

Distant Reading I: hacker les humanités

Hacker les humanités ?

Simon Gabay

Genève, 22 septembre 2020

Utiliser un ordinateur

Nous utilisons quotidiennement des ordinateurs, mais que pourcentage de l'ordinateur laissons-nous inexploitée?

Ce cours propose de montrer la partie immergée de l'iceberg, et de montrer non seulement de révéler la puissance des ordinateurs, mais aussi de mettre cette dernière au profit des sciences humaines.

Regex

Les regex, ou "expression régulière" (*regular expressions*), est une chaîne de caractères, qui décrit, selon une syntaxe précise, un ensemble de chaînes de caractères possibles.

Elles permettent de faire des recherches simples et efficaces dans les données, par exemple:

- `\w` correspond à n'importe quel caractère alphanumérique
- `[abc]` correspond à l'une des trois lettres `a`, `b`, `c`.
- `a$` correspond à un `a` suivi d'une fin de ligne.

Interface en ligne de commande

Une interface en ligne de commande ou CLI (*command line interface*) est une manière d'utiliser l'ordinateur différente de l'utilisation des fenêtres (environnement graphique). On navigue entre les dossiers et on lance les scripts en mode texte (environnement console).

Quoique plus fastidieux au début, ce mode d'utilisation des machines ouvre des possibilités nouvelles, car il est le moyen privilégié d'interaction avec sa machine des développeurs. Nombre de logiciels conçus pour des micro-usages et l'ajout d'une interface de type logicielle rendrait le développement trop fastidieux.

Git

Avec les lignes de commandes, nous serons en mesure d'utiliser le logiciel Git, qui permet la gestion des versions des scripts (version 1, 1.1, 1.2, 3.4...) mais aussi la publication, le partage et le commentaire du code en ligne via des sites comme GitHub ou GitLab.

À cette occasion, nous aborderons rapidement les questions de la documentation des scripts.

Script shell

La programmation shell consiste à utiliser un minilangage de programmation, bien moins puissant que Python, C..., mais extrêmement efficace pour les petits besoins du quotidien.

Les scripts shell reprennent des commandes systèmes utilisables en ligne de commande comme `touch` (créer un fichier), `mkdir` (créer un dossier), `grep` (chercher avec une expression régulière)...

Python

Python est un langage de programmation qui offre un compromis parfait entre puissance et simplicité. Il permet de manipuler, créer, transformer des fichiers de tous ordres (texte, image, tableur, web...). Il permet aussi de créer des visualisations simples et de créer des applications web.

Il nous servira de base à l'introduction de l'algorithmie, et donc à la programmation de scripts plus conséquents que ceux en bash. Un des objectif pourrait être la manipulation simple de fichiers texte, image et web (pour du scrapping par exemple).

On découvrira les conditions, les boucles, les opérateurs...