

TIB LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB.eu

Non-Textual Information:
Software and Services for Science (S3)
Hannover, Germany



fidmath
Fachinformationsdienst Mathematik

fidmath.de

Software as a First-Class Citizen in Web Archives

Helge Holzmann

L3S Research Center, Hannover, Germany

www.HelgeHolzmann.de

05/10/2017

Helge Holzmann (holzmann@L3S.de)



L3S.de



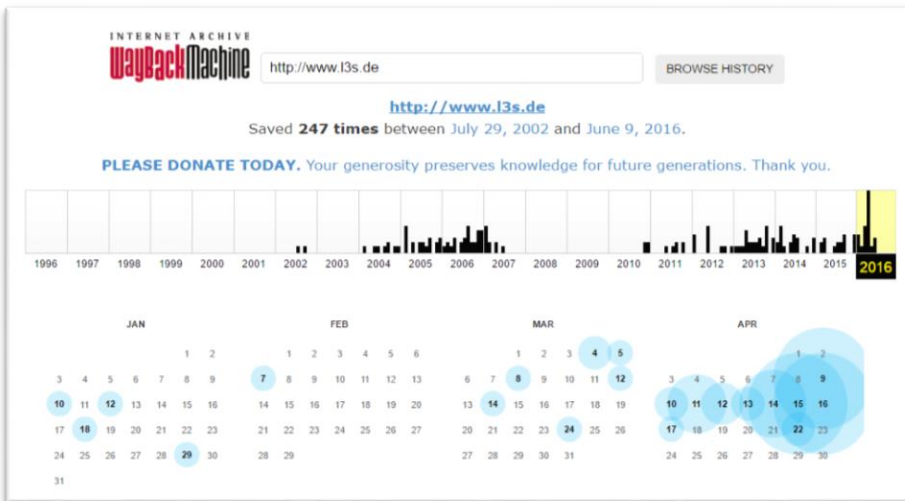
ALEXANDRIA-project.eu



What is a Web archive?

- Web archives preserve our history as documented on the Web
- Consists of all kinds of Web resources
 - i.e., HTML, images, video, scripts, ...
- ... stored in big files in the standardized *WARC* format

Common Crawl



Forschungszentrum L3S Research Center

abida
ASSESSING BIG DATA

Web Science Investigating the Future of Information and Communication.

Startup@L3S

Conferences
WebSci 16 Hannover

TPDL 2016: Call for Participation
TPDL 2016: 20th International Conference on Theory and Practice of Digital



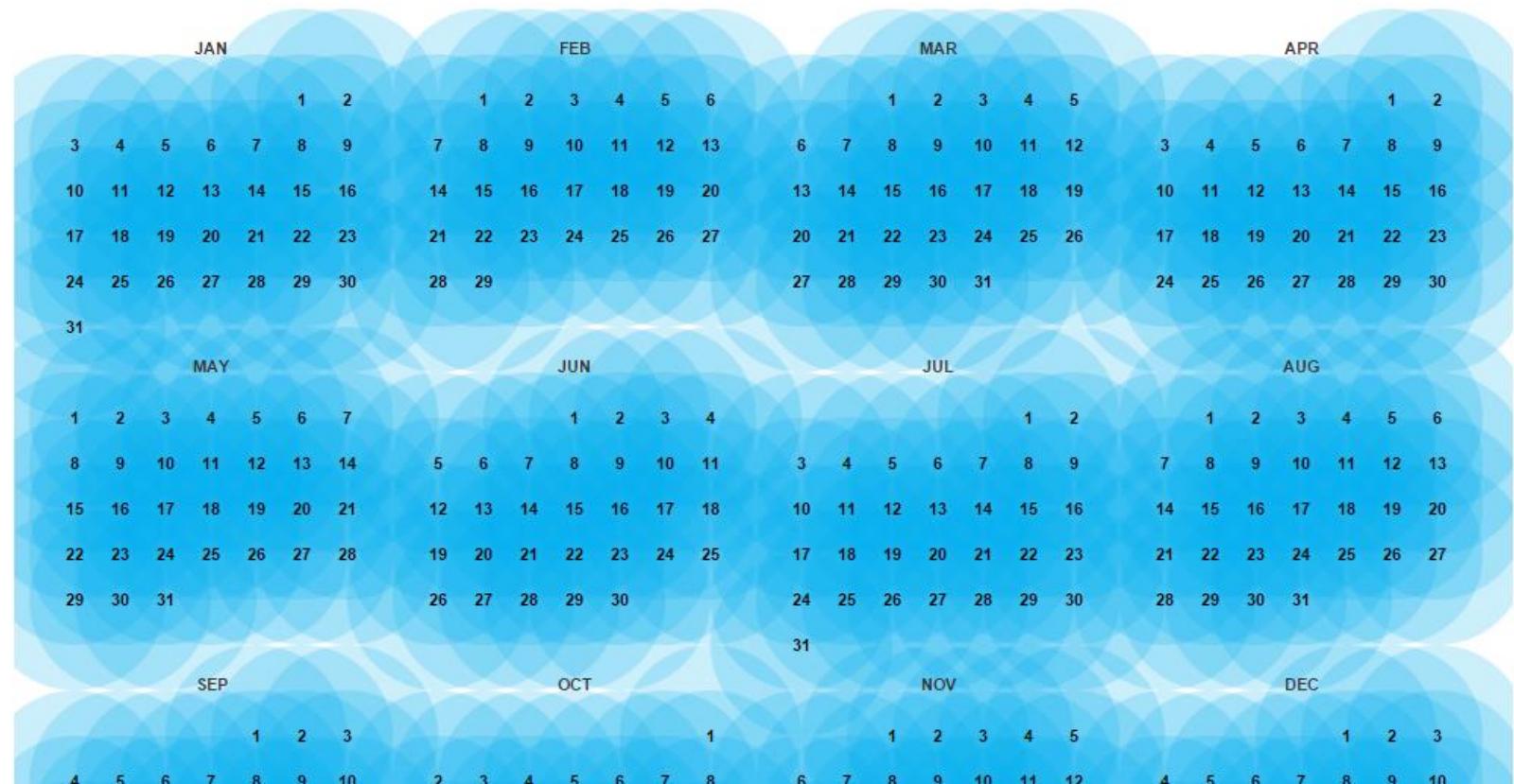
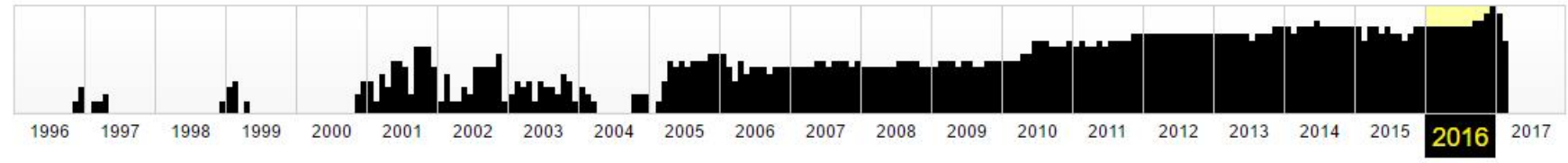
http://nytimes.com

BROWSE HISTORY

<http://nytimes.com>

Saved **49,342 times** between November 12, 1996 and February 12, 2017.

PLEASE DONATE TODAY. Your generosity preserves knowledge for future generations. Thank you.



Subscribe to The Times.



The New York Times

Wednesday, November 7, 2012 Last Update: 8:07 PM ET

Subscribe to The Times.

HOME DELIVERY

DIGITAL SUBSCRIPTION

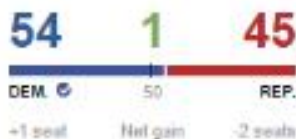
 Search

Follow Us | Personalize Your Weather

President »



Senate »



House »



FULL RESULTS

- President Big Board »
- Exit Polls »
- National Map »

- WORLD
- U.S.
- POLITICS
- NEW YORK
- BUSINESS
- DEALBOOK
- TECHNOLOGY
- SPORTS
- SCIENCE
- HEALTH
- ARTS
- STYLE
- OPINION

- Autos
- Blogs
- Books
- Cartoons
- Classifieds
- Crosswords
- Dining & Wine
- Education
- Event Guide
- Fashion & Style
- Home & Garden
- Jobs
- Magazine
- Movies
- Music
- Obituaries
- Public Editor
- Real Estate
- Special Reports

Obama Wins a Clear Victory, but Balance of Power Is Unchanged in Washington

Boehner Strikes Conciliatory Tone in Talk of Fiscal Cliff

By JONATHAN WEISMAN and JACKIE CALMES 38 minutes ago

The speaker of the House said Wednesday he is ready to accept a budget deal that raises revenues if it is linked to an overhaul of the tax code and entitlements.

Boehner's Offer May Bring Sides to Table 5:03 PM ET

Debate Over Size of Federal Government Left Unresolved

By PETER BAKER 3:15 PM ET

After \$6 billion, two dozen presidential primary days, four general election debates and more TV ads than anyone could watch, the two parties essentially fought to a standstill.

Comments



Today's TimesCast

Republicans search for a path forward. | Changes in store for Obama's cabinet.

Play Video

G.O.P. Factions Grapple Over Meaning of Loss

By MICHAEL COOPER 4:15 PM ET

There was no shortage of theories from inside and outside the party about where it fell short and what to do next.

Day After Election, a Sharp Loss on Wall Street

By NELSON D. SCHWARTZ 4:50 PM ET

Stocks moved sharply lower in New York, with both major indexes down 2-4 percent.

Marriage Gains Cheer Gay Rights Advocates

By ERIK ECKHOLM 12:20 PM ET

Supporters called votes in Maine and Maryland a turning point. Results in Washington State were still being tallied.

MORE ELECTION NEWS

- Californians Back Taxes to Avoid Education Cuts
- Democrats Gain in Senate, Adding to Majority
- Republicans Shore Up Incumbents, Holding House
- Why The Times Was Slower in Calling Election
- Twitter Kills the Fall Whale. One Tweet at a Time

The Opinion Pages

EDITORIAL

Marriage Equality

The nation took a step toward full equality for all Americans with a string of victories for gay marriage.

FIXES

A Change in the Weather on Wall Street

Let's act on the renewed sense of urgency about climate change.



OP-ED COLUMNISTS

- Bruni: Maine and Maryland Say 'We Do'
- Douthat: The Obama Realignment
- Keller: Mo'bama
- Friedman: Hope, Part 2
- Brooks, Collins: Orca, Meet Obama

MORE IN OPINION

- Editorial: Obama's Win
- Op-Ed: Unwelcome at the Party

MARKETS

last week's change

JAPAN	HANG SENG	SHANGHAI	CHINA
Nikkei	Hang Seng	Shanghai	
19,376.93	23,674.98	3,197.33	
Closed for holiday	Closed for holiday	Closed for holiday	

Data delayed at least 15 minutes

GET QUOTES My Portfolio

Stock, ETFs, Funds Go

BIRTHDAY GIFT
THANK YOU GIFT



The Wayback Machine

- Replays Web resources with a temporal dimension
 - Identified by **URL** and **timestamp** (crawl time)

```
http://web.archive.org/web/20121107020708/http://www.nytimes.com/
```

```
http://web.archive.org/web/TIMESTAMP/URL
```

- Challenges
 - A changed URL of a page is a different resource
 - No logical / object identifiers: what does a URL refer to?
 - Search capabilities are limited
 - *'Random'* timestamps, not connected to a meaning

Archived Resource Identifiers

- **Current state:** based on URL and timestamp

```
http://web.archive.org/web/20121107020708/http://www.nytimes.com/
```

```
http://web.archive.org/web/TIMESTAMP/URL
```

- **Desired:** based on objects and related timestamps / events

```
http://web.archive.org/web/OBJECT_ID/OBJECT_EVENT
```

- **Examples:** Entity at event / software at version

```
http://web.archive.org/web/obama_website/election_2012
```

```
http://web.archive.org/software/mathematica/v5.2
```

1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 **2009** 2010 2011 2012 2013

About 109 results (0.383 seconds) [Hide non-archived results](#)

✕ clear times

Mathematica (2009)

<http://wolfram.com/>

Mathematica (2009) ... Wolfram Research: **Mathematica**, Technical and Scientific Software (2009) ... **Mathematica** Homepage (2012) ... Wolfram's **Mathematica** (2012) ... Research, Inc. Makers of **Mathematica** (2011) ... **Mathematica** 8 (2011) ... **Mathematica** Web (2010) ... **Mathematica** (Wolfram Research) (2013) ...
2009 · 2012 · 2011 · 2010 · 2013

Mathematica (2010)

<http://wolfram.com/products/mathematica/index.html>

Mathematica (2010) ... <http://www.wolfram.com/products/mathematica/index.html> (2010) ... <http://www.wolfram.com/products/mathematica/index.html> (2009) ... **Mathematica** Vollversion (2009) ... **Mathematica** (kommerzielle Software) (2011) ... <http://www.wolfram.com/products/mathematica/index.html> (2012) ... Wolfram **Mathematica** 7 (2013) ...
2010 · 2009 · 2011 · 2012 · 2013

Mathematica (2010)

<http://de.wikipedia.org/wiki/mathematica>

Mathematica (2010) ... **Mathematica** – Wikipedia (2010) ... <http://de.wikipedia.org/wiki/Mathematica> (2011) ... **Mathematica** (Wikipedia) (2013) ...
2010 · 2009 · 2011 · 2012 · 2013

Mathematica (2013)

<http://wolfram.com/mathematica>

Mathematica (2013) ... **Mathematica** 8 (2013) ... **Mathematica**-Software (2013) ... **Mathematica** 8.0 (2010) ... www.wolfram.com/mathematica/ (2010) ... <http://www.wolfram.com/mathematica/> (2011) ... Wolfram **Mathematica** 8 (2011) ... Wolfram|**Mathematica** (2012) ... Wolfram **Mathematica** (2012) ... Software **Mathematica** (2009) ...
2013 · 2010 · 2011 · 2012 · 2009



1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 **2009** 2010 2011 2012 2013

About 109 results (0.383 seconds) [Show non-archived results](#)

✕ clear times

Mathematica (2009)

<http://wolfram.com/>

Mathematica (2009) ... Wolfram Research: **Mathematica**, Technical and Scientific Software (2009) ... **Mathematica** Homepage (2012) ... Wolfram's **Mathematica** (2012) ... Research, Inc. Makers of **Mathematica** (2011) ... **Mathematica** 8 (2011) ... **Mathematica** Web (2010) ... **Mathematica** (Wolfram Research) (2013) ...
[2009](#) · [2012](#) · [2011](#) · [2010](#) · [2013](#)

Mathematica (2010)

<http://wolfram.com/products/mathematica/index.html>

Mathematica (2010) ... <http://www.wolfram.com/products/mathematica/index.html> (2010) ... [com/products/mathematica](#) (2009) ... **Mathematica** Vollversion (2009) ... **Mathematica** (kommerzielle Software) (2011) ... <http://www.wolfram.com/products/mathematica/index.html> (2012) ... Wolfram **Mathematica** 7 (2013) ...
[2010](#) · [2009](#) · ~~2011~~ · ~~2012~~ · ~~2013~~

Mathematica (2010)

<http://de.wikipedia.org/wiki/mathematica>

Mathematica (2010) ... **Mathematica** – Wikipedia (2010) ... <http://de.wikipedia.org/wiki/Mathematica> (2011) ... **Mathematica** (Wikipedia) (2013) ...
[2010](#) · [2009](#) · [2011](#) · [2012](#) · [2013](#)

Mathematica (2013)

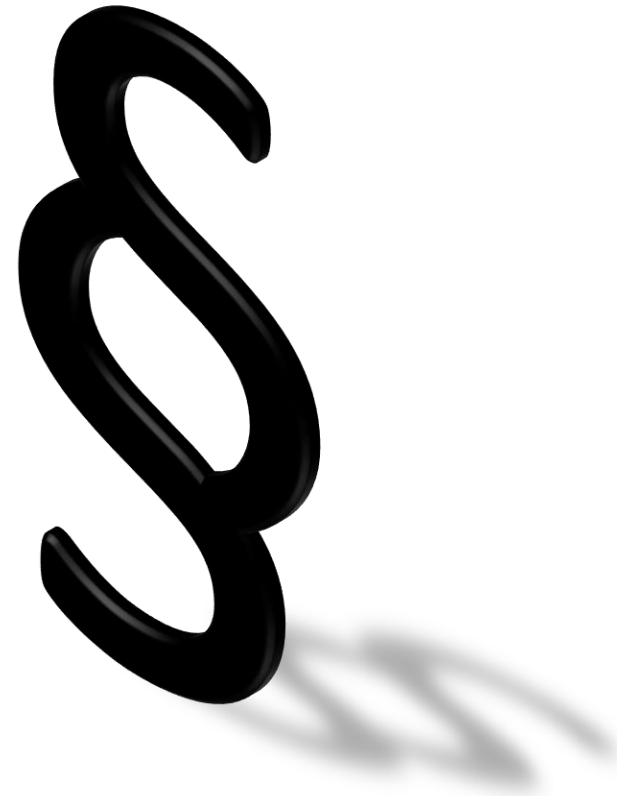
<http://wolfram.com/mathematica>

Mathematica (2013) ... **Mathematica** 8 (2013) ... **Mathematica**-Software (2013) ... **Mathematica** 8.0 (2010) ... www.wolfram.com/mathematica/ (2010) ... <http://www.wolfram.com/mathematica/> (2011) ... Wolfram **Mathematica** 8 (2011) ... Wolfram|**Mathematica** (2012) ... Wolfram **Mathematica** (2012) ... Software **Mathematica** (2009) ...
[2013](#) · [2010](#) · [2011](#) · [2012](#) · ~~2009~~



Why software in Web archives?

- Archiving (scientific) software is crucial
 - Software is used in science across all disciplines
 - as object of research in computer science, maths, ...
 - as tool in humanities and many others ...
 - Software is widely used / cited in publications
 - but rarely published together with them
 - Referenced software / versions may become unavailable
- Archiving software applications is not entirely possible
 - How to obtain the proprietary software / services?
 - Legal questions and issues
 - Are we allowed to archive software without a license?
 - Are we allowed to provide access to archived software?



What is software?

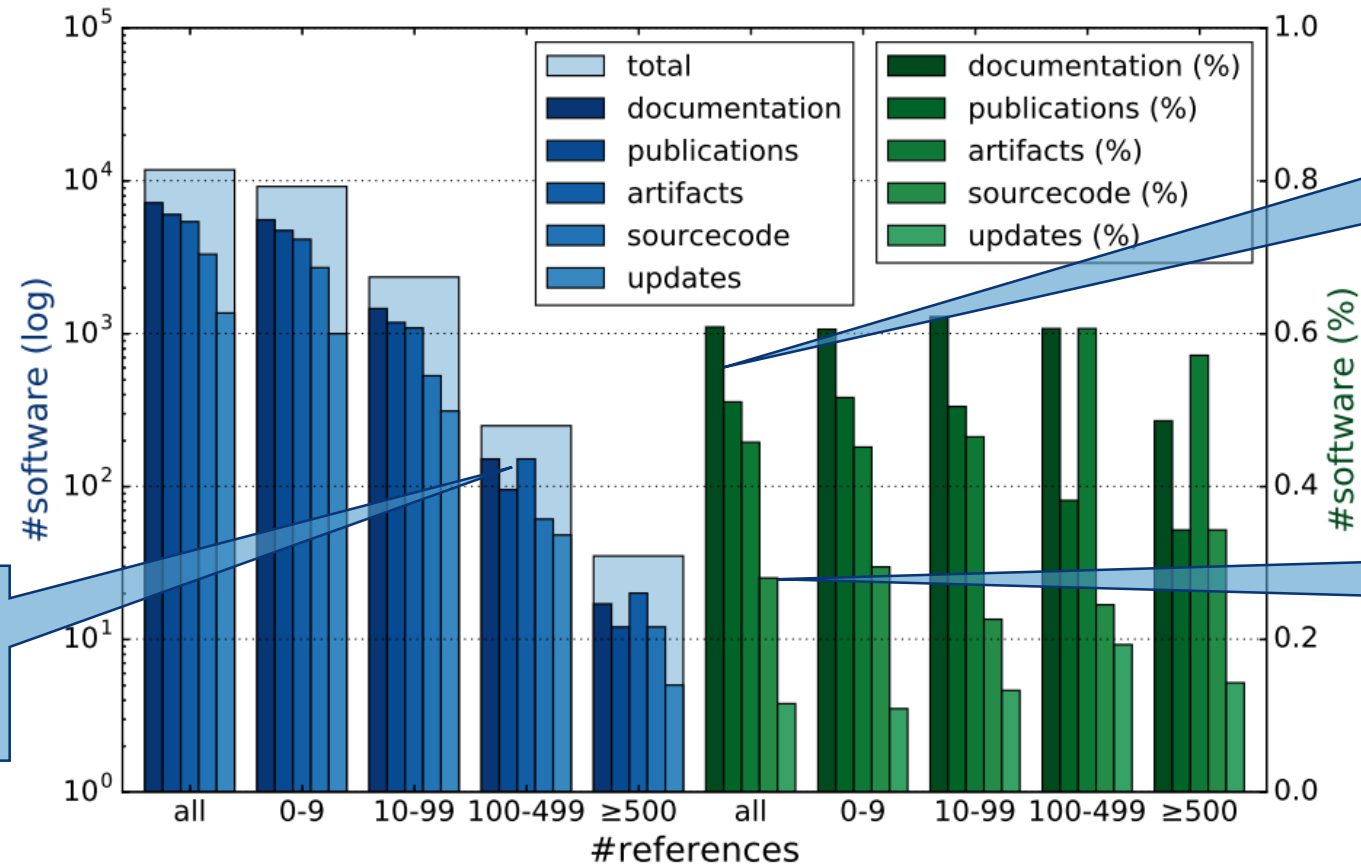
“Software is a comprehensive term used to identify all of the nonhardware components of a computer or communications system. Software includes computer **programs, data** that is used by these programs, and any paper or computer based **documentation** that describes computer systems and how to use them. Software determines **what** a computer does and **how** it does it”

Edmunds (1987)

- What do we actually need?
 - Execution requires the compiled **program** and **data / parameters**
 - To understand the purpose (**what?, how?**), a description may be sufficient
 - To understand the features, **documentation** may be sufficient
 - To comprehend results, **code / algorithms** may be sufficient

What is available on the Web?

- Analysis based on the hyperlinks on mathematical software pages



Artifacts provided for highly referenced articles

~60% link to some sort of documentation

~30% provide source code



Getting temporal...

- **Desired:** Software at version?

[http://web.archive.org/software/*mathematica*/v5.2](http://web.archive.org/software/mathematica/v5.2)

- **Alternative:** Software as used in a publication

[http://web.archive.org/software/*SW*/in/*ARTICLE*](http://web.archive.org/software/SW/in/ARTICLE)


- **Goal:** Linking software / publications with Web archives
 - Based on software catalogs or repositories, such as *swMATH*
 - Best guess about the cited state of a software: publication date
 - to be improved in the future by a pro-active approach as part of the publication process



swMATH.org



- Information service for mathematical software
 - More than 12,000 records, referenced in more than 110,000 articles
 - Following a publication-driven approach
 - Software identified by name, manually checked and complemented (e.g., URLs added)



Search Advanced search Browse

About & Contact Feedback Contribute Help zbMATH

Showing results 1 to 20 of 802. Sorted by year (citations) 20

1 2 3 ... 39 40 41 next

- Biviá-Ausina, Carles; Fukui, Toshizumi: Mixed \Lojasiewiczexponents and log canonical thresholds of ideals (2016)
- Botbol, Nicolás; Dickenstein, Alicia: Implicitization of rational hypersurfaces via linear syzygies: a practical overview (2016)
- Dimca, Alexandru; Sticlaru, Gabriel: Syzygies of Jacobian ideals and weighted homogeneous singularities (2016)
- Dumnicki, M.; Famik, Ł.; Głowska, A.; Lampa-Baczyńska, M.; Malara, G.; Szemberg, T.; Szpond, J.; Tutaj-Gasińska, H.: Line arrangements with the maximal number of triple points (2016)
- Dumnicki, M.; Szemberg, T.; Tutaj-Gasińska, H.: Symbolic powers of planar point configurations. II. (2016)
- Ellis, Graham: Cohomological periodicities of crystallographic groups. (2016)
- Erócal, Burçin; Motsak, Oleksandr; Schreyer, Frank-Olaf; Steenpaß, Andreas: Refined algorithms to compute syzygies (2016)
- Ferčec, Brigita; Giné, Jaume; Romanovski, Valery G.; Edneral, Victor F.: Integrability of complex planar systems with homogeneous nonlinearities (2016)
- Giesbrecht, Mark; Heinle, Albert; Levandovskyy, Viktor: Factoring linear partial differential operators in \mathbb{Z} variables (2016)
- Giné, Jaume; Valls, Claudia: Center problem in the center manifold for quadratic differential systems in \mathbb{P}^3 (2016)
- Hausen, Jürgen; Keicher, Simon; Laface, Antonio: Computing Cox rings (2016)
- Marais, Magdaleen S.; Steenpaß, Andreas: The classification of real singularities using S ingular. II: The structure of the equivalence classes of the unimodal singularities. (2016)
- Marcolla, Chiara; Pellegrini, Marco; Sala, Massimiliano: On the small-weight codewords of some Hermitian codes (2016)
- Margulies, S.; Morton, J.: Polynomial-time solvable #CSP problems via algebraic models and Pfaffian circuits (2016)
- Ma, Yue; Wang, Chu; Zhi, Lihong: A certificate for semidefinite relaxations in computing positive-dimensional real radical ideals (2016)
- Rollenske, Sönke: A new irreducible component of the moduli space of stable Godeaux surfaces (2016)
- Adamus, Janusz; Seyedinejad, Hadi: A fast flatness testing criterion in characteristic zero (2016)
- Afzal, Deeba; Pfister, Gerhard: A classifier for simple isolated complete intersection singularities (2015)
- Albert, Mario; Fetzer, Matthias; Sáenz-de-Cabezón, Eduardo; Seiler, Werner M.: On the free resolution induced by a

Search for articles Clear

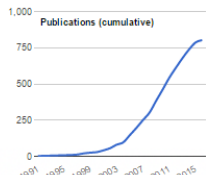
MSC classification

- 13 Commutative algebra
- 14 Algebraic geometry
- 32 Functions of several...
- 34 Ordinary differential...
- 68 Computer science
- Other MSC classes

Publication year

- 2010 - today
- 2005 - 2009
- 2000 - 2004
- before 2000

Chart: cumulative / absolute



SINGULAR

SINGULAR is a Computer Algebra system (CAS) for polynomial computations in commutative algebra, algebraic geometry, and singularity theory. SINGULAR's main computational objects are ideals and modules over a large variety of baserings. The baserings are polynomial rings over a field (e.g., finite fields, the rationals, floats, algebraic extensions, transcendental extensions), or localizations thereof, or quotient rings with respect to an ideal. SINGULAR features fast and general implementations for computing Groebner and standard bases, including e.g. Buchberger's algorithm and Mora's Tangent Cone algorithm. Furthermore, it provides polynomial factorizations, resultant, characteristic set and gcd computations, syzygy and free-resolution computations, and many more related functionalities. Based on an easy-to-use interactive shell and a C-like programming language, SINGULAR's internal functionality is augmented and user-extendible by libraries written in the SINGULAR programming language. A general and efficient implementation of communication links allows SINGULAR to make its functionality available to other programs.

⊗ This software is also referenced in ORMS.

Keywords for this software

Singular, superstring, integrability, Groebner basis, matrix factorizations, MiInor number, algebraic geometry, primary decomposition, decomposition

URL: www.singular.uni-kl.de
 Manual: www.singular.uni-kl.de
 Authors: Wolfram Decker, Gert-Martin Greuel; Gerhard Pfister; Hans Schönemann
 Platforms: ix86-Linux, SunOS-5, IRIX-6, ix86-Win (runs on Windows 95/98/NT4/2000/XP/Vista), FreeBSD, MacOS X, x86_64-Linux (AMD64/Opteron/EM64T), IA64-Linux
 Licence: free and open-source under the GNU General Public Licence.

Add information on this software.

Related software:
 Macaulay2
 CoCoA
 Magma
 Maple
 primdec
 Sage



Tempas TimePortal

- Connecting swMATH and the Wayback Machine



URL: www.singular.uni-kl.de
 InternetArchive: [\[icon\]](#)
 Versions: [Info](#)
 Manual: www.singular.uni-kl.de
 Authors: Wolfram Decker, Gert-Martin Greuel, Gerhard Pfister, Hans Schönemann
 Platforms: ix86-Linux, SunOS-5

886. Lin, Zhiping; Ying, Jiang Qian; Xu, Li: Factorizations for nD polynomial matrices (2001) [\[icon\]](#)
 887. Orecchia, F.: Implicitization of a general union of parametric varieties (2001) [\[icon\]](#)
 888. Schulze, Mathias: Algorithms for the Gauss-Manin connection (2001) [\[icon\]](#)
 889. Selesnick, Ivan W.: Smooth wavelet tight frames with zero moments (2001) [\[icon\]](#)
 890. Siqveland, Arvid: The method of computing formal moduli (2001) [\[icon\]](#)
 891. Zerz, Eva: Extension modules in behavioral linear systems theory (2001) [\[icon\]](#)
 892. Albano, G.; Cioffi, F.; Orecchia, F.; Ramella, I.: Minimally generating ideals of rational parametric curves in polynomial time (2000) [\[icon\]](#)
 893. Bermejo, Isabel; Gimenez, Philippe: On Castelnuovo-Mumford regularity of projective curves (2000) [\[icon\]](#)
 894. Gatermann, Karin: Computer algebra methods for equivariant dynamical systems (2000) [\[icon\]](#)
 895. Greuel, Gert-Martin: Computer algebra and algebraic geometry -- achievements and perspectives (2000) [\[icon\]](#)
 896. Greuel, Gert-Martin; Lossen, Christoph; Shustin, Eugenii: Castelnuovo function, zero-dimensional schemes and singular plane curves. (2000) [\[icon\]](#)
 897. Guerrieri, Anna; Swanson, Irena: Jacobian ideals of trilinear forms: An application of 1-genericity (2000) [\[icon\]](#)
 898. Decker, Wolfram; de Jong, Theo; Greuel, Gert-Martin; Pfister, Gerhard: The normalization: A new algorithm, implementation and comparisons (1999) [\[icon\]](#)
 899. Decker, Wolfram; Greuel, Gert-Martin; Pfister, Gerhard: Primary decomposition: Algorithms and comparisons (1999) [\[icon\]](#)
 900. Martin, Bernd: Computing versal deformations with SINGULAR (1999) [\[icon\]](#)

Software SINGULAR in
 Bermejo, Isabel; Gimenez, Philippe: On Castelnuovo-Mumford regularity of projective cu... (2000)

home | documentation | updates | artifacts

Service & Support
 Documentation & Bibliography

SINGULAR
 A Computer Algebra System for Polynomial Computations

Welcome to SINGULAR

Top News:

- **June 2000:** Release of [SINGULAR version 1.3.8](#), a major upgrade from version 1.2.3 and an alpha vers the upcoming release version 2.0. This version also includes Singular for Windows 95/98/NT/2000
- **February 2000:** SINGULAR has a new WWW home site: <http://www.singular.uni-kl.de>

Use any of the following links to find out more about SINGULAR:

[Overview & Demonstrations](#) | [News & Announcements](#) | [Download & Distribution](#)

Click [here](#) to open this webpage version in the Internet Archive's Wayback Machine



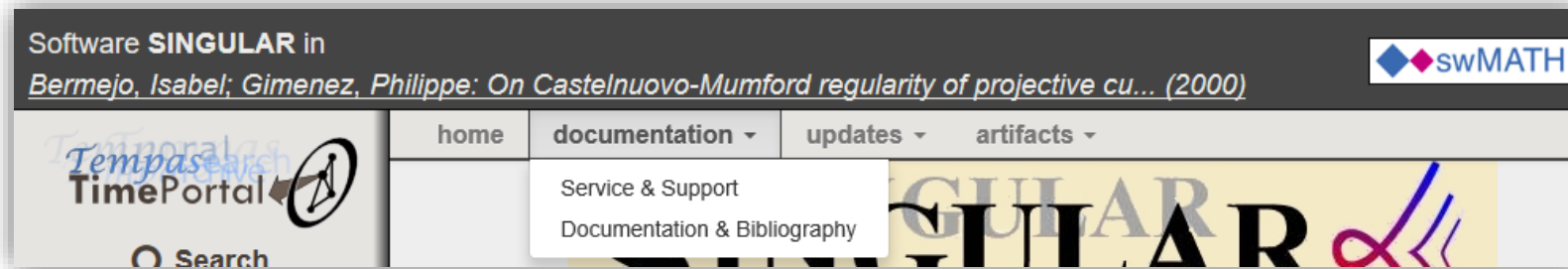
Software as a First-Class Citizen

- Identified by **software** and **publication**

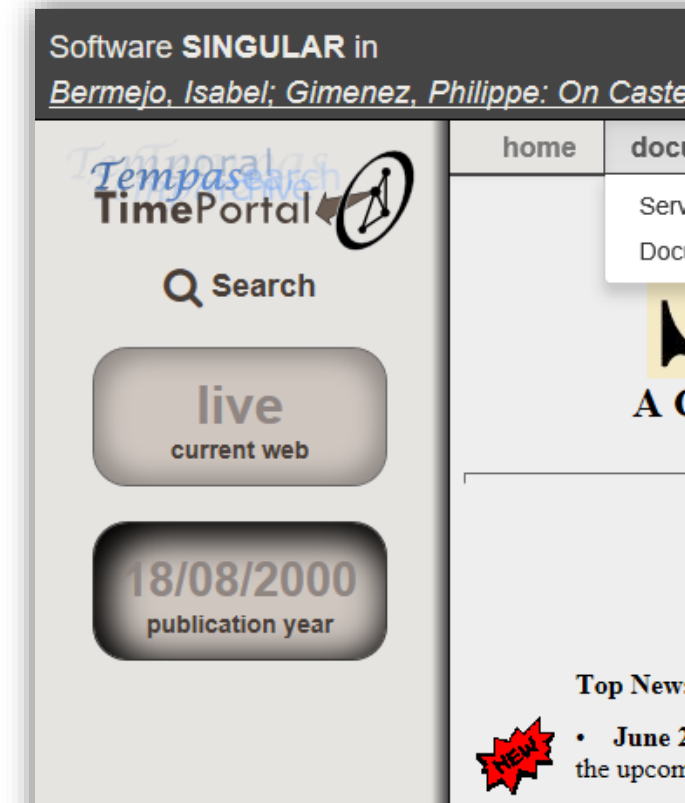


`http://tempas.L3S.de/...?software=866&publication=01415032`

- Focus on the software rather than its webpage

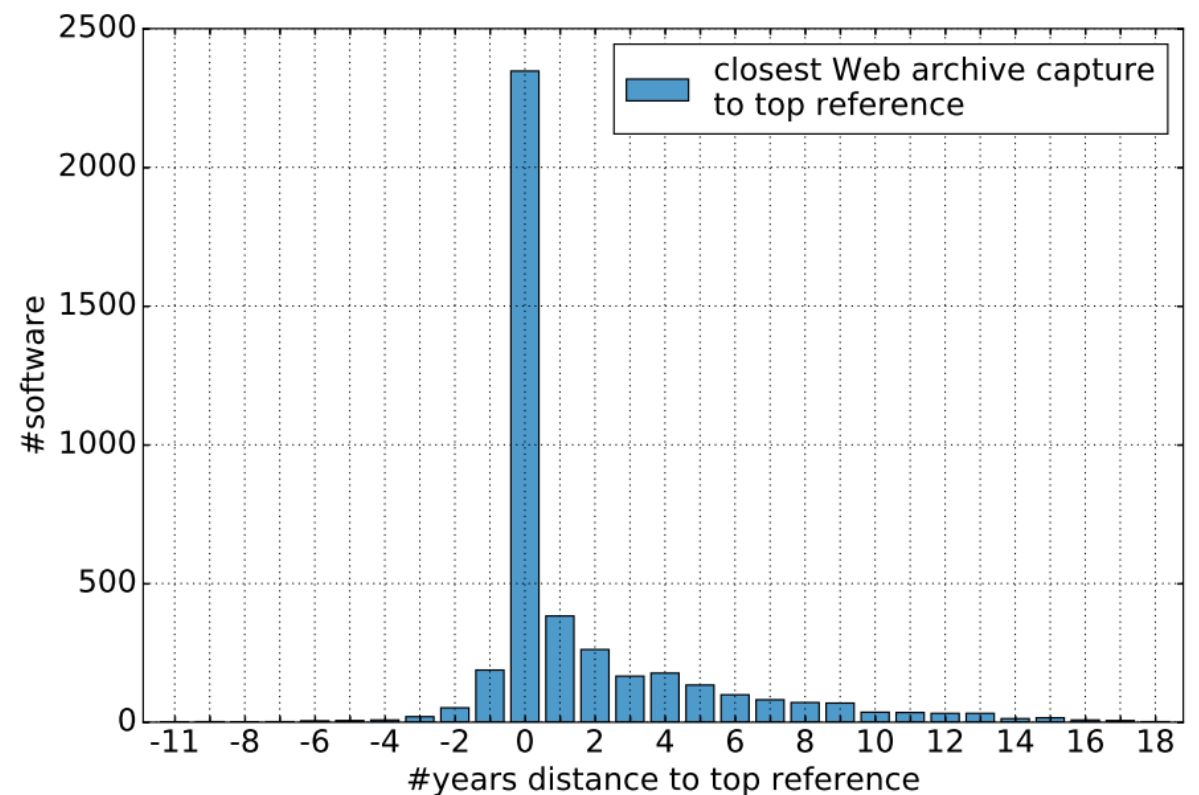
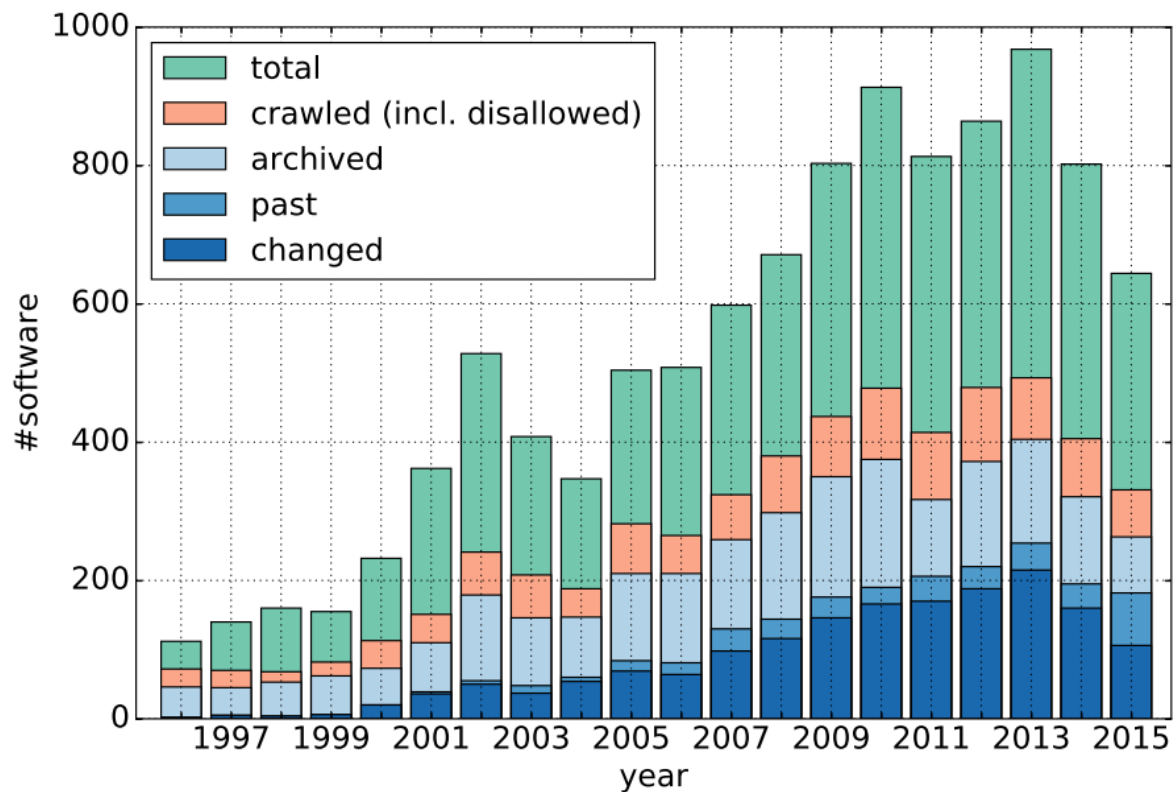


- Automatically augmented with software-specific links
 - here: *documentation, updates, artifacts*
- Meaningful captures rather than random crawl times



What has been archived so far?

- Analysis of software websites in the year of the top referencing article
 - Based on the Internet Archive's Wayback Machine



We can do better (Future Work)

- **Micro archives** representing software (states / versions)
 - Comprising **various resources** on the Web related to a software
 - Archived **on demand at time of use**, not when referenced / published
 - **Tailoring the archiving process** rather than only the replay / access methods
 - e.g., treating source code differently from websites and others
- **Referencing** software micro archives by handle (DOI?)
 - Enable citation of archived software (at time of use) in publications
- **Mining meta data** from these archives
 - Automatic detection of version, features, etc.
 - Automatic labelling of the archived snapshots / timestamps
- **Generalizing** this approach to other entities beyond software

Conclusions

- The **Web provides access to software** quite comprehensively
 - Websites can serve as **surrogates** of the actual software in many aspects
 - A **considerable number** of SW websites include **documentation and artifacts**
- Already **~50%** of software websites are **archived** today
 - We want to create the infrastructure to improve this in the future
- Details on this work:
 - H. Holzmann, W. Sperber and M. Runnwerth. **Archiving Software Surrogates on the Web for Future Reference**. 20th International Conference on Theory and Practice of Digital Libraries (TPDL), September 2016.
- Details on Tempas:
 - H. Holzmann, W. Nejdl and A. Anand. **Exploring Web Archives Through Temporal Anchor Texts**. 9th International ACM Web Science Conference (WebSci'), June 2017. (to appear)

05/10/2017

Helge Holzmann (holzmann@L3S.de)

www.HelgeHolzmann.de



Thank you!

- www.L3S.de
- www.TIB.eu
- www.ALEXANDRIA-project.eu
- www.fidmath.de

Try it on www.swMATH.org



Questions?



TIB LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



fidmath
Fachinformationsdienst Mathematik

