Software Citation, Landing Pages, and the swMATH Service

Wolfgang Dalitz, Helge Holzmann, Wolfram Sperber

FORCE2017, Berlin, 25-27 October 2017

ZUSE INSTITUTE BERLIN

zbMATH the first resource for mathematics

FIZ Karlsruhe Leibniz-Institut für Informationsinfrastruktur

MODAL Mathematical Optimization and Data Analysis Laboratories

Forschungszentrum L3S Research Center

Agenda

- Introduction (W. Sperber):
  - Why we have proposed this meeting?
  - Software Citations and the Software Citation Principles
  - Open problems
- swMATH: The publication based approach and swMATH landing pages (W. Dalitz)
- Live Demo: Software archiving and structurization (H. Holzmann)
- Discussion

Why this meeting?

- Increasing role of scientific (mathematical) software
  Today, 'mathematical modeling and simulation' is an ubiquitous method in natural sciences and engineering and mathematical software is the tool to realize it.

But

- discovering (mathematical) software is difficult,
  an established infrastructure (standards and information services) for scientific (mathematical) software is missing
- less credit for (mathematical) software development
- reproducility and evaluation of scientific results which were achieved by using scientific (mathematical) software is difficult.

Presentation and discussion of an alternative approach for maintaining mathematical software information

Software Citations in biology

Table 1 **Varieties of software mentions in publications, from *Howison & Bullard (2015)*.**

Mention type	Count (n = 286)	Percentage (%)
Cite to publication	105	37
Cite to user's manual	6	2
Cite to name or website	15	5
Instrument-like	53	19
URL in text	13	5
In-text name only	90	31
Not even name	4	1

and in mathematics (for the optimization software SCIP)

Mention Type	Count	Percentage
Cite to publication	84	46,67 %
Cite to user manual	2	1.11 %
Cite to website	19	10.56 %
Cite to other sources	3	1.67 %

The Software Citation Principles

- **Motivation:** citation of software container
- **SCP1. Importance:** Software is a legitimate and citable product of research […]
- **SCP2. Credit and Attribution:** Software citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the software [...]
- **SCP3. Unique identification:** A software citation should include a method for identiciation that is machine actionable, globally unique, interoperable, […]
- **SCP4. Persistence:** Unique identifiers and metadata describing the software and its disposition should persist […]
- **SCP5. Acessibility:** Software citations should facilitate access to the software itself and to its associated metadata, documentation, data, and other materials necessary for both humans and machines […]
- **SCP6. Specificity:** Software citations should facilitate identification of, and access to, the specific version of software that was used. [...]

Remarks

- The Software Citation Principles discuss not only citations but also metadata, persistence of software information, versions etc. They define a general framework for software information as a base for the development of
  - standards for software information (citations, metadata)
  - information services for persistent provision of information
- There is a broad spectrum of use cases for software with different requirements to information about software.

Some open problems for implementation (I)

- Citation standard
- Persistent identifiers (DOIs?)
- Persistent metadata for what? What are ‚software products'? software products are sets of different kind of objects:
  - source code, binary file, service, etc.
  - software documentation
  - software developers
  - software licenses
  - …
- Accessibility: persistent landing pages (metadata plus links) Definition of metadata schemes (CodeMeta)

Some open problems for implementation (II)

- Archiving: persistence of links?
- What is with specificity? A lot of software objects are version-specific

→ high-dimensional metadata schemes for software (CodeMeta)

- Maintenance: Who should (could) maintain software information?

→ diverse developments
for mathematical software especially the swMATH/Tempas concept

```
{
  "@context": {
      "type": "@type",
      "id": "@id",
      "schema":"http://schema.org/",
      "codemeta": "https://codemeta.github.io/terms/",
      "Organization": {"@id": "schema:Organization"},
      "Person": {"@id": "schema:Person"},
      "SoftwareSourceCode": {"@id": "schema:SoftwareSourceCode"},
      "SoftwareApplication": {"@id": "schema:SoftwareApplication"},
      "Text": {"@id": "schema:Text"},
      "URL": {"@id": "schema:URL"},
      "address": { "@id": "schema:address"},
      "affiliation": { "@id": "schema:affiliation"},
      "applicationCategory": { "@id": "schema:applicationCategory", "@type": "@id"},
      "applicationSubCategory": { "@id": "schema:applicationSubCategory", "@type": "@id"},
      "citation": { "@id": "schema:citation"},
      "codeRepository": { "@id": "schema:codeRepository", "@type": "@id"},
      "contributor": { "@id": "schema:contributor"},
      "copyrightHolder": { "@id": "schema:copyrightHolder"},
      "copyrightYear": { "@id": "schema:copyrightYear"},
      "creator": { "@id": "schema:creator"},
      "dateCreated": {"@id": "schema:dateCreated", "@type": "schema:Date" },
      "dateModified":  {"@id": "schema:dateModified", "@type": "schema:Date" },
      "datePublished":  {"@id": "schema:datePublished", "@type": "schema:Date" },
      "description": { "@id": "schema:description"},
      "downloadUrl": { "@id": "schema:downloadUrl", "@type": "@id"},
      "email": { "@id": "schema:email"},
      "editor": { "@id": "schema:editor"},
      "encoding": { "@id": "schema:encoding"},
      "familyName": { "@id": "schema:familyName"},
      "fileFormat": { "@id": "schema:fileFormat", "@type": "@id"},
      "fileSize": { "@id": "schema:fileSize"},
      "funder": { "@id": "schema:funder"},
      "givenName": { "@id": "schema:givenName"},
      "hasPart": { "@id": "schema:hasPart" },
      "identifier": { "@id": "schema:identifier", "@type": "@id"},
      "installUrl": { "@id": "schema:installUrl", "@type": "@id"},
      "isAccessibleForFree": { "@id": "schema:isAccessibleForFree"},
      "isPartOf":  { "@id": "schema:isPartOf"},
      "keywords": { "@id": "schema:keywords"},
      "license": { "@id": "schema:license", "@type": "@id"},
      "memoryRequirements": { "@id": "schema:memoryRequirements", "@type": "@id"},
      "name": { "@id": "schema:name"},
      "operatingSystem": { "@id": "schema:operatingSystem"},
      "permissions": { "@id": "schema:permissions"},
      "position": { "@id": "schema:position"},
      "processorRequirements": { "@id": "schema:processorRequirements"},
      "producer": { "@id": "schema:producer"},
      "programmingLanguage": { "@id": "schema:programmingLanguage"},
      "provider": { "@id": "schema:provider"},
      "publisher": { "@id": "schema:publisher"},
      "relatedLink": { "@id": "schema:relatedLink", "@type": "@id"},
      "releaseNotes": { "@id": "schema:releaseNotes", "@type": "@id"},
      "runtimePlatform": { "@id": "schema:runtimePlatform"},
      "sameAs": { "@id": "schema:sameAs", "@type": "@id"},
      "softwareHelp": { "@id": "schema:softwareHelp"},
      "softwareRequirements": { "@id": "schema:softwareRequirements", "@type": "@id"},
      "softwareVersion": { "@id": "schema:softwareVersion"},
      "sponsor": { "@id": "schema:sponsor"},
      "storageRequirements": { "@id": "schema:storageRequirements", "@type": "@id"},
      "supportingData": { "@id": "schema:supportingData"},
      "targetProduct": { "@id": "schema:targetProduct"},
      "url": { "@id": "schema:url", "@type": "@id"},
      "version": { "@id": "schema:version"},

      "author": { "@id": "schema:author", "@container": "@list" },

      "softwareSuggestions": { "@id": "codemeta:softwareSuggestions", "@type": "@id"},
      "contIntegration": { "@id": "codemeta:contIntegration", "@type": "@id"},
      "buildInstructions": { "@id": "codemeta:buildInstructions", "@type": "@id"},
      "developmentStatus": { "@id": "codemeta:developmentStatus", "@type": "@id"},
      "embargoDate": { "@id":"codemeta:embargoDate", "@type": "schema:Date" },
      "funding": { "@id": "codemeta:funding" },
      "readme": { "@id":"codemeta:readme", "@type": "@id" },
      "issueTracker": { "@id":"codemeta:issueTracker", "@type": "@id" },
      "referencePublication": { "@id": "codemeta:referencePublication", "@type": "@id"},
      "maintainer": { "@id": "codemeta:maintainer" }
  }
}
```

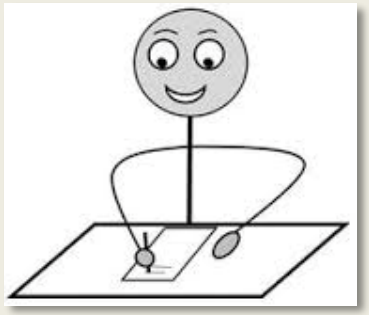
Agenda

- Introduction (W. Sperber):
  - Why we have proposed this meeting?
  - Software Citations and the Software Citation Principles
  - Open problems
- swMATH: The publication based approach and swMATH landing pages (W. Dalitz)
- Live Demo: Software archiving and structurization (H. Holzmann)
- Discussion

Motivation: Scientific Workflow

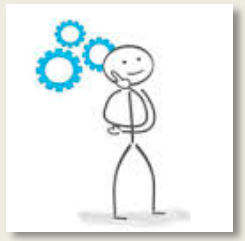
Scientist



Software Developer






Scientific Software in Mathematics

Scientific Software

- plays an important role within the scientific workflow
- produces new scientific results
- is (sometimes) the base of a proof (e.g. four color problem/Vierfarbenproblem)
- is a creative process

Scientific Software Developer

- receives little or no scientific recognition
- gets no or less academic reputation
- gets no credit points in his academic career

Brigding the Gap: www.swmath.org

- make important software visible (and accesible)
- don´t gather all what you can get
- take care of quality

- main idea: publication based approach
- cooperation with Zentralblatt MATH (zbMATH)

What is zbMATH?

zbMATH is a abstracting and reviewing service in pure and applied mathematics

- zbMATH database contains
  - 4 million bibliographic entries with reviews and abstracts
  - drawn from about 3,000 journals and serials and from
  - 180,000 books
- about 7,000 active reviewers from all over the world contribute reviews to zbMATH
- zbMATH covers all available published and peer-reviewed articles, books, conference proceedings as well as other publication formats
- all entries are classified accourding to the Mathematics Subject Classification Scheme (MSC2010)

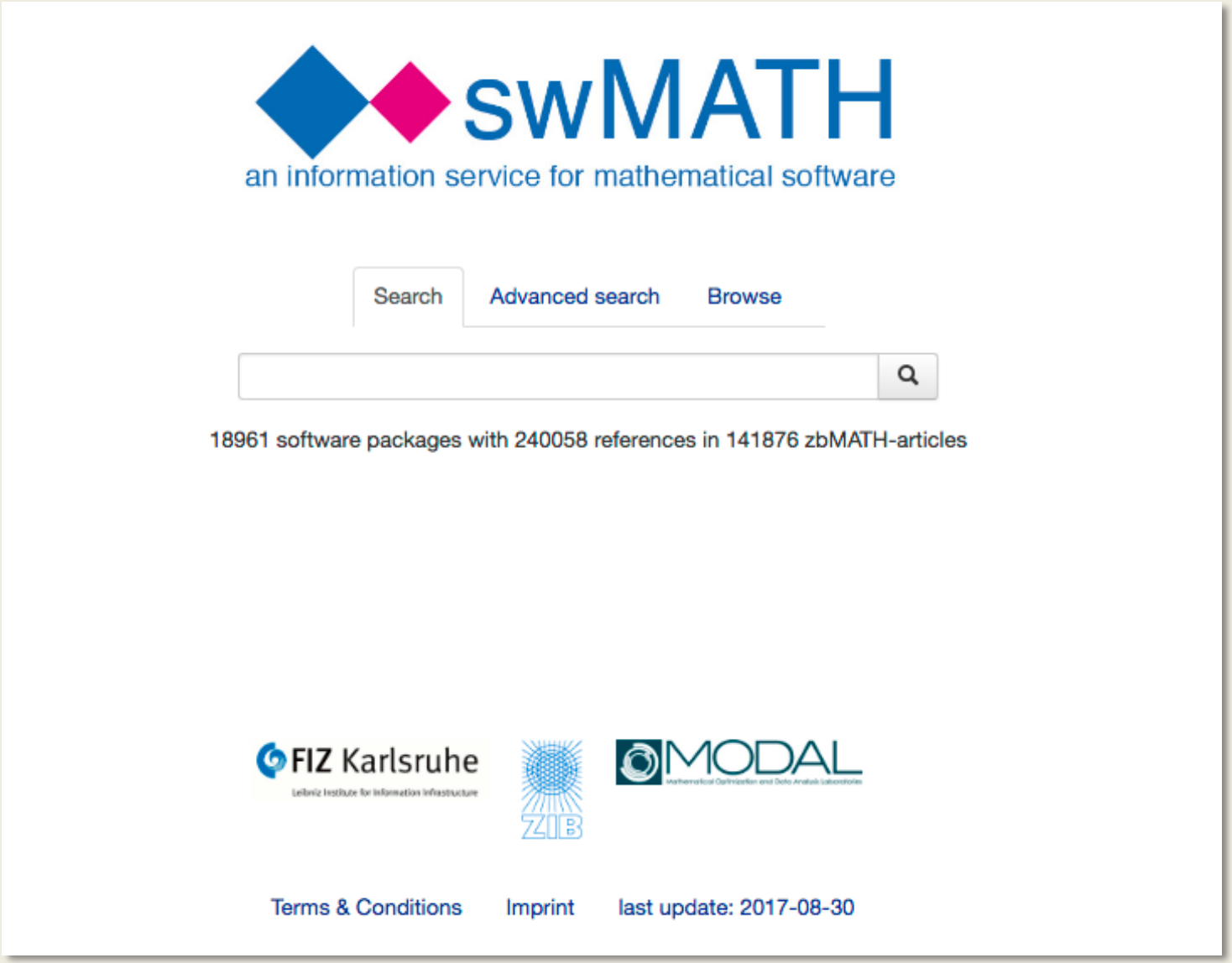
Main Idea: Publication Based Approach

- The intention is to offer a list of all publications that refer to a software recorded in swMATH
- In particular, all articles are given, which are included in Zentralblatt MATH (zbMATH)
- Articles that describes the background and technical details of a program, as well as those publications in which a piece of software is applied or used for research

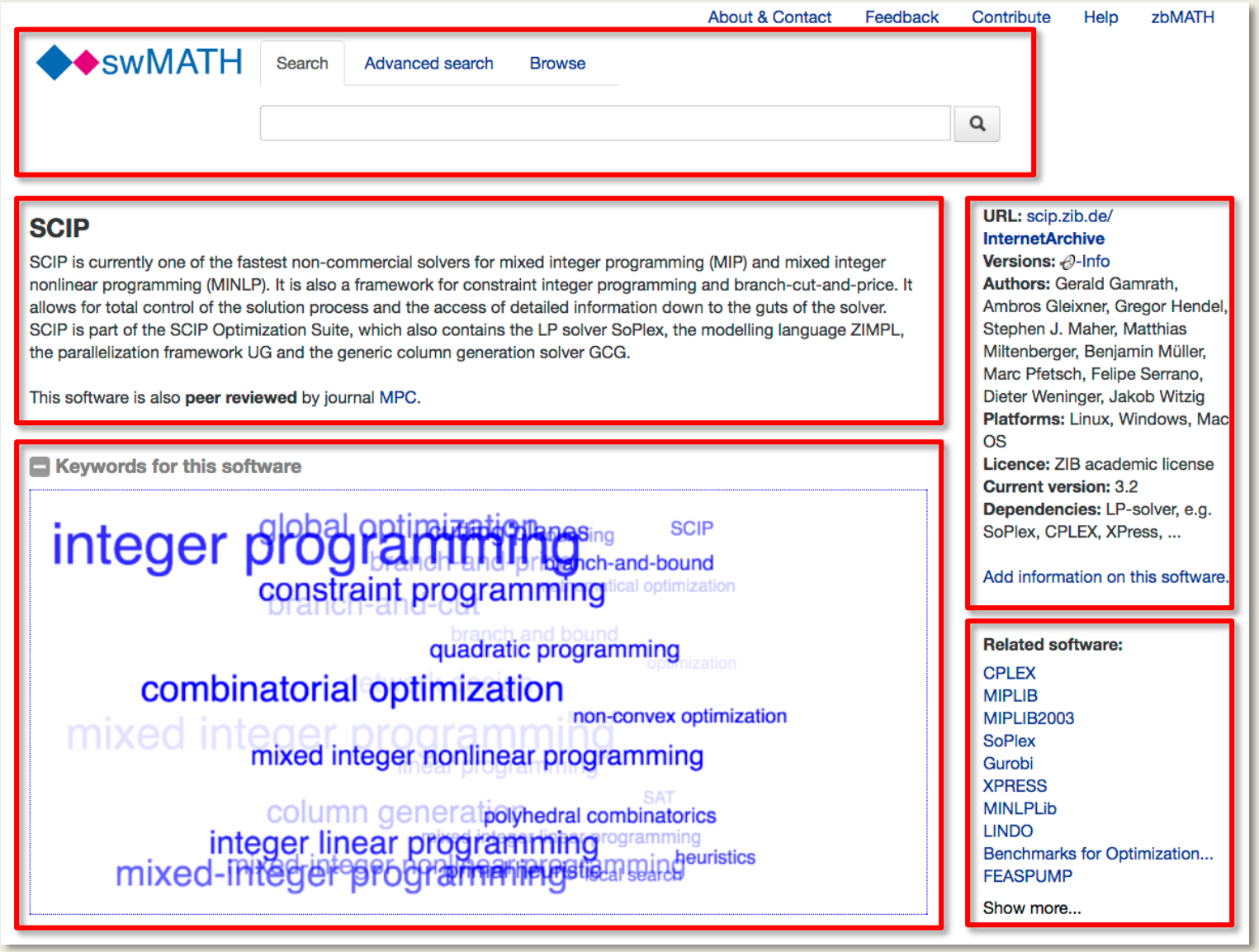
Shortly:

- Which articles refer/review software?
- Store the result into a database

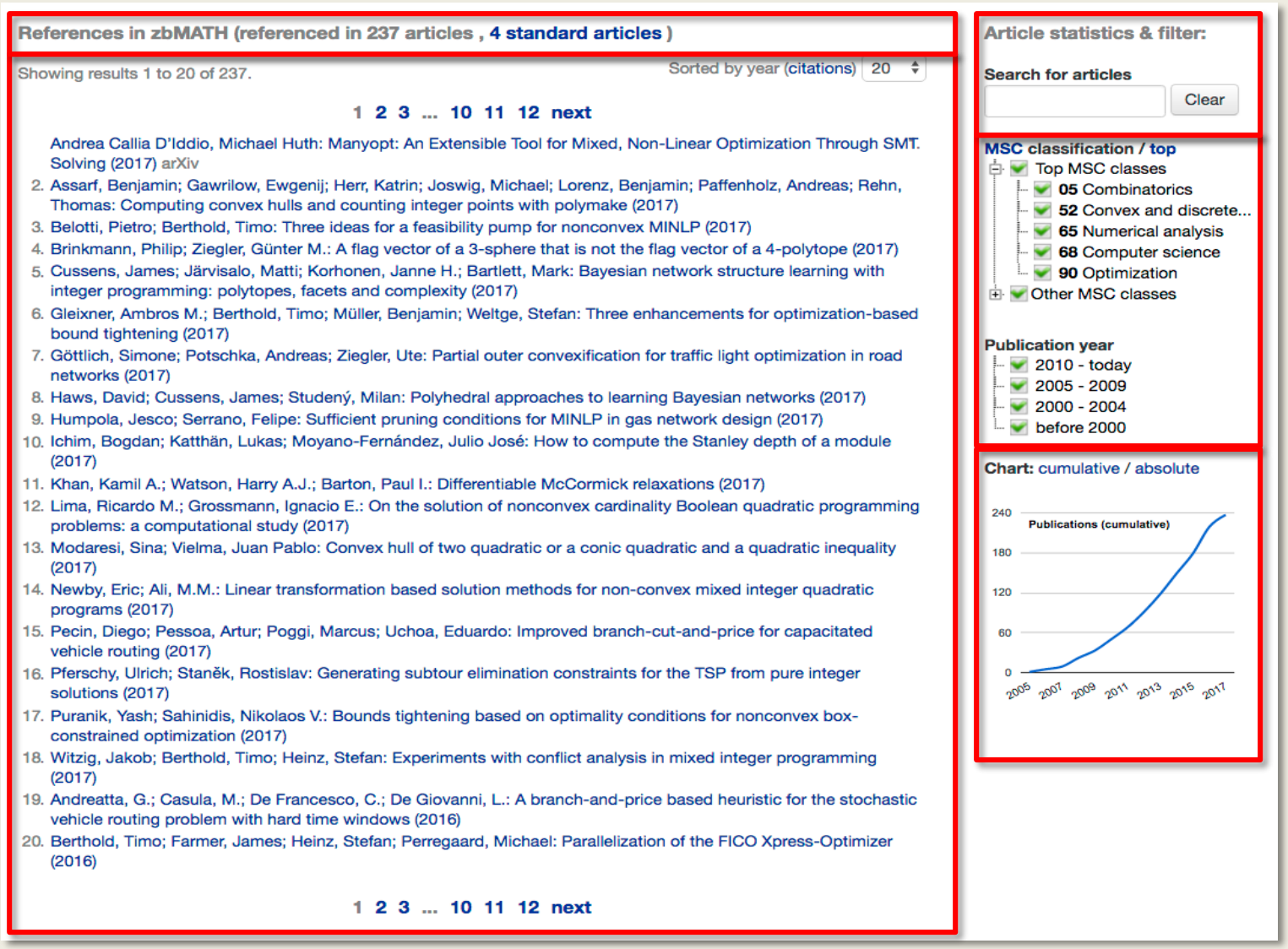
Quick Overview: www.swmath.org



Example


About & Contact
Feedback
Contribute
Help
zbMATH
swMATH
Search
Advanced search
Browse
SCIP
SCIP is currently one of the fastest non-commercial solvers for mixed integer programming (MIP) and mixed integer nonlinear programming (MINLP). It is also a framework for constraint integer programming and branch-cut-and-price. It allows for total control of the solution process and the access of detailed information down to the guts of the solver. SCIP is part of the SCIP Optimization Suite, which also contains the LP solver SoPlex, the modelling language ZIMPL, the parallelization framework UG and the generic column generation solver GCG.
This software is also peer reviewed by journal MPC.
URL: scip.zib.de/
InternetArchive
Versions: -Info
Authors: Gerald Gamrath, Ambros Gleixner, Gregor Hendel, Stephen J. Maher, Matthias Miltenberger, Benjamin Müller, Marc Pfetsch, Felipe Serrano, Dieter Weninger, Jakob Witzig
Platforms: Linux, Windows, Mac OS
Licence: ZIB academic license
Current version: 3.2
Dependencies: LP-solver, e.g. SoPlex, CPLEX, XPress, ...
Add information on this software.
Keywords for this software
integer programming
global optimization
SCIP
branch-and-bound
constraint programming
quadratic programming
combinatorial optimization
non-convex optimization
mixed integer programming
mixed integer nonlinear programming
column generation
polyhedral combinatorics
integer linear programming
heuristics
mixed-integer programming
Related software:
CPLEX
MIPLIB
MIPLIB2003
SoPlex
Gurobi
XPRESS
MINLPLib
LINDO
Benchmarks for Optimization...
FEASPUMP
Show more...

Example cont.



References in zbMATH (referenced in 237 articles , 4 standard articles)

Showing results 1 to 20 of 237. Sorted by year (citations) 20

1 2 3 ... 10 11 12 next

Andrea Callia D'Iddio, Michael Huth: Manyopt: An Extensible Tool for Mixed, Non-Linear Optimization Through SMT. Solving (2017) arXiv

2. Assarf, Benjamin; Gawrilow, Ewgenij; Herr, Katrin; Joswig, Michael; Lorenz, Benjamin; Paffenholz, Andreas; Rehn, Thomas: Computing convex hulls and counting integer points with polymake (2017)
3. Belotti, Pietro; Berthold, Timo: Three ideas for a feasibility pump for nonconvex MINLP (2017)
4. Brinkmann, Philip; Ziegler, Günter M.: A flag vector of a 3-sphere that is not the flag vector of a 4-polytope (2017)
5. Cussens, James; Järvisalo, Matti; Korhonen, Janne H.; Bartlett, Mark: Bayesian network structure learning with integer programming: polytopes, facets and complexity (2017)
6. Gleixner, Ambros M.; Berthold, Timo; Müller, Benjamin; Weltge, Stefan: Three enhancements for optimization-based bound tightening (2017)
7. Göttlich, Simone; Potschka, Andreas; Ziegler, Ute: Partial outer convexification for traffic light optimization in road networks (2017)
8. Haws, David; Cussens, James; Studený, Milan: Polyhedral approaches to learning Bayesian networks (2017)
9. Humpola, Jesco; Serrano, Felipe: Sufficient pruning conditions for MINLP in gas network design (2017)
10. Ichim, Bogdan; Katthän, Lukas; Moyano-Fernández, Julio José: How to compute the Stanley depth of a module (2017)
11. Khan, Kamil A.; Watson, Harry A.J.; Barton, Paul I.: Differentiable McCormick relaxations (2017)
12. Lima, Ricardo M.; Grossmann, Ignacio E.: On the solution of nonconvex cardinality Boolean quadratic programming problems: a computational study (2017)
13. Modaresi, Sina; Vielma, Juan Pablo: Convex hull of two quadratic or a conic quadratic and a quadratic inequality (2017)
14. Newby, Eric; Ali, M.M.: Linear transformation based solution methods for non-convex mixed integer quadratic programs (2017)
15. Pecin, Diego; Pessoa, Artur; Poggi, Marcus; Uchoa, Eduardo: Improved branch-cut-and-price for capacitated vehicle routing (2017)
16. Pferschy, Ulrich; Staněk, Rostislav: Generating subtour elimination constraints for the TSP from pure integer solutions (2017)
17. Puranik, Yash; Sahinidis, Nikolaos V.: Bounds tightening based on optimality conditions for nonconvex box-constrained optimization (2017)
18. Witzig, Jakob; Berthold, Timo; Heinz, Stefan: Experiments with conflict analysis in mixed integer programming (2017)
19. Andreatta, G.; Casula, M.; De Francesco, C.; De Giovanni, L.: A branch-and-price based heuristic for the stochastic vehicle routing problem with hard time windows (2016)
20. Berthold, Timo; Farmer, James; Heinz, Stefan; Perregaard, Michael: Parallelization of the FICO Xpress-Optimizer (2016)

1 2 3 ... 10 11 12 next

Article statistics & filter:

Search for articles

Clear

MSC classification / top

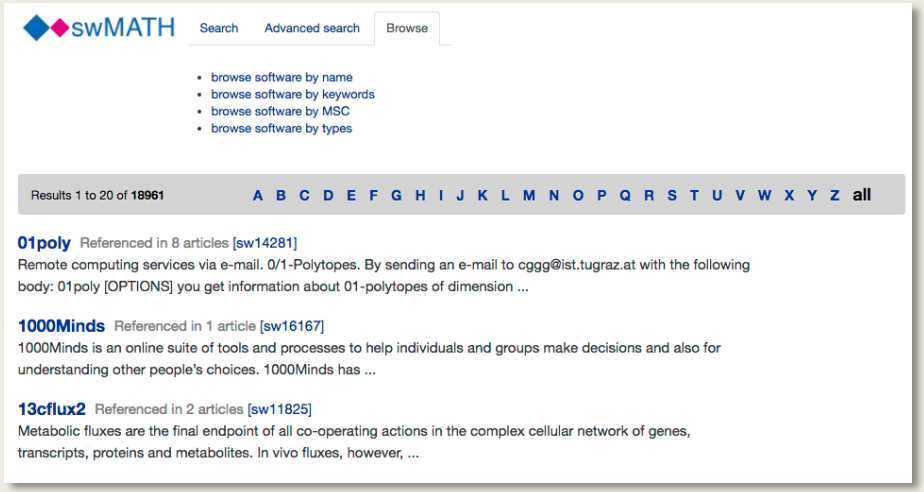
- Top MSC classes
  - **05** Combinatorics
  - **52** Convex and discrete...
  - **65** Numerical analysis
  - **68** Computer science
  - **90** Optimization
- Other MSC classes

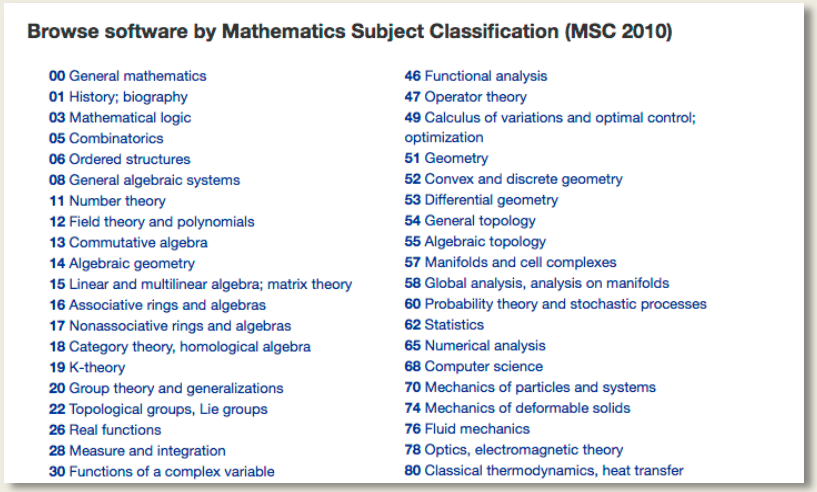
Publication year

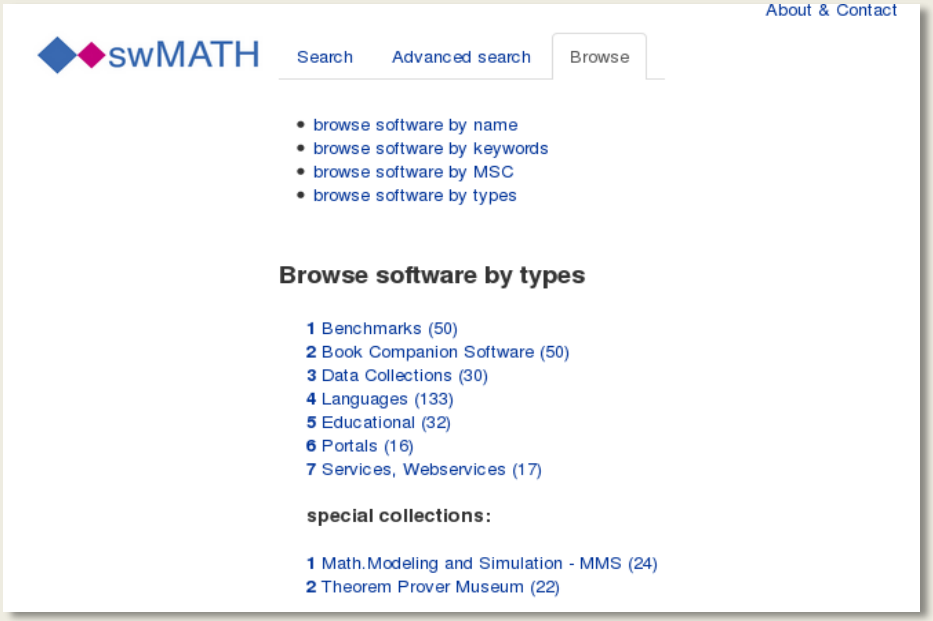
- 2010 - today
- 2005 - 2009
- 2000 - 2004
- before 2000

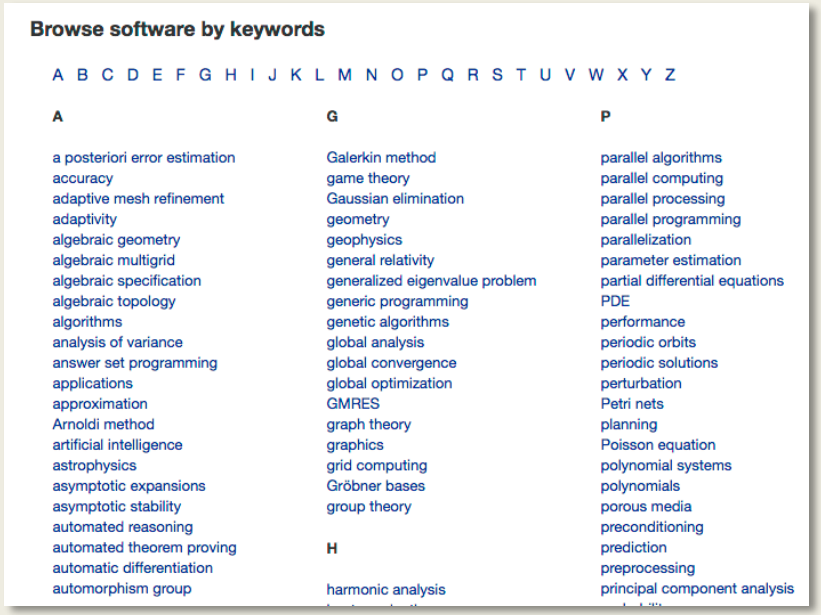
Chart: cumulative / absolute

Features: Browse by Name/MSC/Types/Keyword









Feature: Link to InternetArchive ...

SCIP

SCIP is currently one of the fastest non-commercial solvers for mixed integer programming (MIP) and mixed integer nonlinear programming (MINLP). It is also a framework for constraint integer programming and branch-cut-and-price. It allows for total control of the solution process and the access of detailed information down to the guts of the solver. SCIP is part of the SCIP Optimization Suite, which also contains the LP solver SoPlex, the modelling language ZIMPL, the parallelization framework UG and the generic column generation solver GCG.

This software is also **peer reviewed** by journal MPC.

Keywords for this software

column generation
linear programming
mixed integer programming
MINLP
network design
optimization
branch and bound
branch-and-cut
mathematical optimization
branch-and-price
SCIP

References in zbMATH (referenced in 237 articles , 4 standard articles)

Showing results 1 to 4 of 4. Sorted by: year / citations 20

1. Berthold, Timo; Heinz, Stefan; Vigerske, Stefan: Extending a CIP framework to solve MIQCPs (2012) archived SW
2. Berthold, Timo; Gleixner, Ambros M.; Heinz, Stefan; Vigerske, Stefan: Analyzing the computational impact of MIQCP solver components (2012) archived SW
3. Achterberg, Tobias: SCIP: solving constraint integer programs (2009) archived SW
4. Berthold, Timo: Heuristics of the branch-cut-and-price-framework SCIP (2008) archived SW

URL: scip.zib.de/
InternetArchive
Versions: -Info
Authors: Gerald Gamrath, Ambros Gleixner, Gregor Hendel, Stephen J. Maher, Matthias Miltenberger, Benjamin Müller, Marc Pfetsch, Felipe Serrano, Dieter Weninger, Jakob Witzig
Platforms: Linux, Windows, Mac OS
Licence: ZIB academic license
Current version: 3.2
Dependencies: LP-solver, e.g. SoPlex, CPLEX, XPress, ...

Add information on this software.

Related software:

CPLEX
MIPLIB
MIPLIB2003
SoPlex
Gurobi
XPRESS
MINLPLib
LINDO
Benchmarks for Optimization...
FEASPUMP

Show more...

Article statistics & filter:

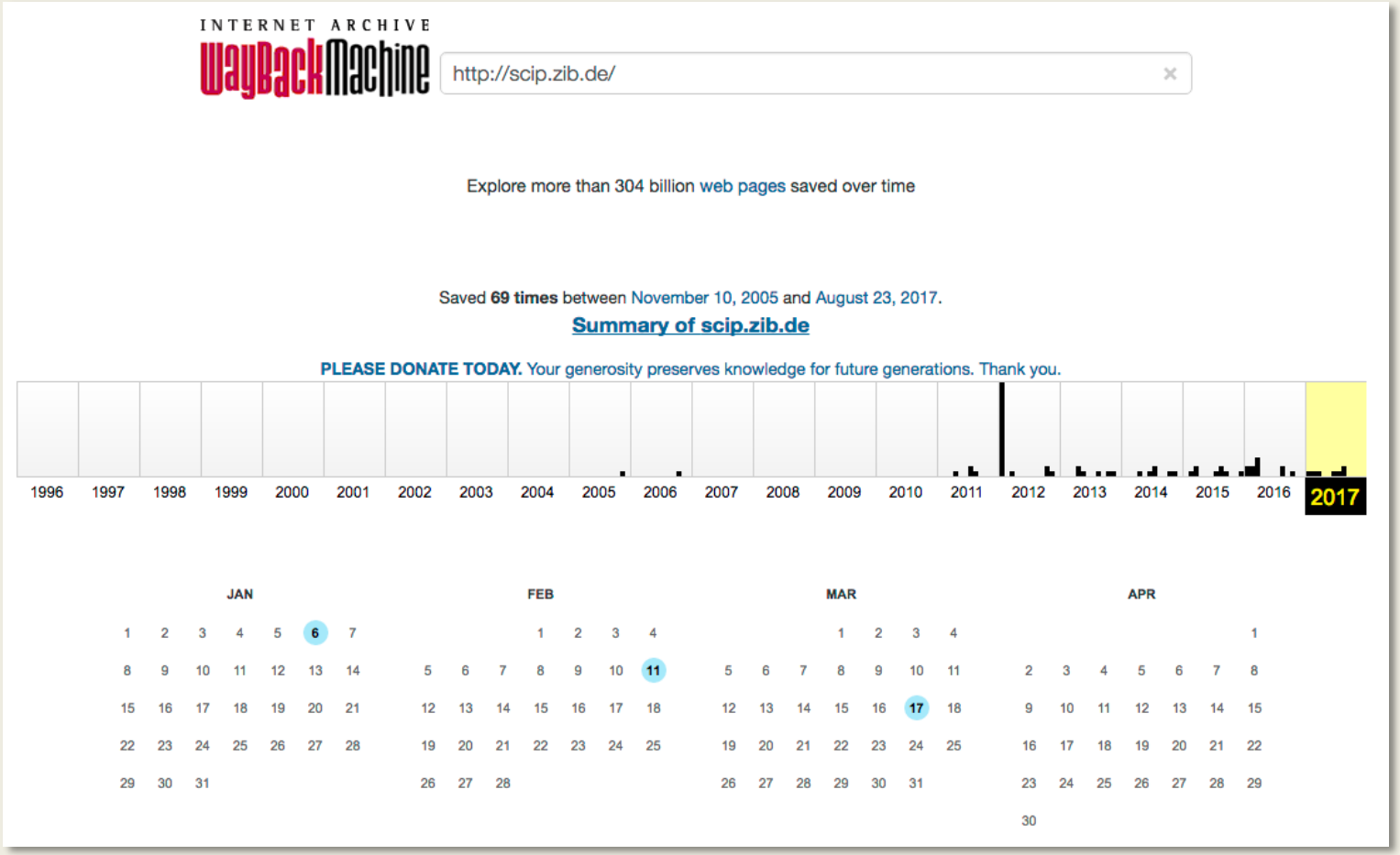
Search for articles

Clear

MSC classification / top

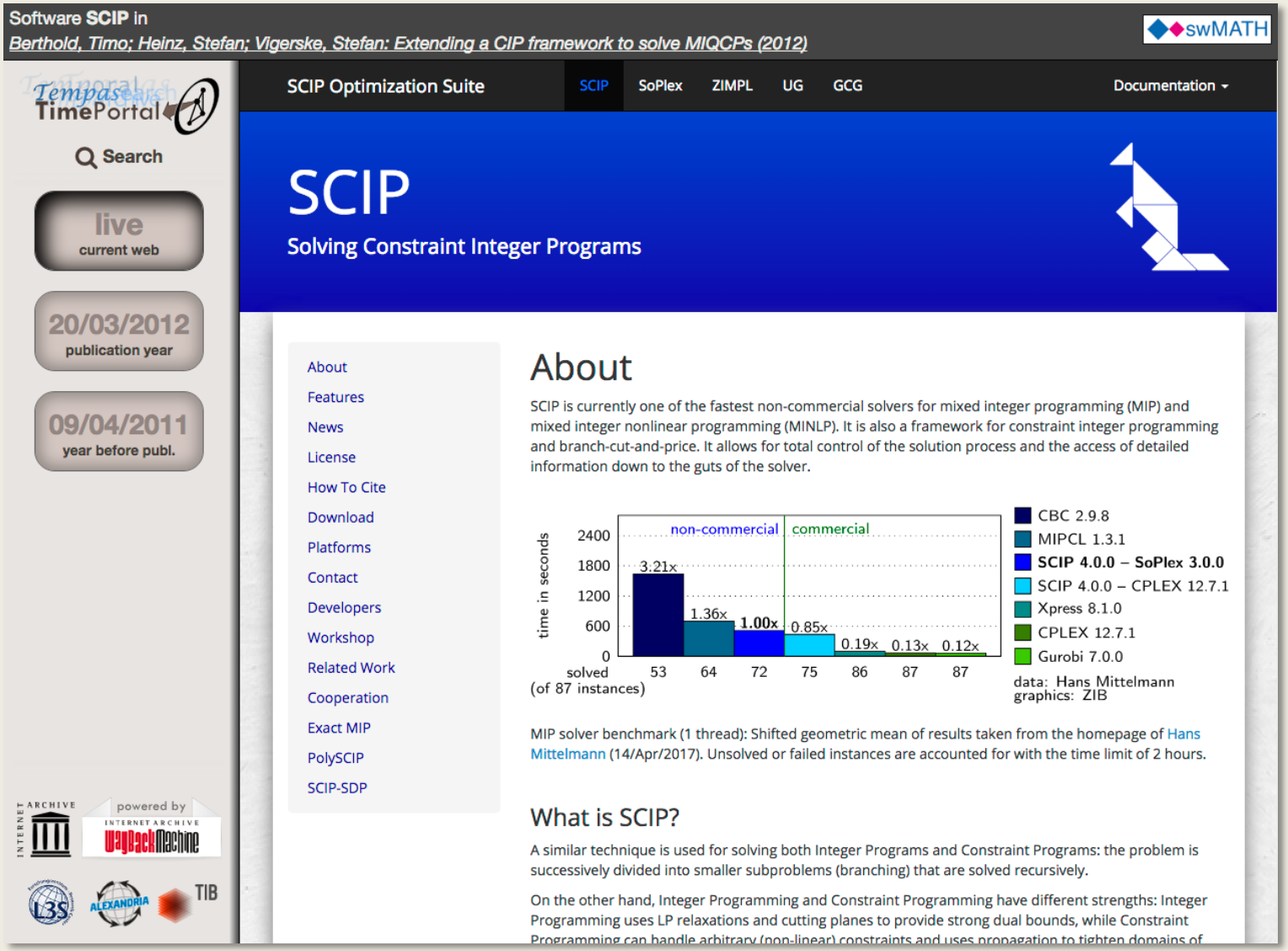
Top MSC classes
05 Combinatorics
52 Convex and discrete...

… leads to the WayBackMachine



Feature: Link to Version History ...

… leads to Tempas TimePortal (L3S)

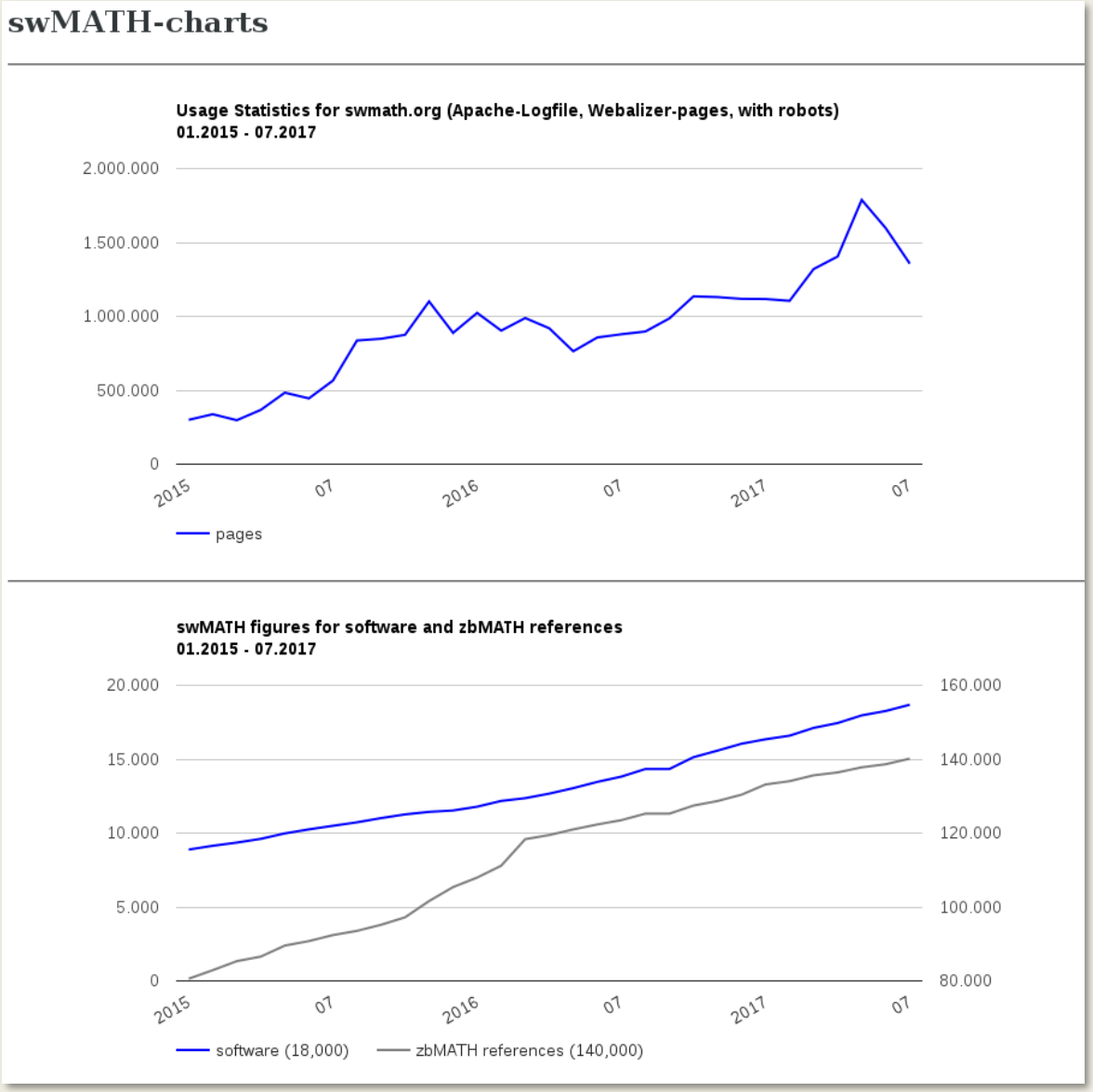


Some Statistics

- swMATH has been started in 2011, a joined project of Research Institute Oberwolfach (MFO) and FIZ Karlsruhe
- currently a project of the BMBF research campus MODAL with FIZ Karlsruhe/zbMATH and Zuse Institute Berlin (ZIB)

- ~19.000 Software Packages
- 240.000+ Software References in
- 140.000+ zbMATH (Scientific Articles)

Usage of swMATH



Some Results

- the numbers of citations listed in swMATH is an indicator of acceptance, spread and quality of the software

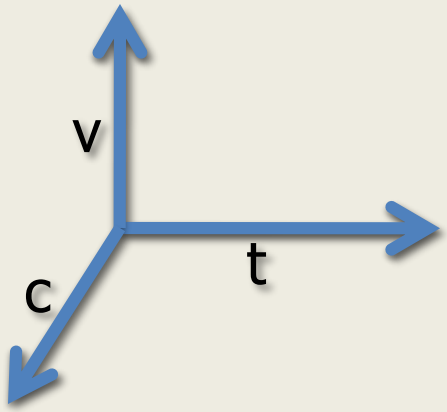
- Software developer
  - receives more scientific recognition
  - gets more academic reputation
  - gets more credit points in his academic career

- swMATH pages can be use as landing pages for software containers

Outlook

- Software Citation Standard
- Software Citation Index
- Access to all relevant Software
  - ArXiv,
  - Google Scholar,
  - Github
  - …
- Archiving
- Access by Time, Category and Version

```
software citation proposal:
@MISC{sw01091,
author = {Gerald Gamrath, Ambros Gleixner, Gregor Hendel, Stephen J. Maher, ... }
title = {SCIP},
note = {[SW] SOI:swmath.01091 (\url{http://www.swmath.org/software/01091})}
}
```



Agenda

- Introduction (W. Sperber):
  - Why we have proposed this meeting?
  - Software Citations and the Software Citation Principles
  - Open problems
- swMATH: The publication based approach and swMATH landing pages (W. Dalitz)
- Live Demo: Software archiving and structurization (H. Holzmann)
- Discussion

Live Demo

http://www.swmath.org

http://tempas.L3S.de/Micrawler

Micrawler

- **What is a micro archive?**
  - A micro archive is a snapshot of a fixed (evolving) set of URLs that are representative for some object or entity (at a given time). Hence, such an archive can be used to describe and / or derive information about its subject at the time of the crawl.
  - In case of software, it is a snapshot of resources on the Web that are related to that software, such as its website, documentation, source code or even binaries [1].
- Micrawler creates such micro archives based on a specification, which can be generated / provided by third-parties, like software repositories or catalogues (swMath)
  - Micrawler is modular and easy to extend / customize
  - Micro archives can be cited through unique identifiers
- Micrawler on GitHub: https://github.com/helgeho/Micrawler

[1] 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). Hannover, Germany. September 2016.

Agenda

- Introduction (W. Sperber):
  - Why we have proposed this meeting?
  - Software Citations and the Software Citation Principles
  - Open problems
- swMATH: The publication based approach and swMATH landing pages (W. Dalitz)
- Live Demo: Software archiving and structurization (H. Holzmann)
- Discussion

Some topics for discussion

- New persistent identifiers for software products (SOIs) and software versions (SVOIs) instead DOIs?
- Is CodeMeta metadata scheme suitable for the variety of use cases? Who should create and maintain metadata?
- Which metadata can be extracted in swMATH?
- Can be the swMATH/Tempas approach also used in other fields?
- Use of Web archives for software
  Do we need special archiving services for software?