

TikaResources

This is a collection of resources that talk about Tika, or provide case studies of using Tika. People should feel free to add any publicly available information they find about Tika.

- [Books](#)
- [Presentations](#)
- [Articles / Blogs](#)
- [Tutorials](#)
- [Podcasts](#)

Books

- [Tika in Action](#) (by Chris A. Mattmann and Jukka Zitting)
(Published November 2011)

Presentations

- [Evaluating Text Extraction: Apache Tika's New tika-eval Module](#) and [video](#) (by Tim Allison)
(Presented: May 18, 2017 at ApacheCon North America 2017)
- [Evaluating Text Extraction, Developing a Toolkit for Apache Tika](#) (by Tim Allison and Paul Herceg)
(Presented: April 15, 2015 at ApacheCon North America 2015)
- [JPL Tika Overview](#) (by Lewis John McGibbney and Tyler Palsulich)
(Presented: August 7, 2014 at JPL)

Articles / Blogs

- [Tika Tuesdays, a series of blogs on Tika](#) (by Eric Pugh)
(Series started in late 2019, continuing in 2020 - blog)
- [Lessons Learned from rtika, a Digital Babel Fish](#) (by Sasha Goodman)
(Published: April 25, 2018 - blog)
- [Apache Tika's Regression Corpus](#) (by Tim Allison)
(Published: October 4, 2016 - blog)
- [Getting Text Out Of Anything \(docs, PDFs, Images\) Using Apache Tika](#) (by Tony Hirst)
(Published: February 9, 2015 - blog)
- [Collecting Data to Improve Tools](#) (by Andy Jackson)
(Published: January 30, 2015 - article)
- [Tika in Action Reading Notes](#) (by Rishi Verma)
(Published: January 21, 2015 - blog)
- [A Tika to ride; characterising web content with Nanite](#) (by William Palmer)
(Published March 21, 2014)
- [The Next Steps for the Digital Babel Fish](#) (by Chris A. Mattmann)
(Published: August 1, 2014 - blog)
- [Content mining with Apache Tika](#) (by Juliet Kemp)
(Published: September 23, 2013 - article)
- [Text feature selection for machine learning – part 2](#) (by Ken Krugler)
(Published: July 21, 2013 - article)
- [Text feature selection for machine learning – part 1](#) (by Ken Krugler)
(Published: July 11, 2013 - article)
- [Using Apache Tika from Python with JNIUS](#) (by Samuele Santi)
(Published: May 13, 2013 - article)
- [Content Detection, Metadata and Content Extraction with Apache Tika](#) (by Micha Kops)
(Published: December 2, 2012 - article)
- [Understanding Information Content with Apache Tika](#) (by Chris A. Mattmann and Oleg Tikhinov)
(Published: June 15, 2010 - article)
- [Content Extraction with Apache Tika and Solr](#) (by Sami Siren)
(Published: January 2009 - article)
- [Using the Tika Java Library In Your .Net Application With IKVM](#) (by Kevin Miller)
(Published: July 02, 2010 - article)

Tutorials

- [TikaOnDotNet 1.4 Released as a Nuget](#) (by Kevin Miller)
(Published: July 12, 2013 - tutorial for using Tika on .NET platform)

Podcasts

- [Apache Tika Feathercast](#) - A podcast interviewing Chris Mattmann about Apache Tika. Interview conducted by Rich Bowen, of [Feathercast.org](#).