

Projet informatique : URL équivoques

Philippe Gambette (gambette@lirmm.fr)

Contexte et objectif

Il s'agit de créer automatiquement des url à double sens (par exemple www.penisland.net en anglais¹ ou www.photosdeputes.fr en français).

Le programme consistera en trois parties principales :

- **concaténation** : concaténer un mot-cible (proposé par l'utilisateur du programme) avec tous les mots possibles du dictionnaire
- **segmentation** : segmenter la chaîne de caractères concaténée à l'étape précédente en trouvant une segmentation alternative (à partir des mots du dictionnaire)
- **choix de l'utilisateur** : toutes les réponses sont fournies, dans un ordre bien choisi, à l'utilisateur qui sélectionnera manuellement les réponses qui auront le plus de sens.

Si la recherche exhaustive est trop longue dans les deux étapes, elle pourra être raccourcie en identifiant certaines règles de simplification : par exemple, pour l'étape de concaténation, si c'est un nom commun qui est proposé par l'utilisateur, essayer uniquement de le concaténer avec des adjectifs accordés de la même façon.

De même, pour l'étape de segmentation, si de nombreuses réponses existent on pourra identifier des patrons grammaticaux qui ont plus de sens que d'autres (par exemple la séquence [verbe à l'infinitif – adjectif – verbe conjugué au passé simple] risque d'avoir peu de sens), et choisir une stratégie d'exploration qui fournisse le plus vite possible à l'utilisateur des résultats variés.

Une méthode de visualisation permettant à l'utilisateur de bien naviguer visuellement dans le résultat, pour pouvoir éventuellement compléter manuellement certaines segmentations pendant que l'algorithme poursuit son exploration, sera également fournie.

```
penisland
pen land
island
is and
...
```

Extensions possibles

L'application web pourra être complétée par une interface d'enregistrement des résultats obtenu, et de vote pour élire les meilleurs résultats.

Le programme pourra être étendu à un dictionnaire phonétique, pour pouvoir générer des holorimes², voire résoudre ou générer des contrepèteries³.

Dans la partie segmentation, on pourra envoyer des requêtes automatiques à un moteur de recherche pour compléter le dictionnaire (vérifier si un mot existe en tant que nom propre).

Étapes

- Choix du langage et des technologies utiles
- Identification des ressources lexicales nécessaires
- Planification de développement
- Réalisation de composants réutilisables très documentés
- Mise à disposition du programme sur une interface web
- Tests et découverte d'URL équivoques

Enseignements connexes

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

1 <http://weirdtechnewshub.blogspot.com/2006/07/top-10-worst-url-flops.html>

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

3 <http://fr.wikipedia.org/wiki/Contrepèterie>