Information Services for Mathematical Research Data

ACA 2016
22nd Conference on Applications on Computer Algebra

Session: Information services for mathematics, services, models, and data

Wolfram Sperber (FIZ Karlsruhe)

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Agenda

- Why such a session on ACA?
- What is mathematical research data?
- CAS Directories (Software)
- The concepts behind swMATH (I): The publication-based approach
- The concepts behind swMATH (II): The Web Archives approach
- Summary

Why this session on the ACA?

Currently, the subject "research data" is a hype in the discussion of scientific infrastructure?

- It would be carrying owls to Athens in this auditorium to make the statement that research results can't be reduced to the content of publications.

Some further more or less platitudes:

- The spectrum of research data has broadened within the computer age.
- Research data are depending on the science.
- Research results cannot be evaluated, repeated or reused without research data.
- Libraries are a powerful provider for maintaining the scientific literature, but information on and maintaining research data is a new challenge.

Do we need an e-infrastructure for our information? How should such a infrastructure be designed

Mathematical Research Data

Also the Symbolic Computation community (mathematical community) has - up to now - no general approach to handle the whole set of mathematical research data.

What is research data (in mathematics)?
"Research data (in mathematics) is data that is collected, observed, (used), or created, for purposes of analysis to produce original research results (in mathematics)."
(Boston University, Library)

What is mathematical research data? A (very rough) landscape

"Non-classical" mathematical research data

- Mathematical software (code)
- Related data
  - Documentations
  - Programming languages or environments
  - Benchmarks and test data
  - Data formats
  - Simulations and visualisations
  - Services (Repositories, Directories, ...)
  - ...

"Classical" mathematical research data

- Publications:
  - Mathematicals models
  - Mathematical terminology
  - Mathematical theories
  - Proofs
  - Algorithms
  - Data
  - Visualizations
  - …

Research data and Symbolic Computation

Symbolic Computation Resources:

- CAS Systems
- Services for CAS Systems (SIGSAM, CA Fachgruppe, Wikipedia, swMATH, ...)
- CAS Data and CAS Models
- Services for CAS Data and CAS Models (Symbolic Data, benchmarks, ...)
- (more general): Virtual Research Environments (OpenDreamKit)
- ...

Development of a suitable infrastructure for Symbolic Computation is not-trivial

It requires the cooperation between the Symbolic Computation community and information experts in this field (~ Mathematical Knowledge Management Initiative)
https://en.wikipedia.org/wiki/Mathematical_knowledge_management

Services for CAS Systems State of the art

- no comprehensive repository for CAS Systems
- but directories for CAS Systems support searching CAS Systems
  - SIGSAM
  - CA Fachgruppe
  - Wikipedia
  - swMATH (focused on software)
  - Symbolic Data (general information about symbolic computation, software plus data)
  - ...

Computer Algebra Software

SIGSAM maintains this collection of references to computer algebra systems, to support our citation policy. Click the name of each system to see further information, links and a citation in BibTeX format. If you have suggestions for changes or additions to this list please contact Infodir_SIGSAM@acm.org.

General purpose commercial systems

- Maple
- Mathematica
- Magma

SIGSAM → Resources → Software
http://www.sigsam.org/Resources/Software.html

Broad purpose *free* computer algebra systems

- Axiom: a general-purpose, strongly typed, computer algebra system.
- CoCoA: a computer algebra system for doing computations in Commutative Algebra.
- Fermat: a computer algebra system oriented towards polynomial and matrix algebra over the rationals and finite fields.
- GAP: a System for Computational Discrete Algebra.
- KASH/KANT: computer algebra system for sophisticated computations in algebraic number fields and global function fields.
- Macaulay2: a system for research in algebraic geometry and commutative algebra.
- Reduce: an interactive system for general algebraic computations of interest to mathematicians, scientists and engineers.
- SageMath: an open-source general purpose computer algebra system.
- SINGULAR: a Computer Algebra System for polynomial computations with special emphasis on the needs of commutative algebra, algebraic geometry, and singularity theory.
- PARI/GP: a computer algebra system designed for for fast computations in number theory.

Special Purpose Systems, Packages and Libraries

- ACE : a Maple library providing tools useful in algebraic combinatorics.
- Albert: an interactive program to assist the specialist in the study of nonassociative algebras.
- ANUNQ: a GAP package for the computation of nilpotent factor groups of finitely presented groups.
- ANUPQ: an interactive interface to the p-quotient, p-group generation and standard presentation algorithms of the ANU pq C program.
- CALI: a REDUCE package for computational commutative algebra.
- CASA: a Computer Algebra System for Algebraic Geometry.
- CHEVIE: a computer algebra system for symbolic calculations with generic character tables of groups.
- EinS: a Mathematica package allowing one to perform symbolic calculations with indexed objects.
- Felix: a special computer algebra system for the computation in commutative and non-commutative rings and modules.
- FeynArts: a Mathematica package for the generation and visualization of Feynman diagrams and amplitudes.
- GiNaC: a system to allow the creation of integrated systems that embed symbolic manipulations together with more established areas of computer science.
- GRAPE: a GAP package for constructing and analysing graphs related to groups, finite geometries, and designs.
- GUAVA: a GAP package for computing with error-correcting codes.
- LiDIA: A C++ Library For Computational Number Theory.
- LiE: A Computer algebra package for Lie group computations.
- MOLGEN: a system for the computation of all structural formulae that correspond to a given molecular formula.
- ORME: a package for equational theoreies.
- SONATA: a system for the construction and the analysis of finite nearrings

Allgemeine Computeralgebrasysteme

axiom

"The Scientific Computation System"
Lizenz: open source

FA Fachgruppe → Computeralgebrasysteme
http://www.fachgruppe-computeralgebra.de/systeme/

Derive

Lizenzinhaber: Texas Instruments
Weiterentwicklung wurde 2007 eingestellt

MAGMA

"Computational Algebra System"
Autoren: The Computational Algebra Group, University of Sydney
Lizenz: kommerziell (Gebühren für Service und Updates)
Ansprechpartner: John Cannon, Allan Steel

Maple

"Mathematics – Modeling – Simulation"
Lizenz: kommerziell
Ansprechpartner: Thomas Richard (mathematisch), Sabine Bormann (Verkauf)

MathCad

"Der globale Standard für Konstruktionsberechnungen"
Lizenz: kommerziell

Mathematica

"Compute – Develop – Deploy"
Veröffentlicht bei Wolfram Research Inc.
Lizenz: kommerziell
Ansprechpartner: Andreas Heilemann (mathematisch), Maryam Karbalai (Verkauf)

MATLAB

"The Language Of Technical Computing"
Seit Herbst 2008 durch Übernahme von MuPAD auch mit einer Symbolic Math Toolbox
Vertrieb über The MathWorks GmbH

List of computer algebra systems

From Wikipedia, the free encyclopedia

The following tables provide a **comparison of computer algebra systems** (CAS).[1][2][3] A CAS is a package comprising a set of algorithms for performing symbolic manipulations on algebraic objects, a language to implement them, and an environment in which to use the language.[4][5] A CAS may include a user interface and graphics capability; and to be effective may require a large library of algorithms, efficient data structures and a fast kernel.[6]

Contents [hide]

Wikipedia → list of computer algebra systems
https://en.wikipedia.org/wiki/List_of_computer_algebra_systems

General [edit]

System	Creator	Development started	First public release	Latest stable version	Latest stable release date	Cost (USD)	License	Notes
Axiom	Richard Jenks	1977	1993 and 2002[7]		August 2014[8]	Free	modified BSD license	General purpose CAS. Continuous Release using Docker Containers
Cadabra	Kasper Peeters	2001	2007	1.42	November 9, 2014	Free	GNU GPL	CAS for tensor field theory
Calcinator	George J. Paulos	2013	2016	2.0	February 2015	Free	Proprietary	Browser-based CAS for desktop and mobile devices
CoCoA-4	The CoCoA Team	1987	1995	4.7.5	2009	Free for non-commercial use	own license	Specialized CAS for commutative algebra
CoCoA-5	Abbott,Bigatti,Lagorio	2000	2011	5.1.1	2014	Free	GNU GPL	Specialized CAS for commutative algebra
Derive	Soft Warehouse	1979	1988	6.1	November 2007	Discontinued	Proprietary	CAS designed for pocket calculators; it was discontinued in 2007
DataMelt (DMelt)	jWork.ORG (Sergei Chekanov)	2005	2015	1.5	May 14, 2016	Free	GNU GPL	Java-based. Runs on the Java platform. Supports Python, Ruby, Groovy, Java and Octave.
Erable (aka ALGB)	Bernard Parisse, Mika Heiskanen, Claude-Nicolas Fiechter	1993	1993	4.20060919	April 21, 2009	Free	LGPL	CAS designed for Hewlett-Packard scientific graphing calculators of the HP 48/49/40/50 series; discontinued in 2009
Fermat	Robert H. Lewis	1986	1993	5.25	July 5, 2016	$70 if grant money available, otherwise $0	Proprietary	Specialized CAS for resultant computation and linear algebra with polynomial entries

Functionality [edit]

Below is a summary of significantly developed *symbolic* functionality in each of the systems.

System	Formula editor	Arbitrary precision	Calculus		Solvers					Graph theory	Number theory	Quantifier elimination	Boolean algebra	Tensors	Probability
			Integration	Integral transforms	Equations	Inequalities	Diophantine equations	Differential equations	Recurrence relations						
Axiom	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Calcinator	Yes	No	Yes	Yes	Yes	No	Yes	No	No	No	No	No	No	No	No
Magma	No	Yes	No	No	Yes	No	Yes	No	No	Yes	Yes	No	No	No	?
Magnus	No	Yes	No	No	No	No	No	No	No	?	?	No	?	No	No
Maple	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Mathcad	Yes	No	Yes	No	Yes	No	No	No	No	No	No	No	No	No	No
Mathematica	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes[20]	Yes
MathHandbook	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	No	Yes	No	Yes
Mathomatic	No	No	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No	No
Symbolic Math Toolbox (MATLAB)	No	Yes	Yes	Yes	Yes	No	No	Yes	No	No	No	No	No	No	No
Maxima	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
SageMath	No	Yes	Yes	Yes	Yes	Yes	Yes[A]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
SymPy	No	Yes	Yes	Yes	Yes	Yes	Yes[21]	Yes	Yes	No	Yes	No	Yes	Yes	Yes
Wolfram Alpha	Pro version only	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	?
GAP	No	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No
Xcas/Giac	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	No	No	No	Yes
Yacas	No	Yes	Yes	No	Yes	No	No	No	No	No	No	No	No	No	?

A. ^ via SymPy

Wikipedia → list of computer algebra systems (II)
https://en.wikipedia.org/wiki/List_of_computer_algebra_systems

Some problems

- No standardized structure
  The existing directories differ in structure and content. The information in SIGSAM and CA Fachgruppe restricts oneself to information on the software product, Wikipedia has also some information about version.

- No standardized metadata scheme
  The directories lists different metadata of a CAS System. More standardized information would be better but requires a higher effort.

- Maintenance
  The presented list are maintained manually. For more information about version and content the effort increases dramatically.

swMATH: The publication-based approach

swMATH is an approach for a comprehensive directory on mathematical software. basing on the close connection between publications and research data:

- (Mathematical) publications cite a (mathematical) software used.
- Hence, publications can be used to identification for software.
- The information of the publication provides a lot of relevant information
- about a software.

Unfortunately, software citations are very rudimentary, in the most cases they contain not more than the name of the software:

Böhm, Janko; Decker, Wolfram; Keicher, Simon; Ren, Yue

Current challenges in developing open source computer algebra systems. (English) Zbl 06585009

Kotsireas, Ilias S. (ed.) et al., Mathematical aspects of computer and information sciences. 6th international conference, MACIS 2015, Berlin, Germany, November 11–13, 2015. Revised selected papers. Cham: Springer (ISBN 978-3-319-32858-4/pbk; 978-3-319-32859-1/ebook). Lecture Notes in Computer Science 9582, 3-24 (2016).

Summary: This note is based on the plenary talk given by the second author at MACIS 2015, the Sixth International Conference on Mathematical Aspects of Computer and Information Sciences. Motivated by some of the work done within the Priority Programme SPP 1489 of the German Research Council DFG, we discuss a number of current challenges in the development of Open Source computer algebra systems. The main focus is on algebraic geometry and the system Singular.

Identification of mathematical software

We use some heuristic methods (searching for characteristic phrases such software/package/module/...) in connection with a name/artificial word in the zbMATH entries. The heuristic methods work surprisingly well.

but:

- Not all software can be identified.
- Most entries are really mathematical software but some belong to other classes of mathematical research data.

Of course, the publication-based approach is limited: Currently we don't get information about versions. But this information is necessary for the verification of research results and reuse of methods.
What can we do?

Development of a citation standard

A citation standard which describes exactly the used software would be a smart and fundamental solution of the problem.

A citation standard for software is discussed intensively in the Web for a long time.
A good summary about the existing practice is the blog of Mike Jackson:
http://www.software.ac.uk/how-cite-and-describe-software?mpw

Citation standard for software (I)

Moreover, he gives some recommendations. He distinguishes four scenarios:

Software purchased off-the shelf
ProductName. Version. Release Date. Publisher. Location

Software downloaded from the web
ProductName. Version. ReleaseDate. Publisher. Location (DOI or URL). DownloadDate

Software checked-out from a public repository
ProductName. (Version). Publisher. CheckoutDate. (Location (URL Repository)). RepositorySpecificCheckoutInformation

Software provided by a researcher
ProductName. (Version). Publisher. Location. ContactDetails. ReceivedDate

Citation standard for software (II)

An agreement on such a standard model would allow a precise identification of the used software.

The next step would be the implementation: In LaTeX, the BibLaTeX/Biber framework can be used. It allows the definition of arbitrary types and their corresponding features
The data model is defined in BibLaTeX in the *.dbx file.
There are some further configuration files, e.g. for the output.)

A first prototypic implementation is shown on the next slide.

The prototype:
A configuration file and the resulting page

```
\ProvidesFile{swmath.dbx}

\DeclareDatamodelEntrytypes{swmath}

\DeclareDatamodelEntryfields[swmath]{
  author,
  prodname,
  creator,
  maintainer,
  version,
  releasedate,
  year,
  provider,
  publisher,
  location,
  doi,
  url,
  downloaddate}

\DeclareDatamodelFields[type=list, datatype=literal]{prodname}
\DeclareDatamodelFields[type=list, datatype=name]{creator}
\DeclareDatamodelFields[type=list, datatype=name]{maintainer}
\DeclareDatamodelFields[type=list, datatype=literal]{version}
\DeclareDatamodelFields[type=field,datatype=literal]{releasedate}
\DeclareDatamodelFields[type=field,datatype=literal]{year}
%\DeclareDatamodelFields[type=field,datatype=verbatim]{publisher}
\DeclareDatamodelFields[type=field,datatype=verbatim]{provider}
\DeclareDatamodelFields[type=list, datatype=verbatim]{location}
\DeclareDatamodelFields[type=list, datatype=verbatim]{doi}
\DeclareDatamodelFields[type=list, datatype=literal]{note}
\DeclareDatamodelFields[type=list, datatype=verbatim]{downloaddate}
\endinput
```

environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity. One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control. R is the base for many R packages listed in https://cran.r-project.org/

References

[swm] **Gonnet, Gaston, Morven Gentleman, and Keith Geddes** (maintained by *Maplesoft Inc.*): **Maple 2016**, Version: **2016**, Date released: 2016-03-02
(Waterloo Maple Inc., Waterloo (Ontario)),
Available at `http://www.maplesoft.com/`.

[swm] **Greyson, Daniel R. and Michael E. Stillman** (maintained by *David Eisenbud*): **Macaulay2**, Version: **1.9**, Date released: 2016-04
(Dept. Mathematics, UIUC, Urbana-Champaign),
Available at `http://www.math.uiuc.edu/Macaulay2`.

[swm] **Sperber, Wolfram** (maintained by *Wolfgang Dalitz and Hagen Chaprary*): **swMATH**, Version: **00:00:99**, Date released: 2014-07-01
(FIZ Karlsruhe, Berlin),
Available at `http://www.swmath.org/`.

[swm] **Wickham, Hadley et al.** (maintained by *R-Project*): **R**, Version: **3.3.1**, Date released: 2016-01-21
(Lucent Technologies, Murray Hill (New Jersey)),
Available at `http://www.r-project.org/`.

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Another solution: Web Archives

The establishment of a BibLaTeX citation standard requires time (I hope, that some communities as the CAS community could play a pioneering role).

What can we do in the meantime?

We could use Web Archives to find out more information on a mathematical software (the method will be explained later).

Analysis of information from the publication-based approach

A lot of the software packages listed in swMATH is referenced in more than one publication.
This allows a lot of conclusions

- What are the mathematical subjects of the software? (description, keywords and MSC codes)
- What are the most important application areas? (keyword and MSC codes)
- How is the acceptance of the software? (number of references)
- What is related (similar) software? (citations plus MSC code)
- Is the software outdated? (citation profile)
- ...

The number of references is also an (heuristic) indicator for the quality, the subjects and the number of references for the granularity, ...

A lot of open questions, e.g., How can we classify the type of the swMATH entries with the aid of publications?

Enhancement of information in swMATH

by using Internet resources, for CAS especially

- search engines
- websites of a software
- mathematical software journals
- Web Archives

to

- identify a URL of websites and the source code of a software
- get more specific information about the available information of a software, especially source code, versions, documentations, authors, license conditions, and further context information (e.g. publications, algorithms, test data, ...)

Web Archives

- Archiving of (selected) web sites with the goal to have a consistent state at any time (This cannot always be achieved).

- Alternative to existing web archives: archiving on demand, e.g. to ensure a consistent state among all information of the software

- Allows preserving descriptions, change logs, documentation, …
  - Source code in case of open source software
  - Even binaries if freely available on the web
  - The website where bought / downloaded the artifact
- Even external resources, such as discussions on forums, tutorials, etc

Web Archives

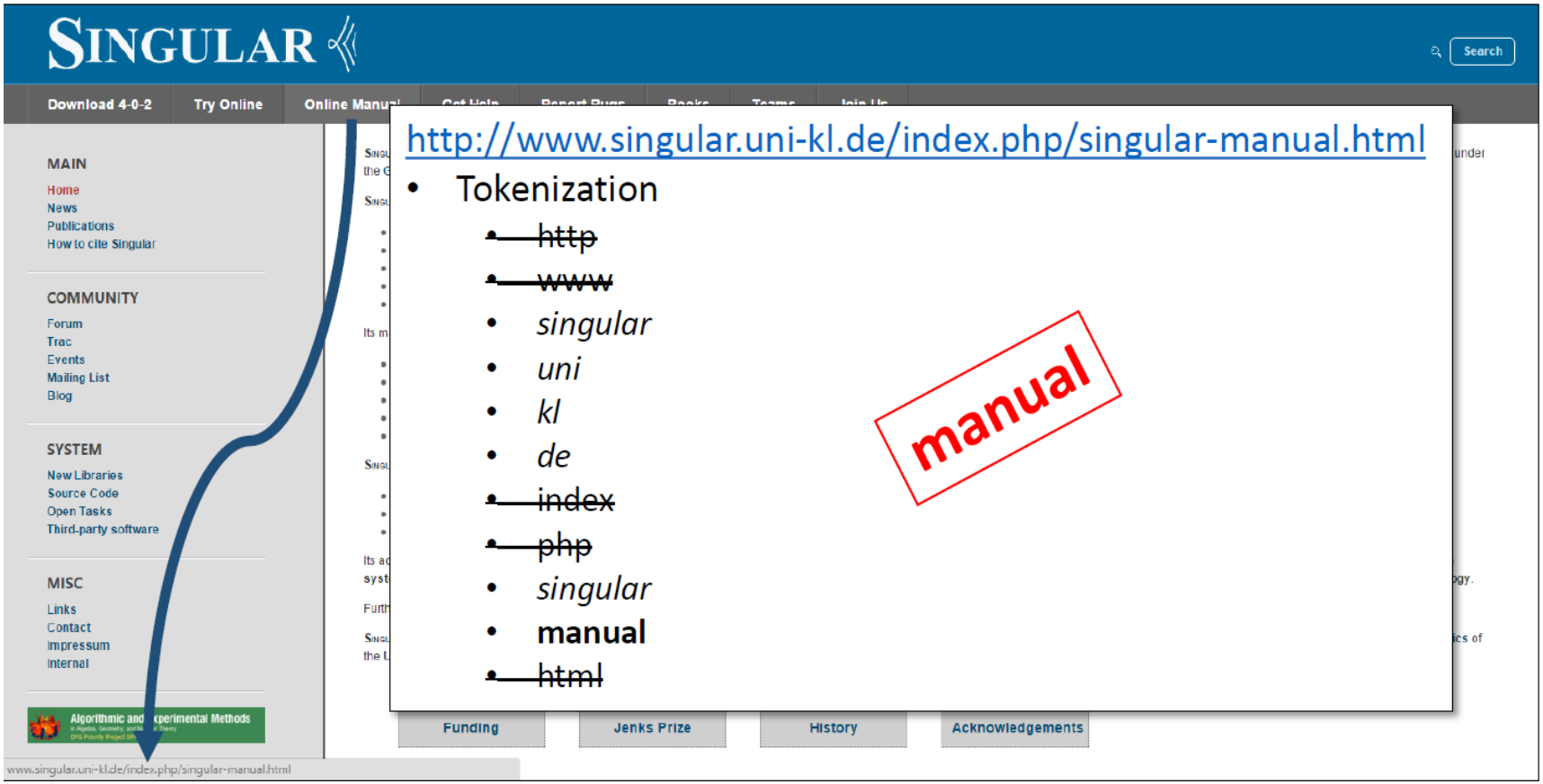
- Challenges
  - Not all pages archived at the exact same time / state / version
  - Mathematical software and its related websites not always easy to discover
    (the list of swMATH resources was used as a seed list)
- Questions
  - How well do websites represent software?
  - What does the web tell us about software?
  - What has already been archived?
  - What can we recover from the past?
  - What are we losing?

The experiments were done by Helge Holzmann (L3S), a cooperation partner of swMATH.

An example: The Singular website of swMATH

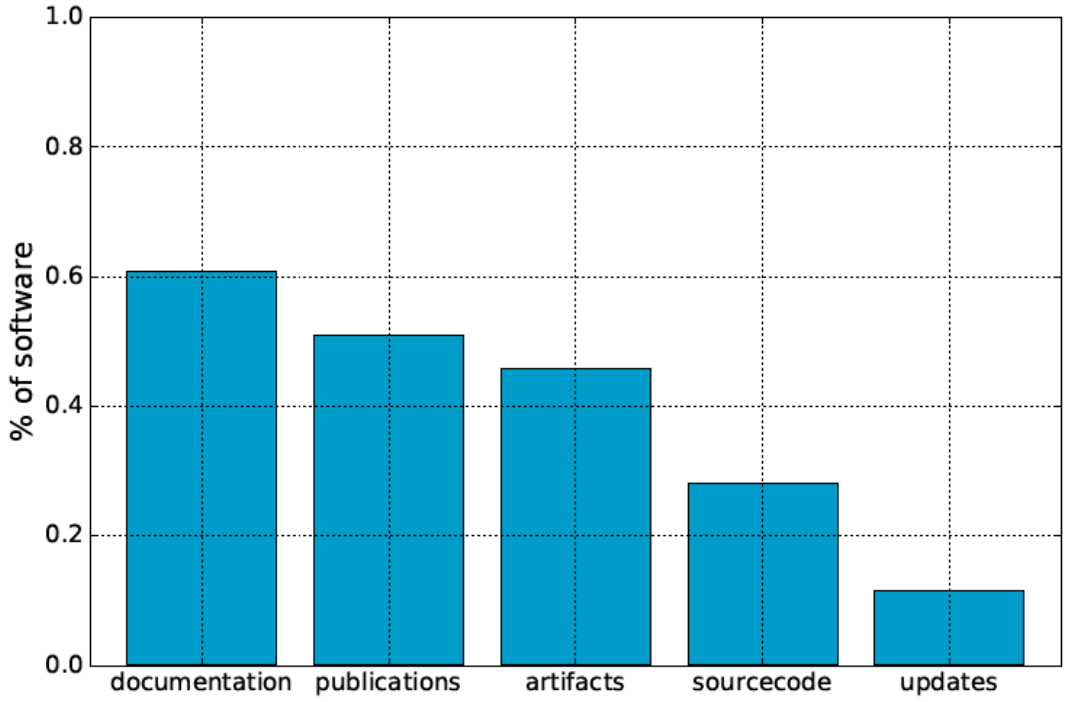
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

An example: Analysis of the archived websites (by some heuristics)



FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

First results: What kind of information can be found on the websites?



gnuplot

Gnuplot is a portable command-line driven graphing utility for Linux, OS/2, MS Windows, OSX, VMS, and many other platforms. The source code is copyrighted but freely distributed (i.e., you don't have to pay for it). It was originally created to allow scientists and students to visualize mathematical functions and data interactively, but has grown to support many non-interactive uses such as web scripting. It is also used as a plotting engine by third-party applications like Octave. Gnuplot has been supported and under active development since 1986.

URL: www.gnuplot.info/
Authors: Williams, T.; Kelley, C.

Add information on this software.

Related software:

R
GridBench
critic
Globus Toolkit
DiPerF
Lua
WIEN2k
Kan
GEPASI
Zebra

Show more...

Keywords for this software

References in zbMATH (referenced in 33 articles)

Showing results 1 to 20 of 33.

Sorted by year (citations) 20

1 **2** **next**

1. Faraway, Julian J.: Linear models with R (2015)
2. Garrido, José M.: Introduction to computational modeling using C and open-source tools (2014)
3. Zok, Tomasz; Popenda, Mariusz; Szachniuk, Marta: MCQ4Structures to compute similarity of molecule structures (2014)
4. Aniszewski, W.; Bogusławski, A.; Marek, M.; Tyliszczak, A.: A new approach to sub-grid surface tension for LES of two-phase flows (2012)
5. Feret, Jerome; Henzinger, Thomas; Koeppl, Heinz; Petrov, Tatjana: Lumpability abstractions of rule-based systems (2012)

Article statistics & filter:

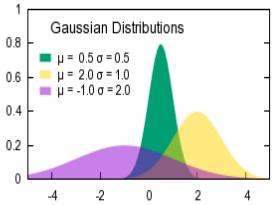
Search for articles

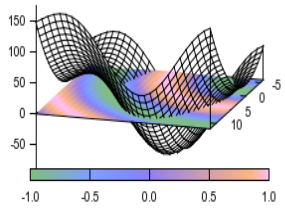
Clear

MSC classification

- Top MSC classes
  - **65** Numerical analysis
  - **68** Computer science
  - **76** Fluid mechanics
  - **81** Quantum Theory
  - **92** Applications of...
- Other MSC classes

gnuplot version 4.6 released!





Gnuplot is a portable command-line driven graphing utility for Linux, OS/2, MS Windows, OSX, VMS, and many other platforms. The source code is copyrighted but freely distributed (i.e., you don't have to pay for it). It was originally created to allow scientists and students to visualize mathematical functions and data interactively, but has grown to support many non-interactive uses such as web scripting. It is also used as a plotting engine by third-party applications like Octave. Gnuplot has been supported and under active development since 1986.

Gnuplot supports many different types of 2D and 3D plots

Please see demos here.

Gnuplot supports many different types of output

interactive screen display:	cross-platform (Qt, wxWidgets, x11) or system-specific (MS Windows, OS/2)
static screen display:	system-specific (OSX(aqua), svga, ...)
direct output to file:	postscript (including eps), pdf, png, gif, jpeg, LaTeX, metafont, emf, svg, ...
mouseable web display formats:	HTML5, svg

Current release is 4.6 (patchlevel 0)

- Download from SourceForge
- Release Notes
- User Manual (PDF)
- version 4.6 demo gallery.

The Development version is gnuplot 4.7 (CVS)

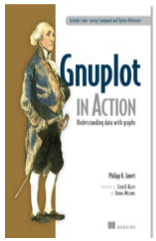
- New features are being added regularly. You are welcome to build gnuplot from the CVS source code. Instructions here.
- Version 4.7 Documentation (PDF), including new features.
- Version 4.7 demo gallery.

News

- 08.03.2012: Release gnuplot 4.6.0.
- 14.11.2011: Release gnuplot 4.4.4.
- 01.03.2011: Release gnuplot 4.4.3.
- 26.09.2010: Release gnuplot 4.4.2.
- 11.09.2010: Release gnuplot 4.4.1.
- 13.03.2010: Release gnuplot 4.4.0.

Copyright/licensing

Gnuplot's copyright.



Now available: A book on gnuplot!

Gnuplot in Action
Understanding Data with Graphs
by Philipp K. Janert

Manning Publications (2009)
ISBN: 1933988398
ISBN-13: 978-1933988399

Conclusions and an invitation for cooperation

Information services are part of the scientific infrastructure
Information services are an important part of the scientific infrastructure. Up to now, a comprehensive concept for the information infrastructure is missing.

The Web provides new opportunities
New concepts from the knowledge management allow the build up and maintain a suitable scientific infrastructure in an efficient way.

But we need your feedback!
The development of a powerful information infrastructure requires the cooperation of the scientific community and experts in information sciences.

Thank you for your patience!

FIZ Karlsruhe

Leibniz-Institut für Informationsinfrastruktur