

D-BDAT

Michel Beigbeder

TP

15 octobre 2018

Rendu du TP

Travail individuel, à déposer une première fois sur `campus.emse.fr` aujourd'hui en fin de séance, puis compléter avec un autre dépôt d'ici le 22 octobre et rendu final le 5 novembre.

Le rendu consiste pour aujourd'hui en un unique fichier source.

Exercice 1

Le but du TP est de déterminer si le graphe (orienté) des catégories de Wikipedia contient des circuits, ce qui serait étonnant vu la sémantique de ce graphe qui relie une catégorie à ses sous-catégories.

Il s'agit donc d'écrire un programme qui lit les données et en fait une représentation en mémoire, puis implémente l'algorithme de Tarjan.

Selon le langage de programmation utilisée, il faudra faire des allocations mémoires judicieuses. À titre d'information le graphe anglophone est composé d'environ 680 000 nœuds, ou un peu plus de 74 millions de nœuds... selon la numérotation de nœud considérée.

1.1 Trois graphes sont fournis, celui des versions anglaises, espagnoles et françaises de Wikipedia en 2011 : chargez le fichier `wikipedia-category_graph-COLL2-en,es,fr.tgz` depuis `campus.emse.fr`.

Explorez ces données pour en comprendre le formatage, compter le nombre de nœuds et le nombre d'arcs de ces graphes.

1.2 Choisissez une structure de données pour représenter le graphe avec des listes d'adjacence.

Pour la structure de listes, veillez à ne pas utiliser une implémentation qui transformerait la lecture en $O(n + m^2)$... Au besoin faites un programme de test indépendant pour vérifier cet aspect.

Vérifiez que cette lecture est correcte en ré-écrivant le graphe.

Faites votre propre fichier de test suffisamment petit mais cependant exhaustif.

1.3 Implémentez l'algorithme de Tarjan. Il est possible de commencer en écrivant un parcours en profondeur tel que décrit dans les polys d'Algorithmique du Tronc Commun Informatique de première année, puis d'y ajouter la gestion de la pile et le calcul des minimum des valeurs `v.numAccessible` selon la terminologie employé sur la page https://fr.wikipedia.org/wiki/Algorithme_de_Tarjan.