



**Emperor-OS**

# **All Special Packages**

**Programming and Data science tools**

Editor: Hussein Seilany

Emperor-has 5 package management system for installing, upgrading, configuring, and removing software. Also providing access over 60,000 software packages of Ubuntu. We pre-installed several tools are available for interacting with Emperor's package management system. You can use simple command-line utilities to a graphical tools. Emperor-OS's package management system is derived from the same system used by the Ubuntu GNU/Linux distribution. We will show you the list of installed packages on Emperor-OS. The packages lists is long. Having a list of installed packages helps system administrators maintain, replicate, and reinstall Emperor-OS systems. Emperor-OS Linux systems install dependencies all the time, hence it is essential to know what is on the system.

In this page you can find a specific package and python modules is installed, count installed packages and find out the version of an installed package. Download lists as PDF file or see them as html page in the following:

### **Are you looking for all in one operating system?**

**70 Packages:** Installed Special Packages

**120 Tools:** Installed Utility

**260 Modules:** Installed Python2 and 3 Modules

**600Fonts:** Installed Fonts

**5 Desktops:** Desktop Manager Packages

**22Tools:** Extra Development Tools

**270 Themes:** Installed Themes

**40 Icons:** Installed Icons

**40Games:** Installed Games

**2533scanners:** supports Scanners

**2500 Cameras:** supports Camera

**4338Packages:** All Installed Packages

## A) Programming tools

### 1) Python 2.7 and 3.7.4

**Description:**

Python is an interpreted, interactive, object-oriented, high-level, general-purpose, versatile and popular programming language. It was created by Guido van Rossum, Python source code is also available under the GNU General Public License (GPL). It's great language, because it is Cross-platform, Free and Open Source, Extensible and easy to Learn and Use. Python has a GUI Programming Support.

**Site:**

<https://www.python.org/>

### 2) pycharm

**Description:**

It is a python IDE for developers. All tools of the python in one IDE. It is developed by the Czech company JetBrains. It can be customize with themes and plugins and extremely useful for Python projects.at the end, the best tool for back end development.

**Site:**

<https://www.jetbrains.com/pycharm/download/>

### 3) Jupiter nootbook

**Description:**

JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data. JupyterLab is flexible, configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning. JupyterLab is extensible and modular.

**Site:**

<https://jupyter.org/>

### 4) Java inteliye community

**Description:**

IntelliJ IDEA (as a Java IDE) is an integrated IDE written in Java for developing computer software. It is developed by JetBrains, and is available as an Apache 2 Licensed community edition, and in a proprietary commercial edition. Both can be used for commercial development.

**Site:**

<https://www.jetbrains.com/idea/download/#section=linux>

### 5) NETBEANS

**Description:**

It is an open-source and easily develop desktop, mobile, and web applications with Java, JavaScript, HTML5, PHP, C/C++ and more. NetBeans is an integrated development environment for Java. NetBeans allows applications to be developed from a set of modular software components called modules.

**Site:**

<https://netbeans.org/>

## 6) golang

**Description:**

Go is an open source programming language that makes it easy to build simple, reliable, and efficient software. It has at least two compilers, gc and gccgo. Go is a statically typed, compiled programming language designed at Google by Robert Griesemer, Rob Pike, and Ken Thompson. Go is syntactically similar to C. It is a robust system-level language used for programming across large-scale network servers and big distributed systems.

**Site:**

<https://golang.org/>

## 7) Mono-develop

**Description:**

Cross platform IDE for C#, F# and more. Mono Develop enables developers to quickly write desktop and web applications on Linux. It also makes it easy for developers to port .NET applications created with Visual Studio to Linux and macOS maintaining a single code base for all platforms.

**Site:**

<https://www.monodevelop.com/>

## 8) Django

**Description:**

Django makes it easier to build better Web apps more quickly and with less code. Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

**Site:**

<https://www.djangoproject.com/>

## 9) codelite

**Description:**

Code Lite is an open source, free, cross platform IDE specialized in C, C++, PHP and JavaScript (mainly for backend developers using Node.js) programming languages which runs best on all major Platforms (OSX, Windows and Linux).

**Site:**

<https://codelite.org/>

## 10) bluefish

**Description:**

Bluefish is a powerful editor targeted towards programmers and web developers, with many options to write websites, scripts and programming code. Bluefish supports many programming and markup languages. Bluefish is an open source development project, released under the GNU GPL license.

Bluefish is a multi-platform application that runs on most desktop operating systems including Linux, FreeBSD, Mac OS-X, Windows, OpenBSD and Solaris.

**Site:**

<http://bluefish.openoffice.nl/index.html>

11) Code blocks

**Description:**

Code::Blocks is a free C, C++ and Fortran IDE built to meet the most demanding needs of its users. It is designed to be very extensible and fully configurable. Finally, an IDE with all the features you need, having a consistent look, feel and operation across platforms.

**Site:**

<http://www.codeblocks.org/>

12) Qt creator

**Description:**

Qt is a free and open-source widget toolkit for creating graphical user interfaces as well as cross-platform applications that run on various software and hardware.

it is a free and open-source widget toolkit for creating graphical user interfaces as well as cross-platform applications that run on various software and hardware platforms such as Linux, Windows, mac OS, Android or embedded systems with little or no change in the underlying codebase while still being a native application with native capabilities and speed.

**Site:**

[www.qt.io](http://www.qt.io)

## B) Gui builder

13) Glade

**Description:**

It is a RAD tool to enable quick & easy development of user interfaces for the GTK+ toolkit and the GNOME desktop environment. The user interfaces designed in Glade are saved as XML, and by using the GtkBuilder GTK+ object these can be loaded by applications dynamically as needed. By using GtkBuilder, Glade XML files can be used in numerous programming languages including C, C++, C#, Vala, Java, Perl, Python and others.

**Site:**

<https://glade.gnome.org/>

14) GTK

**Description:**

Offering a complete set of UI elements, GTK is suitable for projects ranging from small one-off tools to complete application suites. Develop your GTK app with your language of choice by using Language Bindings or wrappers and take full advantage of the official GNOME bindings which guarantee API stability and time-based releases. Glade is a RAD tool that enables quick and easy development of user interfaces for the GTK toolkit and the GNOME desktop environment.

**Site:**<https://www.gtk.org>

## 15) kivy

**Description:**

Kivy is Open source Python library for rapid development of applications that make use of innovative user interfaces, such as multi-touch apps. Kivy is a community project, led by professional software developers. We are responsible for developing and supporting Kivy, alongside of the community. We also work for companies that use Kivy for their professional products.

**Site:**<https://kivy.org/#home>

## 16) pyqt

**Description:**

PyQt is a set of Python v2 and v3 bindings for The Qt Company's Qt application framework and runs on all platforms supported by Qt including Windows, OS X, Linux, iOS and Android. PyQt5 supports Qt v5. PyQt4 supports Qt v4 and will build against Qt v5. The bindings are implemented as a set of Python modules and contain over 1,000 classes.

**Site:**<https://riverbankcomputing.com/news>

## 17) TkInter

**Description:**

Tkinter is Python standard GUI (Graphical User Interface) package. It is a thin object-oriented layer on top of Tcl/Tk. Tkinter is not the only GuiProgramming toolkit for Python. It is however the most commonly used one. CameronLaird calls the yearly decision to keep TkInter "one of the minor traditions of the Python world."

**Site:**<https://wiki.python.org/moin/TkInter>

## 18) wxpython

**Description:**

wxPython is a blending of the WxWidgets (formerly WxWindows) class library for C++ and Python. wxPython is a wrapper for the cross-platform GUI API wxWidgets for the Python programming language. It is one of the alternatives to Tkinter, which is bundled with Python. It is implemented as a Python extension module. Other popular alternatives are PyGTK, its successor PyGObject, and PyQt.

**Site:**<https://wiki.python.org/moin/WxPython>

## 19) pygui

**Description:**

This is a project to develop a cross-platform pythonic GUI API. Document the API purely in Python terms, so that the programmer does not need to read the documentation for another GUI

library, in terms of another language, and translate into Python. The goals of this project are develop a GUI API that is designed specifically for Python, taking advantage of Python's unique language features and working smoothly with Python's data types.

**Site:**

[http://www.cosc.canterbury.ac.nz/greg.ewing/python\\_gui/](http://www.cosc.canterbury.ac.nz/greg.ewing/python_gui/)

20) pyside

**Description:**

PySide is an open source software project providing Python bindings for the Qt framework. Qt is a cross-platform application and UI framework, allowing the developers to write applications once and deploy them across many operating systems without rewriting the source code, while Python is a modern, dynamic programming language with a vivid developer community. Combining the power of Qt and Python, PySide provides the wealth of Qt framework for developers writing software in Python and presents a first-class rapid application development platform available on all major operating systems.

**Site:**

[https://wiki.qt.io/About\\_PySide](https://wiki.qt.io/About_PySide)

21) Vi/Vim Editor

**Description:**

Vim is a highly configurable text editor for efficiently creating and changing any kind of text. It is included as "vi" with most UNIX systems and with Apple OS X. Vim is rock stable and is continuously being developed to become even better. Among its features are persistent, multi-level undo tree, extensive plugin system and support for hundreds of programming languages and file formats, powerful search and replace, integrates with many tools.

**Site:**

<https://www.vim.org/>

22) Nano

**Description:**

GNU Nano is a text editor for Unix-like computing systems or operating environments using a command line interface. It emulates the Pico text editor, part of the Pine email client, and also provides additional functionality. Unlike Pico, Nano is licensed under the GNU General Public License.

**Site:**

<https://www.nano-editor.org/>

23) Geany Editor

**Description:**

Geany is a powerful, stable and lightweight programmer's text editor that provides tons of useful features without bogging down your workflow. It runs on Linux, Windows and Mac OS is translated into over 40 languages, and has built-in support for more than 50 programming languages.

**Site:**

<https://www.geany.org/>

#### 24) Atom

**Description:**

Atom is a hackable text editor, built on Electron, and based on everything we love about our favorite editors. We designed it to be deeply customizable, but still approachable using the default configuration. Atom is a free and open-source text and source code editor for mac OS, Linux, and Microsoft Windows with support for plug-ins written in Node.js, and embedded Git Control, developed by GitHub. Atom is a desktop application built using web technologies.

**Site:**

<https://atom.io/>

#### 25) Sublime Text

**Description:**

Sublime Text is a shareware cross-platform source code editor with a Python application programming interface. It natively supports many programming languages and markup languages, and functions can be added by users with plugins, typically community-built and maintained under free-software licenses.

**Site:**

<https://www.sublimetext.com/>

#### 26) VScode

**Description:**

Visual Studio Code is a streamlined code editor with support for development operations like debugging, task running, and version control. It aims to provide just the tools a developer needs for a quick code-build-debug cycle and leaves more complex workflows to fuller featured IDEs, such as Visual Studio IDE. Visual Studio Code is a source-code editor developed by Microsoft for Windows, Linux and mac OS. It includes support for debugging, embedded Git control and GitHub, syntax highlighting, intelligent code completion, snippets, and code refactoring

**Site:**

<https://code.visualstudio.com/>

#### 27) kdevelop

**Description:**

KDevelop - A cross-platform IDE for C, C++, Python, QML/JavaScript and PHP. KDevelop is a free and open-source integrated development environment for Unix-like computer operating systems and Windows. It provides editing, navigation and debugging features for several programming languages, and integration with build automation and version-control systems, using a plugin-based architecture.

**Site:**

<https://www.kdevelop.org/>

#### 28) Anjuta

**Description:**



Anjuta DevStudio is a versatile software development studio featuring a number of advanced programming facilities including project management, application wizard, interactive debugger, source editor, version control, GUI designer, profiler and many more tools. It focuses on providing simple and usable user interface, yet powerful for efficient development.

Features a number of advanced programming facilities including project management, application wizard, interactive debugger, source editor, version control, GUI designer, profiler and many more tools. Focuses on providing simple and usable user interface, yet powerful for efficient development. It supports the following programming languages: C, C++, Java, Javascript, Python, Vala

**Site:**

<http://anjuta.org/>

29) Lazarus IDE

**Description:**

Lazarus is a Delphi compatible cross-platform IDE for Rapid Application Development. It has variety of components ready for use and a graphical form designer to easily create complex graphical user interfaces. Lazarus is a free cross-platform visual integrated development environment (IDE) for rapid application development (RAD) using the Free Pascal compiler.

**Site:**

<https://www.lazarus-ide.org/>

30) wordpress

**Description:**

WordPress is a free and open-source content management system written in PHP and paired with a MySQL or MariaDB database. Features include a plugin architecture and a template system, referred to within WordPress as Themes. WordPress is open source software you can use to create a beautiful website, blog, or app.

**Site:**

<https://wordpress.org>

31) Gambas

**Description:**

Gambas is a free development environment and a full powerful development platform based on a Basic interpreter with object extensions, as easy as Visual Basic. Gambas is the name of an object-oriented dialect of the BASIC programming language, as well as the integrated development environment that accompanies it. Designed to run on Linux and other Unix-like computer operating systems, its name is a recursive acronym for Gambas Almost Means Basic.

**Site:**

<http://gambas.sourceforge.net/en/main.html>

32) Eric

**Description:**

Eric is a full featured Python editor and IDE, written in Python. It is based on the cross platform Qt UI toolkit, integrating the highly flexible Scintilla editor control. It is designed to be usable as everyday's quick and dirty editor as well as being usable as a professional project management

tool integrating many advanced features Python offers the professional coder. eric includes a plug-in system, which allows easy extension of the IDE functionality with plug-ins downloadable from the net.

**Site:**

<https://eric-ide.python-projects.org>

33) PHP

**Description:**

PHP is a popular general-purpose scripting language that is especially suited to web development. Fast, flexible and pragmatic, PHP powers everything from your blog to the most popular websites in the world.

**Site:**

<https://www.php.net/>

34) Apache

**Description:**

The Apache HTTP Server, colloquially called Apache, is a free and open-source cross-platform web server software, released under the terms of Apache License 2.0. Apache is developed and maintained by an open community of developers under the auspices of the Apache Software Foundation

**Site:**

<https://www.apache.org/>

35) MySql

**Description:**

MySQL is an open-source relational database management system. Its name is a combination of "My", the name of co-founder Michael Widenius's daughter, and "SQL", the abbreviation for Structured Query Language.

**Site:**

<https://www.mysql.com/>

## B) Mobile App Development Frameworks

36) Ionic Framework

**Description:**

Ionic is a complete open-source SDK for hybrid mobile app development created by Max Lynch, Ben Sperry, and Adam Bradley of Drifty Co. in 2013. The original version was released in 2013 and built on top of AngularJS and Apache Cordova.

**Site:**

<https://ionicframework.com/framework>

37) JQuery

**Description:**

jQuery Mobile is a HTML5-based user interface system designed to make responsive web sites and apps that are accessible on all smartphone, tablet and desktop devices. It is built on the rock-solid jQuery and jQuery UI foundation, and offers Ajax navigation with page transitions, touch events, and various widgets. Its lightweight code is built with progressive enhancement, and has a flexible, easily theme able design.

**Site:**

<http://jquerymobile.com/>

38) Angular.js Mobile UI Framework with Bootstrap 3

**Description:**

Build HTML5 Mobile Apps with Bootstrap and Angular JS. Mobile Angular UI is a mobile UI framework just like Sencha Touch or jQuery Mobile. If you know Angular JS and Twitter Bootstrap you already know it.

**Site:**

<http://mobileangularui.com/>

39) NativeScript

**Description:**

NativeScript is an open-source framework to develop apps on the Apple iOS and Android platforms. It was originally conceived and developed by Progress. Native Script apps are built using JavaScript, or by using any language that transpiles to JavaScript, such as Type Script

**Site:**

<https://www.nativescript.org/>

40) Onsen

**Description:**

The most beautiful and efficient way to develop HTML5 hybrid and mobile web apps Onsen UI is an open-source UI framework and components for HTML5 hybrid mobile app development, based on PhoneGap / Cordova. It allows developers to create mobile apps using Web technologies like CSS, HTML5, and JavaScript. Onsen UI has become JavaScript framework-agnostic, meaning developers can create mobile apps with or without any JavaScript framework. Onsen UI also provides comprehensive tools and services through Monaca, both products are developed by the same company.

**Site:**

<https://onsen.io/>

41) reactive

**Description:**

A framework for building native apps with React. React Native brings React's declarative UI framework to iOS and Android. With React Native, you use native UI controls and have full access to the native platform.

**Site:**

<http://www.reactnative.com/>

Database

## 42) Sqlite3

**Description:**

SQLite is an in-process library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine. The code for SQLite is in the public domain and is thus free for use for any purpose, commercial or private. SQLite is the most widely deployed database in the world with more applications than we can count, including several high-profile projects. SQLite is a relational database management system contained in a C library. In contrast to many other database management systems, SQLite is not a client-server database engine. Rather, it is embedded into the end program.

**Site:**

<https://www.sqlite.org/index.html>

## 43) postgres

**Description:**

PostgreSQL is a powerful, open source object-relational database system that uses and extends the SQL language combined with many features that safely store and scale the most complicated data workloads. The origins of PostgreSQL date back to 1986 as part of the POSTGRES project at the University of California at Berkeley and has more than 30 years of active development on the core platform.

**Site:**

<https://www.postgresql.org/>

## 44) mongodb

**Description:**

MongoDB is a general purpose, document-based, distributed database built for modern application developers and for the cloud era. No database makes you more productive. MongoDB is a cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with schema. MongoDB is developed by MongoDB Inc. and licensed under the Server Side Public License.

**Site:**

<https://www.mongodb.com/>

## 45) sqlitebrowser

**Description:**

DB Browser for SQLite (DB4S) is a high quality, visual, open source tool to create, design, and edit database files compatible with SQLite. DB4S is for users and developers who want to create, search, and edit databases. DB4S uses a familiar spreadsheet-like interface, and complicated SQL commands do not have to be learned.

**Site:**

<https://sqlitebrowser.org/>

## 46) mariadb

**Description:**

MariaDB is a community-developed, commercially supported fork of the MySQL relational database management system, intended to remain free and open-source software under the GNU General Public License.

**Site:**

<https://mariadb.org>

47) MariaDB Server

**Description:**

It is one of the most popular database servers in the world. It's made by the original developers of MySQL and guaranteed to stay open source. Notable users include Wikipedia, WordPress.com and Google.

**Site:**

<https://mariadb.org>

48) mysql

**Description:**

MySQL is an open-source relational database management system. Its name is a combination of "My", the name of co-founder Michael Widenius's daughter, and "SQL", the abbreviation for Structured Query Language. It is a database system used on the web, system that runs on a server, ideal for both small and large applications, very fast, reliable, and easy to use, compiles on a number of platforms, free to download, developed, distributed, and supported by Oracle Corporation.

**Site:**

<https://www.mysql.com/>

## C) data science

### Core Libraries & Statistics

#### 49) NumPy

**Description:**

NumPy is the fundamental package for scientific computing with Python. It contains a powerful N-dimensional array object, sophisticated (broadcasting) functions tools for integrating C/C++ and Fortran code, useful linear algebra, Fourier transform, and random number capabilities. Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

**Site:**

<http://www.numpy.org/>

#### 50) SciPy library

**Description:**

The SciPy library is one of the core packages that make up the SciPy stack. It provides many user-friendly and efficient numerical routines, such as routines for numerical integration, interpolation, optimization, linear algebra, and statistics. For tutorials, reference documentation, the SciPy roadmap, and a contributor guide, please see the documentation.

**Site:**

<https://scipy.org/scipylib/>

#### 51) Pandas

**Description:**

is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language. In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.

**Site:**

<https://pandas.pydata.org/>

#### 52) statsmodels

**Description:**

It is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. An extensive list of result statistics are available for each estimator. The results are tested against existing statistical packages to ensure that they are correct.

**Site:**

<http://www.statsmodels.org/devel/>

## D) Visualization

## 53) Matplotlib

**Description:**

It is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. Wikipedia

**Site:**

<https://matplotlib.org/index.html>

## 54) Seaborn

**Description:**

It is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

**Site:**

<https://seaborn.pydata.org/>

## 55) Bokeh

**Description:**

It is an interactive visualization library for modern web browsers. It provides elegant, concise construction of versatile graphics, and affords high-performance interactivity over large or streaming datasets. Bokeh can help anyone who would like to quickly and easily make interactive plots, dashboards, and data applications. Bokeh is an interactive visualization library for modern web browsers. It provides elegant, concise construction of versatile graphics, and affords high-performance interactivity over large or streaming datasets. Bokeh can help anyone who would like to quickly and easily make interactive plots, dashboards, and data applications.

**Site:**

<https://bokeh.pydata.org/en/latest/>

## 56) pydot

**Description:**

It is an interface to Graphviz can parse and dump into the DOT language used by GraphViz, and written in pure Python. This module provides with a full interface to create handle modify and process graphs in Graphviz's dot language and networkx can convert its graphs to pydot. Development occurs at GitHub (under branch dev) , where you can report issues and contribute code.

**Site:**

<https://pypi.org/project/pydot/>

## E) Machine Learning

## 57) Scikit-learn

**Description:**

Simple and efficient tools for predictive data analysis Accessible to everybody, and reusable in various contexts. Built on NumPy, SciPy, and matplotlib . Open source, commercially usable - BSD license. scikit-learn is a Python module for machine learning built on top of SciPy and is distributed under the 3-Clause BSD license. It is a free software machine learning library for the Python programming language.

It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

**Site:**

<http://scikit-learn.org/stable/>

58) ELI5

**Description:**

It is a Python library which allows to visualize and debug various Machine Learning models using unified API. It has built-in support for several ML frameworks and provides a way to explain black-box models. ELI5 is a Python package which helps to debug machine learning classifiers and explain their predictions. It provides support for the following machine learning frameworks and packages. Currently ELI5 allows to explain weights and predictions of scikit-learn linear classifiers and regressors, print decision trees as text or as SVG, show feature importance and explain predictions of decision trees and tree-based ensembles.

**Site:**

<https://eli5.readthedocs.io/en/latest/>

## F) Deep Learning

59) PyTorch

**Description:**

PyTorch is a Python package that provides two high-level features as Tensor computation (like NumPy) with strong GPU acceleration, Deep neural networks built on a tape-based autograd system. You can reuse your favorite Python packages such as NumPy, SciPy and Cython to extend PyTorch when needed. Usually PyTorch is used either as a replacement for NumPy to use the power of GPUs. A deep learning research platform that provides maximum flexibility and speed. PyTorch is not a Python binding into a monolithic C++ framework. It is built to be deeply integrated into Python. You can use it naturally like you would use NumPy / SciPy / scikit-learn etc. You can write your new neural network layers in Python itself, using your favorite libraries and use packages such as Cython and Numba. Our goal is to not reinvent the wheel where appropriate.

**Site:**

<https://pytorch.org/>

60) keras

**Description:**

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast



experimentation. Being able to go from idea to result with the least possible delay is key to doing good research. Use Keras if you need a deep learning library that allows for easy and fast prototyping.

**Site:**

<https://keras.io/>

## G) Natural Language Processing

61) NLTK

**Description:**

It is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum. NLTK is open source software. The source code is distributed under the terms of the Apache License Version 2.0. The documentation is distributed under the terms of the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States license. The corpora are distributed under various licenses, as documented in their respective README files.

**Site:**

<https://www.nltk.org/>

62) spaCy

**Description:**

It is an open-source software library for advanced natural language processing, written in the programming languages Python and Cython. spaCy is a free open-source library for Natural Language Processing in Python. It features NER, POS tagging, dependency parsing, word vectors and more.

**Site:**

<https://spacy.io/>

63) Gensim

**Description:**

Started off as a collection of various Python scripts for the Czech Digital Mathematics Library dml.cz in 2008, where it served to generate a short list of the most similar articles to a given article. Gensim is an open-source library for unsupervised topic modeling and natural language processing, using modern statistical machine learning. Gensim is implemented in Python and Cython.

**Site:**

<https://radimrehurek.com/gensim/>

## H) Data Scraping

64) scrapy

**Description:**

An open source and collaborative framework for extracting the data you need from websites. In a fast, simple, yet extensible way. Description Scrapy is a free and open-source web-crawling framework written in Python. Originally designed for web scraping, it can also be used to extract data using APIs or as a general-purpose web crawler. It is currently maintained by Scrapinghub Ltd., a web-scraping development and services company

**Site:**

<https://scrapy.org/>

## I) Big Data frameworks

65) The Apache Hadoop

**Description:**

Software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.

**Site:**

<https://hadoop.apache.org/>

66) Apache Spark

**Description:**

It is a unified analytics engine for large-scale data processing. Write applications quickly in Java, Scala, Python, R, and SQL. Apache Spark is an open-source distributed general-purpose cluster-computing framework. Spark provides an interface for programming entire clusters with implicit data parallelism and fault tolerance Apache Spark is a unified analytics engine for big data processing, with built-in modules for streaming, SQL, machine learning and graph processing.

**Site:**

<https://spark.apache.org/>

## J) Visualization tools

67) weka

**Description:**

Waikato Environment for Knowledge Analysis, developed at the University of Waikato, New Zealand. It is free software licensed under the GNU General Public License, and the companion software to the book "Data Mining: Practical Machine Learning Tools and Techniques" Weka is tried and tested open source machine learning software that can be accessed through a graphical user interface, standard terminal applications, or a Java API. It is widely used for teaching, research, and industrial applications, contains a plethora of built-in tools for standard

machine learning tasks, and additionally gives transparent access to well-known toolboxes such as scikit-learn, R, and Deeplearning4j.

**Site:**

<https://www.cs.waikato.ac.nz/ml/weka/>

68) knime

**Description:**

KNIME Analytics Platform is available for Windows, Linux and Mac OS. In this chapter, let us look into the steps for installing the platform on the Mac. If you use Windows or Linux, just follow the installation instructions given on the KNIME download page. The binary installation for all three platforms is available at KNIME's page. At KNIME®, we build software for fast, easy and intuitive access to advanced data science, helping organizations drive innovation. Our KNIME Analytics Platform is the leading open solution for data-driven innovation, designed for discovering the potential hidden in data, mining for fresh insights, or predicting new futures. Organizations can take their collaboration, productivity and performance to the next level with a robust range of commercial extensions to our open source platform.

**Site:**

<https://www.knime.com/>

69) orange

**Description:**

Orange is an open-source data visualization, machine learning and data mining toolkit. It features a visual programming front-end for explorative data analysis and interactive data visualization. Open source machine learning and data visualization for novice and expert. Interactive data analysis workflows with a large toolbox.

**Site:**

<https://orange.biolab.si/>

70) DataMelt

**Description:**

DataMelt is a free software for numeric computation, mathematics, statistics, symbolic calculations, and data analysis and data visualization. This multiplatform program combines the simplicity of scripting languages, such as Python, Ruby, Groovy (and others), with the power of hundreds of Java packages. DataMelt, or DMelt, is software for numeric computation, statistics, analysis of large data volumes ("big data") and scientific visualization. The program can be used in many areas, such as natural sciences, engineering, modeling and analysis of financial markets.

**Site:**

<https://datamelt.org/?id=current>

71) RStudio

**Description:**

It is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. RStudio's mission is to create free and open-source software for data science, scientific research, and technical

communication. We do this to enhance the production and consumption of knowledge by everyone, regardless of economic means, and to facilitate collaboration and reproducible research, both of which are critical to the integrity and efficacy of work in science, education, government, and industry.

**Site:**

<https://rstudio.com/products/rstudio/>

## 72) Massive Online Analysis

**Description:**

MOA is the most popular open source framework for data stream mining, with a very active growing community (blog). It includes a collection of machine learning. Massive Online Analysis (MOA) is a free open-source software project specific for data stream mining with concept drift. It is written in Java and developed at the ... Massive Online Analysis is a free open-source software project specific for data stream mining with concept drift. It is written in Java and developed at the University of Waikato, New Zealand.

**Site:**

<https://moa.readthedocs.io/en/latest/install.html>

## 73) The Databionic ESOM Tools

**Description:**

It is a suite of programs to perform data mining tasks like clustering, visualization, and classification with Emergent Self-Organizing Maps (ESOM). Features include Training of ESOM with different initialization methods, training algorithms, distance functions, parameter cooling strategies, ESOM grid topologies, and neighborhood kernels. Visualization of high dimensional dataspace with U-Matrix, P-Matrix, Component Planes, SDH, and more. Animated visualization of the training process. Interactive, explorative data analysis and clustering by linking ESOM to the training data, data classifications, and data Descriptions.

Creation of ESOM classifier and automated application to new data. Creation of non-redundant U-Maps from toroid ESOM.

**Site:**

<http://databionic-esom.sourceforge.net/>

## 74) Theano

**Description:**

Theano is a Python library and optimizing compiler for manipulating and evaluating mathematical expressions, especially matrix-valued ones. In Theano, computations are expressed using a NumPy-esque syntax and compiled to run efficiently on either CPU or GPU architectures. Theano is a Python library that allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently.

**Site:**

[http://deeplearning.net/software/theano/install\\_ubuntu.html](http://deeplearning.net/software/theano/install_ubuntu.html)

## 75) pybrain

**Description:**

PyBrain is a modular Machine Learning Library for Python. Its goal is to offer flexible, easy-to-use yet still powerful algorithms for Machine Learning Tasks and a variety of predefined environments to test and compare your algorithms. While there are a few machine learning libraries out there, PyBrain aims to be a very easy-to-use modular library that can be used by entry-level students but still offers the flexibility and algorithms for state-of-the-art research. We are constantly working on more and faster algorithms, developing new environments and improving usability.

**Site:**

<http://pybrain.org/pages/download>

## K) Scraping tools

### 76) contentgrabber

**Description:**

Web-scraping tools generally use macros or configuration methods, and follow a sequential list of commands. The macro approach is more user friendly and automatically records the actions of a user in a browser. However, there are typically restrictions on accessing the code behind the agent. The configuration approach allows the user to directly configure each part of the agent. They can introduce more code structure, controls, data refinements, or add their own naming conventions.

**Site:**

[https://contentgrabber.com/Manual/understanding\\_the\\_concept.htm](https://contentgrabber.com/Manual/understanding_the_concept.htm)

### 77) ParseHub

**Description:**

It is a free web scraping tool. Turn any site into a spreadsheet or API. As easy as clicking on the data you want to extract. ParseHub is a free and powerful web scraping tool. With our advanced web scraper, extracting data is as easy as clicking on the data you need. Download ParseHub for Free. An error occurred. Try watching this video on [www.youtube.com](http://www.youtube.com), or enable JavaScript if it is disabled in your browser.

**Site:**

<https://www.parsehub.com/>