LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

Preserving information on mathematical software via web archives

Helge Holzmann, Mila Runnwerth

San Diego, January 11th 2018
AMS Special Session on Mathematical Information in the Digital Age of Science, Joint Mathematical Meeting 2018

SUB

TIB

zbMATH
the first resource for mathematics

Forschungszentrum · Research Center
L3S

fidmath
Fachinformationsdienst Mathematik

swMATH
an information service for mathematical software

ALEXANDRIA

Why consider software? (1/2)

- Improving the status of software and its developers in the scientific publication process.
  - Especially in mathematics scientific software comprises sophisticated knowledge.

- The digital revolution provides the logistics to augment all kinds of data to traditional publications and make all data thoroughly accessible.

- Best practices for software in science:
  - **F**indable, **A**ccessible, **I**nteroperable, **R**eusable
  - **R**eplicability, **R**eproducibility, **R**eusability

- Checks & balances for the tools we use in scientific work.

Why consider software? (2/2)

What can possibly go wrong?

- Excel bug(s) – the most commonly used mathematical software in non-mathematical science.
- Knight Capital Group.
- Science alert: “A Bug in FMRI Software Could Invalidate 15 Years of Brain Research”

- Google: disasters caused by software / mathematical errors

AC1 fx =AA1*AB1

	AA	AB	AC
1	425	154.2	100000
2	850	77.1	100000
3	1700	38.55	100000
4	6375	10.28	100000
5	6425	10.2	100000
6	12750	5.14	100000
7	12850	5.1	100000
8	25500	2.57	100000
9	25700	2.55	100000

Goal: Make software visible in science

- If the software code is available on a version controlled repository, always cite the SHA value.

- If the software code is freely available in a less standardised environment
  - Download the software for yourself;
  - Follow the citing instructions on the webpage;
  - Ask the authors to move the software to a version controlled repository;
  - Create a web archive.

- If the software code is not available
  - Follow the citing instructions on the webpage;
  - Create a web archive.

Evaluation of the situation in maths

Q1: How well is software represented by its surrogate on the web?
Q2: Which information about software is available on the web?
Q3: How many websites of software are archived?
Q4: For how many of these can we recover referenced versions?

Data basis: swMATH – an information service for mathematical software

A1: High correlation between references in literature and in-links on the web. Good representation.
A2: Software pages are well structured with i. e. a documentation, download, update section.

Result 1: It is worthwhile considering software surrogates on the web.

A3: About half of the webpages have been archived.
A4: Only about 20% of these can be linked to a given referenced version in literature.

Result 2: There is work to be done.

For details see „Archiving Software Surrogates on the Web for Future Reference" by Holzmann, Sperber, Runnwerth (TPDL 2016).

Coupling swMATH and web archives

References in zbMATH (referenced in 1311 articles , 4 standard articles)

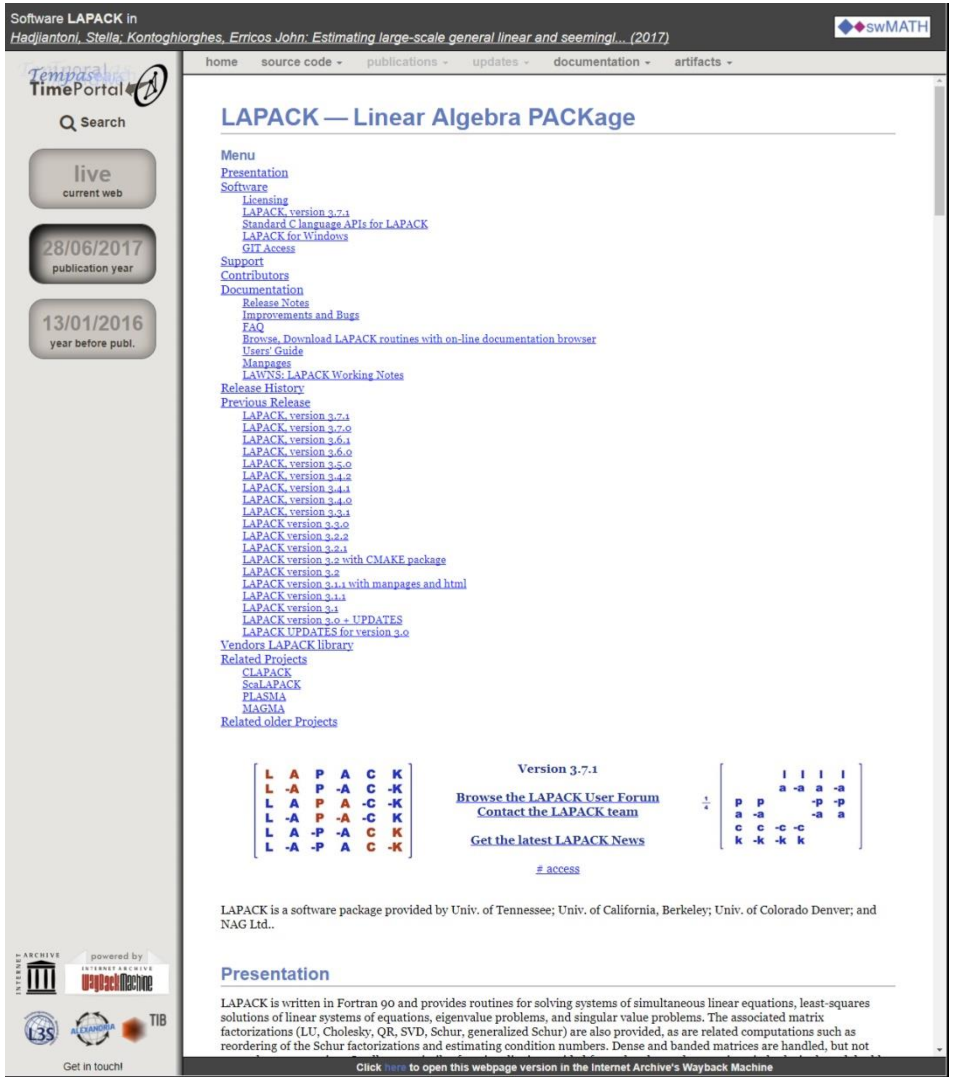
Showing results 1 to 20 of 1311.

Sorted by year (citations) 20

1 2 3 ... 64 65 66 next

1. Bylina, Beata: The block WZ factorization (2018) archived SW
2. González-Calderón, Alfredo; Vivas-Cruz, Luis X.; Herrera-Hernández, Erik César: Application of the Θ-method to a telegraphic model of fluid flow in a dual-porosity medium (2018) archived SW
3. Wang, Xuezhong; Che, Maolin; Wei, Yimin: Partial orthogonal rank-one decomposition of complex symmetric tensors based on the Takagi factorization (2018) archived SW
4. Amestoy, Patrick; Buttari, Alfredo; L'Excellent, Jean-Yves; Mary, Theo: On the complexity of the block low-rank multifrontal factorization (2017) archived SW
5. Bezanson, Jeff; Edelman, Alan; Karpinski, Stefan; Shah, Viral B.: Julia: a fresh approach to numerical computing (2017) archived SW
6. Birgin, E.G.; Martínez, J.M.: The use of quadratic regularization with a cubic descent condition for unconstrained optimization (2017) archived SW
7. Boiko, Andrey V.; Demyanko, Kirill V.; Nechepurenko, Yuri M.: On computing the location of laminar-turbulent transition in compressible boundary layers (2017) archived SW
8. Bosner, Nela; Karlsson, Lars: Parallel and heterogeneous m-Hessenberg-triangular-triangular reduction (2017) archived SW
9. Crivellaro, Alberto; Perotto, Simona; Zonca, Stefano: Reconstruction of 3D scattered data via radial basis functions by efficient and robust techniques (2017) archived SW
10. Diao, Huai-An: On condition numbers for least squares with quadric inequality constraint (2017) archived SW
11. Gianluca Frison, Dimitris Kouzoupis, Andrea Zanelli, Moritz Diehl: BLASFEO: Basic linear algebra subroutines for embedded optimization (2017) arXiv archived SW
12. Glinskiy, Boris; Kuchin, Nikolay; Kostin, Victor; Solovyev, Sergey: Parallel computations for solving 3D Helmholtz problem by using direct solver with low-rank approximation and HSS technique (2017) archived SW
13. Hadjiantoni, Stella; Kontoghiorghes, Erricos John: Estimating large-scale general linear and seemingly unrelated regressions models after deleting observations (2017) archived SW
14. Huang, Wen; Absil, P.-A.; Gallivan, K.A.: Intrinsic representation of tangent vectors and vector transports on matrix manifolds (2017) archived SW
15. Jandron, Michael A.; Ruffa, Anthony A.; Baglama, James: An asynchronous direct solver for banded linear systems (2017) archived SW
16. Maréchal, Alexandre; Périn, Michaël: Efficient elimination of redundancies in polyhedra by raytracing (2017) archived SW
17. Martinsson, Per-Gunnar; Quintana Ortí, Gregorio; Heavner, Nathan; van de Geijn, Robert: Householder QR factorization with randomization for column pivoting (HQRRP) (2017) archived SW
18. M. N. Gevorkyan, A. V. Demidova, A. V. Korolkova, D. S. Kulyabov, L. A. Sevastianov: The Stochastic Processes Generation in OpenModelica (2017) arXiv archived SW
19. Ostanin, Igor A.; Zorin, Denis N.; Oseledets, Ivan V.: Fast topological-shape optimization with boundary elements in two dimensions (2017) archived SW
20. Pouransari, Hadi; Coulier, Pieter; Darve, Eric: Fast hierarchical solvers for sparse matrices using extended sparsification and low-rank approximation (2017) archived SW

1 2 3 ... 64 65 66 next

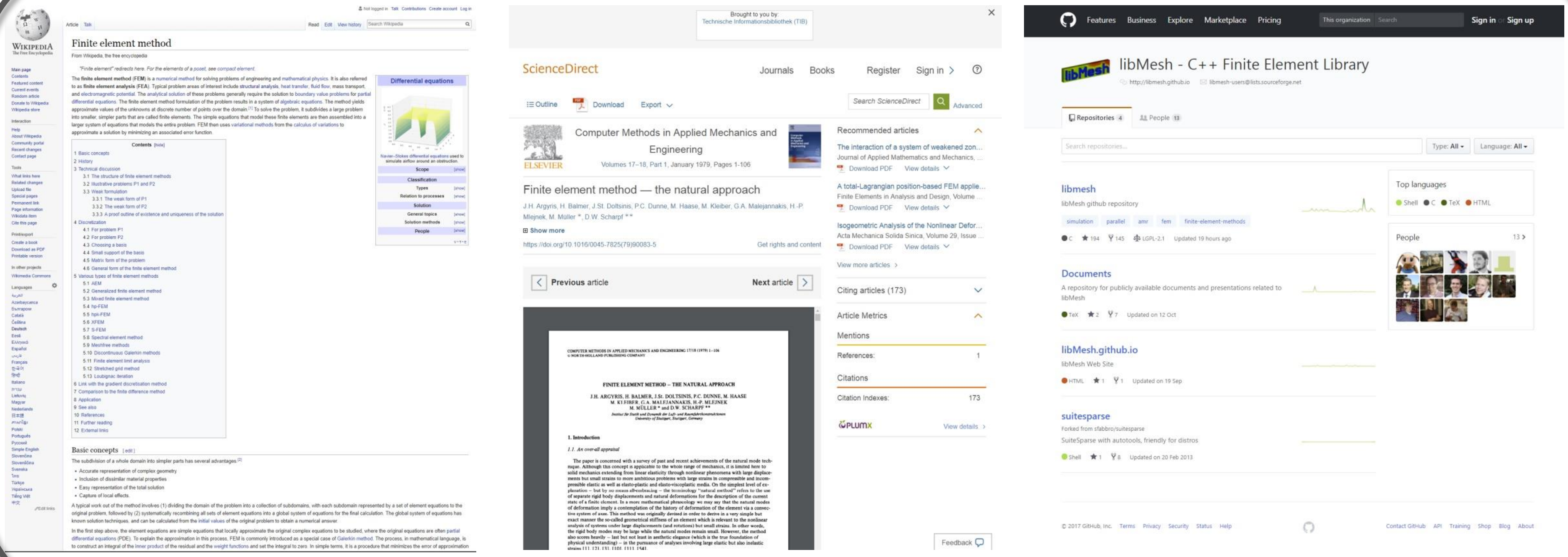


Live Demo

Make your own (micro) web archive for referencing software: in theory

MyFiniteElementMethod

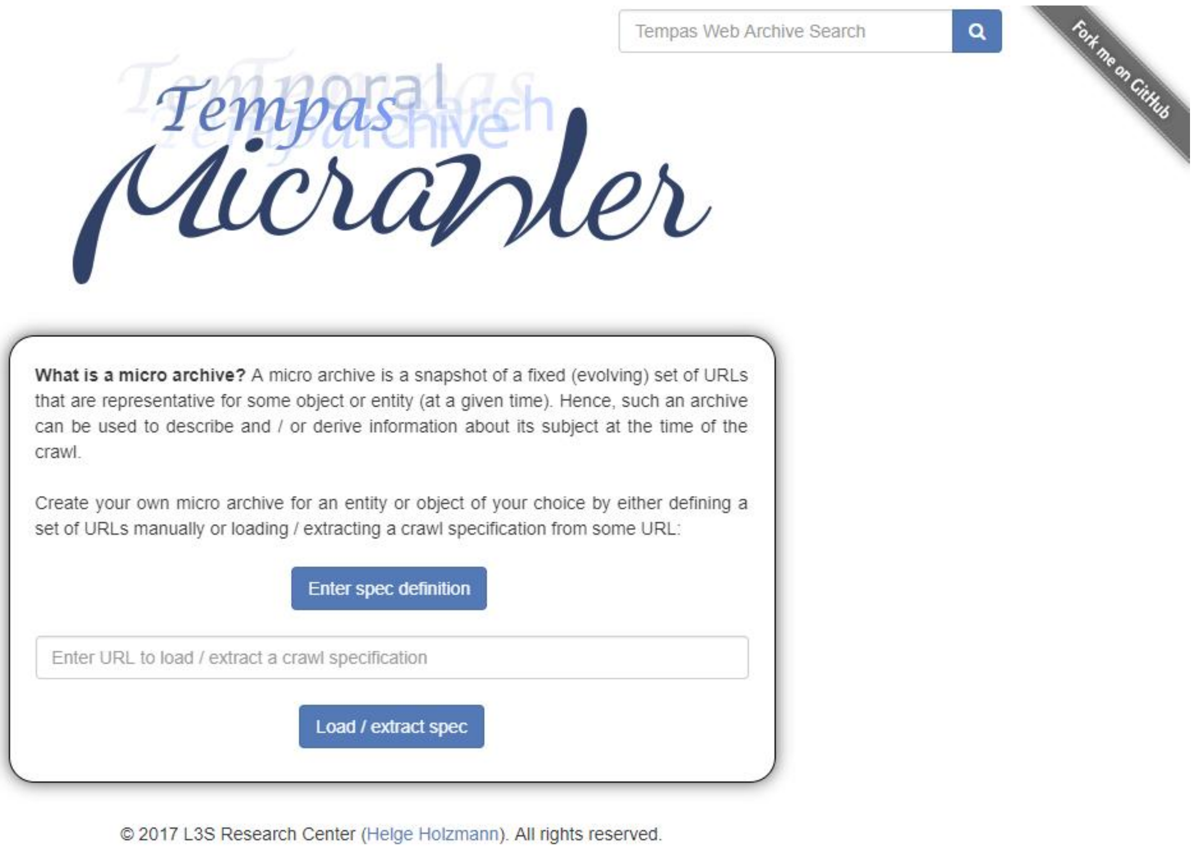
Entity



Micro archive

Make your own (micro) web archive for referencing software: in practice

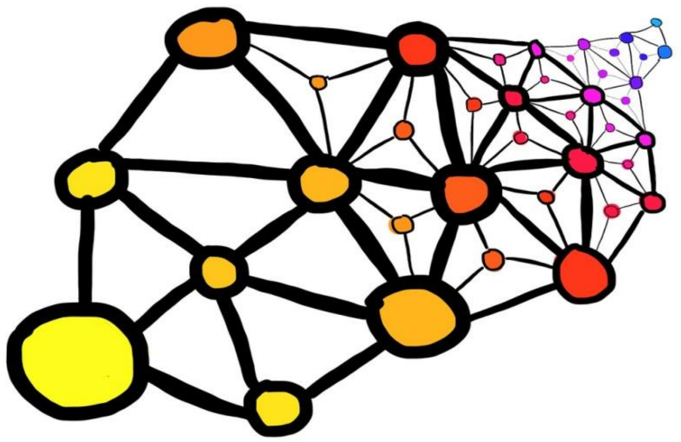
http://tempas.l3s.de/micrawler/



Live Demo

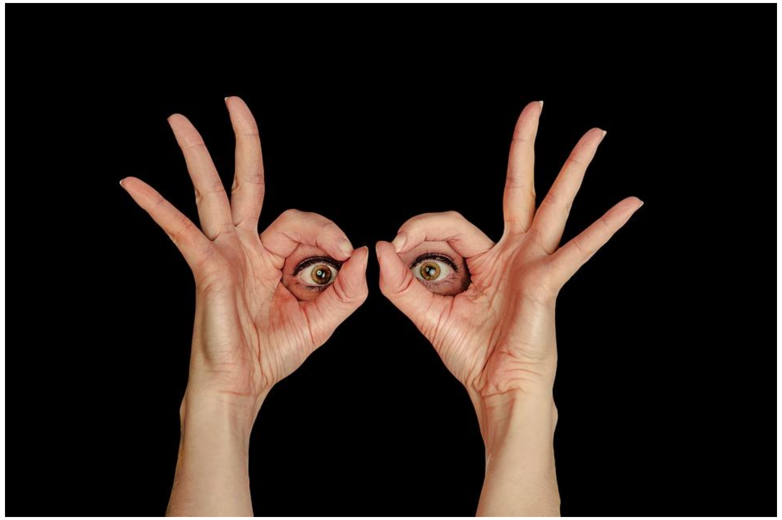
What you gain with Micrawler

- The information from the web is archived precisely as you processed it.
- Archive the website when you use its content, not when you reference / cite it.
- All your information from the web is verifiable.
- You create an archive according to your semantic specification.
- Your data is retraceable through time.



Outlook

- Enable referencing micro archives by handle (i. e. DOI)
  - Pointing to the micro archive from a publication
- Mining meta data from those archives automatically
  - Detect versions, features, etc.
- Consider scientifically relevant web content for web archives



LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

To be continued…

Contact:
Helge Holzmann
Mila Runnwerth