VisaTM étude

Recensement d'outils
de fouille de textes

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
ABBYY Solutions	ABBYY technologies and platforms for document recognition, data capture, and language processing.	Commercial	Classification de textes ; Reconnaissance d'entités nommées ; Découverte de connaissances ; Traduction automatique ; Recherche d'information ;	Non	https://www.abbyy.com/en-eu/	Russie ;
ABNER	ABNER is a software tool for molecular biology text analysis.	Libre	Reconnaissance d'entités nommées ;	Oui	http://pages.cs.wisc.edu/~bsettles/abner/	États-Unis ;
Abzooba	Social media and text analytics software	Commercial	Analyse de sentiments ; Classification ; Reconnaissance d'entités nommées ;	Non	http://www.abzooba.com/	États-Unis ;
ADAM	Data Mining and Image Processing Toolkits	Libre	Classification ; Clustering ; Reconnaissance de forme ; Règles d'association ; Optimisation ; Traitement d'images ;	Non	http://projects.itsc.uah.edu/datamining/adam/	États-Unis ;
AdaMSoft	ADaMSoft is a free and Open-Source System for Data Management, Data and Web Mining, statistical Analyis and more.	Libre	Classification ; Clustering ; Analyse de régression ;	Non	http://adamsoft.sourceforge.net/	Italie ;
ai-one	Transforms big data into opportunity using machine learning that mimics the biological brain's ability to find patterns and relationships.	Commercial	Apprentissage automatique ;	Non	http://www.ai-one.com/	États-Unis ; Suisse ; Allemagne ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Aika	Java library that automatically extracts and annotates semantic information into text	Libre	Annotation ; Désambiguïsation ; Catégorisation de textes ; Reconnaissance d'entités nommées ; Extraction d'information ;	Non	http://www.aika-software.org	Allemagne ;
Alceste	logiciel d'analyse de données textuelles, ou statistique textuelle	Commercial	Classification ;	Non	http://www.image-zafar.com/Logiciel.html	France ;
AllenNLP	An open-source NLP research library, built on PyTorch.	Libre	Etiquetage sémantique ; Reconnaissance d'entités nommées ; Q&A ; Résolution de coréférence ; Textual entailment ; Constituency parsing ;	Non	http://allennlp.org/	États-Unis ;
Alteryx Project Edition	Predictive for Project Edition	Commercial	Prétraitement ; Apprentissage automatique ;	Non	https://www.alteryx.com/fr/predictive-project-edition	États-Unis ;
Alveo	Alveo connects HCS (Human Communication Science) researchers, their desks, computers, labs, and universities and accelerates HCS research to produce emergent knowledge that comes from novel application of previously unshared tools to analyse previously difficult to access data sets.	?	Recherche d'informations ; Reconnaissance de la parole ; Annotation ;	Non	http://alveo.edu.au/	Australie ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Alvis NLP	A pipeline framework for Natural Language Processing	Libre	Annotation ; Classification ; Clustering ; Q&A ; Traduction ; Recherche d'information ; Analyse de sentiments ; Analyse d'opinions ; Reconnaissance d'entités nommées ; Racinisation ; PoS tagging ;	Oui	http://www.quaero.org/module_technologique/alvis-nlp-alvis-natural-language-processing/	France ; Allemagne ;
Amazon Comprehend	Détection d'informations et de relations dans un texte	Commercial	Extraction de phrases clés ; Analyse de sentiments ; Analyse syntaxique ; Reconnaissance d'entités nommées ; Extraction d'informations ; Détection de la langue ; Classification ; Topic modeling ;	Non	https://aws.amazon.com/comprehend/	États-Unis ;
AMI	Intégrateur de solutions logicielles de pointe pour la Cybersécurité, la Cyber Intelligence, la Veille Stratégique et le Traitement Automatique de la Parole.	Commercial	Traitement automatique de la parole ; Recherche d'informations ; Extraction de mots clés ; Annotation ;	Non	https://www.bertin-it.com/	France ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Anaconda	Python data science platform	Libre/ Commercial	Apprentissage automatique ;	Non	https://www.anaconda.com/	États-Unis ;
Analec	Annotation et analyse de corpus écrits	Libre	Annotation ;	Non	http://www.lattice.cnrs.fr/ Telecharger-Analec	France ;
Annomarket	Cloud-based Text Annotation	Commercial	Annotation ;	Oui	https://github.com/annomarket/	?
AntConc	Freeware text analysis and concordance tool kit	Libre	Concordancier ;	Non	http://www.laurenceanthony.net/ software/antconc/	Japon ;
Apache cTAKES	Natural language processing system for extraction of information from electronic medical record clinical free-text.	Libre	Extraction d'information ; Annotation ;	Non	http://ctakes.apache.org/	États-Unis ;
Apache Mahout	Environment for quickly creating scalable performant machine learning applications.	Libre	Architecture logicielle ; Analyse régression ; Clustering ;	Non	http://mahout.apache.org/	États-Unis ;
Apache OpenNLP	The Apache OpenNLP library is a machine learning based toolkit for the processing of natural language text.	Libre	Apprentissage automatique ; Lemmatisation ; Parsing ; Chunking ; Tokenisation ; PoS tagging ; Sentence splitting ; Reconnaissance d'entités nommées ; Résolution de coréférence ; Détection de la langue ; Classification ;	Oui	https://opennlp.apache.org/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Apache UIMA	Component software architecture for the development, discovery, composition, and deployment of multi-modal analytics for the analysis of unstructured information and integration with search technologies.	Libre	Architecture logicielle ;	Oui	https://uima.apache.org/	États-Unis ;
Argo	Argo is a workbench for building and running text-analysis solutions. It facilitates the development of custom workflows from a selection of elementary analytics.	Libre	Annotation ; Recherche d'informations ; Reconnaissance d'entités nommées ;	Non	http://argo.nactem.ac.uk/	Royaume-Uni ;
Ascribe	Accelerate ROI via our surveys, vast sample and advanced text analytics	Commercial	Classification ; Analyse d'opinions ; Analyse de sentiments ; Clustering ; Extraction d'informations ;	Non	https://goascribe.com/	États-Unis ;
ats	Regression and clustering analysis	Libre	Clustering ; Analyse de régression ;	Non	http://www.mepx.org/	?
Averbis	Averbis provides leading text mining and machine learning solutions for your business. We convert text into information, automate cognitive processes, and make meaningful predictions.	Commercial	Découverte de connaissances ; Extraction terminologique ; Reconnaissance d'entités nommées ; Classification de textes ; Analyse de sentiments ; Analyse d'opinions ; Recherche d'informations ;	Non	https://averbis.com/en/	Allemagne ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Aylien	AI-driven content analysis solutions that bring the power of NLP to the masses. We help developers, data scientists, and marketers understand human-generated textual content at scale.	Commercial	Analyse de sentiments ; Classifiation ; Résumé automatique ; Reconnaissance d'entités nommées ; Extraction d'informations ; Annotation ;	Non	https://aylien.com/	Irlande ;
Babel X	Babel X® is a multi-lingual, geo-enabled, text-analytics, social media and web-monitoring platform designed to meet the needs of our customers by fully leveraging publicly available information in this era of overwhelming quantities of geographically diverse, multi-lingual data.	Commercial	Analyse de sentiments ; Clustering ; Reconnaissance d'entités nommées ; Extraction de relations ; Recherche d'informations ;	Non	https://www.babelstreet.com/	États-Unis ;
BANNER	Named entity recognition system	Libre	Reconnaissance d'entités nommées ;	Oui	http://banner.sourceforge.net/	États-Unis ;
BioCreative	BioCreative: Critical Assessment of Information Extraction in Biology is a community-wide effort for evaluating text mining and information extraction systems applied to the biological domain	Libre	Reconnaissance d'entités nommées ; Extraction de relations ; Annotation ;	Oui	http://www.biocreative.org/events/biocreative-v/CFP/	?
BioLemmatizer	The BioLemmatizer is a domain-specific lemmatization tool for the morphological analysis of biomedical literature.	Libre	Lemmatisation ;	Non	http://biolemmatizer.sourceforge.net/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
BioNLP	BioNLP is an initiative by the Center for Computational Pharmacology at the University of Colorado Denver to create and distribute code, software, and data for applying natural language processing techniques to biomedical texts.	Libre	Parsing ; Lemmatisation ; Annotation ; Classification de textes ; Extraction d'informations ;	Oui	http://bionlp.sourceforge.net/	États-Unis ;
BioTagger	These pages describe briefly Penn's BioTagger software suite. Currently the tagger supports three types of entities: gene entities, genomic variations entities and malignancy type entities.	Libre	Reconnaissance d'entités nommées	Non	https://www.seas.upenn.edu/~strctlrn/BioTagger/BioTagger.html	États-Unis ;
Biotex	Biomedical term extraction	Libre	Extraction terminologique ;	Non	http://tubo.lirmm.fr/biotex/index.jsp	France ;
Bitext	NLP Plateform	Commercial	Lemmatisation ; PoS tagging ; Détection de la langue ; Extraction de phrases ; Reconnaissance d'entités nommées ; Analyse de sentiments ;	Non	https://www.bitext.com/	Espagne ;
Bluima	Natural language processing toolkit for neuroscience	Libre	Reconnaissance d'entités nommées ;	Non	https://github.com/BlueBrain/bluima	Suisse ;
Brainspace	Brainspace creates breakthrough machine learning software that intelligently detects and relates unique phrases in massive unstructured datasets	Commercial	Clustering ;	Non	https://www.brainspace.com/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Brat	Online environment for collaborative text annotation.	Libre	Extraction d'information ; Annotation ; Reconnaissance d'entités nommées ;	Oui	http://brat.nlplab.org/	Japon ; Royaume-Uni ;
Bulstem	Stemming for Bulgarian	Libre	Racinisation ;	Oui	https://github.com/peio/PyBulStem	États-Unis ;
Caffe2	A new lightweight, modular, and scalable deep learning framework	Libre	Apprentissage profond ;	Non	https://caffe2.ai	États-Unis ;
Calliope	Logiciel d'analyse des tendances et de "fouille de textes"	Libre/ Support technique et formation payants	Extraction terminologique ; Analyse de tendances ; Analyse des co-occurrences ;	Non	https://www.calliope-textmining.com/	France ;
Canopy	Enthought Canopy provides a proven scientific and analytic Python package distribution plus key integrated tools for iterative data analysis, data visualization, and application development. Users have the ability to extend and innovate with scripting and open platform APIs, driving the creation and sharing of innovative workflows, tools, and applications.	Commercial	Appentissage automatique ;	Non	https://store.enthought.com/basket/	États-Unis ;
Carrot2	Carrot2 organizes your search results into topics. With an instant overview of what's available, you will quickly find what you're looking for.	Libre	Clustering ; Recherche d'informations ;	Non	http://search.carrot2.org/stable/search	Pologne ; Royaume-Uni ;
Chemicalize	Chemicalize is a powerful online platform for chemical calculations, search, and text processing.	Libre	Annotation ; Reconnaissance d'entités nommées ;	Non	https://chemicalize.com/welcome	Hongrie ;
CiteSpace	Visualizing Patterns and Trends in Scientific Literature	Libre	Clustering ;	Non	http://cluster.cis.drexel.edu/~cchen/citespace/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Clarabridge CX Suite	CX Analytics is the backbone of the world's most complex Customer Experience Management Programs, providing the industry's most accurate Natural Language Processing (NLP), sentiment and data categorization, making issues transparent—and next steps clear. CX social : Social listening, rapid social media engagement, and social media analytics that empower teams of all sizes to wow customers and have a big impact.	Commercial	Clustering ; Analyse de sentiments ; Parsing ; Extraction de relations ;	Non	https://www.clarabridge.com/	États-Unis ;
ClearNLP	The ClearNLP project provides fast and robust NLP components implemented in Java	Libre	Tokenisation ; Sentence splitting ; Parsing ; Etiquetage de rôles sémantiques ; PoS tagging ;	Oui	http://clearnlp.wikispaces.com/	États-Unis ;
ClearTK	ClearTK is a framework for developing machine learning and natural language processing components within the Apache Unstructured Information Management Architecture.	Libre	Classification ; Clustering ; Parsing ; Racinisation ; Tokenisation ; PoS tagging ; Feature extraction ;	Non	https://cleartk.github.io/cleartk/about.html	États-Unis ;
Clementine	Clementine packages a number of tools with a GUI which simplifies the process of performing a data mining project. In particular the Clementine workbench supports a number of data mining algorithms through a simple linked node interface supporting the entire business process of data mining using the CRISP-DM model.	Commercial	Clustering ; Classification ; Analyse de régression ;	Non	http://datamining.togaware.com/survivor/Summary20.html	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
CLUTO	CLUTO is a software package for clustering low- and high-dimensional datasets and for analyzing the characteristics of the various clusters. CLUTO is well-suited for clustering data sets arising in many diverse application areas including information retrieval, customer purchasing transactions, web, GIS, science, and biology.	Libre	Clustering ;	Non	http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview	États-Unis ;
CMSR Data Miner	Integrated environment for predictive modeling, segmentation, data visualization, statistical data analysis, and rule-based model evaluation	Libre	Classification ; Clustering ;	Non	http://www.roselladb.com/starprobe.htm	Australie ;
CogComp-NLP	CogComp-NLP provides a suite of state-of-the-art Natural Language Processing (NLP) tools that allows you to annotate plain text inputs.	Libre	Lemmatisation ; PoS tagging ; Parsing ; Extraction de relations ; Reconnaissance d'entités nommées ; Etiquetage de rôles sémantiques ; Analyse de co-référence ;	Non	http://nlp.cogcomp.org/	États-Unis ;
Cogito	Multilingual text analytics, cognitive technology software that understands the meaning of words in context.	Commercial	Reconnaissance d'entités nommées ; Classification ; Recherce d'informations ; Désambiguisation ; Extraction de mots clés ; Découverte de connaissances ;	Non	http://www.expertsystem.com/fr/	Italie ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Cognitive Computation Group NLP Tools		Libre	Reconnaissance d'entités nommées ; PoS tagging ; Chunking ; Lemmatisation ; Etiquetage de rôles sémantiques ; Résolution de coréférence ;	Non	http://cogcomp.org/page/software/	États-Unis ;
Coheris SPAD integral	Pour l'analyse de données et le traitement de toute l'information, notamment l'information textuelle.	Commercial	Tokenisation ; Lemmatisation ; Classification ; Clustering ; Arbre de décision	Non	https://www.coheris.com/produits/analytics/logiciel-data-mining/analyse-de-donnees/	France ;
ConcQuest	Concordancier dédié à la recherche d'expressions complexes à travers des corpus monolingues et multilingues alignés.	Libre	Concordancier ; Rercherche d'informations ;	Non	http://olivier.kraif.u-grenoble3.fr/index.php?option=com_content&task=view&id=36&Itemid=55	France ;
Content Annotation Manager	Create relevant metadata based on vocabularies and rules	Commercial	Annotation ; Reconnaissance d'entités nommées ;	Non	http://www.mondeca.com/content-annotation-manager/	France ;
ContentMine	Text and data mining tools	Libre	Extraction de connaissances ;	Non	http://www.contentmine.org/text-and-data-mining-tools/	Royaume-Uni ;
CORICO	Outil de visualisation de données multifactorielles sans équivalent. A partir d'un tableau de données, "L'Iconographie des Corrélations" élimine les "fausses bonnes corrélations" (celles qui sont dues à une tierce variable), et révèle les corrélations "masquées" lorsqu'une variable dépend de plusieurs variables.	Commercial	Découverte de connaissances ;	Non	http://www.coryent.com/corico.html	France ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Cortex Manager	CorText proposes a full ecosystem of modeling and exploratory tools for analyzing text corpora.		Reconnaissance d'entités nommées ; Topic modeling ; Extraction terminologique ; Apprentissage profond ; Clustering ; Analyse de sentiment ; Indexation ; Plongement de mots ;	Non	https://www.cortext.net/projects/cortext-manager/	France ;
Databionic ESOM Tools	The Databionic ESOM Tools is a suite of programs to perform data mining tasks like clustering, visualization, and classification with Emergent Self-Organizing Maps	Libre	Classification ; Clustering ;	Non	http://databionic-esom.sourceforge.net/	Allemagne ;
Dataiku	Dataiku DSS is the collaborative data science software platform for teams of data scientists, data analysts, and engineers to explore, prototype, build, and deliver their own data products more efficiently.	Libre/ Commercial	Clustering ;	Non	https://www.dataiku.com/	États-Unis ;
DataMelt	Free mathematics software for scientists, engineers and students. It can be used for numeric computation, statistics, symbolic calculations, data analysis and data visualization.	Libre/ Commercial	Classification ; Clustering ; Analyse de régression ;	Non	http://jwork.org/dmelt/	Allemagne ;
DataPreparator	DataPreparator is a free software tool designed to assist with common tasks of data preparation (or data preprocessing) in data analysis and data mining.	Libre	Prétraitement ;	Non	http://www.datapreparator.com/	Australie ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Datumbox	The Datumbox Machine Learning Framework is an open-source framework written in Java which allows the rapid development of Machine Learning and Statistical applications.	Libre	Classification ; Clustering ; Analyse de régression ; Analyse de sentiments ; Détection de la langue ; Extraction de mots clés ; Extraction de textes ;	Non	http://www.datumbox.com/machine-learning-framework/	Grèce ;
DBpedia Spotlight	It is a tool for automatically annotating mentions of DBpedia resources in text, providing a solution for linking unstructured information sources to the Linked Open Data cloud through DBpedia.	Libre	Annotation ;	Non	https://www.dbpedia-spotlight.org	Allemagne ;
Deeplearning4j	Eclipse Deeplearning4j is the first commercial-grade, open-source, distributed deep-learning library written for Java and Scala. Integrated with Hadoop and Apache Spark, DL4J brings AI to business environments for use on distributed GPUs and CPUs.Deep learning	Libre	Apprentissage profond	Non	https://deeplearning4j.org/	États-Unis ;
DeLFT	DeLFT (Deep Learning Framework for Text) is an Open Source Keras framework for text processing, covering sequence labeling (e.g. named entity tagging) and text classification (e.g. comment classification)	Libre	Apprentissage profond ; Reconnaissance d'entités nommées ; Classification ; Analyse de sentiments ; Analyse d'opinions	Non	http://science-miner.com/deep-learning/	France ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Diction	DICTION 7 is a computer-aided text analysis program for determining the tone of a verbal message	Commercial	Analyse de sentiments ;	Non	https://www.dictionsoftware.com/	États-Unis ;
Digimind	Social media analytics	Commercial	Clustering ; Traduction automatique ; Annotation ; Analyse de sentiments ; Recherche d'informations ;	Non	http://www.digimind.com/fr/	France ;
Discovertext	With dozens of powerful text analytics, data science, human coding, and machine-learning features, including instant access to the Gnip PowerTrack 2.0 for Twitter and the free Twitter Search API, DiscoverText provides cloud-based software tools to quickly evaluate large amounts of text, survey, and Twitter data.	Commercial	Classification ;	Non	https://discovertext.com/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
DKPro Core	A collection of software components for natural language processing (NLP) based on the Apache UIMA framework.	Libre	PoS tagging : Tokenisation ; Parsing ; Lemmatisation ; Etiquetage des rôles sémantiques ; Segmentation ; Reconnaissance d'entités nommées ; Chunking ; Racinisation ; Identification de langue ; Résolution de coréférence ; Apprentissage profond ; Analyse morphologique ; Clustering ; Annotation ;	Oui	https://dkpro.github.io/dkpro-core/	Allemagne ;
Dlib	Dlib is a modern C++ toolkit containing machine learning algorithms and tools for creating complex software in C++ to solve real world problems.	Libre	Apprentissage profond ; Classification ; Clustering ;	Non	http://dlib.net/	États-Unis ;
Dtm-Vic	Statistique Exploratoire Multidimensionnelle pour données complexes comprenant des données numériques et textuelles.	Libre	Classification ;	Non	http://www.dtmvic.com/05_SoftwareF.html	France ;
Egas	Collaborative biomedical text annotation.	Libre	Annotation ; Extraction de concepts ; Extraction de relations	Non	https://demo.bmd-software.com/egas/	Portugal ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
ELAN	ELAN is a professional tool for the creation of complex annotations on video and audio resources.	Libre ?	Annotation ;	Non	https://tla.mpi.nl/tools/tla-tools/elan/	Pays-Bas ;
ELKI	ELKI is an open source (AGPLv3) data mining software written in Java. The focus of ELKI is research in algorithms, with an emphasis on unsupervised methods in cluster analysis and outlier detection.	Libre	Clustering ;	Non	https://elki-project.github.io/	Allemagne ; Danemark ;
Elsevier Text Mining	Enables the retrieval of highly specified information from unstructured content, providing more meaningful answers to complex research questions.	Commercial	Recherche d'informations ; Reconnaissance d'entités nommées ; Extraction de relations ;	Non	https://www.elsevier.com/solutions/professional-services/text-mining	Pays-Bas ;
Enju	A deep syntactic parser for English	Libre	Parsing ;	Oui	http://www.nactem.ac.uk/enju/index.html	Royaume-Uni ;
EnjuParser	Enju is a syntactic parser for English	Libre	Parsing ;	Oui	http://pubannotation.org/annotators/EnjuParser	Japon ;
Entity fishing	entity-fishing is an open source tool dedicated to the automatic identification and disambiguation of Wikidata entities in multilingual text and PDF documents. The tool is based on machine-learning techniques (Gradient Tree Boosting, CRF, embeddings) exploiting Wikipedia as training source.	Libre	Reconnaissance d'entités nommées ; Apprentissage automatique ;	Non	http://science-miner.com/entity-disambiguation/	France ;
Etuma	Etuma text analysis service turns all your open-ended customer feedback into consistent and actionable information.	Commercial	Classification ;	Non	http://www.etuma.com/home	Finlande ;
EventMine	Event extraction system for biomedical text	Libre	Extraction d'évènements ;	Non	http://nactem.ac.uk/EventMine/	Royaume-Uni ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Expernova	Expernova utilise des algorithmes sophistiqués, s'appuyant sur le Big Data et le Machine Learning, pour connecter les réseaux d'innovation et dessiner un panorama global.	Commercial	Apprentissage automatique ;	Non	https://fr.expernova.com/	France ;
Fastr	Automatic indexing	Libre	Indexation ;	Non	https://perso.limsi.fr/jacquemi/FASTR/	France ;
FastText	Library for efficient text classification and representation learning	Libre	Classification de textes ; Apprentissage prodond ;	Non	https://fasttext.cc/	États-Unis ;
FreeLing	An Open-Source Suite of Language Analyzers	Libre	Reconnaissance d'entités nommées ; Annotation ;	Oui	http://nlp.lsi.upc.edu/freeling/demo/demo.php	Espagne ;
Galaxy	Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research.	Libre	Architecture logicielle ;	Oui	https://galaxyproject.org/	États-Unis ;
Gargantex	A web platform to explore text-mining.	Libre	Textométrie ;	Non	https://gargantext.org/	France ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
GATE	Suite of tools written in Java, used for human language processing, analysis, and information extraction.	Libre	Tokenisation ; Segmentation ; Chunking ; Résolution de coréférence ; Reconnaissance d'entités nommées ; Sentence splitting ; Annotation ; Analyse morphologique ; PoS tagging ; Classification de textes ; Apprentissage automatique ;	Oui	https://gate.ac.uk/	Royaume-Uni ;
Genia tagger	Part-of-speech tagging, shallow parsing, and named entity recognition for biomedical text	Libre	PoS tagging ; Reconnaissance d'entités nommées ; Parsing ;	Oui	http://www.nactem.ac.uk/GENIA/tagger/	Royaume-Uni ;
Gensim	Scalable statistical semantics ; Analyze plain-text documents for semantic structure ; Retrieve semantically similar documents	Libre	Clustering ;	Non	https://radimrehurek.com/gensim/	République Tchèque ;
GibbsLDA	GibbsLDA++ is a C/C++ implementation of Latent Dirichlet Allocation (LDA) using Gibbs Sampling technique for parameter estimation and inference. It is very fast and is designed to analyze hidden/latent topic structures of large-scale datasets including large collections of text/Web documents	Libre	Clustering ;	Non	http://gibbslda.sourceforge.net/	Japon ; Vietnam ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
GLoVe	GloVe is an unsupervised learning algorithm for obtaining vector representations for words.	Libre	Apprentissage profond ; Plongement de mots ;	Non	https://nlp.stanford.edu/projects/glove/	États-Unis ;
Glozz	Plateforme d'annotation	Libre	Annotation ;	Non	http://www.glozz.org/	France ;
GNU PSPP	GNU PSPP is a program for statistical analysis of sampled data	Libre		Non	https://www.gnu.org/software/pspp/	?
Google Cloud Natural Language API	L'API Google Cloud Natural Language révèle la structure et la signification des textes grâce à des modèles de machine learning puissants, dans une API REST conviviale.	Commercial	Reconnaissance d'entités nommées ; Analyse de sentiments ; Parsing ; Classification ; Apprentissage profond ;	Non	https://cloud.google.com/natural-language/	États-Unis ;
Google Prediction API	L'API Prediction de Google propose des fonctionnalités de filtrage par motif et de machine learning.	Commercial	Classification ;	Non	https://cloud.google.com/prediction/	États-Unis ;
GROBID	GROBID is an Open Source tool for parsing and extracting structured information from technical and scientific documents in raw format like PDF	Libre	Prétraitement ; Normalisation ; Structuration de documents ; Reconnaissance d'entités nommées ;	Non	http://science-miner.com/document-engineering/	France ;
Heart of Gold	Middleware architecture for the integration of deep and shallow natural language processing components. It provides a uniform and flexible infrastructure for building applications that use Robust Minimal Recursion Semantics (RMRS) and/or general XML standoff annotation produced by natural language processing components.	Libre	Annotation ;	Non	http://heartofgold.dfki.de/	Allemagne ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
HunPos tagger	Hunpos is an open source reimplementation of TnT, the well known part-of-speech tagger by Thorsten Brants.	Libre	PoS tagging ;	Oui	http://mokk.bme.hu/resources/hunpos/	Hongrie ;
Hyperbase	Logiciel hypertexte pour le traitement documentaire et statistique des corpus textuels	Libre	Textométrie ; Concordancier ; Lemmatisation ; Classification ; Clustering ;	Non	http://bcl.cnrs.fr/article69?redirected_from=www%252eunice%252efr%252fbcl%252farticle69	France ;
IBM SPSS Modeler	Versatile data and text analytics workbench that helps you build accurate predictive models quickly and intuitively, without programming.	Commercial	Extraction d'informations ; Reconnaissance d'entités nommées ; Extraction de relations ;	Non	http://www.spss.com.hk/software/modeler/	États-Unis ;
IBM Watson Natural Language Understanding	Analyze text to extract meta-data from content such as concepts, entities, keywords, categories, relations and semantic roles. Returns both overall sentiment and emotion for a document, and targeted sentiment and emotion towards keywords in the text for deeper analysis.	Commercial	Extraction d'informations ; Reconnaissance d'entités nommées ; Extraction de relations ; Etiquetage de rôles sémantiques ; Analyse de sentiments ; Détection de la langue ;	Non	https://www.ibm.com/watson/services/natural-language-understanding/	États-Unis ;
ILSP NLP	Natural Language Processing services developed by the NLP group of the Institute for Language and Speech Processing	Libre	PoS tagging ; Lemmatisation ; Parsing ; Chunking ; Sentence splitting ; Reconnaissance d'entités nommées ; Tokenisation ;	Oui	http://nlp.ilsp.gr/soaplab2-axis/	Grèce ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
ILSP NLP Web Services	Natural Language Processing services developed by the NLP group of the Institute for Language and Speech Processing	Libre	Parsing ; Chuncking ; Lemmatisation ; Reconnaissance d'entités nommées ;	Oui	http://nlp.ilsp.gr/soaplab2-axis/	Grèce ;
INCEpTION	Semantic annotation	Libre	Annotation ;	Oui	https://inception-project.github.io	Allemagne ;
Indico	Text and image analysis to create transformative tools.	Commercial	Apprentissage automatique ;	Non	https://indico.io/	États-Unis ;
Intellexer	Based on the use of Natural Language Processing and Machine Learning technologies, tools for text analytics solutions that can be used as standalone applications as well as integrated in the existing systems.	Commercial	Recherche d'informations ; PoS tagging ; Segmentation ; Parsing ; Extraction de relations ; Analyse de sentiments ; Reconnaissance d'entités nommées ; Résumé automatique ; Q&A ; Classification ; Clustering ; Détection de la langue ; Vérification de l'orthographe ;	Non	https://www.intellexer.com/products.html	États-Unis ;
Iramuteq	Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires.	Libre	Classification ;	Non	http://iramuteq.org/	France ;
Java Automatic Term Extraction	Java Automatic Term Extraction toolkit - a library of state-of-the-art term extraction algorithms and framework for developing term extraction algorithms	Libre	Extraction terminologiques ;	Non	https://code.google.com/archive/p/jatetoolkit/	Royaume-Uni ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Jazzy	Java Spell Check API	Libre	Vérification de l'orthographe ;	Oui	https://sourceforge.net/projects/jazzy/	?
JCoRE	The JULIE Lab Component Repository (JCoRe) is an open software repository for full-scale natural language processing based on the UIMA middleware framework.	Libre	Sentence segmentation ; Tokenisation ; PoS tagging ; Reconnaissance d'entités nommées ;	Non	http://julielab.github.io/	Allemagne ;
Jubatus	Online distributed machine learning on the data streams of Big Data	Libre	Classification ; Clustering ; Analyse de régression ;	Non	http://jubat.us/en/overview.html	Japon ;
KAF annotator	Stand-alone application for annotating KAF files with any set of tags to any level. This annotator is used to create gold-standard data for evaluating the Kybot output.	Libre	Annotation ;	Non	http://kyoto-project.eu/xmlgroup.iit.cnr.it/kyoto/index2091.html?option=com_content&view=article&id=	Europe ; Japon ;
KEA	KEA is an algorithm for extracting keyphrases from text documents. It can be either used for free indexing or for indexing with a controlled vocabulary.	Libre	Extraction de mots clés ;	Oui	http://community.nzdl.org/kea/	Nouvelle-Zélande ;
Keatext	Keatext is an AI – driven text analytics technology that makes it easy for you to analyze large volumes of unstructured customer feedback	Commercial	Analyse de sentiments ; Analyse d'opinions ;	Non	https://www.keatext.ai/	Canada ;
KEEL	KEEL (Knowledge Extraction based on Evolutionary Learning) is an open source (GPLv3) Java software tool that can be used for a large number of different knowledge data discovery tasks.	Libre	Découverte de connaissances	Non	http://www.keel.es/	Espagne ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Keyterm	rank extracted terms and concepts according to a relevance score estimating how well they capture the most important and discriminant subject matter of a document given a background collection of technical and scientific documents	Libre	Extraction de phrases clés	Non	http://science-miner.com/term-and-concept-relevance/	France ;
KH coder	KH Coder is a free software for quantitative content analysis or text mining. It is also utilized for computational linguistics. You can analyze Japanese, English, French, German, Italian, Portuguese and Spanish text with KH Coder. Also, Chinese (simplified), Korean and Russian language data can be analyzed with the latest Alpha release (Version 3).	Libre	Clustering ; Concordancier ; Réseau de co-occurrences ;	Non	http://khc.sourceforge.net/en/	Japon ;
KIWI	Keyword extractor	Libre	Extraction de mots clés ;	Non	http://www.quaero.org/module_technologique/kiwi-keyword-extractor/	France ;
KNIME	Open source data analytics, reporting and integration platform.	Libre	Apprentissage automatique ;	Non	https://www.knime.com/	Suisse ;
KnowledgeREADER	Integrated customer intelligence by combining visual text discovery and sentiment analysis with the power of predictive analytics	Commercial	Analyse de sentiments ;	Non	http://www.angoss.com/	Canada ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
LanguageComputer (Cicero, Ferret)	Understanding the information stored in any large collections of text.	Commercial	Reconnaissance d'entités nommées ; Résolution de coréférence ; Annotation ; Extraction de relations ; Extraction d'évènements ; Q&A ;	Non	http://www.languagecomputer.com/	États-Unis ;
LAPPS Grid	An open, interoperable web service platform for natural language processing (NLP) research and development	Libre	Architecture logicielle ; Tokenisation ; Reconnaissance d'entités nommées ; PoS tagging ; Sentence splitting ; Parsing ; Chunking ;	Non	http://www.lappsgrid.org/	États-Unis ;
Lavastorm analytics engine	Visual data discovery solution	Libre	Prétraitement ;	Non	http://www.lavastorm.com/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Le Trameur	Le Trameur est un programme d'analyse comportant de nombreuses fonctionnalités pour l'analyse automatique, statistique et documentaire de textes en vue de leur profilage sémantique, thématique et de leur interprétation. Ce logiciel est à l'origine un outil de textométrie : il intègre les fonctionnalités classiques de ce type d'outils dans ce domaine. Il dispose aussi des fonctionnalités particulières qui permettent d'annoter dynamiquement des corpus ou d'explorer des ressources richement annotées (treebanks monolingues/multilingues ou des alignements).	Libre	Textométrie ; Annotation	Non	http://www.tal.univ-paris3.fr/trameur/#p4	France ;
Lexalytics (Semantria)	Natural language processing	Commercial	PoS tagging ; Extraction de relations ; Classification ; Tokenisation ; Extraction de relations ; Analyse d'opinions ; Analyse de sentiments ; Résolution d'anaphore ; Racinisation ; Reconnaissance d'entités nommées ;	Non	https://www.lexalytics.com/	États-Unis ;
Lexi-co	Analyses textometriques	Libre	Textométrie ;	Non	http://www.lexi-co.com/index.html	France ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Leximancer	Leximancer automatically analyses your text documents to identify the high level concepts in your text documents, delivering the key ideas and actionable insights you need with powerful interactive visualisations and data exports.	Commercial	Extraction d'informations ;	Non	https://info.leximancer.com/	États-Unis ;
Liblinear	Library for large linear classification	Libre	Classification ;	Non	http://www.csie.ntu.edu.tw/~cjlin/liblinear/	Taïwan ;
LibShortText	A Library for Short-text Classification and Analysis	Libre	Classificationde textes ; Tokénisation ; Racinisation ;	Non	https://www.csie.ntu.edu.tw/~cjlin/papers/libshorttext.pdf	Taïwan ;
LIBSVM	Library for support vector machine	Libre	Classification ;	Oui	http://www.csie.ntu.edu.tw/~cjlin/libsvm/	Taïwan ;
LingPipe	LingPipe is tool kit for processing text using computational linguistics.	Libre/ Commercial	Reconnaissance d'entités nommées ; Détection de la langue ; Classification ; Sentence splitting ; Tokenisation ; PoS tagging ;	Oui	http://alias-i.com/lingpipe/index.html	États-Unis ;
Linguistic Inquiry Word Count	LIWC2015 is the gold standard in computerized text analysis. Learn how the words we use in everyday language reveal our thoughts, feelings, personality, and motivations.	Commercial	Annotation ; Classification ;	Non	http://liwc.wpengine.com/	États-Unis ;
LIONoso	Integrated tool for Machine Learning and Intelligent Optimization	Commercial	Apprentissage automatique ;	Non	http://lionoso.com/	Italie ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
LPU	LPU (which stands for Learning from Positive and Unlabeled data) is a text learning or classification system that learns from a set of positive documents and a set of unlabeled documents (without labeled negative documents). This type of learning is different from classic text learning/classification, in which both positive and negative training documents are required.	Libre	Classification ;	Non	https://www.cs.uic.edu/~liub/LPU/LPU-download.html	États-Unis ;
Luminoso	Understand, measure and act on large amounts of unstructured text.	Commercial	Apprentissage automatique ; Classification ; Analyse de tendances ;	Non	https://luminoso.com/	États-Unis ;
Magaputer	Data and text mining solutions	Commercial	Classification ; Clustering ; Analyse de régression ; Reconnaissance d'entités nommées ; Extraction de relations ; Détection de la langue ; PoS tagging ; Extraction de mots clés ; Parsing ; Analyse de sentiments ; Résolution d'anaphore ;	Non	http://megaputer.com/site/index.php	États-Unis ;
MALLET	MALLET is a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text.	Libre	Clustering ; Classification ;	Oui	http://mallet.cs.umass.edu/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
MaltParser	MaltParser is a system for data-driven dependency parsing, which can be used to induce a parsing model from treebank data and to parse new data using an induced model	Libre	Parsing ;	Oui	http://www.maltparser.org/	Suède ;
Massive Online Analysis (MOA)	Open source framework for data stream mining	Libre	Clustering ; Classification ; Analyse de régression ;	Non	https://moa.cms.waikato.ac.nz/	Nouvelle-Zélande ;
MATE	Multilevel Annotation, Tools Engineering	Libre	Annotation ;	Oui	http://xml.coverpages.org/mate.html	Royaume-Uni ; Allemagne ; Danemark ;
MatheoSoftware	Patents search and analysis, technological trends, data analysis	Commercial	Recherche d'informations ; Analyse de tendances ;	Non	https://www.matheo-software.com/	France ;
MATLAB	Analyse de données, développement d'algorithmes et création de modèles mathématiques, deep learning	Commercial	Clustering ; Classification ; Analyse de régression ; Apprentissage profond ;	Non	https://fr.mathworks.com/?s_tid=gn_logo	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Meaning Cloud	Extract the meaning of all kind of unstructured content: social conversation, articles, documents...	Libre/ Commercial	Analyse de sentiments ; Analyse de tendance ; Clustering ; Classification de textes ; Résumé automatique ; Détection de la langue ; Lemmatisation ; PoS tagging ; Etiquetage morphosyntaxique ;	Non	https://www.meaningcloud.com/	États-Unis ;
MeCab		Libre	Parsing ;	Oui	https://taku910.github.io/mecab/libmecab.html	Japon ;
MER	Minimal name entity recognizer	Libre	Reconnaissance d'entités nommées ;	Non	https://github.com/lasigeBioTM/MER	Portugal ;
Merlin	Deep learning	Libre	Apprentissage profond ;	Non	http://www.cstr.ed.ac.uk/projects/merlin/	Royaume-Uni ;
MetaMAp	Tool for recognizing UMLS concepts in texts	Libre	Annotation ;	Non	https://metamap.nlm.nih.gov/	États-Unis ;
MicroFocus	Big Data and analytics software	Commercial	Recherche d'informations ; Découverte de connaissances ; Apprentissage automatique ;	Non	https://software.microfocus.com/en-us/software/big-data-analytics-software	Royaume-Uni ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Microsoft Azure	Détectez le sentiment, les phrases clés, les sujets et la langue du texte	Commercial	Analyse de sentiments ; Extraction de mots clés ; Détection de la langue ; Annotation sémantique ;	Non	https://azure.microsoft.com/fr-fr/services/cognitive-services/text-analytics/	États-Unis ;
Microsoft Cognitive ToolKit	The Microsoft Cognitive Toolkit (CNTK) is an open-source toolkit for commercial-grade distributed deep learning. It describes neural networks as a series of computational steps via a directed graph. CNTK allows the user to easily realize and combine popular model types such as feed-forward DNNs, convolutional neural networks (CNNs) and recurrent neural networks (RNNs/LSTMs). CNTK implements stochastic gradient descent (SGD, error backpropagation) learning with automatic differentiation and parallelization across multiple GPUs and servers.	Libre	Apprentissage profond ;	Non	https://docs.microsoft.com/en-us/cognitive-toolkit/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Microsoft Distributed Machine Learning Toolkit	Distributed machine learning has become more important than ever in this big data era. Especially in recent years, practices have demonstrated the trend that more training data and bigger models tend to generate better accuracies in various applications. However, it remains a challenge for common machine learning researchers and practitioners to learn big models from huge amount of data, because the task usually requires a large number of computation resources. In order to tackle this challenge, we release the Microsoft Distributed Machine Learning Toolkit (DMTK), which contains both algorithmic and system innovations.	Libre ?	Topic modeling ; Apprentissage automatique ; Apprentissage profond	Non	http://www.dmtk.io/index.html	États-Unis ;
Microsoft SQL Server Analysis Services	Ensemble d'outils pour la fouille de données	?	Classification ; Clustering ; Anayse de régression ;	Non	https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/data-mining-tools	États-Unis ;
MiningMart	Prétraitement des données	Libre	Prétraitement ;	Non	http://mmart.cs.uni-dortmund.de/research/index.html	Allemagne ;
MiniPar		Libre	Parsing ;	Oui	https://webdocs.cs.ualberta.ca/~lindek/minipar.htm	Canada ;
ML-Flex	an open-source software package designed to enable flexible and efficient processing of disparate data sets for machine-learning (classification) analyses	Libre	Classification ;	Non	http://mlflex.sourceforge.net/	États-Unis ;
MLPACK	Scalable machine learning library, written in C++	Libre	Classification ; Clustering ; Analyse de régression ;	Non	http://mlpack.org/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
mlpy	mlpy provides a wide range of state-of-the-art machine learning methods for supervised and unsupervised.	Libre	Classification ; Clustering ;	Non	http://mlpy.sourceforge.net/	Italie ;
Modalisa	Création de questionnaires en ligne ou papier, diffusion et recueil de données, transformation de variables, codification de textes, analyses univariées et multivariées, indicateurs spécifiques, régressions, rapports dynamiques exportables sous PowerPoint, export et import des données sous format Excel et Texte...	Commercial	Classification	Non	https://modalisa.com/logiciel/modalisa.php	France ;
Modular toolkit for Data Processing	Modular toolkit for Data Processing (MDP) is a library of widely used data processing algorithms that can be combined according to a pipeline analogy to build more complex data processing software.	Libre	Classification ; Clustering ;	Non	https://pypi.python.org/pypi/MDP/2.4	États-Unis ; Allemagne ;
Monkeylearn	Text Analysis with machine learning	Libre/ commercial	Reconnaissance d'entités nommées ; Analyse de sentiments ; Extraction de topics ; Apprentissage automatique ; Prétraitement ;	Non	https://monkeylearn.com/	États-Unis ;
MorphAdoerner	MorphAdorner is a Java command-line program which acts as a pipeline manager for processes performing morphological adornment of words in a text	Libre	Tokenisation ; PoS tagging ; Reconnaissance d'entités nommées ; Sentence splitting ; Lemmatisation ; Annotation ;	Non	http://morphadorner.northwestern.edu/morphadorner/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
MSTParser	MSTParser is a non-projective dependency parser that searches for maximum spanning trees over directed graphs. Models of dependency structure are based on large-margin discriminative training methods. Projective parsing is also supported.	Libre	Analyse en dépendances	Non	https://www.seas.upenn.edu/~strctlrn/MSTParser/MSTParser.html	États-Unis ;
MutationFinder	MutationFinder is a biomedical natural language processing (NLP) system for extracting mentions of point mutations from free text. MutationFinder achieves high performance (99% precision, 81% recall on blind test data) as an information extraction system	Libre	Extraction d'informations ;	Oui	https://sourceforge.net/projects/mutationfinder/	États-Unis ;
mutext	Analytics and decision science solutions.	Commercial	Parsing ; Classification de textes ; Clustering ;	Non	https://www.mu-sigma.com/	Inde ;
MXNet	A flexible and efficient library for deep learning	Libre	Apprentissage profond ;	Non	https://mxnet.incubator.apache.org	États-Unis ;
NaCTeM Software Tools	The National Centre for Text Mining bases its service systems on a number of text mining software tools.	Libre	Reconnaissance d'entités nommées ; PoS tagging ; Parsing ; Sentencfe splitting ; Paragraph splitting ; Extraction d'évènements ; Annotation ;	Oui	http://nactem.ac.uk/software.php	Royaume-Uni ;
Narrative Science Quill	Quill transforms data into automated, human-sounding Intelligent Narratives that empower your people with insights to improve every aspect of your business.	Commercial	Génération de langage naturel ;	Non	https://narrativescience.com/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
NaturalText	For Life sciences: NaturalText's Machine Learning Algorithms can process Scientific Papers, Bio Sequences to find patterns and help scientists, researchers to advance their research. For financial sector : NaturalText's Machine Learning Algorithms can combine various data formats, cross verify for incorrect information, help companies to know more from the data	Commercial	Extraction d'informations ; Extraction de relations ; Apprentissage automatique ; Découverte de connaissances ;	Non	http://naturaltext.com/	Inde ;
NCBI Text Mining Tools	Ensemble de applications web ou de bureau pour la fouille de textes dans le domaine biomédical	Libre	Annotation ; Recherche d'informations ; Reconnaissance d'entités nommées ; Normalisation ; Désambiguisation ;	Non	https://www.ncbi.nlm.nih.gov/research/bionlp/Tools/	États-Unis ;
NERD	Named entity recognition and desambiguation	Libre	Reconnaissance d'entités nommées ;	Non	http://nerd.eurecom.fr/	France ;
NERSuite	Named Entity Recognition toolkit	Libre	Reconnaissance d'entités nommées ;	Non	http://nersuite.nlplab.org/	Japon ;
Netowl	Text and Entity Analytics Products	Commercial	Reconnaissance d'entités nommées ; Analyse de sentiments ;	Non	https://www.netowl.com/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Neural designer	Neural Designer is a software tool for advanced analytics. It includes tools for descriptive, diagnostic, predictive and prescriptive analytics. It allows you to get actionable insights resulting in smarter decisions and better business outcomes. Neural networks are the most powerful method to discover intricate relationships, recognize complex patterns or predict current trends in your data.	Libre/commercial	Apprentissage automatique ;	Non	https://www.neuraldesigner.com/	Espagne ;
NeuroNER	A Named-Entity Recognition program based on neural networks	Libre	Reconnaissance d'entités nommées ;	Non	http://neuroner.com/	États-Unis ;
NLPCube	Natural Language Procecssing Toolkit with support for tokenization, sentence splitting, lemmatization, tagging and parsing for more than 60 languages	Libre	Tokenisation ; Segmentation ; Lemmatisation ; Parsing ; POS tagging ; Analyse en dépendances : Reconnaissance d'entités nommées ;	Non	https://pypi.org/project/nlpcube/	États-Unis ;
NLTK	Platform for building Python programs to work with human language data	Libre	Parsing ; Chunking ; Concordiancier ; Classification ; Clustering ; Extraction de relations sémantiques ; Anlyse de sentiments ; Racinisation ; Tokenisation ; Traduction automatique ;	Non	http://www.nltk.org/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
NooJ	Linguistic development environment software as well as a corpus processor.	Libre	Annotation ;	Non	http://www.nooj4nlp.net/	France ;
Noopsis	Noopsis automatise la collecte d'informations stratégiques par la fouille de documents textuels.	Commercial	Recherche d'informations ; Analyse sémantique ;	Non	http://www.noopsis.fr/index.fr.html	France ;
ntent	Semantic search technology to determine user intent and the contextual meaning of words	Commercial	Recherche d'informations ; Reconnaissance d'entités nommées ; Détection de la langue ; Classification ; Lemmatisation ; Toklenisation ; Extraction d'informations ; Apprentissage automatique ;	Non	http://www.ntent.com/	États-Unis ;
Nvivo	NVivo est un logiciel qui supporte des méthodes de recherches qualitatives et combinées. Il est conçu pour vous permettre d'organiser, analyser et trouver du contenu perspicace parmi des données non structurées ou qualitatives telles que des interviews, des réponses libres obtenues dans le cadre d'un sondage, des articles, des médias sociaux et des pages Web.	Commercial	Classification ; Visualisation	Non	http://www.qsrinternational.com/nvivo-french	Australie ;
Odintext	Text analytics software	Commercial	Analyse de sentiments ; Apprentissage automatique ;	Non	http://odintext.com/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
OGER	Biomedical entity recognizer	Libre	Reconnaissance d'entités nommées ;	Oui	http://www.ontogene.org/resources/oger	Suisse ;
OntoText	Ontotext provides a complete set of Semantic Technology enabling better content management, knowledge discovery and semantic search.	Commercial	Annotation sémantique ; Extraction de relations ; Désambiguisation ; Apprentissage automatique ; Classification ;	Non	https://ontotext.com/	Bulgarie ;
Open Calais	Way to tag the people, places, companies, facts, and events in your content to increase its value, accessibility and interoperability.	Commercial	Annotation ; Reconnaissance d'entités nommées ; Extraction de relations ; Extraction d'évènements ; Clustering ;	Oui	http://www.opencalais.com/about-open-calais/	États-Unis ;
Open semantic search	Free Software for your own Search Engine, Explorer for Discovery of large document collections, Media Monitoring, Text Analytics, Document Analysis & Text Mining platform based on Apache Solr or Elasticsearch open-source enterprise-search and Open Standards for Linked Data, Semantic Web & Linked Open Data integration	Libre	Recherche d'informations ; Annotation ;	Non	https://www.opensemanticsearch.org/	Allemagne ;
OpenCCG	OpenCCG, the OpenNLP CCG Library, is an open source natural language processing library written in Java, which provides parsing and realization services based on Mark Steedman's Combinatory Categorial Grammar (CCG) formalism	Libre	Parsing ; PoS tagging ;	Oui	http://openccg.sourceforge.net/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
OpenMinTED	OpenMinted sets out to create an open, service-oriented e-Infrastructure for Text and Data Mining (TDM) of scientific and scholarly content. Researchers can collaboratively create, discover, share and re-use Knowledge from a wide range of text-based scientific related sources in a seamless way	Libre			http://openminted.eu/	Europe ;
OpenNER	OpeNER's main goal is to provide a set of ready to use tools to perform some natural language processing tasks, free and easy to adapt for Academia, Research and Small and Medium Enterprise to integrate them in their workflow. More precisely, OpeNER aims to be able to detect and disambiguate entity mentions and perform sentiment analysis and opinion detection on the texts, to be able for example, to extract the sentiment and the opinion of customers about certain resource (e.g. hotels and accommodations) in Web reviews.	Libre	Reconnaissance d'entités nommées ; Analyse de sentiment ; Analyse d'opinion	Non	https://www.opener-project.eu/index.html	Italie ; Espagne ; Pays Bas ;
OpenNN	OpenNN is an open source class library written in C++ programming language which implements neural networks, a main area of machine learning research.	Libre	Apprentissage profond ;	Non	http://www.opennn.net/	Espagne ;
OpenRefine	Nettoyage, mise en forme et transformation de données	Libre	Prétraitement ;	Non	http://openrefine.org/	États-Unis ;
OpenText	Digitize processes and discover the value in information using analytics and Artificial Intelligence.	Commercial	Recherche d'informations ; Classification ;	Non	http://www.opentext.com/	Canada ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Oracle Data Mining	Oracle Data Mining (ODM), a component of the Oracle Advanced Analytics Database Option, provides powerful data mining algorithms that enable data analytsts to discover insights, make predictions and leverage their Oracle data and investment.	Commercial	Classification ; Clustering ; Analyse de régression ; Détection d'anomalies ; Règle d'associations ;	Non	http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index.html	États-Unis ;
Orange Data Mining Toolbox	Open source machine learning and data visualization for novice and expert.	Libre	Clustering ; Classification ; Analyse de régression ;	Non	https://orange.biolab.si/	Slovénie ;
Overview	Overview is a document mining application originally built for investigative journalists. It's also used for legal work, training machine learning models, and research of all types. It's a visualization and analysis tool designed for sets of documents, from dozens to millions of pages of materia	Libre	Clustering ; Recherche d'informations ; Annotation ; Reconnaissance d'entités nommées ;	Non	https://www.overviewdocs.com/	?
Pagelyser	Pagelyzer is a tool which compares two web pages versions and decides if they are similar or not.	Libre	Analyse de pages web ;	Non	http://pagelyzer.openpreservation.org/	France ;
PANDAS	Library providing high performance, easey-to-use data structure and data analysis tools for the Python programming language	Libre	Analyse de régression ;	Non	http://pandas.pydata.org/	États-Unis ;
Pattern	Pattern is a web mining module for the Python programming language.	Libre	PoS tagging ; Analyse de sentiments ; Apprentissage automatique ; n-gram ; Clustering ;	Non	https://www.clips.uantwerpen.be/pages/pattern	Pays-Bas ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Penelope	Penelope is a cloud-based, open and modular platform that consists of tools and techniques for mapping landscapes of opinions expressed in online (social) media. The platform is used for analysing the opinions that dominate the debate on certain crucial social issues, such as immigration, climate change and national identity.		Tokenisation ; Lemmatisation ; Chunking ; PoS tagging ; Reconnaissance d'entités nommées ; Analyse en dépendances ; apprentissage automatique ; plongement lexical ; Sentencisation ;	Non	https://penelope.vub.be	Europe ;
Philologic	PhiloLogic™ is the primary full-text search, retrieval and analysis tool developed by the ARTFL Project and the Digital Library Development Center (DLDC) at the University of Chicago. This is a Free Software implementation of PhiloLogic for large TEI-Lite document collections.	Libre	Recherche d'information ;	Non	https://sites.google.com/site/philologic3/home	États-Unis ;
Pingar	Pingar DiscoveryOne Content Enrichment automatically tags and categorizes content. Typically, it is used to improve findability of information in an Electronic Content Management System (ECMS) by enabling faceted search.	Commercial	Classification de textes ; Recherche d'informations ; Découverte de connaissances ;	Non	http://pingar.com/	États-Unis ;
PlaidML	Framework for making deep learning work everywhere	Libre	Apprentissage profond ;	Non	https://github.com/plaidml/plaidml	?

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
PoolParty Semantic Suite	PoolParty is a world-class semantic technology suite that offers sharply focused solutions to knowledge organization and content business.	Commercial	Reconnaissance d'entités nommées ; Annotation ; Recherche d'informations ; Extraction de relations ; Extraction terminologique ; Classification de textes ;	Non	https://www.poolparty.biz/	Autriche ;
PrediCX	Automated Text Analytics for Voice of Customer (VoC) data, chatbots, service desks, complaint handling, call center automation and early warning of issues. Quick and Easy to Deploy, High-Impact, AI and Machine Learning for Text Analysis.	Commercial	Apprentissage automatique ; Analyse de sentiments ;	Non	https://warwickanalytics.com/predicx/	Royaume-Uni ;
Prosuite	ProSuite is an integrated collection of Provalis Research text analytics tools that allow one to explore, analyze and relate both structured and unstructured data. Provalis Research Text Analytics Software allows one to perform advanced computer assisted qualitative coding on documents and images using QDA Miner, to apply the powerful content analysis and text mining features of WordStat on textual data, and to perform advanced statistical analysis on numerical and categorical data using SimStat	Commercial	Clustering ; Concordancier ; Annotation ; Analyse de tendances ; Extraction de mots clés ; Classification de textes ;	Non	https://provalisresearch.com/products/prosuite/	Canada ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Proxem Studio	Transforme les données textuelles en prise de decision	Commercial	Classification de textes ; Recherche d'informations ; Découverte de connaissances ; Annotation ; Analyse de sentiments ;	Non	https://www.proxem.com/	France ;
PubRunner	A framework for keeping biomedical text mining result up-to-date	Libre	Architecture logicielle ;	Oui	https://github.com/jakelever/pubrunner	États-Unis ;
PubTator	PubTator is a Web-based tool for accelerating manual literature curation.	Libre	Annotation ;	Oui	https://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/PubTator/index.cgi?user=User284144660	États-Unis ;
PyTorch	PyTorch is a deep learning framework for fast, flexible experimentation.	Libre	Apprentissage profond ;	Non	https://pytorch.org/	États-Unis ; France ;
Qlucore Omics Explorer	Identify patterns and structure when exploring biological data	Commercial	Classification ; Clustering ;	Non	https://www.qlucore.com/	Suède ;
QWAM	Solutions logicielles métier répondant aux besoins de gestion des flux d'information (documentaire, textuelle, multimedia), de moteur de recherche, de veille et d'analyse et d'enrichissement sémantique	Commercial	Extraction d'informations ;	Non	http://www.qwamci.com/	France ;
R	Free software environment for statistical computing and graphics.	Libre	Tokenisation ; Racinisation ; PoS tagging ; Parsing ; Reconnaissance d'entités nommées ; Analyse de sentiments ; Classification ; Clustering ; Apprentissage profond ;	Non	https://www.r-project.org/	Nouvelle-Zélande ; Canada ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
R.TeMIS	Environnement graphique de travail sous R permettant de créer, manipuler et analyser des corpus de textes.	Libre	Racinisation ; Chunking ; Clustering ;	Non	http://rtemis.hypotheses.org/	France ;
RapidMiner	RapidMiner is code free data science platform that unifies data prep, machine learning, and model deployment.	Libre/ commercial	Tokenisation ; Racinisation ; Reconnaissance d'entités nommées ; Analyse de sentiments ; Classification ; Clustering ;	Non	https://rapidminer.com/	Allemagne ;
Rasp	The RASP system includes state-of-the-art modules for finding sentence boundaries, finding individual words, analyzing words to identify the word root and any suffixes, assigning part-of-speech labels to words in running text, and analyzing the grammatical relations between words and larger units within sentences.	Libre/ commercial	Tokenisation ; Lemmatisation ; PoS tagging ; Parsing ;	Oui	https://www.ilexir.co.uk/rasp/index.html	Royaume-Uni ;
Rattle	GUI for data mining using R	Libre	Classification ; Clustering ;	Non	https://rattle.togaware.com/	États-Unis ;
RepKnight	Software platform provides real-time cyber intelligence to keep people, companies and assets safe from internal and external threats	Commercial	Recherche d'informations ;	Non	https://www.repknight.com/	Royaume-Uni ;
Resoomer	Résumé automatique	Libre/ Commercial	Résumé automatique	Non	https://resoomer.com	France ;
Rocket Folio	Automated content search and publishing for desktop, digital media, web, or intranet	Commercial	Recherche d'informations ;	Non	http://www.rocketsoftware.com/products/rocket-folionxt	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Rosetta	ROSETTA is a toolkit for analyzing tabular data within the framework of rough set theory. ROSETTA is designed to support the overall data mining and knowledge discovery process: From initial browsing and preprocessing of the data, via computation of minimal attribute sets and generation of if-then rules or descriptive patterns, to validation and analysis of the induced rules or patterns.	Libre	Découverte de connaissances ;	Non	http://bioinf.icm.uu.se/rosetta/	Suède ;
Rosette Text Analytics	Multilingual Text Analytics Solution	Commercial	Classification de textes ; Clustering ; Analyse de sentiments ; Reconnaissance d'entités nommées ; Entity linking ; Extraction de relations ; Détection de la langue ; Traduction automatique ; Tokenisation ; PoS tagging ; Lemmatisation ;	Non	https://www.rosette.com/	États-Unis ;
S4	S4 (Structured and Semantic Search Service) is a generic end-to-end infrastructure to rapidly build and deploy a search service providing state-of-the-art structured and semantic searches in collections of technical and scientific documents.	Libre	Recherche d'informations ;	Non	http://science-miner.com/structured-and-semantic-search-infrastructure/	France ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
SANSA	SANSA is a big data processing engine for scalable processing of large-scale RDF data. SANSA uses Spark and Flink which offer fault-tolerant, highly available and scalable approaches to process massive sized datasets efficiently. SANSA provides the facilities for Semantic data representation, Querying, Inference, and Analytics	Libre	Classification ; Clustering ; Règlesd'association ; Détection d'anomalie ;	Non	http://sansa-stack.net/	Allemagne ;
SAP predictive analytics	Predictive modeling suite	Commercial		Non	https://www.sap.com/products/analytics/predictive-analytics.html	Allemagne ;
SAS Enterprise Mining	Create accurate predictive and descriptive models for large volumes of data.	Commercial	Prétraitement ; Apprentissage automatique ;	Non	https://www.sas.com/en_id/software/analytics/enterprise-miner.html	États-Unis ;
SAS Text Miner	Text mining software from SAS automatically finds information buried in unstructured text data.	Commercial	Découverte de connaissances ; Reconnaissance d'entités nommées ; Clustering ; Extraction de relations ; Apprentissage automatique ;	Non	https://www.sas.com/en_us/software/text-miner.html	États-Unis ;
Scikit-learn	Simple and efficient tools for data mining and data analysis. Accessible to everybody, and reusable in various contexts. Built on NumPy, SciPy, and matplotlib	Libre	Classification ; Clustering ; Analyse de régression ;	Non	http://scikit-learn.org/stable/index.html	France ;
Scrapy	An open source and collaborative framework for extracting the data you need from websites. In a fast, simple, yet extensible way.	Libre	Indexation ;	Non	https://scrapy.org	?

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
SDL MultiTerm Extract	Computer assisted translation	Commercial	Traduction automatique ;	Non	https://www.sdltrados.com/	Allemagne ;
Semdee	Comprendre et exploiter de gros volumes de données textuelles	Commercial	Apprentissage automatique ;	Non	http://www.semdee.com/	France ;
SemRep	SemRep is a UMLS-based program that extracts three-part propositions, called semantic predications, from sentences in biomedical text.	Libre	Extraction de relations	Non	https://semrep.nlm.nih.gov/	États-Unis ;
SENNA	ENNA is a software distributed under a non-commercial license, which outputs a host of Natural Language Processing (NLP) predictions: part-of-speech (POS) tags, chunking (CHK), name entity recognition (NER), semantic role labeling (SRL) and syntactic parsing (PSG).	Libre	PoS tagging ; Chunking ; Reconnaissance d'entités nommées ; Etiquetage de rôle sémantique ; Parsing ;	Non	http://ronan.collobert.com/senna/	États-Unis ;
Sentic API	Sentic API provides the semantics and sentics (i.e., the denotative and connotative information) associated with the concepts of SenticNet 4, a semantic network of commonsense knowledge that contains 50,000 nodes (words and multiword expressions) and thousands of connections (relationships between nodes).	Libre	Analyse de sentiments ;	Non	http://sentic.net/api/	États-Unis ;
SentiStrength	SentiStrength estimates the strength of positive and negative sentiment in short texts, even for informal language	Libre	Analyse de sentiments ;	Non	http://sentistrength.wlv.ac.uk/	Royaume-Uni ;
Shogun	The Shogun Machine learning toolbox offers a wide range of efficient and unified Machine Learning methods.	Libre		Non	http://shogun-toolbox.org/	?

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Simple Extractor	Software application oriented to extracting terminology from texts. Some of its main features are its simple use and its intuitive interfaces. This tool allows the setting of different extraction criteria and exporting files.	Commercial	Extraction terminologique ;	Non	http://www.dail-software.com/help/9_en/index.html	Espagne ;
Sisense	Business intelligence tool for simplifying complex data preparation and analysis.	Commercial	Prétraitement ; Apprentissage automatique ;	Non	https://www.sisense.com/get/pricing/	États-Unis ; Israël ;
Sketch Engine	Sketch Engine is the ultimate tool to explore how language works. Its algorithms analyze authentic texts of billions of words (text corpora) to identify instantly what is typical in language and what is rare, unusual or emerging usage. It is also designed for text analysis or text mining applications	Commercial	Collocation ; Concordancier ; Extraction terminologique ;	Non	https://www.sketchengine.eu	République Tchèque ;
SMART Text Miner	The SMART Text Miner is a sophisticated software tool that can extract hidden knowledge from legacy texts.	?	Extraction terminologique ;	Non	http://www.smartny.com/miner.htm	États-Unis ;
Smartlogic	Semantic platform that allows organizations to realize the business value of their information. By leveraging a common vocabulary and sophisticated semantic techniques Semaphore:	Commercial	Découverte de connaissances ; Reconnaissances d'entités nommées ; Extraction de relations ; Classification ; Recherche d'informations ; Désambiguisation ;	Non	https://www.smartlogic.com/	États-Unis ;
SoftLaw	Analyse de documents juridiques	Commercial	Extraction d'informations ;	Non	https://www.softlaw.digital/	France ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
SpaCy	Software library for NLP.	Libre	Apprentissage profond ; Tokenisation ; PoS tagging ; Segmentation de phrases ; Parsing ; Reconnaissance d'entités nommées ; classification de textes ;	Non	https://spacy.io/	Australie ;
Stanbol	Set of reusable components for semantic content management.	Libre	Détection de la langue ; Tokenisation ; PoS tagging ; Chunking ; Lemmatisation ; Reconnaissance d'entités nommées ;	Non	https://stanbol.apache.org	États-Unis ;
Stanford CoreNLP	Stanford CoreNLP provides a set of human language technology tools. It can give the base forms of words, their parts of speech, whether they are names of companies, people, etc., normalize dates, times, and numeric quantities, mark up the structure of sentences in terms of phrases and syntactic dependencies, indicate which noun phrases refer to the same entities, indicate sentiment, extract particular or open-class relations between entity mentions, get the quotes people said, etc.	Libre	POS tagging ; Reconnaissance d'entités nommées ; Résolution de coréférence ; parsing ; analyse en dépendances ; Analyse de sentiments ; Extraction d'informations	Non	https://stanfordnlp.github.io/CoreNLP/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Stanford NLP	An integrated suite of natural language processing tools for English and (mainland) Chinese, including tokenization, part-of-speech tagging, named entity recognition, parsing, and coreference.	Libre	PoS Tagging ; Résolution de coréférence ; Tokenisation ; Reconnaissance d'entités nommées; Parsing ; Extraction de relations ; Classification ;	Oui	https://nlp.stanford.edu/	États-Unis ;
Stanford Topic Modeling Toolbox	The Stanford Topic Modeling Toolbox (TMT) brings topic modeling tools to social scientists and others who wish to perform analysis on datasets that have a substantial textual component	Libre	Clustering ; Topic modeling ;	Non	https://nlp.stanford.edu/software/tmt/tmt-0.4/	États-Unis ;
Statistica Text Miner	Statistica Text Miner is an optional extension of Statistica Data Miner, ideal for translating unstructured text data into meaningful, valuable clusters of decision-making "gold."	Commercial	Lemmatisation ; Clustering ;	Non	https://www.statsoft.fr/logiciels/textminer.php	États-Unis ;
Stratifyd	Stratifyd's data analytics platform allows users to integrate, analyze, and visualize data in a single platform, empowering analysts through a holistic view of both structured and unstructured data.	Commercial	Clustering ; Analyse de sentiments ; Extraction d'nformations ;	Non	https://www.stratifyd.com/	États-Unis ;
streamDM	streamDM is a new open source software for mining big data streams using Spark Streaming	Libre	Classification ; Clustering ; Analyse de régression ;	Non	http://huawei-noah.github.io/streamDM/	Chine ;
Synaptica	Software solution for knowledge organization and discovery.	Commercial	Annotation ; Classification ;	Non	http://www.synaptica.com/	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Sysomos	Sysomos is a unified, insights-driven social platform that gives marketers the easiest way to Search, Discover, Listen, Publish, Engage, and Analyze at scale across earned, owned, and paid media.	Commercial	Recherche d'informations ; Découverte de connaissances ;	Non	https://sysomos.com/	Canada ;
Systran	Understanding, analyze and act in over 50 languages	Commercial	Analyse de sentiments ; Traduction automatique ; Reconnaissance d'entités nommées ; Détection de la langue ; Segmentation ; Tokenisation ; PoS tagging ; Analyse morphologique ;	Non	http://www.systran.io/	Corée du Sud ;
TACIT	Text Analysis,Crawling and Interpretation Tool	Libre	Classification de textes ; Clustering ;	Non	http://tacit.usc.edu/	États-Unis ;
Tagtog	Biomedical annotation tool	Libre	Annotation ;	Non	https://www.tagtog.net/	Pologne ;
Talismane	NLP framework: sentence detector, tokeniser, pos-tagger and dependency parser	Libre	Tokenisation ; POS tagging ; Analyse en dépendances ; Détection de phrases ;	Non	https://github.com/joliciel-informatique	France ;
TAMS	TAMS stands for Text Analysis Markup System. It is a convention for identifying themes in texts (web pages, interviews, field notes). It was designed for use in ethnographic and discourse research.	Libre	identification de thèmes	Non	http://tamsys.sourceforge.net	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
TANAGRA	TANAGRA is a free DATA MINING software for academic and research purposes. It proposes several data mining methods from exploratory data analysis, statistical learning, machine learning and databases area.	Libre	Apprentissage automatique ;	Non	https://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html	France ;
TapoRware	TAPoRware is a set of text analysis tools that enables users to perform text analysis on HTML, XML and plain text files, using documents from the users' machine or on the web.	Libre	Concordancier ; Tokenisation ; Extraction de liens ; Résumé automatique ;	Non	http://taporware.ualberta.ca/~taporware/about.shtml	Canada ;
TBXTools	TBXTools allows easy and rapid Terminology Extraction and Management. This tool implements both statistical and linguistic methods, along with several utilities to create and manage terminological databases. It is written in Python and uses NLTK (Natural Language Toolkit)	Libre	Extraction terminologique ;	Non	https://sourceforge.net/projects/tbxtools/	Espagne ;
TensorFlow	Machine learning library	Libre	Apprentissage automatique ;	Non	https://www.tensorflow.org/	États-Unis ;
Termsuite	Outil d'extraction terminologique et d'alignement multilingue de termes.	Libre	Extraction terminologique ;	Oui	https://termsuite.github.io/fr/	France ;
Text2data	Advanced text analytics	Commercial	Analyse de sentiments ; Résumé automatique ; Classification de textes ; Reconnaissance d'entités nommées ; Clustering ; Extraction de mots clés ;	Non	http://text2data.org/	?

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Textable	Free open source software to analyze and process texts visually	Libre	Concordancier ; Collocation ; Lemmatisation ; POS tagging ; Clustering ; Classification ; Visualisation ; Pétraitement ;	Non	http://textable.io/	Suisse ;
Textalytics	Textalytics is a meaning extraction service that produces meaningful data from social media content, contracts, news, and other documents	Commercial	Annotation ; PoS tagging ; Parsing ; Lemmatisation ; Analyse de sentiments ; Clustering ;	Oui	https://textalytics.io/	?
Textblob	TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.	Libre	Extraction terminologique ; PoS tagging ; Analyse de sentiments ; Classification ; Traduction automatique ; Détection de la langue ; Tokenisation ; Parsing ; Correction orthographique ;	Non	https://textblob.readthedocs.io/en/dev/	États-Unis ;
TextCat	http://odur.let.rug.nl/~vannoord/TextCat/	Libre	Classification de textes ;	Oui	http://odur.let.rug.nl/~vannoord/TextCat/	Pays-Bas ;
TextObserver	Outil de d'observation et d'exploitation des données textuelles multidimensionnelles.	Libre	Textométrie ;	Non	http://textopol.u-pec.fr/textobserver/	France ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
TextRazor	The TextRazor API helps you extract and understand the Who, What, Why and How from your news stories with unprecedented accuracy and speed	Libre	Reconnaissance d'entités nommées ; Classification ; Annotation ; Désambiguisation ; Extraction de relations ; Extraction de mots clés ;	Oui	https://www.textrazor.com/	Royaume-Uni ;
Theano	Deep learning	Libre	Apprentissage profond ;	Non	http://www.deeplearning.net/software/theano/	Canada ;
Thematic	AI text analytics and visualizations	Commercial	Clustering ;	Non	https://getthematic.com/	?
Theysay	Emotional AI and advanced data analytics to stream, interpret, and bring together opinions, moods, and feelings across the Web.	Commercial	Analyse de sentiments ; Analyse d'opinions ;	Non	http://www.theysay.io/	Royaume-Uni ;
Think Analytics	The ThinkAnalytics Search and Recommendations Engine provides a powerful, scalable, real-time and comprehensive multi-content/multi-platform Recommendations Engine supporting across content delivery of recommendations and search to multiple platforms such as the set top box, mobile, web, smart TV, games consoles, and others.	Commercial	Apprentissage automatique ;	Non	https://thinkanalytics.com/	Royaume-Uni ;
TnT	Statistical Part-of-Speech Tagging	Libre	POS tagging	Non	http://www.coli.uni-saarland.de/~thorsten/tnt/	Allemagne ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
Torch	Torch is a scientific computing framework with wide support for machine learning algorithms that puts GPUs first. It is easy to use and efficient, thanks to an easy and fast scripting language, LuaJIT, and an underlying C/CUDA implementation.	Libre	Apprentissage automatique ;	Non	http://torch.ch/	?
TreeTagger	The TreeTagger is a tool for annotating text with part-of-speech and lemma information	Libre	PoS tagging ; Lemmatization ; Chunking ;	Oui	http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/	Allemagne ;
Tropes	Analyse sémantique de textes	Libre	Textométrie ;	Non	http://www.tropes.fr/	France ;
Tweet NLP	We provide a tokenizer, a part-of-speech tagger, hierarchical word clusters, and a dependency parser for tweets, along with annotated corpora and web-based annotation tools.	Libre	Tokenisation ; PoS tagging ; Parsing ; Clustering ; Annotation ;	Non	http://www.cs.cmu.edu/~ark/TweetNLP/	États-Unis ;
TwinWord	Text analysis APIS to understand and associate words	Commercial	Analyse de sentiments ; Clustering ; Classfication de textes ; Lemmatisation ; Extraction de mots clés ;	Non	https://www.twinword.com/	États-Unis ;
TXM	Analyses textométriques	Libre	Textométrie	Non	http://textometrie.ens-lyon.fr/	France ;
U-compare	U-Compare is an integrated text mining/natural language processing system based on the UIMA Framework	Libre	Annotation ; Reconnaissance d'entités nommées ; Tokenisation ; PoS Tagging ; Lemmatisation ; Parsing ; Extraction d'évènements ;	Non	http://u-compare.org/	Japon ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
UAIC NLP	The Natural Language Processing (NLP) Group at UAIC-FII has been involved in many national and European projects dealing with: morphology, information retrieval, dialogue systems, anaphora resolution, WordNet, discourse parsing and summarization, question-answering, textual entailment, etc.	Libre	PoS tagging ; Chunking ; Reconnaissance d'entités nommées ; Parsing; Sentence splitting ; Résolution d'anaphore ;	Oui	http://nlptools.info.uaic.ro/Resources.jsp	Roumanie
UDPipe	Trainable pipeline for tokenizing, tagging, lemmatizing and parsing Universal Treebanks and other CoNLL-U files	Libre	Lemmatisation ; POS tagging ; parsing ; Analyse en dépendances ; Annotation ; Tokenisation ;	Non	https://ufal.mff.cuni.cz/udpipe	République Tchèque ;
UltiPro Perception	Understand what employees are saying and how they truly feel about the workplace, with surveys and sentiment analysis.	Commercial	Apprentissage automatique ; Analyse d'opinions ;	Non	https://www.ultimatesoftware.com/UltiPro-Solution-Features-Employee-Surveys	États-Unis ;
Unitex	Open Source Corpus Processing Suite.	Libre	Reconnaissance d'entités nommées ; Désambiguïsation ;	Non	http://unitexgramlab.org/	France ;
Vertica	Data mining and analysis	Commercial	Apprentissage automatique ;	Non	https://www.vertica.com/	États-Unis ;
VisualText	Integrated development environment for building information extraction systems, natural language processing systems, and text analyzers.	Libre/Commercial	Extraction d'informations ; Résumé automatique ; Clustering ; Reconnaissance d'entités nommées ; Recherche d'informations ; Annotation ;	Non	http://www.textanalysis.com/Products/products.html	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
VizTrails	Open-source scientific workflow and provenance management system that supports data exploration and visualization.	Libre	Analyse de données ; Visualisation ;	Non	https://www.vistrails.org/index.php/Main_Page	États-Unis ;
VosViewer	VOSviewer is a software tool for constructing and visualizing bibliometric networks. These networks may for instance include journals, researchers, or individual publications, and they can be constructed based on citation, bibliographic coupling, co-citation, or co-authorship relations. VOSviewer also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature.	Libre	Extraction terminologique ; Co-occurrence ;	Non	http://www.vosviewer.com/	Pays-Bas ;
Vowpal WAbbit	The Vowpal Wabbit (VW) project is a fast out-of-core learning system	Libre	Classification ; Analyse de régression ;	Non	https://github.com/JohnLangford/vowpal_wabbit/wiki	États-Unis ;
Voyant	Environnement en ligne de lecture et d'analyse de textes numériques.	Libre	Clustering ; Concordancier ;	Non	http://voyant-tools.org/	Canada ;
Voziq	Unify every customer experience data source, and apply combined power of text analytics and predictive algorithms for strategic customer intelligence.	Commercial	Analyse de sentiments ; Apprentissage automatique ;	Non	http://voziq.com/	États-Unis ;
WarpLDA	Cache efficient implementation for Latent Dirichlet Allocation	Libre	Topic modeling ;	Non	https://github.com/thu-ml/warplda	Chine ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
WebAnno	WebAnno is a general purpose web-based annotation tool for a wide range of linguistic annotations including various layers of morphological, syntactical, and semantic annotations.Additionaly, custom annotation layers can be defined, allowing WebAnno to be used also for non-linguistic annotation tasks.	Libre	Annotation sémantique ;	Non	https://webanno.github.io/webanno/	Allemagne ;
Weblicht	WebLicht is an execution environment for automatic annotation of text corpora. Linguistic tools such as tokenizers, part of speech taggers, and parsers are encapsulated as web services, which can be combined by the user into custom processing chains. The resulting annotations can then be visualized in an appropriate way, such as in a table or tree format.	Libre ?	Tokenisation ; PoS tagging ; Parsing ;	Oui	https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page	Allemagne ;
Weka	Collection of machine learning algorithms for data mining tasks.	Libre	Classification ; Clustering ; Analyse de régression ; Arbre de décision ; Règle d'association ;	Oui	http://www.cs.waikato.ac.nz/ml/weka/	Nouvelle-Zélande ;
Word2Vec	Word embeddings	Libre	Apprentissage profond ; Plongement de mots	Non	https://github.com/dav/word2vec	États-Unis ;

Outil	Description	Licence	Tâche(s)	OMTD	Source	Pays
WordFreak	WordFreak is a java-based linguistic annotation tool designed to support human, and automatic annotation of linguistic data as well as employ active-learning for human correction of automatically annotated data.	Libre	Annotation ;	Non	http://wordfreak.sourceforge.net/	États-Unis ;
YaTeA	Term extraction	Libre	Extraction terminologique ;	Oui	http://search.cpan.org/~thhamon/Lingua-YaTeA-0.5/	France ;