avril 2021 – 2 heures

Note importante : les parties I et II sont à rédiger sur des copies séparées.

Partie I

Exercice 1 – Généralités sur les modèles et systèmes de recherche d'information

Question 1 : Rappeler pourquoi on utilise des fichiers inverses dans les systèmes de recherche d'information. (5 lignes max)

Question 2 : Expliquer l'intérêt de l'étape de normalisation lors de la construction des tableaux de rappel/précision normalisés pour l'évaluation des systèmes de recherche d'information. (5 lignes max)

Exercice 2 : Obtention des termes d'indexation

Considérons les 2 documents suivants :

D₁ = "la recherche d'information s'exerce constitutionnellement"

D₂ = "les exercices des informaticiens"

Question 1 : En ne considérant aucun anti-dictionnaire, indiquer les termes qui indexent ces 2 documents.

Question 2 : En considérant l'anti-dictionnaire suivant : {à, au, d, de, des, du, elle, elles, en, est, je, il, ils, la, le, les, lui, s, sa, ses, son, tu, un, une}, indiquer les termes qui indexent ces documents.

Question 3 : En considérant, en plus de la question 2, l'utilisation d'une version simplifiée de l'algorithme de Porter ayant les règles suivantes :

1. S → /

2. IC → /

3. IEN → /

4. ION → /

5. E → /

6. MENT → /

7. *double consonne* → *la consonne*

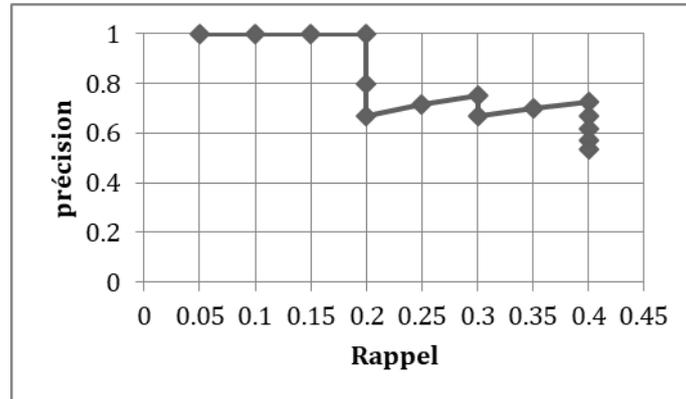
Indiquer les termes qui indexent les documents (en décrivant les règles de réécritures utilisées dans chaque cas).

Question 4 : Prenons le cas des mots « information » et « informaticien » avec les règles de réécritures de la question 3. Expliquer le problème qui survient dans ce cas. Proposer une solution pour régler ce problème.

Exercice 3 : Évaluation de système de recherche d'information

Partie A

Considérons, à partir du résultat d'un système pour une requête Q1, la courbe de rappel précision non normalisée suivante :



Question 1 : D'après ce graphique, indiquez votre estimation du nombre de documents pertinents (la valeur $|P|$ du cours) pour Q1.

Question 2 : Indiquer la précision maximale approximative obtenue à partir de la courbe ci-dessus, après que le système ait retrouvé 5 documents pertinents pour Q1.

Partie B

Supposons maintenant qu'un système de recherche d'information renvoie 11 documents pour une requête Q2 quelconque. Supposons de plus qu'un humain a déterminé que 5 documents sont pertinents pour cette requête, et que dans la réponse du système les documents en position 3, 4, 6 et 9 sont pertinents.

Question 3 : Donner le tableau non normalisé et le tableau normalisé de rappel/précision pour ce système sur cette requête.

Il existe des mesures qui caractérisent par une seule valeur numérique l'évaluation d'un système. Une telle mesure est par exemple la précision moyenne *AveP* (en anglais : *Average Precision*). Cette précision est une valeur réelle unique pour une requête, elle est définie par :

$$AveP = \frac{\sum_{k=1}^n prec(k).rel(k)}{|P|}$$

Avec :

- P : l'ensemble des documents pertinents pour la requête ($|P|$ est sa cardinalité);
- n : le nombre de documents renvoyés par le système (égal à $|R|$ d'après le cours);
- $rel(k)$: une fonction binaire égale à 1 si le document au rang k est pertinent, et égale à 0 si le document en rang k est non-pertinent pour la requête;
- $prec(k)$: la valeur de précision pour le rang k (calculée à partir du tableau de rappel/précision non normalisé).

Cette formule se base sur les valeurs du tableau non normalisé.

Question 4 : En se basant sur les explications ci-dessus, donner les valeurs de précision moyenne *AveP* du système pour la requête Q2.

Partie 2

Exercice 1. Filtres de convolution

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Image I

-1	1	-1
2	1	3
5	-2	4

Filtre F1

+1/2	0	-1/2
------	---	------

Filtre F2

+1
-2
+1

Filtre F3

On veut convoluer l'image I avec les filtres F1, F2 et F3.

Question 1.1 : Si l'on effectue une convolution de I sans "padding", quelle sera la taille de l'image résultat avec les filtres F1, F2 et F3 ?

Question 1.2 : Même question avec un "padding" de (1,1) pour F1, de (1,0) pour F2, et de (1,1) pour F3.

Question 1.3 : Dans le cas sans "padding", donner la valeur du pixel en haut à gauche pour l'image I convoluée, respectivement avec les filtres F1, F2 et F3.

Question 1.4 : Combien d'opérations sont-elles nécessaires pour calculer un élément de l'image de sortie dans la convolution avec le filtre F1 ?

Question 1.5 : À quel genre d'opérateur correspondent les filtres F2 et F3 ?

Exercice 2. Histogrammes

0	2	3	3	3	4
1	2	2	3	4	4
1	3	3	4	2	4
5	5	4	2	2	4
3	2	2	4	4	4
0	1	3	3	3	4

Image

Question 2.1 : Calculer l'histogramme en niveau de gris de l'image ci-dessus.

Exercice 3. Apprentissage profond (deep learning)

On considère une couche de convolution qui prend en entrée 128 plans de taille 56×56 et qui produit en sortie 256 plans de la même taille avec des filtres de taille 3×3 .

Question 3.1 : Faut-il prévoir un padding ? Si oui, de quelle taille ?

Question 3.2 : Combien y a-t-il de paramètres dans la fonction de convolution ?

On considère une couche complètement connectée à 4096 entrées et 1000 sorties.

Question 3.3 : Combien y a-t-il de paramètres dans la fonction correspondante ?

Question 3.4 : Combien d'opérations flottantes sont effectuées pour chaque image d'entrée ?

Question 3.5 : À quoi servent les couches "ReLU" dans les réseaux de neurones ? Par quoi peut-on les remplacer ?

Question 3.6 : À quoi servent les couches "Max Pool" dans les réseaux de neurones ? Par quoi peut-on les remplacer ?