

## Projet PCA "follow"

Stéphane Rubini

23 octobre 2018

Le but du projet est de développer un logiciel qui sera baptisé *follow*, pour le suivi *a posteriori* des modifications documentaires. Ce logiciel a pour objectif d'aider à la comparaison de deux textes, normalement relativement proches, en faisant apparaître leurs différences. Un alignement des deux textes est recherché afin de minimiser le nombre d'opérations à réaliser pour passer d'un texte à l'autre. Par opération, on entend des opérations d'édition de type *remplacement*, *insertion* ou *suppression* de mots. Le logiciel ne fait pas apparaître les modifications de la mise en page (retour à la ligne, blanc, tabulation). En revanche, la ponctuation et les caractères spéciaux sont pris en compte.

La figure 1 montre l'interface graphique du logiciel.

Les fonctionnalités du logiciel sont les suivantes :

- Chargement du texte à comparer au texte de référence, qualifié de "nouveau" par la suite, et affichage dans la zone de texte à droite de la fenêtre. Le nom du fichier chargé apparaît en dessous de cette zone.

Le nom du fichier est fourni sur la ligne de commande (second argument) au moment du lancement du logiciel, ou obtenu à partir d'un sélecteur de fichiers déclenché par le menu déroulant *Fichier*, option *Ouvrir*.

- Chargement du texte de référence.

Si le nouveau texte n'est pas encore chargé, alors le texte de référence est affiché tel quel dans la zone de texte à gauche de la fenêtre. Dans le cas contraire, les différences avec le nouveau texte sont affichées dans la zone de texte à gauche par une présentation en couleur :

- un mot affiché en rouge apparaît dans la référence, mais a été supprimé dans le nouveau texte,
- un mot affiché en vert a été inséré dans le nouveau texte,
- un couple de mots affiché en rouge et bleu indique le remplacement du mot rouge dans le texte de référence par le mot bleu dans le nouveau texte.

La mise en page est celle du nouveau texte. Le nom du fichier de référence apparaît en dessous de la zone de texte.

Le nom du fichier est fourni sur la ligne de commande (premier argument) au moment du lancement du logiciel, ou obtenu à partir d'un sélecteur de fichiers déclenché par le menu déroulant *Fichier*, option *Ouvrir référence*.

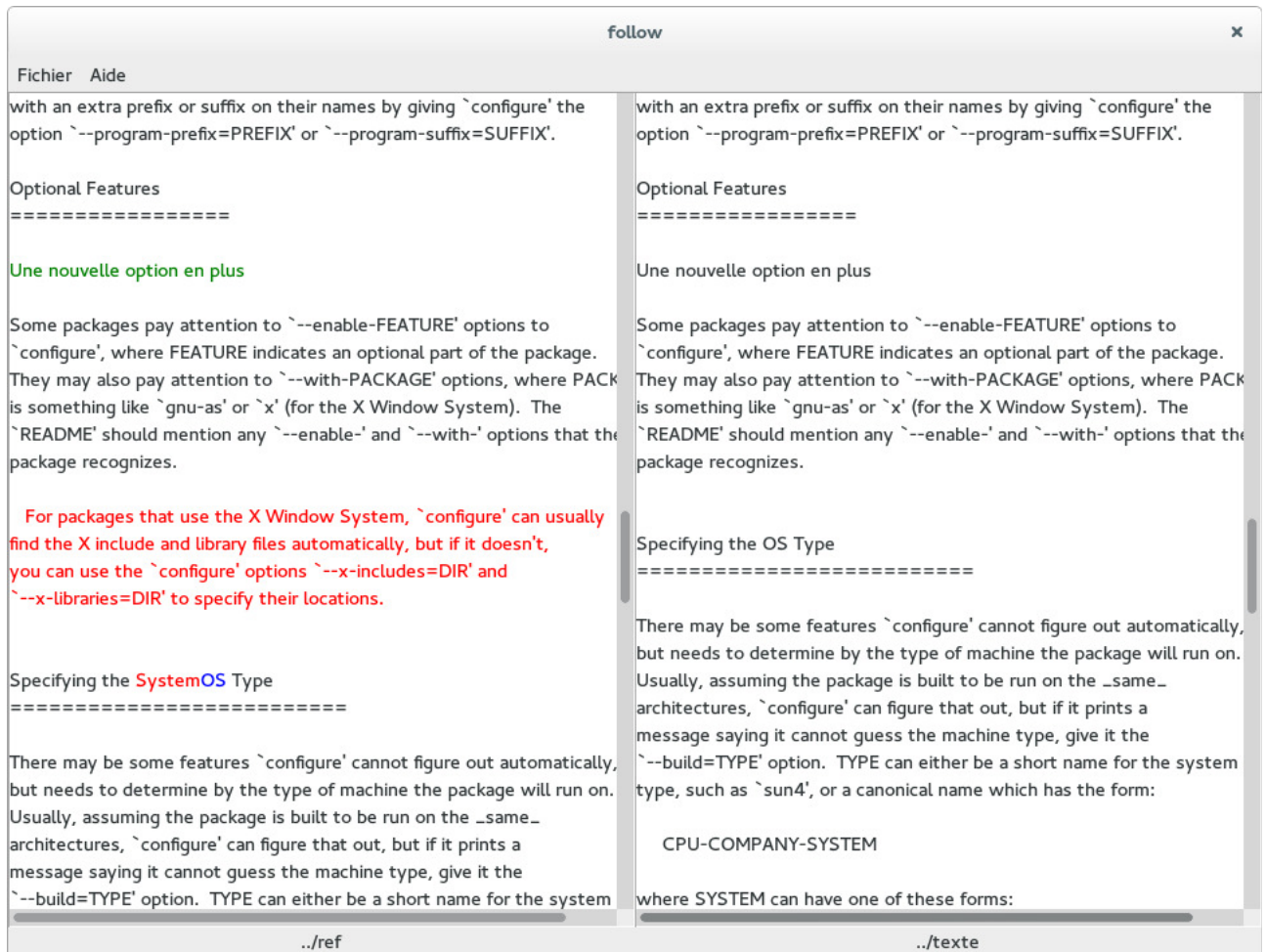


FIGURE 1 – Interface graphique du logiciel follow

- Affichage d'une fenêtre "A propos" qui décrit le nom, la version et l'auteur du logiciel. Son affichage est déclenché par le menu déroulant *Aide*, option *À propos*. La fenêtre est fermée par l'activation d'un bouton *OK*.
- Arrêt du logiciel, déclenché par le menu déroulant *Fichier*, option *Quitter* ou par un clic sur le bouton de destruction de la fenêtre graphique.

## Planning prévisionnel, organisation et notation

Le planning prévisionnel de déroulement du projet est décrit dans le tableau ci-dessous.

Séance (2H)	Taches	Durée
1 et 2	Présentation du projet Développement et test du module <i>liste</i>	30 min 3H
3 et 4	Table de hashage (cours) Développement et test du module <i>strhash</i>	30 min 3H30
5 et 6	Construction de la représentation interne des textes Algorithme de recherche des PLSC (cours) Recherche de l'alignement des textes	1H 30 min 4H00
7 et 8	GTK, présentation Conception de l'IHM	30 min 2H
9, 10 et 11	Intégration de l'ensemble	6H
12	Démonstration finale	2H
	Total	24H

Les trois premières étapes de développement seront validées par un test unitaire.

De plus, chaque étape fera l'objet d'une évaluation par l'enseignant. Le passage à l'étape suivante du développement est conditionné à une évaluation positive de la part de ce dernier. Il est précisé que vous devez présenter à cette occasion une mise en œuvre que vous avez vous-même développée.

Pour les 2 premières étapes uniquement, et dans le cas où vous considérez que vous y avez consacré trop de temps, il est possible "d'acheter"<sup>1</sup> la mise en œuvre du module à votre enseignant. Cet achat aura évidemment un coût au niveau de la notation de l'étape correspondante.

---

1. symboliquement bien entendu !