

Getting Started: Open Source Patent Software Tools

Paul Oldham

24 March 2015

Basic Tools for Open Source Patent Analysis

In developing the Open Source Manual we will be using a number of tools. To get started it you might want to download and sign up for some or all of the the following:

- ▶ Open Office
- ▶ Google Spreadsheets and Google Charts (Google Account)
- ▶ Open Refine
- ▶ Gephi
- ▶ RStudio
- ▶ Python
- ▶ Tableau Public

Open Office

- ▶ Download and Install Apache Open Office
- ▶ Open Office is an open source alternative to Microsoft Office
- ▶ You will mainly be using the Open Office Spreadsheets.
- ▶ Tip: When saving files save as **.csv** to avoid situations where a programme can't read **.odt** files.
- ▶ Excel is not open source, but if you are very familiar with it you may want to carry on. Note that Excel introduces extra invisible characters that will cause problems at various points.

Google Charts

- ▶ Create a Google Account to Access Google Spreadsheets and other Google programmes
- ▶ Take a look at the Google Charts Gallery and API
- ▶ As part of the sign up you may want to download and install Google Drive to store and share files. As an alternative use Dropbox

Github

Github is a platform for collaborating on open source projects using online repositories. There are two aspects to installing Github (git and Github itself).

- ▶ The Github website will lead you through the instructions.
- ▶ Github bootcamp is useful as an introduction.

Open Refine website

Used to assist with data cleaning and manipulation. Formerly known as Google Refine and now an open source tool.

- ▶ [Download Open Refine for your Platform here](#)

Network Visualization

Gephi is a powerful Java network visualization programme. It is eternally in beta but can be used to visualise concepts, inventors and applicants.

- ▶ Download gephi
- ▶ Note that installing gephi can cause problems for Mac users with respect to the Java version (Apple disables Java by default). Follow the help pages on this. Work is ongoing to address this problem on Macs.
- ▶ Cytoscape is another network visualisation programme
- ▶ Pajek is another (windows based) programme that is quite popular.
- ▶ Vos Viewer is also quite popular.

Dashboarding and Interactivity

- ▶ Tableau Public. A free version of the Tableau Desktop visualization and analytics software that allows for the creation of interactive dynamic graphs and geocoded maps. The Gallery shows what can be done. Not for use with sensitive data.
- ▶ Shiny apps for R users allows for the creation of online interactive apps (upto 5 for free). See the Gallery for examples. See RBloggers for more examples and tutorials

R and RStudio

1. “R” is a programming language that is mainly used for statistics. It sounds intimidating but thinking of it as “Arrrrr... me hearties” and Captain Jack Sparrow may help. RStudio has made R much more accessible to non-programmers.
2. One of the strengths of R is that it is possible to import, clean, graph, visualize and map patent data in one place.
3. There is plenty of support for using R, such as Rbloggers
4. John Hopkins University offers excellent free courses on programming with R through Coursera. The courses are tough but rewarding.
5. The open source patent analysis manual will walk through some of the basic steps in using R to get you started.

Python

1. Python is an easy to learn programming language that some researchers are already using for patent research.
2. Existing tools for working with patents include George Song's EPO OPS Code.
3. Patentsnapper for USPTO patents.
4. Patent 2 Net is an open source Python package to allow researchers, small companies and developing countries to access patent data using the EPO OPS service.

Text Mining

This list will grow quite a lot in the coming weeks and months. But here are some we found interesting.

17. Jigsaw Visual Analytics for exploring and understanding document collections.
18. Weka Java based text mining software.
19. Word Trees: claims analysis, various options. See for example Google Word Trees and [Jason Davies(<https://www.jasondavies.com/wordtree/>)] tree creator.

Other Tools and Resources

This is a mixed list of other tools we have heard about but need to take a proper look at.

1. KH Coder: free software allowing quantitative content analysis/text mining
2. LabPlot: KDE-application for interactive graphing and analysis of scientific data, allowing the creation, managing and editing of plots, based on data from a spreadsheet or on data imported from external files. Plots can be exported to several pixmap and vector graphic formats. creating and editing plots
3. imDEV: application of RExcel, integrating Excel and R for multivariate data visualization, exploration, and analysis. Interactive modules for dimensional reduction (imPCA), prediction (imPLS), feature selection, analysis of correlation (imCorrelations) and generation of networked structures (imGraph) provide an integrated environment for systems level analysis of multivariate data.

Other Tools (continued)

4. Tulip: data visualization software
5. SigmaJS: JavaScript library dedicated to graph drawing. It allows the creation of interactive static and dynamic graphs
6. Kendo UI: Create widgets for responsive visualisations.
7. Timeline: A KnightLab (northwestern university) is a tool allowing for the creation of interactive timelines and is available in 40 languages

Other Tools (continued)

8. Miso Project: open source toolkit facilitating the creation of interactive storytelling and data visualization
9. Sci2: A toolset for studying science.
10. IBM Many Eyes visualizations. More of a website than a tool but worth a look around?