

Projet informatique : Chemins mnémotechniques

Philippe Gambette (gambette@lirmm.fr)

Contexte et objectif

Il s'agit de créer un outil pour aider à trouver des “chemins mnémotechniques” pour retenir du vocabulaire dans une langue étrangère. Il s'agit donc de retenir l'association entre deux mots, par exemple “travel” et “voyage”, en construisant un chemin de mots entre les deux.

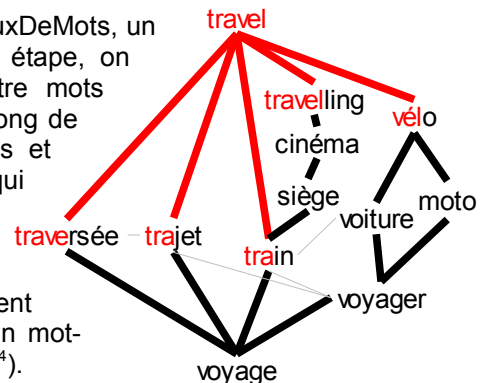
Ce chemin de mots commence par associer au mot anglais un mot français dont la graphie est proche. Par exemple pour “travel” : un “travelling” ou une “traversée”. Puis le chemin procède par associations d'idées en français : *travel* → *travelling* → *cinéma* → *siège* → *train* → *voyage*.

Comme source de mots français et de relations entre ces mots par association d'idées, on utilisera les données de JeuxDeMots¹ : un graphe de mots reliés par des arêtes quand ils ont été mis en relation par les internautes qui ont joué à JeuxDeMots. Par exemple, on peut avoir un aperçu des mots associés à “voyage” de la base de JeuxDeMots sur la page <http://www.lirmm.fr/jeuxdemots/rezo.php>.

Jeux de mots

Ainsi, le programme consistera en deux parties principales, illustrées sur l'exemple de l'association travel-voyage :

- **proximité graphique** : trouver, parmi tous les mots de JeuxDeMots, un ensemble de mots proches du mot “travel”. Pour cette étape, on pourra utiliser une mesure classique de distances entre mots (distance de Hamming ou de Levenshtein²). S'il est trop long de calculer la distance entre tous les mots de JeuxDeMots et “travel”, on pourra se limiter seulement à des mots qui contiennent des sous-chaînes de “travel”, ou bien examiner seulement les mots qui ont un préfixe ou un suffixe commun avec “travel”. Ceci demandera l'utilisation de structures de données adaptées pour trouver rapidement des mots ayant un préfixe ou un suffixe commun avec un mot-cible (arbre des préfixes communs³ voire arbre des suffixes⁴).
- **chemins lexicaux** : parcourir le graphe de JeuxDeMots à la recherche du terme cible “voyage”. Le graphe étant très grand, il faudra peut-être définir des stratégies d'exploration pour trouver des chemins intéressants entre “voyage” et “traversée” (un parcours en largeur depuis “voyage” et “traversée” en parallèle par exemple). Si les temps de calcul d'une solution complètement automatique sont trop longues, on pourra envisager une interface permettant à l'utilisateur de diriger la requête en choisissant les mots du voisinage de “voyage” ou “traversée” qui lui semblent les plus prometteurs pour trouver un chemin.



En fonction du mot français choisi, le graphe pourra fournir de nombreux chemins différents ou au contraire très peu de possibilités. Il faudra donc penser à utiliser une stratégie adaptée au mot choisi, par exemple en fonction du nombre de voisins du mot français, ou bien du nombre de mots à distance 2 du mot français dans le graphe.

En effet, on peut penser que la stratégie qui commence par l'étape de proximité graphique puis celle du chemin lexical est adaptée au mot “voyage” qui a beaucoup de voisins dans le graphe de JeuxDeMots. En revanche, pour un mot ayant peu de voisins dans le graphe, il faudra plutôt effectuer la recherche de chemin lexical “à l'aventure” dans le graphe, et calculer, pour tous les mots rencontrés, la proximité graphique avec le mot cible en langue étrangère.

A cause de ces stratégies très différentes en fonction de la requête, il pourra être utile de laisser à l'utilisateur la possibilité de guider l'exploration, et d'interagir avec la recherche, et donc de lui fournir des résultats partiels. De plus, les moyens mnémotechniques sont souvent très personnels, il faudra donc faire plusieurs

1 <http://www.lirmm.fr/jeuxdemots>

2 http://fr.wikipedia.org/wiki/Distance_de_Levenshtein

3 <http://www.iro.umontreal.ca/~felipe/IFT2030-Automne2002/TravauxPratiques/devoir2/tp2.html>

4 http://fr.wikipedia.org/wiki/Arbre_des_suffixes

propositions de chemins à l'utilisateur, quand les données le permettent.

Proposer le programme sous forme d'interface web permettra une utilisation simplifiée ne nécessitant pas de télécharger la base de données de JeuxDeMots (d'environ 50 Mo).

Etapes

- Familiarisation avec les données JeuxDeMots
- Choix du langage et des technologies utiles
- Planification de développement
- Réalisation de composants réutilisables très documentés
- Mise à disposition du programme sur une interface web
- Test des stratégies d'exploration du graphe (dans quels cas vaut-il mieux commencer par l'étape de proximité graphique plutôt que la recherche de chemin lexical "à l'aventure")
- Tests et évaluation des moyens mnémotechniques trouvés (longueur moyenne du chemin trouvé)

Enseignements connexes

- Algorithmes de graphes
- Langages et automates
- Architecture des applications de la toile