

Sous-Titrage Intelligent ¹

Résumé

Dans ce projet inspiré de [sma], on se propose de développer un outil qui automatise toute la gestion des sous-titres et propose un apprentissage progressif d'une langue étrangère via des sous-titres intelligents.

1 Description du Projet

Par *sous-titre intelligent* on entend un mélange des sous-titres dans la langue parlée et dans la langue étudiée. Selon le niveau de l'utilisateur, les mots compliqués sont traduits dans la langue parlée, les mots simples sont omis ou laissés dans la langue étudiée. L'efficacité de l'apprentissage par ce genre de technique a été scientifiquement confirmé [KM14]. On pourra d'ailleurs s'inspirer de cet article (ainsi que de [sma]) pour l'interface.

Le programme prendrait soit la forme d'un module VLC soit d'un programme tournant en local. Au lancement d'un film ou d'une série, le module nous demanderait le nom/le compte de l'utilisateur qui s'apprête à regarder (car le niveau est personnalisé). Le programme va ensuite automatiquement télécharger les sous-titres dans la langue de l'utilisateur et dans la VO de la vidéo. En fonction du niveau de l'utilisateur, le module crée un sous-titre intelligent et le cale automatiquement sur la vidéo. Le module se souvient des mots déjà rencontrés et adapte le sous-titre intelligent en conséquence. Il propose également deux types d'aide à la compréhension : avant le visionnage, l'utilisateur a le choix de voir un aperçu global des difficultés linguistiques qu'il va rencontrer ; après le visionnage, le module propose un résumé des nouveaux mots et expressions rencontrés et les ajoute au dictionnaire lié à l'utilisateur. Enfin, le module permet de réviser les mots déjà rencontrés (et donc traduits) via un QCM qui met à jour la liste des mots appris, peut-être appris, rencontrés une seule fois, *etc.*

2 Ojectifs à remplir

2.1 Fonctionnalités attendues

Todo.

2.2 Objectifs Techniques

Vidéo, Son & Reconnaissance Il faut pouvoir extraire de la vidéo et du son les débuts de phrases pour pouvoir caler le sous-titre sur la vidéo.

API, Web Scraping Il faut aller télécharger une liste de bons sous-titres potentiels.

Langage naturel & Apprentissage Il faut se documenter sur les façons efficaces de classer les mots d'un sous-titre en plusieurs catégories de difficulté en fonction de dictionnaire & autre base de données et surtout des mots que l'utilisateur a déjà rencontrés lors de visionnages passés.

1. On ne propose ici qu'une idée ; la fiche est à compléter.

3 Remarques

Pour aller plus loin, de nombreuses pistes sont à explorer en faisant un peu l'état de l'art des outils existants.

Références

[KM14] Geza Kovacs and Robert C Miller. Smart subtitles for vocabulary learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 853–862. ACM, 2014.

[sma] <http://www.smartsubs.fr/>.