Article

Science Mapping and Visualization Tools used in Bibliometric & Scientometric Studies: An Overview

Pallab Pradhan, Scientist B (LS)

Science mapping and visualization techniques are used for data analysis in bibliometric and scientometric studies. Science mapping and visualization helps to explore the scientific knowledge, make it livelier by visualizing and showcasing the impact of research by displaying the structural relationship and dynamics of scientific research domains. Various software tools have been developed incorporating different science mapping and visualization techniques. This article briefly presents features and functions of such software tools used for data analysis in bibliometric and scientometric studies.

1. Introduction

In the fastest growing technological world of information communication technology & scientific research and development, an overwhelming amount of information / data in various formats is generated directly or indirectly. As far as academic and scientific community is concerned, a large number of scholarly articles are being published on daily basis by research scholars and academician across the world. Bibliometric and scientometric analysis, science mapping and visualization techniques are used to understand the evolution of scientific knowledge, intellectual concepts, social structures, inter-connections and inter-relationships of thousands of, millions of scholarly articles. Bibliographies, references and citation data retrieved from databases like Google Scholar, Web of Science and Scopus are used to conduct such analysis and studies to measure the scientific productivity, and to analyse impact of scientific research embedded in research articles.

According to Encyclopaedia of Information Science and Technology, IGI Global: "Science mapping is the development and application of computational techniques to the visualization, analysis, and modelling of a broad range of scientific and technological activities as a whole". Mapping of science is nothing but to represent complex, abstract or raw data in a visually understandable format. Mapping of science helps to clearly understand the concepts of science, making them more lively visible and concrete. The fundamental principle of science mapping is to represent the body of scientific literature in a tangible form so that one can handle it more effectively. A science map provides an overview of the scientific landscape which can be used to support exploration, description, or explanation of the state and development of scientific knowledge and practices. Citations provide insights into value of published article by the scientific community.

The workflow of science mapping consists of a number of steps including data retrieval, pre-processing, network extraction, normalization, mapping, analysis and visualization. At the end of this process, the analyst has to interpret and obtain some conclusions from the results (Cobo et al., 2011).According to Tech Target, "Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easily with data visualization software.

Bibliometrics / Scientometrics involve quantitative studies of scientific literatures. These studies include

topics like data analysis, science mapping, and visualization, etc. The Online Dictionary of Library and Information Science (ODLIS) defines Scientometrics software as set of computer based programs, designed and developed to analyse citation based bibliographic data as input to perform the specific tasks i.e. structural analysis of scholarly communication, mapping of scientific research, creation of metrics based social maps, information representation and organisation, visualisation of research, micro level analysis (co-word, co-author, cited references, bibliographic coupling, co-citation) etc. as output (Kumar, Shivarama, &Choukimath, 2015).

There are a number of science mapping and visualization software tools specifically used in bibliometric &scientometric studies available free of cost on the Internet. Most of these software tools are based on modern mathematical algorithms, statistical methods, graphs theory, sophisticated networks theory and visualization techniques, etc. This article briefly elaborates features and functions of software tools available for science mapping and visualization.

2. Bibliometric & Scientometric Analysis, Science Mapping, and Visualization Software Tools

2.1 BibExcel

Bibexcel is a versatile bibliometric toolbox developed by Olle Persson, Department of Sociology, Umea University, Umea, Sweden. BibExcel is specifically designed to assist a user in analysing bibliographic data, or any data of textual nature formatted in a similar manner. The idea is to generate data files that can be imported to Excel, or any program that takes tabbed data records, for further processing. BibExcel can read data retrieved from different bibliographic sources, such as ISI Web of Science (WoS), Scopus in Procite export format, also it can convert to other formats. In Bibexcel, it is possible to do most types of bibliometric analysis, e.g. bibliometric, citation analysis, co-citation, shared references, bibliographic coupling, cluster analysis, prepare bibliometric maps for mapping with Pajek, NetDraw.

This tool-box includes a number of tools, some of them visible in the window interface and others hide behind the menus. Many of the tools can be used in combination to achieve the desired result. The program offers a high degree of flexibility to the users in both data management and analysis which is one of the real strengths of this tool. Bibexcel allows easy interaction with other software, e.g. Pajek, Excel, SPSS, etc. BibExcel can produce net-files for co-authorships, co-citations etc, and then convert these files for further analysis and visualization with Pajek. BibExcel can also make clu-files and vec-files for Pajek which is freely available on the net.

BibExcel is a free-ware for academic and non-profit use. The software can be downloaded from http://homepage.univie.ac.at/juan.gorraiz/bibexcel/.

2.2 BiblioTool

BiblioTools 2.2, is a set of python scripts developed by Sebastian Grauwin for transforming raw bibliographic data as extracted from the Web of Science into “maps of science” gathering relevant information about millions of articles in a single picture. These scripts are designed to be used as a black box, i.e. no prior knowledge of python is required for using them. However, user is free to change the codes if they want to which should help them to generate their own desired “maps of science”.

BiblioTools1.0 was first released in 2011 based on SQL queries. Further, in July 2012, BiblioTools 2.1 was released based on python scripts with updates and corrections in years 2012 and 2014. User can download BiblioTools 2.2 file, unzip it and place the unzipped folder wherever they want to install. The BiblioTools scripts have been heavily tested on Unix, Mac and Windows. In order to run them, user need to be able to

run python and need to have the following packages installed: argparse, numpy, matplotlib, networkx. BiblioTools 2.2 can be downloaded from: http://www.sebastian-grauwin.com/?page_id=427.

User need to have an access to the Web of Science to extract bibliographic data, gephi to open and visualize the different maps produced by the scripts, and a latex compiler to visualize some tables produce by the scripts in pdf format. BiblioTools 2.2 can perform various bibliometrics related exercises, e.g. data parsing, filtering, detecting bibliographic coupling communities, co-occurrence maps, etc.

2.3 CiteSpace

CiteSpace is a freely available Java application developed by Chaomei Chen, College of Information Science and Technology, Drexel University, Philadelphia, USA for progressive knowledge domain visualization and analyzing trends and patterns in scientific literature. It helps to conduct interactive visual analytic studies of scientific literature concerning a scientific field, a discipline, or an institution, and identify and interpret salient patterns and trends.

CiteSpace provides various functions to facilitate the understanding and interpretation of network patterns and historical patterns, including identifying the fast-growth topical areas, finding citation hotspots in the land of publications, decomposing a network into clusters, automatic labelling clusters with terms from citing articles, geospatial patterns of collaboration, and unique areas of international collaboration. CiteSpace supports structural and temporal analyses of a variety of networks derived from scientific publications, including collaboration networks, author co-citation networks, and document co-citation networks. It also supports networks of hybrid node types such as terms, institutions, and countries, and hybrid link types such as co-citation, co-occurrence, and directed citing links.

The primary source of input data for CiteSpace is the Web of Science. CiteSpace also provides some simple interfaces for obtaining data from PubMed, arXiv, ADS, and NSF Award Abstracts. CiteSpace can be used to generate geographic map overlays viewable in Google Earth map based on the locations of authors.

CiteSpace software is freely available. The current version of the software is 5.0.R3 SE. As CiteSpace is a Java application, user need to make sure that their computer supports Java, including Windows, Linux or Mac, and Java Runtime (JRE) is installed on computer. It is currently optimized for Windows 64-bit Java 8 (i.e. Java 1.8) and can be downloaded from: http://cluster.ischool.drexel.edu/~cchen/citespace/download/.

2.4 CitNet Explorer

CitNetExplorer is a Java based software tool developed by Nees Jan van Eck and LudoWaltman at Centre for Science and Technology Studies (CWTS), Leiden University, Netherlands. It is a software tool for visualizing and analyzing citation networks of scientific publications. The tool allows citation networks to be imported directly from the Web of Science database and can be exported in the popular Pajek file format. Citation networks can be explored interactively by using zoom and scroll functionality, for instance by drilling down into a network and by identifying clusters of closely related publications, direct and higher-order indirect citation relations.

CitNetExplorer is freely available for non-commercial research and teaching purposes, and distributed under a non-commercial research and teaching license. CitNet Explorer version 1.0.0, released on March 10, 2014, can be downloaded from: http://www.citnetexplorer.nl/download. To run CitNet Explorer, user need to install Java version 6 or higher in their system.

2.5 VOSviewer

VOSviewer is a free Java based program, primarily developed by Nees Jan van Eck and Ludo Waltman at Centre for Science and Technology Studies (CWTS), Leiden University, Netherlands. It is a software tool for constructing and visualizing bibliometric networks. These networks may for instance include journals, researchers, or individual publications, and they can be constructed based on co-citation, bibliographic coupling, or co-authorship relations. VOSviewer also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature.

VOSviewer provides a number of advanced features for creating bibliometric networks (e.g., co-authorship, bibliographic coupling, and co-citation networks). For instance, the influence of publications with many authors, many citations, or many references can be reduced using a fractional counting approach. Data cleaning can be performed using thesaurus files. Co-authorship networks, co-occurrence networks, and citation-based networks can be created directly based on Web of Science, Scopus, PubMed, and RIS files. Networks can be imported from and exported to Pajek network files and GML files. Natural language processing techniques are inbuilt in the software for creating term co-occurrence networks based on English-language textual data. Also, state-of-the-art techniques for network layout and network clustering are available in the software. Visualizations of bibliometric networks can be explored in full detail using zoom and scroll functionality.

The software is freely available to use for any purpose. VOSviewer version 1.6.5, released on September 28, 2016, can be downloaded from: http://www.vosviewer.com/download. To run CitNetExplorer, user need to have Java version 6 or higher installed in their system.

2.6 CopalRed

CoPalRed is a Knowledge System developed by Rafael Bailón-Moreno at EC^3 Research Group at Department of Chemical Engineering. University of Granada, Spain in 2003. CoPalRed collects the information contained in the databases, e.g. WOS, Scopus, MedLine, ProCite, etc. and transforms it into new knowledge, which was not explicit in the information contained in the databases, but is generated from the information contained therein. Since July 2011, CopalRed has changed its name to Techne Co Word, since it continues its development under the cover of the research group "Techne, Knowledge and Product Engineering". Its current stable version is 2.0 released in July, 2011. It works only on Windows operating system.

CopalRed, as a knowledge system, consists of mainly four modules, i.e. information capture, information debugging, knowledge base generation, and knowledge management module.

CoPalRed is mainly able to perform three types of analysis:

- Structural analysis: It reveals the network structure of the scientific field under study, defining the actors and their relationships.
- Strategic analysis: It places each actor in a relative position within the network, defining it according to the intensity of its external relations (centrality) and according to its internal cohesion (density).
- Dynamic analysis: It analyzes the transformations (translations-translations) of the actors over time. Identifies approaches, bifurcations, appearances and disappearances of the actors.

CoPalRed generates two types of outputs of knowledge:

- Specific outputs that can be displayed on screen, printed or saved; and
- Compatible standard outputs which are a set of reports with content similar to the specific outputs,

but with the particularity of being presented in formats compatible to Microsoft Word and Microsoft Excel.

CopalRed's knowledge management module is capable of performing various bibliometrics functions including drawing networks and exporting them to any image editor, rotating them, changing thresholds, Bradford distributions, Lotka distributions, affiliations, bibliographies, ranking of documents by relevance, and so on in any combination of themes or subnets.

2.7 CRExplorer

The CRExplorer is a new software based on the programs provided at Loet Leydesdorff's homepage athttp://www.leydesdorff.net/software/rpys/.The developers of CRExplorer are Andreas Thor, Werner Marx, Lutz Bornmann (Germany) and Loet Leydesdorff, (The Netherlands). The CRExplorer uses data from Web of Science or Scopus databases as input. User needs to download publication sets including the references cited to analyse in CRExplorer. The program focusses on the analysis of the cited references, in particular on the referenced publication years. When the aggregated citations are plotted along the time axis, one obtains a "spectrogram" with distinct peaks. CRExplorer visualizes this spectrogram, cleans the cited references (called "disambiguation"), and uses a smoothening algorithm to suppress the noise. The CRExplorer uses the Reference Publication Year Spectroscopy (RPYS) method which was developed by Werner Marx, who used it for the first time in the field of meteorology.

CRExplorer requires a system with Java 8 support (Java Run Time) to run. There are two options to run CRExplorer:

- Java Web Start: Click the CRExplorer Start link to launch CRExplorer directly from this web page using Java Web Start Launcher.
- Download: You can download a runnable JAR file. On most systems, a double click on the JAR file will start CRExplorer.

The last version of CRExplorer 1.6.8 was released on August 29, 2016. This software has different features, such as:

- It reads files from Scopus. The file format "CSV" (including citations, abstracts and references) downloaded from Scopus.
- CRExplorer exports files in the Web of Science and Scopus format. These files can be imported in other bibliometric programs (e.g. VOSviewer).
- Internal file format: working files are/can be saved in the internal file format "*.cre".
- Co-citation for cited references which are co-cited in a specific publication and data clustering.

2.8 Inter Disciplinary Research (IDR) Toolkit

The IDR toolkit offers a novel tool, the overlay maps of science, as a method to explore the degree of interdisciplinary of a set of publications. It was developed by Dr. Alan Porter andDr. Ismael Rafols.The overlay technique visualizes the spread of publications over the global map of science, i.e. the structure of science as obtained from the analysis of cross-citations between disciplines. It follows the method introduced in Rafols & Meyer (2010) to create the overlay map on the basis of a global map of science. It helps to visually locate bodies of research within the sciences, both at each moment of time and dynamically. A user has to rely on access to the Web of Science to obtain the set of Web of Science Categories (WCs) for a given set of articles and the files available in mapping kit (also available at http://users.fmg.uva.nl/lleydesdorff/overlaytoolkit/). The objective is to obtain the set of Web of Science Categories (WCs) for a given set of articles; provide this to network software; and output overlay information to add to a suitable basemap using Pajek and/or VOSviewer.

The analysis can be carried out at different units of aggregation in IDR: for example, for a university or corporation, for a research topic, or for a research programme or funding agency. By locating the publications over the map of science, user can gain an understanding of the diversity of disciplines involved. Since attribution of publications to disciplines is problematic and controversial, the overlay maps are only reliable with large numbers. In IDR, a user need to have a set of at least 70 publications for an exploratory map, but for accurate representations, it is recommended to have above 1,000 publications.

The maps allow one to intuitively perceive various aspects of disciplinary diversity. First, the number of disciplines involved. Second, the balance of disciplines, i.e. whether publications are evenly distributed or some disciplines are predominant. Third, and crucially, the cognitive distance between the disciplines involved –whether the research investigated covers disparate or cognate areas of science. This aspect of disparity is a key advantage of the maps: they differentiate between short-range interdisciplinarity (e.g. chemistry and physics), or long-range interdisciplinarity (e.g. social science and biology). Measures of interdisciplinarity can be associated with these maps.

2.9 IN-SPIRE

IN-SPIRE™ Visual Document Analysis, a powerful information visualization software is developed by Pacific Northwest National Laboratory with assistance from U.S. Department of Energy. It is designed to help analyst to uncover relationships, trends, and themes hidden within data which can lead to new knowledge and new insights that could be used to assess terrorist threats, determine how to treat a medical condition, or gather market research on the competition and many more.

IN-SPIRE™ can quickly and automatically convey the gist of large sets of unformatted text documents such as technical reports, web data, newswire feeds and message traffic. IN-SPIRE™ can handle real-time data by adding new documents as they arrive. It also processes foreign language data and provides robust support for translation. By clustering similar documents together. This Windows-based software unveils common themes and reveals hidden relationships within the collection.

IN-SPIRE™ analyses a multitude of text files and determines key topics or themes in each to create a signature for each document in the collection. IN-SPIRE's two main visualizations display representations of the documents in which those with similar or related topics appear closer together. The Galaxy visualization uses the metaphor of the stars in the night sky with each star representing an individual document. The ThemeView™ visualization uses a 3-dimensional terrain map display to provide a high-level overview of the data. The Search tools in IN-SPIRE™ support simple queries, phrase queries, and queries with example text. Other tools help to explore trends over time and relationships between concepts.

IN-SPIRE™ has in-built algorithms to perform various analyses, such as:

- Entity Extraction: IN-SPIRE comes bundled with an entity extractor that extracts People, Organizations, Locations, E-mail addresses, web addresses, and more.
- Automatic Keyword Extraction: Essential keywords and phrases are now automatically identified and extracted when documents are processed by IN-SPIRE, enabling keywords and phrases such as "united states", "supreme court", and "pope benedict."
- Themes: Keywords and phrases are grouped into themes based on their co-occurrence within documents in a dataset.
- Surprising Terms: The Time tool's terms view uses a new algorithm to analyze change in term

frequency over time. All terms in the document content are analysed to find those with interesting spikes in frequency.

IN-SPIRE™ is a licensed commercial software. It mostly operates on Windows workstations and Servers. A typical IN-SPIRE computer would include a 1.5 Ghz processor, 2Gb of memory, and 200 Gb disk drive in which about 20-50 Gb should be available to store the program and associated datasets. The latest version of IN-SPIRE is version 5.9.

2.10 HistCite

HistCite is a software package used for bibliometric analysis and information visualization. It was developed by Eugene Garfield, the founder of the Institute for Scientific Information (ISI). It is a software developed to allow the users to aid researchers in visualizing the results of literature searches in the Web of Science. HistCite let the users analyze and organize the results of a search to obtain various views of the topic's structure, history, and relationships. It is easy, fast, and provides perspectives and information not otherwise available.

HistCite helps the user to analyse:

- How much literature has been published in a specific field? When and in what countries has, it been published? What countries are the major contributors to that field? What are the languages most frequently used by the items published in that field?
- What journals cover the literature of the field? Which are the most important?
- Who are the key authors in that field? What institutions do these authors represent?
- Which articles are the most important?
- How have the various contributors to the field influenced each other?

In regards to information visualization, HistCite performs one specific application that it converts bibliographies into diagrams called historiographs.

HistCite operates on Windows computers with Internet Explorer. The latest version of HistCite is version 12.3. A free copy of HistCite software can be downloaded from: http://interest.science.thomsonreuters.com/forms/HistCite/ by agreeing to an End User License Agreement provided by Thomson Reuters, now a part of Clarivate Analytics.

2.11 Loet Leydesdorff's Software

Leydesdorff's software is a set of command-line DOS based programs to analyse and evaluate bibliometrics data obtained from the data sources such as Scopus, Web of Science, and Google Scholar. The set of software / programs were developed by Loet Leydessorff at Science & Technology Dynamics, University of Amsterdam, Amsterdam School of Communications Research (ASCoR), The Netherlands.

The set of programs is freely available for academic and research community. The different programs allow the user to perform several bibliometric analyses, such as: co-word analyses, co-authorship, author bibliographic coupling, journal bibliographic coupling, author co-citation, animation of network data, co-word mapping of texts (lines), visualization of various networks and organization of Google Scholar files into files for relational database management (MS Access, dBase), etc. The institutional and international collaboration, even the collaboration at the level of cities can be analysed with the help of this program. The results can be visualized using external software, such as Pajek, UCINET, Network Workbench Tool or the Sci2 Tool. Also, the visualization of these collaboration networks can be done using Google Maps and external software.

2.12 Publish or Perish (PoP)

Publish or Perish (PoP) is a software program that retrieves and analyses academic citations. It was

developed by Anne-WilHarzing, Professor of International Management, Middlesex University, London in the year 2006. Publish or Perish is designed to empower individual academics to present their case for research impact to its best advantage based on citations. It uses Google Scholar and Microsoft Academic Search to obtain the raw citations, then analyses these and presents the following metrics:

- Total number of papers and total number of citations,
- Average citations per paper, citations per author, papers per author, and citations per year,
- Hirsch's h-index, Zhang's e-index and related parameters,
- Egghe's g-index,
- The contemporary h-index,
- Three variations of individual h-indices,
- The average annual increase in the individual h-index,
- The age-weighted citation rate(AWCR) and AW-index,
- An analysis of the number of authors per paper and Multi-authored h-index.

The results from the analysis are available / displayed on-screen and can also be copied to the Windows clipboard (for pasting into other applications) or saved to a variety of output formats (for future reference or further analysis). Publish or Perish includes a detailed help file with search tips and additional information about the citation metrics.

Publish or Perish works on system based on Windows, OS X and GNU/Linux operating systems. Its latest version is 5.26.2 released on 8th February 2017.

2.13 SciMAT

SciMAT (Science Mapping Analysis Software Tool) is an open source science mapping software tool which incorporates methods, algorithms, and measures for all the steps in science mapping workflow, from preprocessing to the visualization of the results. It was developed by M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, and F. Herrera, Soft of Computing and Intelligent Information Systems (Sci2s) Research Group, University of Granada, Spain. SciMAT has been supported by the Project of Spanish Ministry of Education and Science. SciMAT allows the user to carry out studies based on several kinds of bibliometric networks. Different normalization and similarity measures can be used over the data. Several clustering algorithms can be chosen to cut up the data.It is based on a longitudinal science mapping approach.

SciMAThas different modules that help the analyst to carry out the steps of the science mapping workflow, such as: a) knowledge base and document Management module; b) science mapping and analysis module; and c) visualization module.

The main characteristics of SciMAT are:

- Loaders: It helps for acquisition and handling of data in ISI Web of Knowledge format and RIS format.
- Bibliometric networks: SciMAT has incorporated methods to build several kinds of bibliometric networks based on co-word analysis, co-author, co-citation, bibliographic coupling, and authors' clusters, etc.
- Preprocessing: SciMAT implements a wide range of preprocessing tools such as detecting duplicate and misspelled items, time slicing, data reduction and network reduction.
- Normalization: SciMAT has different normality measures to perform normalization, i.e. Association Strength, Equivalence Index, Inclusion Index, Jaccard's Index and Salton's Cosine.

- Mapping (through clustering): For science mapping, SciMAT has incorporated Simple Centers Algorithm, Single-linkage, Complete-linkage, Average-linkage and Sum-linkage clustering algorithms.
- Analysis: SciMAT allows the analyst to perform: network analysis (Callon's density and centrality), performance and quality analysis (sum, minimum, maximum and average citations, and bibliometric measures based on citations, such as: h-index, g-index, hg-index or q2-index), and temporal analysis. Also, it helps to track the conceptual, intellectual or social evolution of a research field through the course of time periods.
- Visualization: To visualize the result, strategic diagram, cluster network, overlapping map, evolution map can be created in SciMAT.
- Report: In SciMAT, outputs / reports are generated in HTML and LaTeX format.

SciMAT has been developed using Java as programing language.As such, it run in the majority of the operating systems i.e. (Windows, Linux, MacOS, etc. To run SciMAT, version 8 of Java has to be installed in the system. It is free software available under GPLv3 license which can be downloaded, redistributed or modified under the terms of the GNU General Public License. The latest version of SciMATis v1.1.04 released on 12th July, 2016.

2.14 Network Workbench (NWB) Tool

The Network Workbench (NWB) project at Indiana University, USA developed a large-scale network analysis, modelling and visualization cyber infrastructure toolkit for biomedical, social science and physics research. The NWB can perform network analysis, modelling, and visualization with the most effective algorithms and a wide variety of reference datasets. The NWB tool supports network science research across scientific boundaries.

The NWB tool is an algorithm integration framework that supports easy addition and dissemination of existing and newly created algorithms. The tool uses the Cyber infrastructure Shell (CIShell) written in Java, an OSGI-based software architecture, to facilitate easy plug and play of diverse algorithms. The NWB tool has a menu driven interface that supports network/graph load, view and save operations, etc. The NWB tool includes GUESS, a powerful and flexible network visualization tool. The tool can load, process, and save various network file formats including NWB (*.nwb), GraphML (*.xml or. graphml), XGMML (*xml), Pajek (*.net), Pajek (*.mat), TreeML (*.xml), Scopus (*.scopus), NSF (*.nsf), Endnote (*.enw), Bibtex (*.bib), CSV (*.csv), ISI (*.isi) and two-column edge lists (*.edge). It also supports viewing and saving plain text files (*.txt) generated by algorithms.

The Network Workbench tool provides significant functionality for scientometricians, including the ability to analyze, clean, and create networks derived from scholarly data sources, such as ISI, Scopus, the NSF grant database, BibTeX and Endnote reference formats.

Some of important scientometrics analysis features of NWB tool are:

- Removes ISI duplicate records and rows with multitudinous fields;
- Detects duplicate nodes;
- Update network by merging nodes; and
- Extract directed network, paper citation network, author paper network, co-occurrence network of word and reference, co-author network, document co-citation network and co-citation similarity network.

The NWB tool is freely available to download at http://nwb.cns.iu.edu/download.html. The latest official release version of NWB tool is 1.0.0 released on 15th September, 2009. The NWB tool works on Windows, Linux and Mac systems.

2.15 Sci2 Tool

The Science of Science (Sci2) is a modular toolset specifically designed for the study of science. It supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels. The Sci2 tool was developed by Cyber Infrastructure for Network Science Center Team, Indiana University, USA.

Sci2 is OSGi/ Cyber Infrastructure Shell (CIShell) powered, which means it can easily be extended to support and integaret new algorithms, input formats, and more through the OSGi/CIShell plugin architecture. Sci2 tool hosts many tools to aid in every step of the data preparation, analysis, and visualization process. A number of data formats are supported on Sci2 platform, such as: NWB (*.nwb), GraphML (*.xml or. graphml), XGMML (*xml), Pajek (*.net), Pajek (*.mat), TreeML (*.xml), Scopus (*.scopus), NSF (*.nsf), Endnote (*.enw), Bibtex (*.bib), CSV (*.csv), ISI (*.isi), and edge lists (*.edge).

Some of the main functionalities of Sci2 tool are:

- Loading Data: Load a supported file format for preparation, analysis or visualization.
- Data Preparation: Extract networks from raw data or update currently existing networks by merging nodes and removing duplicates.
- Processing: Clean data for analysis and visualization.
- Analysis: Employ a variety of advanced analysis algorithms for temporal, topical, geospatial, and network data.
- Modelling: Graph generation with aging, scaling, random, and other specifications.
- Visualization: Visualize temporal, topical, geospatial, and network data.

Similar to the NWB tool, Sci2 tool can read data files from different databases such as Web of Science and Scopus and different export formats such as Bibtex and EndNote. It can also import plain text exports direct from Web of Knowledge. The data preparation function cleans the bibliographic data and creates different networks and tables that can be used in preprocessing, analysis, and visualization. Principally, the networks that can be extracted are as follows: co-author, co-PI (Principal Investigator), co-word, document co-citation, journal co-citation, author co-citation, author bibliographic coupling, document bibliographic coupling, and journal bibliographic coupling. Moreover, the tool can build different direct linkage networks such as author-citation, document citation, source-citation paper, and, finally, author-document (consumed/produced) network.

The Sci2 tool is available freely and can be downloaded upon registration at: https://sci2.cns.iu.edu/user/welcome.php.

2.16 SitKis

Sitkis is a free Java and MS Access based software tool developed exclusively for bibliometric analysis. It helps researchers during computation process of research analysis and evaluation of scientific information. It provides tools for extremely streamlined analysis of bibliometric networks. User can calculate large amount of data in few minutes utilizing Sitkis.

Sitkis incorporates the following functionality:

- Import ISI Web of Science data on articles and other publications into Microsoft Access Database for easy access and manipulation;
- Create analyses of citing and cited articles, including time series trends in citation classics within a specific discourse;
- Export UCINET-compatible networks of citing and cited articles. Export functionality includes

automatic normalization, selection of year ranges, threshold levels for minimum number of citations, etc;

- ❖ Group articles into independent dense groups, and analyze different theoretical streams within a selected scientific discourse; and
- ❖ Generate statistics based on the country or university of the original or citing authors.

Advanced users can also accomplish the following tasks much more efficiently by utilizing Sitkis:

- ❖ Manage knowledge of a topic area by categorizing and prioritizing articles,
- ❖ Create bibliometric analyses by coding articles according to criteria of their choice,
- ❖ Cluster and partition co-citation networks to reveal the structure of citation and co-citation patterns; and
- ❖ Use abstracts of articles in a text analysis software.

Sitkis requires Microsoft Access 2000 (or newer, part of Microsoft Office) and Java 1.3 (or newer) run time environment (freely available from Sun Microsystems) to function. Sitkis runs on all PC computers, but a relatively powerful computer is recommended for analysis dealing with large number (over 5000) of articles.User can download the new version of SitKis from: https://www.dropbox.com/s/ukl1624t02r31sg/sitkis2012.zip.

2.17 VantagePoint

VantagePoint is a powerful text-mining tool for discovering knowledge in search results from patent and literature databases. VantagePoint helps the user to rapidly understand and navigate through large search results, giving them a better perspective on their information. Vantage point is a commercial software tool developed by Search Technology, Inc., USA.

VantagePoint works with search results from text databases. It is most helpful when we need to work with thousands of records instead of a few dozens. VantagePoint is a menu-driven platform. Major features and functionalities of Vantage Point are as follows:

- ❖ Extracting and Indexing Data: After the user imports the data into VantagePoint, fields are mined from the text using pattern matching, rule-based, and natural language processing techniques to obtain: a) Fields found in most bibliographic databases (for example, authors, inventors, affiliations, assignees, dates, descriptors, index terms, and classification codes); b) Normalized values that are sometimes buried in the basic fielded data or presented in mixed formats, such as dates, classification codes, or patent numbers; c) Meaningful words and phrases from the titles, abstracts, patent claims, and other free text; and d) Entities defined in user-managed dictionaries. VantagePoint comes with access to their in-built library of import filters, thesauri and macros. It also comes with tools to build user's own import filters, thesauri, and macros for proprietary data sources and business processes. User can also import Excel or Access data using VantagePoint's "Import Database Table" wizard.
- ❖ Top 10 Lists: User can browse their data from many perspectives. The simplest tool is a list. Using a list of affiliations (or authors or countries), user can see the top affiliations (or authors or countries) and browse the records for each one. User can sort and search his/her lists in an instant, and create groups in their lists in many different ways. "List Comparison" lets the user compare any two lists from any set of VantagePoint data and identify the common or unique items in the lists. With time-sliced data sets, this can help the user to identify new or emerging concepts, authors, institutions or countries. "Using Group" using Thesaurus, user

can apply a thesaurus to own list and create groups of list items. "Find and Group" searches the list for user's target text and adds matching items to a group. "Find" provides for three search strings and several Boolean operators (for example, "and," "adjacent," "near4," "or", etc.).

- Co-occurrence Matrices: Using a Co-occurrence matrix, user can spread a list across any other list. For example, using a co-occurrence matrix of "author" and "year of publication," user can see an author's publication trend over time and browse their publications by year. This way, user can see which authors have been working in an area for the longest time as well as whether they are still active. Further, data from lists and co-occurrence matrices can be easily exported to other applications for visualization or customized analysis.

- Maps/Networks: Beyond one-dimensional (lists) and two-dimensional (co-occurrence matrices) analyses, VantagePoint performs multidimensional statistical analysis to identify clusters and relationships among concepts, people, organizations and countries.

- Data Cleaning: VantagePoint uses fuzzy matching techniques to identify, associate, and clean up data. For example, it handles misspellings, alternative hyphenation and capitalization, and differing conventions for listing the names of people and companies, providing the user with higher-quality data that is essential for good analysis.

- User Managed Thesauri: With VantagePoint user can create, edit, and apply thesauri, which can be used for specialized data reduction or conversion. Using a thesaurus, user can easily combine all variants of a data element (for example, United States, US, U.S., USA, etc.). For other types of analysis, user can combine data elements into broad categories (for example, "aluminum alloys," "magnesium alloys," "carbon fiber reinforced plastics," and "copper alloys" into "Materials"; or "USA," "Canada," and "Mexico" into "North America").

- Scripting: User can automate sequences of actions using VantagePoint's extensions to Microsoft's VBScript. User can create and distribute scripts and make his/her own customized analysis techniques and processes easily repeatable.

VantagePoint only works on Windows 7, 8, and 10 platforms. VantagePoint's installation on Windows Server is not supported and also on other operating systems, e.g., Macintosh OS X or Linux. VantagePoint is commercial, licensed on a per-seat basis. The standard commercial single seat VantagePoint license is perpetual and requires no annual fees after the initial purchase. However, user can get a 30-day trial free copy of VantagePoint.

2.18 Bibliometrix R Package

Bibliometrix provides a set of tools for quantitative research in scientometrics and bibliometrics. Bibliometrix works with data extracted from the two main bibliographic databases: Scopus and Clarivate Analytics's ISI Web of Knowledge. Bibliometrics turns the main tool of science, quantitative analysis, on itself. Essentially, bibliometrics is the application of quantitative analysis and statistics to publications such as journal articles and their accompanying citation counts. Bibliometrix package provides various routines for importing bibliographic data from these two databases, performing bibliometric analysis.

Bibliometrix supports scholars in three key phases of analysis:

- Data importing and conversion to R format;
- Bibliometric analysis of a publication dataset; and

- Building matrices for co-citation, coupling, collaboration, and co-word analysis. Matrices are the input data for performing network analysis, multiple correspondence analysis, and any other data reduction technique.

The stable version of Bibliometrix can be downloaded directly from RCRAN Package Repository available at https://cran.r-project.org/web/packages/bibliometrix/index.html. The latest version of Bibliometrix package is 1.4 which was released on 24-01-2017. The package is licensed and distributed under GPL-3 license.

2.19 metaknowledge

metaknowledge, a Python3 package is basically used for doing computational research in bibliometrics, scientometrics, and network analysis. It is also being used to simplify the process of doing systematic reviews in any disciplinary context. metaknowledge is developed by NetLab research group headed by Dr. John McLevey's at the University of Waterloo, Canada.

metaknowledge is a Python3 library that simplifies bibliometric research using data from various sources. It reads a directory of plain text files containing meta-data on publications and citations, and writes to a variety of data structures that are suitable longitudinal research, computational text analysis (e.g. topic models and burst analysis), Reference Publication Year Spectroscopy (RPYS), and network analysis (including co-citation, co-authorship, multi-modal, multi-level, and dynamic). metaknowledge can easily handles large datasets (e.g. several million records) efficiently.

metaknowledge currently handles data from the Web of Science, Scopus, PubMed, Proquest Dissertations & Theses, administrative data from the National Science Foundation, and the Canadian tri-council granting agencies: SSHRC, CIHR, and NSERC.

There are different modules available in metaknowledge, such as:

- contour: A nicer matplotlib graph visualizer and contour plot;
- WOS: The backend functions and classes associated with the Web of Science;
- medline: The backend functions & classes associated with Medline, the format used by Pubmed;
- proquest: The backend functions and classes associated with ProQuest;
- scopus: The backend functions and classes associated with records from scopus; and
- journalAbbreviations: Handles the abbreviated journal names used by WOS.

The metaknowledge comes with a command-line application named [metaknowledge]. This provides a simple interface to the python package which allows the generation of most of networks along with ways to manage the records themselves. Datasets created with metaknowledge can be analyzed using NetworkX and the standard libraries for data analysis in Python. It is also easy to write data to csv or graphml files for analysis and visualization in R, Stata, Visone, Gephi, or any other tools for data analysis. metaknowledge also has a simple command line tool for extracting quantitative datasets and network files from Web of Science files. This makes the library more accessible to researchers who do not know Python, and makes it easier to quickly explore new datasets.

metaknowledge is a free and open source software, distributed under the GPL license. Its latest version is metaknowledge 3.1.2 which can be downloaded from https://pypi.python.org/pypi/metaknowledge#downloads.

3. Conclusion

In this article, a brief profile of different science mapping and visualization software tools used in bibliometric and scientometric studies is provided. A total number of nineteen software tools have been covered in the article which are widely used by the academic and scientific research community worldwide. It is very difficult to choose the best among all as each of these software tools have their own set of features, characteristics and functions to perform specific task. Most of these tools are freely available to the users over the Internet to download and use. The students, the research scholars and the professionals who have interest to work in bibliometric and scientometric related studies would find these tools useful in their studies.

References

Aria, M. &Cuccurullo, C. (2017) A brief introduction to bibliometrix. Available at http://www.bibliometrix.com/documentation/bibliometrix-vignette.html (Accessed on 12/02/2017).

BIBLIOMETRIX R Package: Bibliometric and Co-Citation Analysis Tool. Available at http://www.bibliometrix.org/ (Accessed on 12/02/2017).

Chen, C. CiteSpace: Visualizing Patterns and Trends in Scientific Literature. Available at http://cluster.cis.drexel.edu/ ~ cchen/citespace/ (Accessed on 12/02/2017).

Cited References Explorer. Available at http://www.crexplorer.net/ (Accessed on 12/02/2017).

CiteSpace101. CiteSpace: Visualizing Trends and Patterns in Scientific Literature. Available at https://sites.google.com/site/citespace101/home (Accessed on 12/02/2017).

CitNetExplorer: Analyzing citations patterns in scientific literature. Available at http://www.citnetexplorer.nl/ (Accessed on 12/02/2017).

Cobo, M., López-Herrera, A., Herrera-Viedma, E., & Herrera, F. (2011). Science Mapping Software Tools: Review, Analysis, and Cooperative Study among Tools. Journal of the American Society for Information Science and Technology, 62(7), 1382-1402. Available at doi:10.1002/asi.21525 (Accessed on 12/02/2017).

CopalRed. Available at https://es.wikipedia.org/wiki/Copalred (Accessed on 12/02/2017).

EC3. CopalRed version 1.0 Espanol (beta). Available at http://ec3.ugr.es/copalred/index.htm (Accessed on 12/02/2017).

Encyclopaedia of Information Science and Technology, Third Edition. Science Mapping. Available at http://www.igi-global.com/chapter/science-mapping/112859 (Accessed on 12/02/2017).

Fernandez, V. Sitkis: Literature Review. Available at http://en.vicenc-fernandez.eu/sitkis-literature-review/ (Accessed on 12/02/2017).

Grauwin, S. BiblioTools. Available at http://www.sebastian-grauwin.com/?page_id = 427 (Accessed on 12/02/2017).

Harzing, Anne-Wil. Publish or Perish. Available at http://www.harzing.com/resources/publish-or-perish (Accessed on 12/02/2017).

HistCite. Available at https://en.wikipedia.org/wiki/Histcite (Accessed on 12/02/2017).

IDR: Measuring and Mapping Interdisciplinary Research. Available at http://www.idr.gatech.edu/overlay.php (Accessed on 12/02/2017).

Jamali, H. R. Scientometric Portal. Available at https://sites.google.com/site/hjamali/scientometric-portal (Accessed on 12/02/2017).

Kumar, A., Shivarama, J., & Choukimath, P. A. (2015). Popular Scientometric Analysis, Mapping and Visualisation Softwares: An Overview. In 10th International CALIBER-2015, HP University and IIAS, Shimla, Himachal Pradesh, India (pp. 157-170). Gandhinagar, Gujarat: INFLIBNET Centre. Available at http://ir.inflibnet.ac.in/handle/1944/1855 (Accessed on 12/02/2017).

Leydesdorff, L. Loet Leydesdorff's Software. Available at http://www.leydesdorff.net/software.htm (Accessed on 12/02/2017).

NWB Team. (2006). Network Workbench Tool. Indiana University, Northeastern University, and University of Michigan. Available at http://nwb.slis.indiana.edu (Accessed on 12/02/2017).

Pacific Northwest National Laboratory. IN-SPIRE™ Visual Document Analysis. Available at http://in-spire.pnnl.gov/ (Accessed on 12/02/2017).

Persson, O., R. Danell, & J. Wiborg Schneider. (2009). How to use Bibexcel for various types of bibliometric analysis. In celebrating scholarly communication studies: A Festschrift for OllePersson at his 60th Birthday, ed. F. Astrom, R. Danell, B. Larsen, J. Schneider, p 9-24. Leuven, Belgium: International Society for Scientometrics and Informetrics. Available at http://homepage.univie.ac.at/juan.gorraiz/bibexcel/ (Accessed on 12/02/2017).

Reid McIlroy-Young, John McLevey, and Jillian Anderson. 2015. metaknowledge: open source software for social networks, bibliometrics, and sociology of knowledge research. URL: http://www.networkslab.org/metaknowledge.

Sci2 Tool: A Tool for Science of Science Research & Practice. Available at https://sci2.cns.iu.edu/user/index.php (Accessed on 12/02/2017).

Science of Science Tool (Sci2). Available at https://en.wikipedia.org/wiki/Science_of_Science_Tool_(Sci2) (Accessed on 12/02/2017).

SciMAT: Science Mapping Analysis Tool. Available at http://sci2s.ugr.es/scimat/ (Accessed on 12/02/2017).

Search Technology. VantagePoint: Serious Software for Serious Professionals. Available at https://www.thevantagepoint.com (Accessed on 12/02/2017).

VOSviewer: Visualizing scientific landscapes. Available at http://www.vosviewer.com/ (Accessed on 12/02/2017).

Visitors

Prof. Anamik Shah, Vice-Chancellor, Gujarat Vidyapith, Ahmedabad, Gujarat

Dr. Nabi Hasan, University Librarian, Aligarh Muslim University, Aligarh, Uttar Pradesh

Dr. Kumud Sharma, Department of Hindi, University of Delhi, New Delhi

Dr. (Mrs.) Pankaj Mittal, Additional Secretary, UGC, New Delhi

Prof. Ram Bux, Department of Hindi, Jawaharlal Nehru University, New Delhi

Prof. Krishna Amin-Patel, Centre Head, National Institute of Design (NID), Gandhinagar, Gujarat

Dr. Ram Gopal Singh, Professor & Research Director, Department of Hindi, Gujarat Vidyapith, Ahmedabad, Gujarat