

Rapport projet PIM : PageRank

Décembre 2023 - Janvier 2024

Groupe EF03

THEVENET Louis

MORISSEAU Albin

Table des matières

1. A FAIRE	2
1.1. Infos	2
1.2. A faire	2
2. Résumé du rapport	2
3. Introduction	2
3.1. Paramètres de la ligne de commande	2
3.2. Valeurs par défaut des paramètres	3
3.3. Spécification des paramètres	3
4. Architecture de l'application	3
5. Choix techniques / Structure de données	3
5.1. Cas Matrices Pleines	3
5.2. Cas Matrices Creuses	4
6. Performances des implantations	4
7. Difficultés rencontrées et solutions apportées	4
7.1. Représentation par matrices creuses	4
8. Répartition des tâches	4
9. Conclusion et avancement	5
10. Annexes :	5
10.1. Apport personnel du projet	5

1. A FAIRE

1.1. Infos

Remise du rapport complet Mardi

Remise du code, présentation et script Mercredi

Présentation + démo Jeudi

1.2. A faire

- virer fichiers d'exemples du gitlab si j'ai oublié
- finir raffinages (vérifier que ça colle avec le code (peu probable))
- Finir le rapport
 - faire les tests de temps
- faire présentation et script

2. Résumé du rapport

Ce rapport contient les éléments essentiels de la conception d'un programme défensif ayant pour but de mesurer la popularité des pages Internet. On envisage deux principales implantations de l'algorithme, par des tableaux simples puis en exploitant le fait que la matrice d'adjacence est très creuse.

Le choix des matrices creuses est beaucoup plus efficace en termes de complexité et devra être privilégié (valeur par défaut) d'autant plus si le graphe d'entrée possède de nombreux sommets.

Ce rapport contient l'architecture complète de l'application et des différents modules utilisés ainsi qu'un comparatif des performances des deux implantations.

3. Introduction

Chaque jour, plus de 7 milliards de recherches sont effectuées via des moteurs de recherche soit presque 81000 par seconde. Chaque recherche peut renvoyer des millions de résultats. Ainsi, il est essentiel pour faciliter la vie des utilisateurs et minimiser les coûts, de trier ces différents résultats de la manière pertinente et efficace.

Ce projet consiste à implanter l'algorithme PageRank qui mesure la popularité de pages Internet en utilisant les référencements qui mènent d'une page à une autre.

On pourra choisir les différents paramètres de l'algorithme ainsi que l'implantation choisie lors de la saisie de la commande d'appel.

3.1. Paramètres de la ligne de commande

- A** Le *Dumping Factor* (α), est un paramètre de pondération, plus sa valeur approche 1, plus la convergence est lente
- K** Le nombre d'itération (k) de l'algorithme à réaliser
- E** La précision (ϵ) permet d'interrompre l'algorithme lorsque la variation du vecteur poids d'une itération à l'autre lui est inférieure
- R** Le préfixe utilisé pour les fichiers de sortie (<prefixe>.pr et <prefixe>.prw)

- C Permet de choisir la version matrices creuses
- P Permet de choisir la version matrices pleines

3.2. Valeurs par défaut des paramètres

Les paramètres du programme lorsque non spécifiés ont pour valeurs :

$$\begin{aligned}\alpha &:= 0.85 \\ k &:= 150 \\ \varepsilon &:= 0.0\end{aligned}$$

3.3. Spécification des paramètres

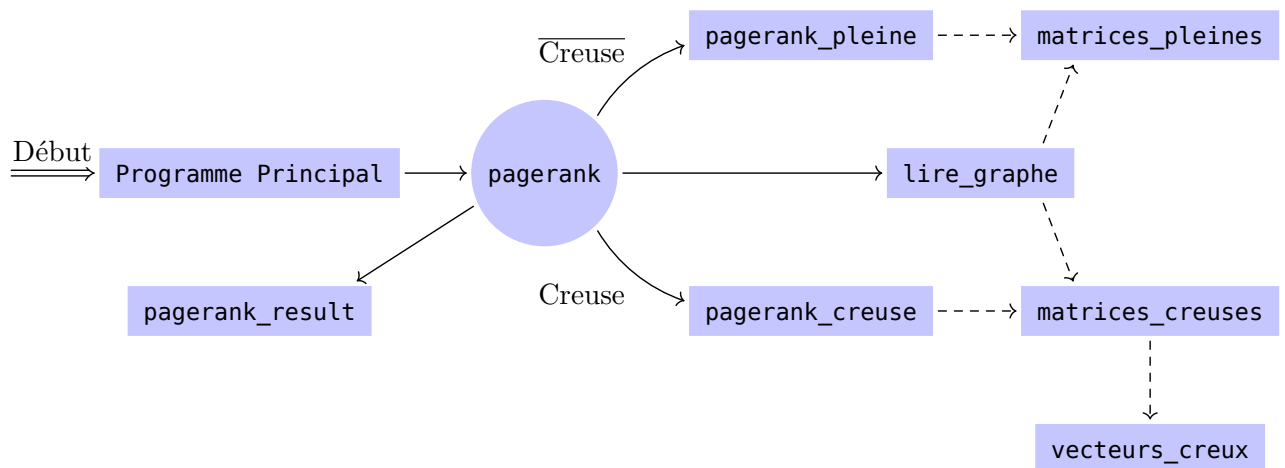
Les paramètres doivent respecter certaines conditions pour que le programme exécute l'algorithme :

$$\begin{aligned}\alpha &\in [0, 1] \\ \varepsilon &\geq 0\end{aligned}$$

Un seul algorithme choisi (Creuse ou Pleine)

Préfixe non vide

4. Architecture de l'application



Légende :

\Rightarrow point d'entrée

\rightarrow appelle

\dashrightarrow utilise

5. Choix techniques / Structure de données

5.1. Cas Matrices Pleines

```

1  type T_Matrice_Pleine is array(1..N, 1..P) of Long_Float;
2
3  type T_Matrice is record
4      Mat : T_Matrice_Pleine;
5      Lignes : Integer;
6      Colonnes : Integer;
7  end record;
```

5.2. Cas Matrices Creuses

```
1  -- matrices_creuses.adb
2  type T_Matrice is array(1..Taille) of T_Vecteur_Creux;
3  type T_Facteurs is array(1..Taille) of Long_Float;
4
5  -- vecteurs_creux.adb
6  type T_Cellule;
7  type T_Vecteur_Creux is access T_Cellule;
8
9  type T_Cellule is
10     record
11         Indice : Integer;
12         Valeur : Long_Float;
13         Suivant : T_Vecteur_Creux;
14     end record;
```

6. Performances des implantations

Fichier graphe Test	Implantation Pleine	Implantation Creuse
sujet.net		
worms.net		
brainlinks.net		

7. Difficultés rencontrées et solutions apportées

7.1. Représentation par matrices creuses

Nous avons initialement choisi de réaliser un vecteurs creux de vecteurs creux mais les performances n'étaient pas très bonnes alors que l'implémentation était compliquée.

Nous avons ensuite choisi d'utiliser un tableau de vecteurs creux afin que l'opération d'accès aux têtes de colonnes se fasse en temps constant (les vecteurs représentent les colonnes).

8. Répartition des tâches

Modules	Spécifier	Program- mer	Tester	Relire
programme_principal	A	A	A	L
lire_graphe	A	A	A	L
pagerank	A	A	A	L
pagerank_creuse	L	L	L	A
pagerank_pleine	L	L	L	A
matrices_creuses	L	L	L	A
matrices_pleines	L	L	L	A
vecteurs_creux	A	A	A	L
pagerank_result	A	A	A	L

9. Conclusion et avancement

Nous avons réussi à implanter de manière robuste les deux versions de l'algorithme.

Nous avons pour perspective d'améliorer la précision de nos calculs de poids en introduisant un paramètre générique de précision dans nos différents programmes.

10. Annexes :

Raffinages complets, Grilles d'auto évaluation : Raffinages projet Pagerank groupe EF03

10.1. Apport personnel du projet

Albin Morisseau Ce projet, m'as permis de

Louis Thevenet ...