

beCP

2023

Tâche 2.1: Entraînement à la traduction (translation)

Auteur: l'équipe beCP Préparation: Robin Jadoul

Note: Cette tâche est output only. Cela veut dire que vous devez uniquement soumettre les résultats de vos calculs. Vous ne devez pas soumettre de code. Il y a plusieurs fichiers d'input, et vous pouvez soumettre vos solutions ensemble ou séparément. Bien entendu, cela signifie que vous n'êtes pas obligés d'utiliser le même programme pour résoudre chaque input, et vous pouvez même essayer d'en résoudre quelques-uns à la main.

Vous vous entraînez à la traduction pour pouvoir parler la langue locale quand vous vous rendez à des concours de programmation internationaux. Les devoirs demandent généralement beaucoup d'efforts et vous préféreriez vous entraîner pour la partie programmation des concours. Vous avez donc l'habitude de tout donner à un programme de traduction utilisant l'intelligence artificielle.

Catastrophe! Votre connexion internet tombe en panne et votre devoir est à rendre dans trois heures. Votre tâche est de classifier un tas de documents et d'indiquer pour chaque document dans quelle langue il est écrit, en vous basant sur quelques exemples par langue.

La liste des langues avec lesquelles vous allez travailler est la suivante :

- anglais (**en**)
- espéranto (**eo**)
- maori (**mi**)
- indonésien (**id**)
- lojban (**jbo**)

Input

Dans le téléchargement pour cette tâche, vous recevez un ensemble de 5 dossiers. Le premier dossier, `reference`, contient les exemples connus par langue, nommés `[language]_[i].txt`. Dans chacun des autres dossiers `set_[subtask]`, vous trouverez un certain nombre de fichiers `F`, simplement nommés `[i].txt`. Chacun de ces fichiers est un document pour lequel vous devez déterminer la langue correcte.

Output

Pour chacun des quatre dossiers, vous devez soumettre un seul fichier. La i -ème ligne du fichier doit contenir un seul mot : le code de langue (2 ou 3 lettres) de la langue pour ce fichier. Un squelette de code est fourni pour vous aider à lire les fichiers donnés.

Score

Pour un ensemble de fichiers (une sous-tâche) avec F fichiers à classer, si vous classez correctement x des F fichiers, vous recevez $25 \cdot \frac{x}{F}$ points.

Contraintes supplémentaires

Sous-tâche	Points	Contraintes
A	25	$F = 20$
B	25	$F = 100$
C	25	$F = 1\,000$
D	25	$F = 5\,000$

Exécuter votre code

Dans les téléchargements pour cette tâche, vous trouverez un squelette de code `translation.cpp`, où vous pouvez compléter le code pour trouver la langue d'un texte. Vous pouvez le faire en implémentant la fonction `string classify(string)` qui prend le contenu d'un fichier en entrée, et renvoie un des cinq codes de langue possibles. Une fonction `bool contains(string haystack, string needle)` est également fournie comme exemple de ce que vous pourriez faire dans votre code. Vous pouvez utiliser cette fonction, mais ce n'est pas une obligation.

Pour accompagner le code, vous trouverez un fichier `driver.cpp`. Ce fichier fournit la fonction `main` pour le programme et s'occupera d'appeler votre fonction `classify` sur un seul fichier ou sur tous les fichiers d'un dossier, dans le bon ordre.

Pour compiler un programme, vous pouvez utiliser la commande suivante :

```
g++ -std=c++11 -Wall -Wextra -Wshadow translation.cpp -o translation.  
Pour exécuter votre code sur un seul fichier, par exemple reference/en_1.txt,  
vous pouvez alors utiliser la commande  
./translation reference/en_1.txt
```

et pour l'exécuter sur chaque fichier dans un dossier, par exemple `set_A`, vous pouvez utiliser la commande

`./translation set_A`. Ces commandes afficheront les langues que votre programme détecte à l'écran, et vous avertiront si l'une d'entre elles n'est pas un code de langue valide. Pour sauver la sortie dans un fichier que vous pouvez soumettre, vous pouvez ajouter `> nom_du_fichier_a_sauver.txt` à la fin de ces commandes. Par exemple :

```
./translation set_A > set_A_solutions.txt.
```

Si vous affichez de l'information de débogage, assurez-vous d'utiliser `cerr` au lieu de `cout`.

Vous avez le droit de demander de l'aide technique pour faire tourner votre code et l'exécuter sur les données de la tâche en demandant à un surveillant ou en envoyant une question via le CMS.