* A software might have several functionalities.
  + This means that a software purpose classification task is a **multi-class classification task**[1]. For example: MATLab software might have classification purpose as:
    - Modelling
    - Simulation
    - Visualization
    - Programming
  + When a software belongs to one or more than one class it is said to be a general purpose software.
  + Some software might thus belong to one or more classification class. If a software belongs to only one class, thus it is said to be a domain specific software or custom software.

[1] Wikipedia contributors, "Multiclass classification," *Wikipedia, The Free Encyclopedia,* <https://en.wikipedia.org/w/index.php?title=Multiclass_classification&oldid=1023970866> (accessed November 9, 2021).

What is a research software ?

Software that is used to generate, process or analyze results that you intend to appear in a publication (either in a journal, conference paper, monograph, book or thesis). Research software can be anything from a few lines of code written by yourself, to a professionally developed software package. Software that does not generate, process or analyze results - such as word processing software, or the use of a web search - does not count as research software for the purposes of this survey. ( [arXiv:1507.03989](https://arxiv.org/abs/1507.03989))

# Notes: The role of software in a research

Orduña-Malea, E., Costas, R. Link-based approach to study scientific software usage: the case of VOSviewer. Scientometrics 126, 8153–8186 (2021). <https://doi.org/10.1007/s11192-021-04082-y>

Scientific software is a fundamental player in modern science, participating in all stages of scientific knowledge production. Software occasionally supports the development of trivial tasks, while at other instances it determines procedures, methods, protocols, results, or conclusions related with the scientific work.

Software is an essential component in the ecosystem of modern Science, particularly in those disciplines that follow a data-driven paradigm, guided by the ongoing generation, availability, and consumption of high volumes of scientific data (Hey, Tansley & Tolle 2009; Li & Yan, 2018).

Scientific software can play important roles in processes related to data collection, management, formatting, analysis, modelling, simulation, prediction, visualization, and dissemination (Howison et al., 2015; Pan et al., 2017), becoming essential in the scientific discovery process (Pradal et al., 2013).

Thus, scientific software has a direct effect on the validity of scientific results, since replacing the software could in turn lead to replacing an underlying procedure or logic assumption (Hannay et al., 2009; Howison & Herbsleb, 2011; Li et al., 2017; Yang et al., 2018).

This rising importance of software in the scientific process prompted the perception of scientific-purpose software ( research software) as a research product of its own ( the software to be considered as a research output)

Computer Software in Science and Mathematics , Stephen Wolfram Vol. 251, No. 3 (September 1984), pp. 188-203 (16 pages) <https://www.jstor.org/stable/24920353>

Computer experiments are not limited by processes that occur in nature. For example, a computer program can describe the motion of magnetic monopoles in magnetic fields, even though magnetic monopoles have not been detected in physical experiments. i.e. ***Software allows experiments to be made beyond constraints of the physical world.***

The computer can also be used to study the properties of *abstract mathematical systems*.

Tikhonchev, M Yu, Shimansky, G A, Lebedeva, E E, Lichadeev, V V, Ryazanov, D K, and Tellin, A I. 2001. "The role of computer simulation in nuclear technologies development." Slovakia.

Employment of computer simulation has essentially increased the efficiency of scientific researches and made these researches faster and cheaper. Also it has expanded a sphere of objects under research. i.e ***The use of software has made researches to become more effective, faster and cheaper.***

Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature <https://doi.org/10.1002/asi.23538>

Biology is a leading domain for the importance of software in science, given the importance of computerized data analysis and the rise of bioinformatics. Some of the most well-cited papers of any kind in any science are biology software related (Science Watch, 2003)

# Note: Purpose of software use

Orduña-Malea, E., Costas, R. Link-based approach to study scientific software usage: the case of VOSviewer. Scientometrics 126, 8153–8186 (2021). <https://doi.org/10.1007/s11192-021-04082-y>

software is used in all stages of academic work (Howisonet al., 2015), from annotating preliminary ideas to processing large volumes of data or disseminating research results. Among the vast amount of software available in the scientific endeavor, we can distinguish between general-purpose1 and scientific-purpose software.

Duck, G., Nenadic, G., Brass, A. *et al.* bioNerDS: exploring bioinformatics’ database and software use through literature mining. *BMC Bioinformatics* **14,**194 (2013). <https://doi.org/10.1186/1471-2105-14-194>

# Software = BioNerDs

* A named entity recogniser for the recovery of bioinformatics databases and software from primary literature.
* A dictionary and rule-based resource recognition system.
* Uses textual clues to recognise both **old and new database and software names** within the full-text literature.

Results:

* After applying bioNerDS to full-text articles from BMC Bioinformatics and Genome, it was found that Biology *Bioinformatics’s emphasis on* ***new tools*** and Genome Biology’s greater emphasis on ***data analysis***.
* The data also illustrates some shifts in resource usage. Some software like R and Gene Ontology became popular.

**Potential purpose**:

* ***Text Mining*** *– discovery of database and software from scientific literature*. Aautomatically identifying resource names on a large-scale from the scientific literature.
  + Determines what software and data resources are available and used in computational analyses.
  + BioNerDs is designed and developed as an NER tool that aims to recognise database and software mentions in literature, and to provide a document-level “list” of resources mentioned in a given article.
  + BioNerDS can recognize mentions of bioinformatics’ databases and software in primary literature with a reasonable accuracy

Duck G, Nenadic G, Filannino M, Brass A, Robertson DL, Stevens R (2016) A Survey of Bioinformatics Database and Software Usage through Mining the Literature. PLoS ONE 11(6): e0157989. <https://doi.org/10.1371/journal.pone.0157989>

# Software = bioNerDS Version 2

Potential purpose of the software: ***Text Mining***

* Provides an evaluation of ***levels of usage*** of database and software resources.
* provides an audit of the resources and a comparison of their relative usage in a biomedical literature (sub-disciplines of bioinformatics, biology and medicine.)

Result:

* there is a striking imbalance in a resource usage. Only 5% of the resources account for 47% of software / DB usage and Over 70% of resources are mentioned only once.
* Bioinformatics is more dynamic – more dev of novel resources and larger number of software mentions but in Biology and medicine software usage is more stable ( no novel software being developed or introduced ). Bioinformatics has highest proportion of software mentions (30.8 av mentions per doc) , followed by biology (12.9) then medicine (4.4).

**Observation: From BioNerDs what did you observe ?**

* BioNerDs version 2 - Large amount of text file, research publications, has been automatically analyzed to discover insights for instance to discover the level of novel software usage across disciplines like Bioinformatics, Biology and medicine.
* In BioNerDs - Automatically recovered bioinformatics database and software entities with some acceptable level of measure.

Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature <https://doi.org/10.1002/asi.23538>

What was done here ( Software Purpose) : **Text Analysis**

* Content analysis of 90 biology articles
* Identify software mentions and classify them according to their ability to realize the functions of citations.
* Developed a content analytic to identify mentions of software in the literature

The result:

* 31% - 43% mentions involve formal citation.
* Verry common informal citations even in high impact journals.
* Software is frequently inaccessible:
  + 15% - 29% inaccessible packages
  + 90% - 98% inaccessible specific versions
  + Only 24% - 40% provide a source code.

Assessing the impact of software on science: **A bootstrapped learning of software entities in full-text papers** <https://doi.org/10.1016/j.joi.2015.07.012>

Little is known about the impact of software in a research / science. To fill this gap:

**What was done here: AUTOMATIC EXTRACTION OF SOFTWARE MENTIONS FROM FULL-TEXT CORPORA**

* ***Extraction*** of software entities from full text papers using an ***improved bootstrapping*** method.

**Result**:

* The entity extraction system outperforms three baseline methods on extracting software mentions from all published papers on PLoS ONE in 2014.
  + 25,997 mentions extracted , of which 2342 are unique software
  + 7405 citations

Greuel, Gert-Martin & Sperber, Wolfram. (2014). **swMATH – An Information Service for Mathematical Software**. 10.1007/978-3-662-44199-2\_103.

**Software** : **swMATH**

**Related software: zbMATH, MathSciNet**

**Software type:** a mathematical software.

**Software Purpose** :

* providing systematic collection of references
* information service
* liking of software-related mathematical publications

*Mathematical software* is the connecting bridge between mathematical theories and concrete applications.

*Mathematical software* implements mathematical objects which are used to ANALYZE , SOLVE , SIMULATE , or MODEL a mathematical problem. Deep down a SOFTWARE is also a representation of mathematical knowledge because of its inherent mathematical reasoning and conclusions.

*Mathematical software* has the following properties:

* MS is often experimental
* Usually used for MODELLING; SIMULATING and SOLVING mathematical problems
* Mathematical software is dynamic , always through continuous improvement and extension.

# Purpose of software – pre-thesis literature

Montroull LE, Rothbard DE, Kanal HD, D'Mello V, Dodson V, Troy CM, Zanin JP, Levison SW, Friedman WJ. Proneurotrophins Induce Apoptotic Neuronal Death After Controlled Cortical Impact Injury in Adult Mice. ASN Neuro. 2020 Jan-Dec;12:1759091420930865. doi: 10.1177/1759091420930865. PMID: 32493127; PMCID: PMC7273561.

**Software**: [ImageJ](https://en.wikipedia.org/wiki/ImageJ)

**Purpose**: Image processing

**Task**: A traumatic brain injury can cause mortality and disability worldwide. Scientists used ImageJ software to analyze a brain image of a mice to determine the effect of regulating a special kind of receptor (p75NTR).

**Digital Image processing**

Scientists use image processing software to process images in various areas of research like agriculture, military, industry, medical science, space exploration … etc. Digital image processing is a branch of digital signal processing. Digital image processing has several advantages such as:

* Flexibility on choice of wide range of algorithms that can be applied on the image
* Better handling of noise and distortion of the image …etc.

Image processing is usually applied in wide areas of research mainly in satellite imagery, medical imaging, character recognition, photo enhancement, etc. (<https://en.wikipedia.org/wiki/Digital_image_processing> )

In a satellite imagery, satellite images are captured by imaging satellites. Satellite images have many uses in meteorology, agriculture, geology, education, cartography etc. However, satellite images have relatively high resolution and require image processing to create meaningful images from the data and to remove noise. (<https://en.wikipedia.org/wiki/Satellite_imagery> )

**Image processing vs Image analysis**

**Image processing –** input is image and output is image. The task is enhancing a noisy image. Tasks are Denoise, Deblur, destriping, …etc

**Image Analysis –** input is image and output is description about the image or its content like color, properties, …etc.

Vogelgsang, J., Vukovich, R., Wedekind, D., & Wiltfang, J. (2019). Higher Level of Mismatch in APOEε4 Carriers for Amyloid-Beta Peptide Alzheimer's Disease Biomarkers in Cerebrospinal Fluid. *ASN neuro*, *11*, 1759091419845524. <https://doi.org/10.1177/1759091419845524>

**Software**: Prism [GrapPad](https://www.graphpad.com/) 8 – Statistical Data analysis and Data Visualization software.

**Purpose**: Statistical analysis software that combines scientific graphing, comprehensive curve fitting (nonlinear regression), understandable statistics, and data organization. Designed for biological research applications in pharmacology, physiology, and other biological fields for data analysis, hypothesis testing, and modeling.

Huang Y, Song YJ, Isaac M, Miretzky S, Patel A, Geoffrey McAuliffe W, Dreyfus CF. Tropomyosin Receptor Kinase B Expressed in Oligodendrocyte Lineage Cells Functions to Promote Myelin Following a Demyelinating Lesion. ASN Neuro. 2020 Jan-Dec;12:1759091420957464. doi: 10.1177/1759091420957464. PMID: 32927995; PMCID: PMC7495938.

**Software**: [Quantity One 1-D Analysis Software](https://www.bio-rad.com/de-de/product/quantity-one-1-d-analysis-software?ID=1de9eb3a-1eb5-4edb-82d2-68b91bf360fb)

**Purpose**: Software used for Bio-Rad imaging systems to acquire, quantitate, and analyze a variety of data. The software allows automatic configuration of imaging systems with appropriate filters, lasers, LEDs, and other illumination sources. It also contains tools for automated analysis of tests and assays such as 1-D electrophoretic gels, western blots, and colony counts.

**Software** : [SPSS](https://www.ibm.com/uk-en/analytics/spss-statistics-software)

**Purpose**: Software package used for interactive, or batched, statistical analysis in social science, health sciences and marketing. Software platform offers advanced statistical analysis, a library of machine-learning algorithms, text analysis, open-source extensibility, integration with big data and deployment into applications.

# Notes: Software citation

Orduña-Malea, E., Costas, R. Link-based approach to study scientific software usage: the case of VOSviewer. Scientometrics 126, 8153–8186 (2021). <https://doi.org/10.1007/s11192-021-04082-y>

While it is commonly accepted that there is no need to mention/cite general-purpose software (Pan et al., 2019), it is recommended that scientific publications using scientific-purpose software should mention it (Niemeyer et al., 2016). The main reasons to encourage the citation of scientific-purpose software include credit allocation, reproducibility, transparency, and discovery (Smith, Katz, & Niemeyer 2016).

Previous literature has provided a significant body of knowledge about the lack of formal mentions of software in scientific publications.

Since formal software citation practices are not very matured, the **identification of textual mentions of software** in the text of scientific papers has been a quite common approach to capture the impact of software on science (Pan et al., 2016)

Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature <https://doi.org/10.1002/asi.23538>

* Software citations give visibility.
* We have different types of citations, formal and informal.
  + Informal
    - footnoted URL to webpages maintained by Software projects
    - simply mentioned in text
    - non-explicit mentions – apparently a software is used but not mentioned
  + Formal – RRID
* a random sample of 90 articles in the biology literature found seven different ways that software was mentioned, including simple names in the full-text, URLs in footnotes, and different kinds of mentions in reference lists: project names or websites, user manuals or publications that describe or introduce the software