

Catégorisation de données textuelles multilingues pour la détection d'opinions

Encadré par Rogovschi Nicoleta



L 3 A A 1

MANUEL D'INSTALLATION

Référence du document : L3AA1_2021_MI

Version du document : 1.00

Date du document : 18/04/2021

Auteurs : Laforge Johan, Tanriverdi Messie, Thay Jacky, Trividic Yann

Validé par : Trividic Yann

Validé le : 18/04/2021

Soumis le : 18/04/2021

Type de diffusion : Document électronique (PDF)

Confidentialité : Standard – Étudiants et corps enseignant de l'UFR Mathématiques et Informatique de l'Université de Paris

Mots-clés : manuel d'installation, configuration, installation, descartes, université, paris, math-info, licence, L3AA1, catégorisation, données, textuelles, multilingues, détection opinion, opinion mining, artificial intelligence, sentiment analysis, Python, nlp.

Table des matières

Introduction	3
Guide de lecture	3
2.1 Maîtrise d'œuvre et maîtrise d'ouvrage	3
2.1.1 Responsables	3
2.1.2 Personnels administratif et technique	4
Configuration minimum requise	5
Installation du logiciel	5
4.1. Ubuntu	5
4.2 Windows	7
4.3. Exécuter le programme à partir du code source avec un interpréteur Python	10
4.4 Compiler sa propre version du logiciel	11
Paramétrage du logiciel et clés d'API	12
Autres informations	12
Glossaire	13

1. Introduction

Ce document a pour but de définir les attentes du maître d'ouvrage envers le maître d'œuvre durant la réalisation du projet. Le non-respect de ces attentes pourra donc engendrer des pénalités. Par ailleurs, le cahier des charges constitue une pièce de références du contrat pour les deux parties (client et fournisseur) en levant toutes les ambiguïtés possibles. Cela permet donc de fournir une livraison au plus proche des attentes du client.

Les termes en gras dans ce texte disposent d'une entrée dans le glossaire, section 10 de ce document.

2. Guide de lecture

2.1 Maîtrise d'œuvre et maîtrise d'ouvrage

Les différents acteurs de la maîtrise d'œuvre et de la maîtrise d'ouvrage étant identiques, nous nous sommes permis dans ce document, par souci de concision, de fusionner ces différentes catégories.

2.1.1 Responsables

Rogovschi Nicoleta

Contact : nicoleta.rogovschi@u-paris.fr

Mme Rogovschi est enseignante-chercheuse à l'Université de Paris dans l'UFR de Mathématiques et d'Informatique. Elle est la maîtresse d'ouvrage, la maîtresse d'œuvre et l'encadrante principale de ce projet.

Janiszek David

Contact : david.janiszek@u-paris.fr

M. Janiszek est enseignant-chercheur à l'Université de Paris dans l'UFR de Mathématiques et d'Informatique, dont il est aussi le directeur.

Il pilote le cours "Projet tutoré" (IF06M030) dans lequel s'inscrit ce projet (voir section 4).

2.1.2 Personnels administratif et technique

Laforge Johan

Contact : laforgejohan7@gmail.com

Étudiant en 3^e année de licence MAGE à l'Université de Paris.

A intégré en 2017 l'UFR de Mathématiques et d'Informatique.

Développeur Python dans le cadre de ce projet.

Tanriverdi Messie

Contact : tanriverdi.messie@gmail.com

Étudiante en 3^e année de licence d'Informatique et Applications à l'Université de Paris.

A intégré en 2017 l'UFR de Mathématiques et d'Informatique.

Développeuse Python dans le cadre de ce projet.

Thay Jacky

Contact : jacky.thay@yahoo.com

Apprenant en 3^e année de licence d'Informatique et Applications à l'Université de Paris.

A intégré en 2020 l'UFR de Mathématiques et d'Informatique.

Développeur Python dans le cadre de ce projet.

Trividic Yann

Contact : yann.trividic@hotmail.fr

Étudiant en 3^e année de licence d'Informatique et Applications à l'Université de Paris.

A intégré en 2017 l'UFR de Mathématiques et d'Informatique.

Développeur Python dans le cadre de ce projet.

3. Configuration minimum requise

Processeur : Un processeur avec quatre cœurs et une fréquence de 2,5 GHz.

RAM : 2 Go de RAM disponible.

Système d'exploitation : Windows 10 64 bits / Ubuntu 18.04 64 bits ou supérieur.

Espace sur le disque dur : Sans compter l'installateur, il faut au minimum 1,5 Go de disponible pour une installation sous Ubuntu et 2,5 Go de disponible sous Windows 10.

Écran : Un affichage d'une résolution minimum de 1280x720 (**HD Ready**) pour afficher correctement l'application.

Navigateur Web : Une installation récente de Mozilla Firefox ou de Google Chrome.

Divers : Une connexion internet pour les recherches sur **Twitter** et **IMDb**, ainsi que pour la traduction automatique (**TA**) et la détection de langues.

La recherche locale sans TA, puis la classification par polarité sur des fichiers CSV ne demande pas de connexion internet.

4. Installation du logiciel

4.1. Ubuntu

Prérequis spécifiques : avoir installé une version 18.04 d'Ubuntu ou ultérieure en tant que système d'exploitation courant.

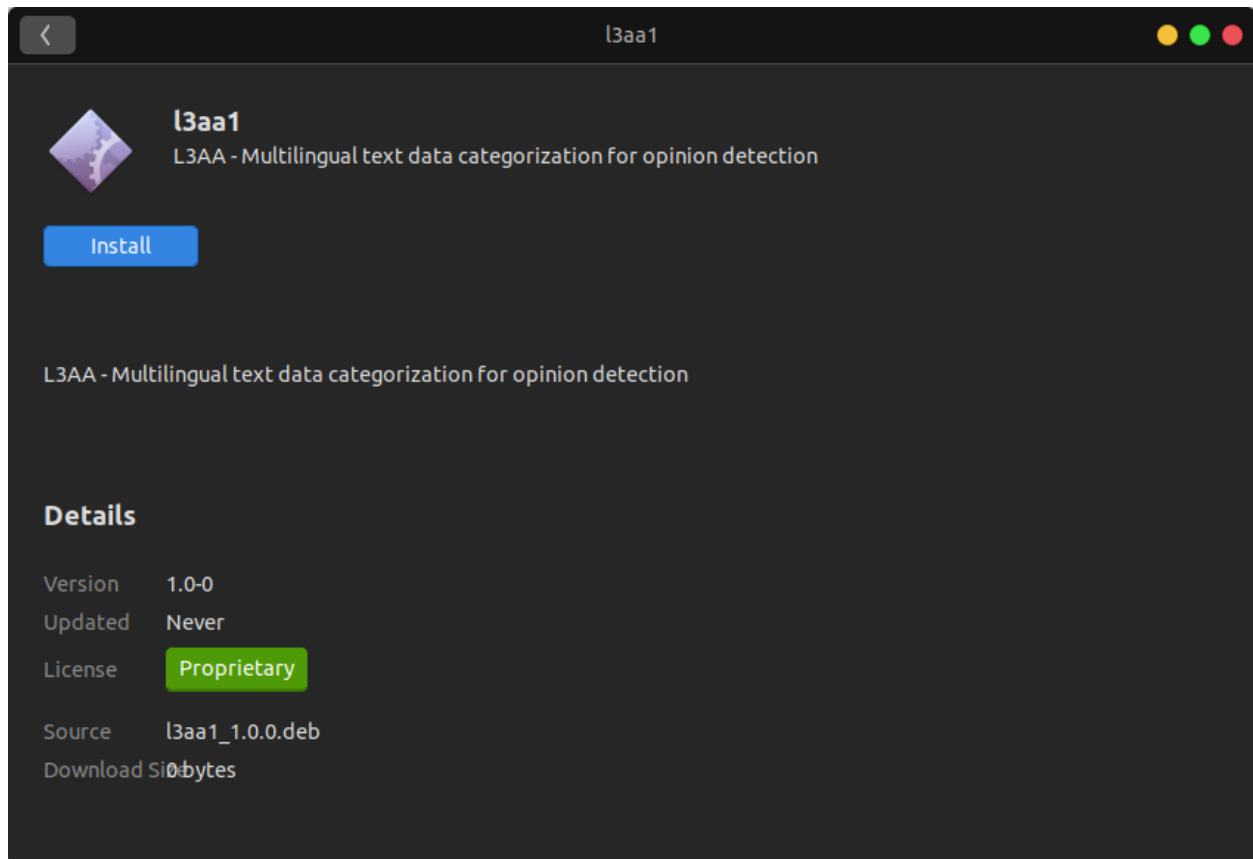
Pour installer le logiciel, commencez par télécharger la dernière version du logiciel (actuellement v1.0.0). Une fois le fichier téléchargé, deux options sont possibles :

- Ouvrez un terminal et rendez-vous dans le répertoire où est situé le paquetage Debian (fichier **DEB**). Une fois dans ce répertoire, veuillez exécuter la commande suivante :

```
dpkg -i l3aa1_1.0.0.deb
```

Une fois que le terminal vous redonne la main, si aucun message d'erreur n'est apparu, alors votre programme est installé.

- Effectuez un double-clic sur le fichier DEB. Un assistant d'installation s'ouvre, puis après chargement, il vous est possible d'installer le paquet en cliquant sur le bouton **Install**. Après un certain temps d'installation, le programme peut être lancé.



Avec la plupart des configurations Ubuntu, le programme n'a pas à être lancé avec les droits d'administrateurs et peut être lancé comme n'importe quel autre programme en recherchant **l3aa1** dans la barre de recherche de votre système d'exploitations. Cependant, dans certains environnements, il est possible que l'application nécessite les droits d'administrateurs pour être lancée. Pour ce faire, ouvrez un terminal et entrez la commande suivante :

```
sudo l3aa1
```

Le programme se lance.

Pour désinstaller le programme, vous pouvez ouvrir un terminal et exécuter la commande suivante :

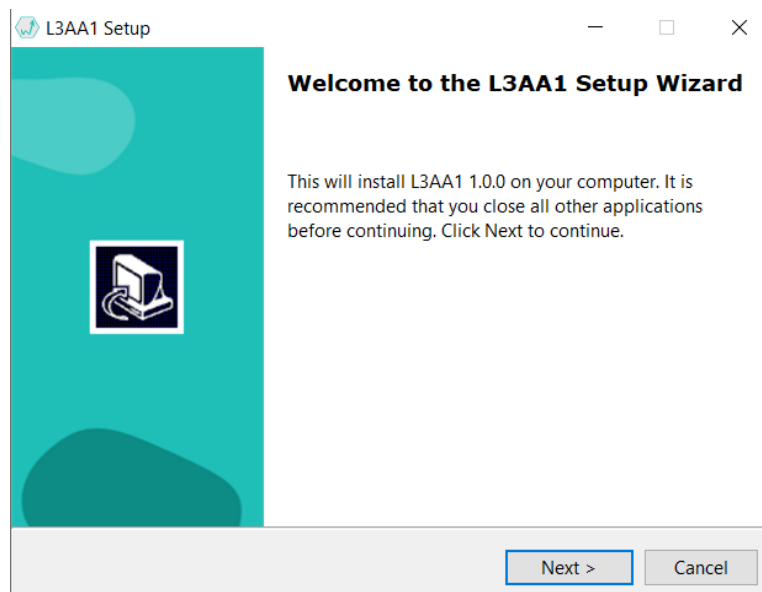
```
sudo apt-get remove l3aa1
```

Alternativement, il vous est possible de désinstaller le programme avec l'aide d'une interface graphique. Cherchez le programme l3aa1 dans le dashboard de votre système d'exploitation, faites un clic droit sur l'icône du programme, puis Show details. La fenêtre du gestionnaire de paquets Debian s'ouvre à nouveau, et il vous est possible depuis celle-ci de désinstaller votre paquet.

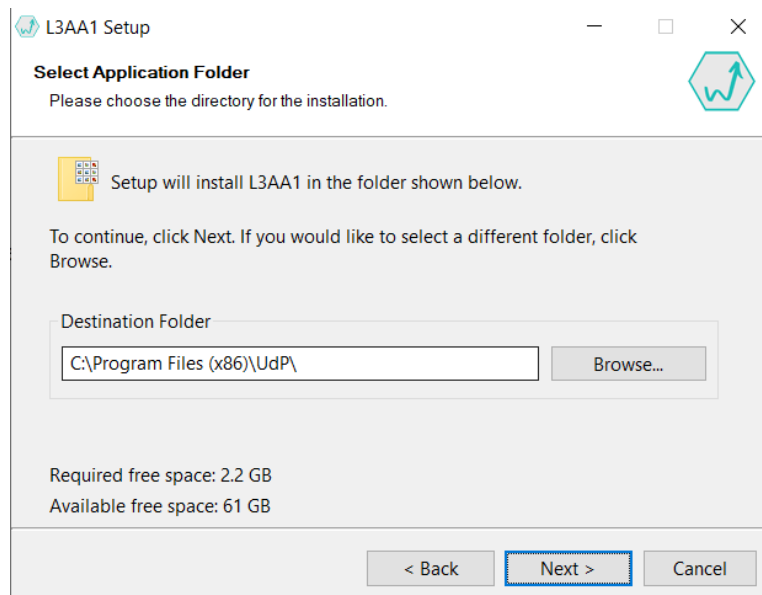
4.2 Windows

Attention : cette application a été développée sous Windows 10, sa compatibilité avec des versions antérieures de Windows ne peut être garantie.

Pour installer le logiciel, commencez par télécharger la dernière version du logiciel (actuellement v1.0.0). Une fois le fichier téléchargé, veuillez lancer l'installeur :

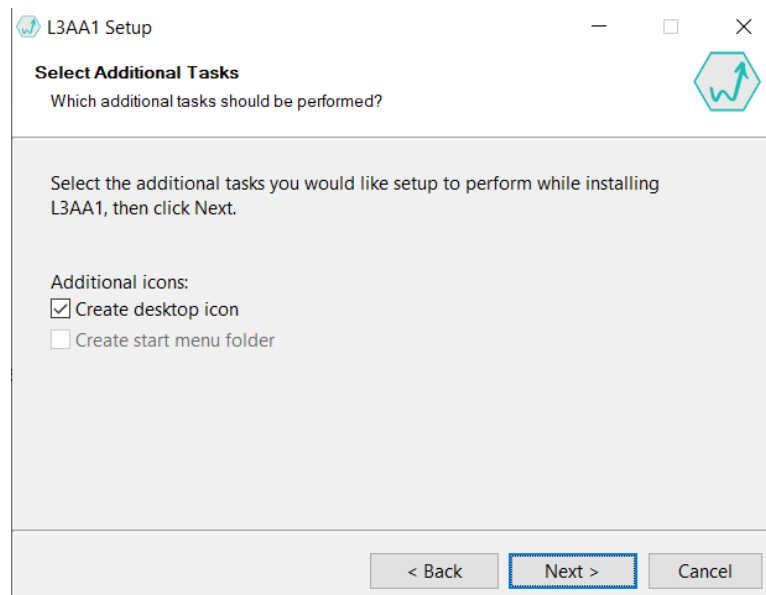


Cliquez sur Next > :

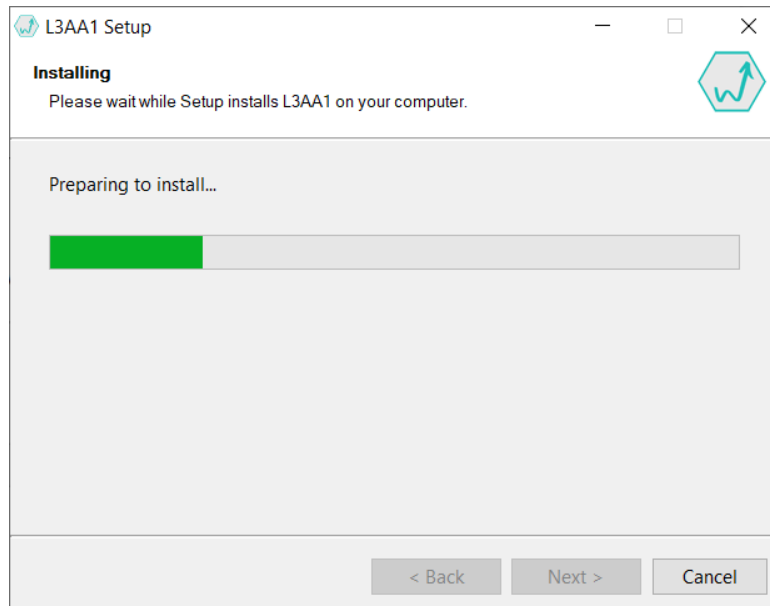


Environ 2,2 Go d'espace libre sur votre disque dur seront nécessaires pour installer l'application.

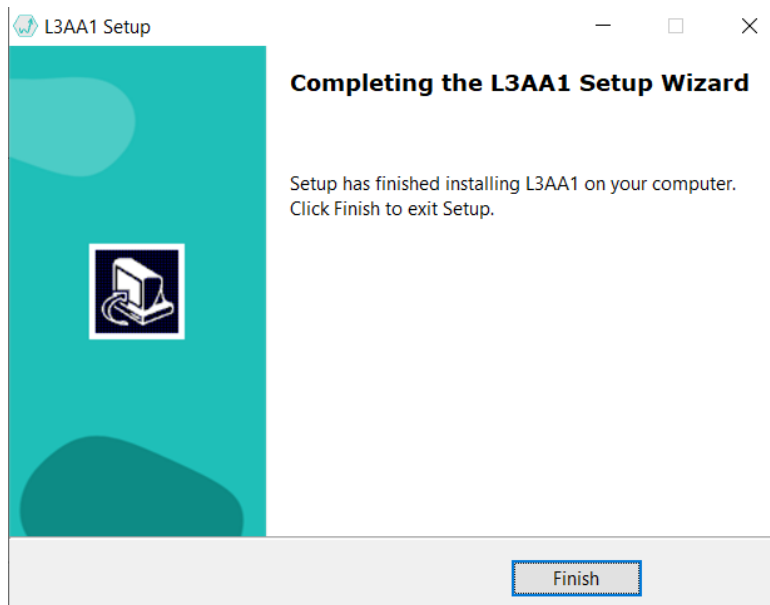
Modifiez le chemin de l'installation si cela est nécessaire et cliquez sur Next > :



Cochez la case Create desktop icon si vous souhaitez avoir un raccourci sur votre bureau et cliquez sur Next > :



L'installation de l'application commencera une fois que vos choix auront été enregistrés.



Une fois l'installation terminée, une fenêtre de confirmation s'affichera.

Pour la désinstallation, un fichier `Uninstall_L3AA1.exe` est disponible dans le chemin d'installation que vous avez indiqué précédemment.

4.3. Exécuter le programme à partir du code source avec un interpréteur Python

Prérequis spécifiques : avoir une version 3.7 de Python ou ultérieure avec `pip` et `setuptools` installés (commande pour installer `setuptools` : `pip install -U pip setuptools`, potentiellement à adapter selon votre environnement).

Attention : avec `python3.9` et les versions ultérieures, il peut être nécessaire de mettre à jour votre version de `pip`. Pour cela, entrez la commande `python3.9 -m pip install --upgrade pip`.

Attention : ce tutoriel n'est complet que pour Ubuntu et Windows. Sur macOS, des étapes supplémentaires peuvent être nécessaires pour mener à bien le tutoriel en paramétrant plus en détails votre `pip` ou en utilisant Homebrew (le fichier `brew.txt` sera alors à utiliser en remplacement du fichier `requirements.txt`).

Téléchargez le dossier `src`. Une fois téléchargé, vous pouvez renommer ce dossier et le placer où vous le souhaitez dans votre ordinateur. Ouvrez un terminal et rendez vous à l'intérieur du dossier téléchargé.

Dans un premier temps, il faut installer les dépendances liées au programme. Celles-ci sont listées dans le fichier `requirements.txt`, situé dans le dossier `setup`. Pour installer ces dépendances, veuillez entrer l'une des commandes suivantes dans un terminal ouvert sur le dossier `setup` :

- `pip3 install -r requirements.txt`

Si la version par défaut utilisée par votre OS est ultérieure à 3.7 ou 3.7.

- `python3.x -m pip install -r requirements.txt`

Si votre version par défaut est antérieure à 3.7 et que vous disposez d'une version 3.x >= 3.7 installée sur votre machine.

Attention : Les versions de PyQt5 (la bibliothèque de gestion d'interfaces graphiques utilisée dans le cadre du projet) ne sont pas toutes compatibles avec l'installateur `pip`. Si vous

rencontrez des problèmes au moment de l'installation de la bibliothèque PyQt5, essayez de l'installer séparément, et mettez à jour le fichier `requirements.txt` en supprimant la ligne destinée à PyQt5.

Une fois que le terminal vous a rendu la main, vous pouvez lancer le programme en utilisant la même version de Python que celle utilisée pour installer les dépendances, à savoir en exécutant :

`python 133a1.py` ou `python3.x 133a1.py` selon votre environnement et votre interpréteur.

Le programme se lance à la suite de cette dernière commande. Lors du premier démarrage, il est possible que des fichiers aient à être téléchargés, la fenêtre peut donc mettre du temps avant d'apparaître sur votre écran.

Si vous souhaitez vous approprier ou modifier ce programme, vous le pouvez en respectant les conditions décrites dans la licence GPLv3. Une documentation au format hmtl est disponible dans le dossier `src`.

4.4 Compiler sa propre version du logiciel

Attention : Cette section est destinée aux personnes ayant déjà un niveau d'expérience avancé en informatique. Il est possible que des spécificités propres à votre environnement de travail impliquent de modifier le fichier **SPEC** ou l'énoncé de la commande utilisés dans ce tutoriel. À moins que cela ne soit nécessaire pour vous, nous vous recommandons d'utiliser l'un des modes d'installation énoncés dans l'une des sections précédentes, ceux-ci étant beaucoup simples.

Prérequis spécifiques : **1)** avoir pu mener à terme toutes les étapes de la section 4.3 avec succès. **2)** avoir installé sur sa machine une version compatible du module `pyinstaller`¹ avec la version de Python utilisée lors du prérequis 1).

Ouvrez un terminal et rendez-vous dans le dossier `setup` en haut de l'arborescence du projet, là où figure le fichier `13aa1.spec`. Exécutez la commande suivante :

```
pyinstaller 13aa1.spec
```

¹ <https://pyinstaller.readthedocs.io/en/stable/installation.html>

Une fois que le terminal vous a rendu la main, si la compilation a pu s'effectuer sans problème, alors vous pouvez lancer le programme. L'exécutable varie selon les systèmes d'exploitation :

- **Ubuntu** : exécuter `l3aa1` (dans le dossier `./dist/l3aa1`)
- **Windows** : exécuter `l3aa1.exe` (dans le dossier `.\dist\l3aa1`)

5. Paramétrage du logiciel et clés d'API

Une fois le logiciel lancé, aucun paramétrage particulier n'est nécessaire. Toutefois, si vous avez décidé d'installer le logiciel en utilisant le code source du programme, alors il vous est possible de changer les clés d'API actuellement utilisées dans le logiciel pour vos propres clés d'API. Cela peut se faire en modifiant le fichier `config.py` situé dans le dossier `setup`. Ainsi, les questions de limitations du logiciel concernant l'extraction de données Twitter et la détection de langues ne se poseront plus.

6. Autres informations

Ce programme étant distribué sous licence GLPv3, il vous est possible d'utiliser et de modifier ce programme dans le respect des conditions citées par cette licence.

7. Glossaire

API : en informatique, une interface de programmation d'application ou interface de programmation applicative (souvent désignée par le terme API pour Application Programming Interface) est un ensemble normalisé de classes, de méthodes, de fonctions et de constantes qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels (source : Wikipedia).

CSV : *Comma-separated values*, connu sous le sigle CSV, est un format texte ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules (source : Wikipedia).

DEB : Un fichier DEB est un package permettant d'installer une application sous les systèmes Linux Debian. La distribution Debian propose un outil de gestion de package permettant d'automatiser l'installation, la configuration et la mise à jour des logiciels installés par ce biais (source : www.commentcamarche.net).

HD Ready : « HD Ready » (en français « prêt pour la haute définition ») est la désignation commerciale (abandonnée) relative à un label et à un visuel (logotype), lancé au mois de juin 2005; Conçu pour identifier un téléviseur ou équipement TV compatible avec les signaux de télévision HD. L'appareil « HD Ready » ne permet pas de restituer la pleine résolution des images à haute définition mais adapte le signal à la résolution moyenne affichée à l'écran (généralement 720 lignes maximum) (source : Wikipedia).

IMDb : Internet Movie Database (littéralement « Base de données cinématographiques d'Internet »), abrégé en IMDb, est une base de données en ligne sur le cinéma mondial, sur la télévision, et plus secondairement les jeux vidéo (source : Wikipedia).

sentiment analysis : voir *opinion mining*.

SPEC : Fichier de spécification utilisé par pyinstaller pour paramétrer pyinstaller en fonction de la structure du code source du projet.

TA : La traduction automatique désigne la traduction brute d'un texte entièrement réalisée par un ou plusieurs programmes informatiques. Un traducteur humain n'intervient pas pour corriger les erreurs du texte au cours de la traduction, mais seulement avant et/ou après. On la distingue de la traduction assistée par ordinateur où la traduction est en partie manuelle, éventuellement de façon interactive avec la machine (source : Wikipedia).

Twitter : réseau social de microblogage géré par l'entreprise Twitter Inc. Il permet à un utilisateur d'envoyer gratuitement des micromessages, appelés tweets, sur internet, par messagerie instantanée ou par SMS. Ces messages sont limités à 280 caractères (source : Wikipedia).