

Apache Nutch

	Fill in	Note
Name of the software	Apache Nutch	
Developer	Apache Software Foundation	
Version	Apache Nutch 1.16 Apache Nutch 2.4	
Year of publication	First release V1: 06.06.2010 Newest version V1: 11.10.2019 First release V2: 07.07.2012 Newest version V2: 11.10.2019	
Description of the software	<p>Apache Nutch is a highly extensible and scalable open source web crawler software project. Nutch is coded entirely in the Java programming language, but data is written in language-independent formats. It has a highly modular architecture, allowing developers to create plug-ins for media-type parsing, data retrieval, querying and clustering.</p> <p>The fetcher ("robot" or "web crawler") has been written from scratch specifically for this project.</p>	
Link to the software	nutch.apache.org	GitHub: https://github.com/apache/nutch
Is it open-source software?	YES	
Is it free software?	YES	
License	Apache License 2.0	
Does it cost to use?	NO	

Does it collect any personal data?	It sends and receives cookies like a normal browser would do, but no form of log in or personal information is required.	
Does it respect the right to forget?		Software should have a function to delete all data. The deletion should be permanent and not reversible.
Application logs	Scrapy uses Java's logging system.	
Does it contain an online user guide?	https://cwiki.apache.org/confluence/display/NUTCH/Home#Home-Tutorials	
Hardware configuration	https://cwiki.apache.org/confluence/display/NUTCH/HardwareRequirements	
Software configuration	Runs on Linux, Windows, Mac OS X, BSD.	
Text data visualisation	NO	
Does it analyse pictures/photos?	NO	
Multiple languages	NO	
Third-party software products integration	NO	
Cross-platform support for desktop, tablet and mobile users	NO	
Scalability	Scalable	
Automatically creates database	NO	
Uses AI	NO	

