

Apache OpenNLP 2017 in Review

A look back at OpenNLP's 2017.

Summary

Open NLP got off to a quick start in 2017 thanks to a 1.7.0



release on December 31, 2016. This version added support for Java 8 and set the tone for OpenNLP's 2017. In total, there were 7 releases in 2017. OpenNLP also got a new logo and website in 2017 with an updated look and easier navigation. OpenNLP also released its first model, a language detection model capable of identifying 103 languages. OpenNLP moved to [GitHub](#) for source management greatly simplifying the process of reviewing and merging pull requests.

Some features and improvements that were added to OpenNLP in 2017 include:

- A new language model CLI tool.
- Moses format support.
- CONLL-U format support.
- Language codes now are ISO 639-3 compliant.
- Many more unit tests.
- Prefix and suffix feature generators are now configurable.
- Learnable lemmatizer now returns all possible lemmas for a given word and part-of-speech tag.
- A new language detection component and trained language model.
- Evaluation tests now support ISO-639-3 language codes.
- Fixed handling of xml parsers used through out the package.
- New experimental API for word vectors and support for GloVe vector files.
- Added annotator notes to BratAnnotator.
- Add 20Newsgroups format support to the doccat component.
- Resolved concurrency issue in POS tagger.

Community Development

Apache OpenNLP has added 6 new committers and PMC members in 2017.

Talks and Presentations

Apache OpenNLP was presented at several events in 2017 and there will be more OpenNLP talks in 2018 across the world.

Title	Media
Deriving Actionable Insights from High Volume Media Streams by Peter Thygesen and Jörn Kottmann	https://www.youtube.com/watch?v=ZkInPRApV60
Embracing Diversity: Searching over multiple languages Tommaso Teofili and Suneel Marthi, Berlin Buzzwords, Berlin Germany, June 12, 2017	https://www.youtube.com/watch?v=ZrWxySF-9KY&index=34&list=PLq-odUc2x7i-9Nijx-WfoRMOAfHC9XzTt
A Deep Text Analysis System based on OpenNLP Boris Galitsky, ApacheCon Europe 2016, Seville Spain, November 2016	https://feathercast.apache.org/2017/03/17/apachecon-seville-2016-a-deep-text-analysis-system-based-on-opennlp-boris-galitsky/
It takes a Village to solve a Problem in Data Science Daniel Russ, Data Science Maryland Meetup, Laurel Maryland, June 19, 2017	
Large Scale Processing of Text Suneel Marthi, Hadoop Summit/DataWorks Summit, San Jose California, June 15, 2017	

Releases

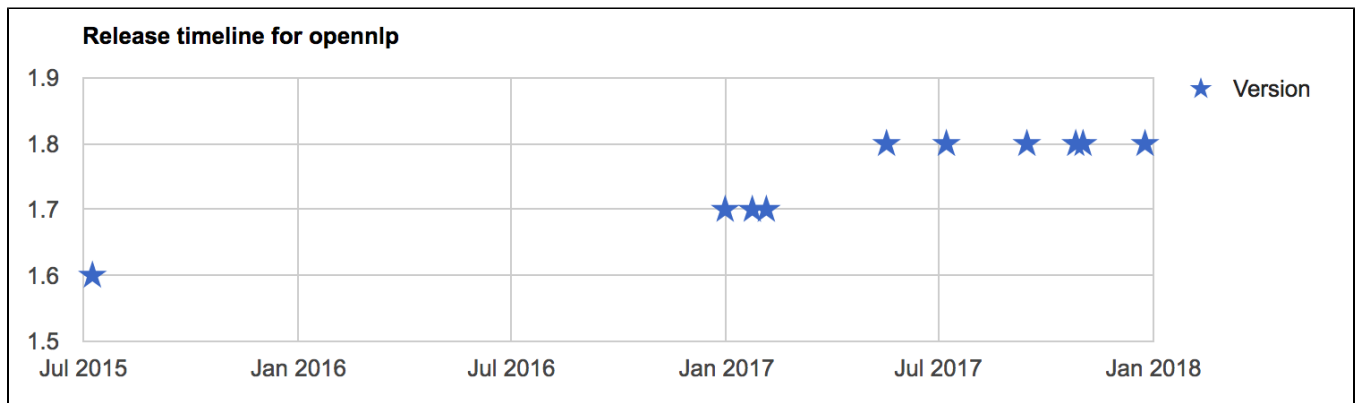
OpenNLP had 7 releases in 2017. They were:

- [1.8.4](#) - December 25, 2017
- [1.8.3](#) - October 26, 2017
- [1.8.2](#) - September 15, 2017
- [1.8.1](#) - July 8, 2017
- [1.8.0](#) - May 18, 2017
- [1.7.2](#) - February 4, 2017
- [1.7.1](#) - January 23, 2017
- [1.7.0](#) - December 31, 2016

Models

The OpenNLP team was very excited to announce the language detection model's [release](#) on November 2, 2017. This model is capable of identifying 103 languages. The model is available for [download](#) from the OpenNLP website.

OpenNLP Release Timeline



Activity

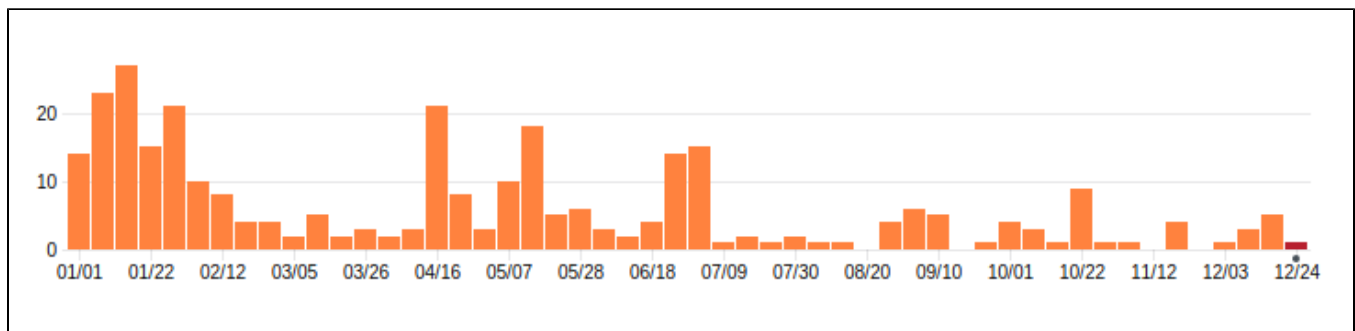
OpenNLP added 6 new committers and PMC members in 2017. There are currently 21 [committers](#) and 15 PMC members.

Tasks

- 289 JIRA tasks were closed in 2017.
- 346 JIRA tasks were opened in 2017.

Code

- There were [269](#) closed pull requests.
- There were 323 git commits throughout the year:



Notable Use of OpenNLP



OpenNLP powers an Air New Zealand Oscar chat bot.

"Air New Zealand uses OpenNLP to power its chatbot, Oscar. Launched in February 2017, Oscar provides a conversational interface for customers to ask questions about flights, amenities and policies. Using OpenNLP, we've been able to consistently provide over 50% conversational success and support hundreds of intents."

[See other uses of OpenNLP.](#)