

LexPredict ContraxSuite Documentation

Software and Data Dependencies

Release 1.0.1 - September 1, 2017

How to Get Support	1
Software	2
Google Tesseract	2
Apache Tika	2
NLTK	3
Stanford NLP	3
ElasticSearch	4
jqWidgets	4
Data and Media	5
LexPredict Knowledge Sets	5
Canvas	5

How to Get Support

For support or help in setting up the application, please contact support@contraxsuite.com.

Software

Google Tesseract

Tesseract is the world's most-tested OCR engine, first developed by HP in the 1980s, then taken over by Google in the 2000's. Tesseract has been used to process more pages of material than any other tool in the history of time. Tesseract has UTF-8 support, ships over 100 languages with pre-trained models, and can be trained to recognize any document or language. Tesseract remains under active development and support, and new versions are obtaining even better performance.

Tesseract itself is licensed under an Apache License Version 2.0, and therefore can be distributed and re-used commercially. While ContraxSuite does not include Tesseract itself, many operating systems or other packages may distribute Tesseract.

Tesseract models may be obtained for over 100 languages. Most Linux operating systems such as Debian/Ubuntu or Red Hat make these languages available through operating system package managers. However, they can also be obtained here:

- <https://github.com/tesseract-ocr/langdata>
- <https://github.com/tesseract-ocr/tessdata>

More information about Tesseract can be found here:

- <https://github.com/tesseract-ocr/>
- <https://github.com/tesseract-ocr/tesseract>
- <https://github.com/tesseract-ocr/tesseract/wiki>
- <https://github.com/tesseract-ocr/tesseract/wiki/TrainingTesseract>

Apache Tika

ContraxSuite relies on the Apache Tika library to assist in the extraction of metadata and textual content from a wide variety of file formats. Tika is maintained by the Apache Foundation, and has been under active development since 2007. Apache Tika is available under the Apache License Version 2.0, and can be used by organizations without a commercial license.

More information about Apache Tika can be found below:

- <https://tika.apache.org/>
- <https://wiki.apache.org/tika/>

NLTK

ContraxSuite's core NLP functions rely on NLTK, which is the most mature and featureful Python library for natural language processing. NLTK is an open source project licensed under the Apache License Version 2.0.

In addition to NLTK's core software, some functionality may rely on corpora that are available to be downloaded through NLTK. These corpora are licensed through a range of permissive to restrictive licenses. All corpora used or installed as part of ContraxSuite's default configuration are suitable for organizations to use internally.

All NLTK data and licenses can be found here:

https://github.com/nltk/nltk_data

More information can be found at the NLTK Documentation below:

- <http://www.nltk.org/>
- <http://www.nltk.org/install.html>
- <http://www.nltk.org/book/>

Stanford NLP

Many ContraxSuite implementations rely on Stanford NLP libraries for built-in or customized functionality. In the Stanford NLP Group's own words, they provide:

“statistical NLP, deep learning NLP, and rule-based NLP tools for major computational linguistics problems, which can be incorporated into applications with human language technology needs. These packages are widely used in industry, academia, and government.”

This library, like ContraxSuite, is available under a dual-license model with GPL v2/3 or commercial through the Stanford Office of Technology Licensing. In general, internal usage or continued adherence to GPL terms should allow organizations to use the library under the open license model.

As is commonly the case with interpreted languages like Java and Python, a separate, MIT-licensed client library is available to enable organizations to avoid integrating the core library. This client can be found here:

<https://github.com/stanfordnlp/python-stanford-corenlp/blob/master/LICENSE>

More information can be found at the Stanford NLP Group below:

- <https://nlp.stanford.edu/software/>
- <https://stanfordnlp.github.io/CoreNLP/>
- <https://nlp.stanford.edu/software/lex-parser.html>

- <https://nlp.stanford.edu/software/tagger.html>
- <https://nlp.stanford.edu/software/CRF-NER.html>
- <https://nlp.stanford.edu/software/segmenter.html>
- <https://nlp.stanford.edu/software/sutime.html>

ElasticSearch

ContraxSuite relies on ElasticSearch to provide deep search within documents. ElasticSearch is distributed, highly-available, multi-type, and near-real-time. ElasticSearch can be easily installed on Windows using an MSI or Linux using operating system package managers like `apt` or `yum`.

While ElasticSearch is not distributed with ContraxSuite, ElasticSearch is licensed under the open source Apache License Version 2.0. No commercial license is required to utilize ElasticSearch.

jqWidgets

ContraxSuite's default UI relies on jqWidgets to provide rich, interactive, and cross-device widgets like tables and charts. While you can use ContraxSuite via API alone or develop your own user interface to replace jqWidgets, it is recommended that most organizations begin by utilizing jqWidgets.

As of July 2017, jqWidgets license guidance permits the usage of jqWidgets for the purpose of testing applications with jqWidgets.

Production usage of ContraxSuite generally requires the purchase of a commercial license for jqWidgets. As of August 2017, licenses for most deployments cost \$199 USD and include support.

Under certain conditions, some other organizations can use jqWidgets non-commercial free of charge. Please refer to the jqWidgets license page for more details:

- <http://www.jqwidgets.com/>
- <http://www.jqwidgets.com/license/>

Data and Media

LexPredict Knowledge Sets

ContraxSuite relies on a number of “Knowledge Sets” that LexPredict has assembled. These knowledge sets are distributed through Github under Creative Commons Attribution Share Alike 4.0 license. More information about available knowledge sets can be found in the Knowledge Sets documentation and on Github at <https://github.com/LexPredict/lexpredict-legal-dictionary>.

Canvas

ContraxSuite’s default UI relies on the Canvas HTML5 theme to provide a modern, cross-platform interface. While you can use ContraxSuite via API alone or develop your own user interface to replace Canvas, it is recommended that most organizations begin by utilizing Canvas.

The Canvas theme can be licensed for \$16, including support. The theme is available for purchase at the Envato Marketplace here:

<https://themeforest.net/item/canvas-the-multipurpose-html5-template/9228123>

Please refer to the Canvas license page for more details.