Reference Management: BibTeX and Beyond

1 December 2005
Barbara Rauch
Jochen Leidner (CiteULike & LSV BibTeX Collector slides)
IGK Colloquium

Spoken Language Systems Group (LSV)
Saarland University

{barbara.rauch},{jochen.leidner}@lsv.uni-saarland.de

Agenda

- Reference management in general
- Review of BibTeX
- Three less well-known reference management systems
 - Note: there are many others, here focus on OSindependent ones

Reference Management in General

What is Reference Management?

- Collecting and organising your references (citation data + maybe more).
- RM software should support at least retrieval and automatic bibliography creation,
- but maybe also data input (editors but also from online sources such as Citeseer), sharing, consistency checks, filing of off-prints or local copies, ...

Why Reference Management?

- Reuse: save time; consistency; fewer errors
- Retrieval: find quickly information on documents you have read in the past
- Automatic bibliographies: can benefit from standard style sheets, easy to change formatting and sorting

Why shared RM?

- Potential benefits of **sharing** references with others:
 - avoid data entry
 - recommendations
 - local systems can save disk space by sharing electronic copies
- But: privacy issues

Possible Main Requirements

- 1) Underlying citation data format:
 - · common exchange format? Certain flexibility to accommodate local choices? Support import and export from/to other common formats (bibtex, XML, MODS?, ...)
- 2) Automatic bibliography creation
 - · LaTeX compatibility? [or your favourite editor]
- 3) Retrieval: powerful search etc.
- 4) Support for sharing:
 - · distinction of public and private data fields; possibility to retrieve the references you added, restrict search to them etc.

Requirements: Some Practical Issues

- OS-independence?
- Easy to use
- Low maintenance for administrator if needed, little work needed for set-up
- Stable & working correctly
- Support: documentation, mailing lists etc.
- Must be free?

Possible Solutions

- Create your own customised database
 - Completely flexible, but more work, and no standard
- If you work with LaTeX: BibTeX
- Existing RM systems:
 - Single-user systems (e.g. JabRef)
 - Multi-user systems:
 - Local systems (e.g. **RefDB**)
 - Web-based systems (e.g. CiteULike)

BibTeX

Bibliography Creation in LaTeX

Two ways of creating bibliographies in LaTeX:

- 1) Manually: thebibliography environment
- 2) Automatically: with BibTeX

Manual process:

- create thebibliography environment with list of references in the document, each with unique ID
- to create a citation, use \cite{ID}
- run latex as usual

```
%%%%%%%%%%% mydoc.tex %%%%%%%%
\documentclass{article}
\begin{document}
```

Manual Example

Refer to Latex guides for more information (e.g. \cite{MG}).

\begin{thebibliography}

\bibitem[Goo94]{MG} Goossens, M., Mittelbach, F., Samarin, \emph {A LaTeX Companion}, Addison-Wesley, Reading, MA, 1994.

\bibitem[Kop99]{HK} Kopka, H., Daly P.W., \emph{A Guide to LaTeX}, Addison-Wesley, Reading, MA, 1999. bibitem{Pan} Pan, D., `A Tutorial on MPEG/Audio Compression," \emph{IEEE Multimedia}, Vol.2, pp.60-74, Summer 1998.

\end{thebibliography}
\end{document}

latex

Refer to Latex guides for more information (e.g. [Goo94]).

References

[Goo94] Goossens, M., Mittelbach, F., Samarin, A LaTeX Companion, Addison-Wesley, Reading, MA, 1994.

[Kop99] Kopka, H., Daly P.W., A Guide to LaTeX, Addison-Wesley, Reading, MA, 1999. bibitemPan Pan, D., "A Tutorial on MPEG/Audio Compression," IEEE Multimedia, Vol.2, pp.60-74, Summer 1998.

Automatic Process with BibTeX

- create *mydatabase*.bib text file containing all your reference entries, again each with unique id
- in your LaTeX document
 - as before, use \cite command with ID to cite
 - specify style to use, e.g. \bibliographystyle{plain}
 - insert \bibliography{mydatabase} where you want the bibliography
- then run latex, bibtex, latex. This will extract only the cited references from your .bib file, and format them according to the specified style.

BibTeX entries

• Each entry in BIBTeX follows a standard format. The format changes slightly based on the entry type, but the syntax remains the same. The syntax is:

```
@entry_type {uniqueID,
  fieldname= "value"
  fieldname= "value" }
```

• Some fields are required, others optional.

BibTeX Entry Types

@article

An article from a journal or magazine.

Required fields: author, title, journal, year.

Optional fields: volume, number, pages, month, note.

@book

A book with an explicit publisher.

Required fields: author or editor, title, publisher, year. Optional fields: volume or number, series, address, edition, month, note.

Others: @booklet, @conference, @inbook, @incollection, @inproceedings, @manual, @mastersthesis, @misc, @phdthesis, @proceedings, @techreport, @unpublished.

BibTeX style files

- .bst file
- determines how bibliography entries will appear at the output, such as if they are sorted or not, or how they are labeled etc.
- You can write them yourself
- but many standards exist, part of Latex standard package or provided by publishers

\bibliographystyle{alpha}
\bibliography{mybib}

\end{document}

BibTeX Example

latex & bibtex

%%% alpha.bst %%%% %%% Style File %%%

Refer to Latex guides for more information (e.g. [GMS94]).

References

[GMS94] M. Goossens, F. Mittelbach, and A. Samarin. A LaTeX Companion. Addison-Wesley, Reading, MA, 1994.

Why go beyond BibTeX?

- You might not use LaTeX
- Even if you do, you might prefer a neutral database without LaTeX markup
- No input editor as such, only format + compilation (though emacs mode)
- There are RM systems which act as (G)UI to bibtex databases, and others with a different data format but can export in bibtex format

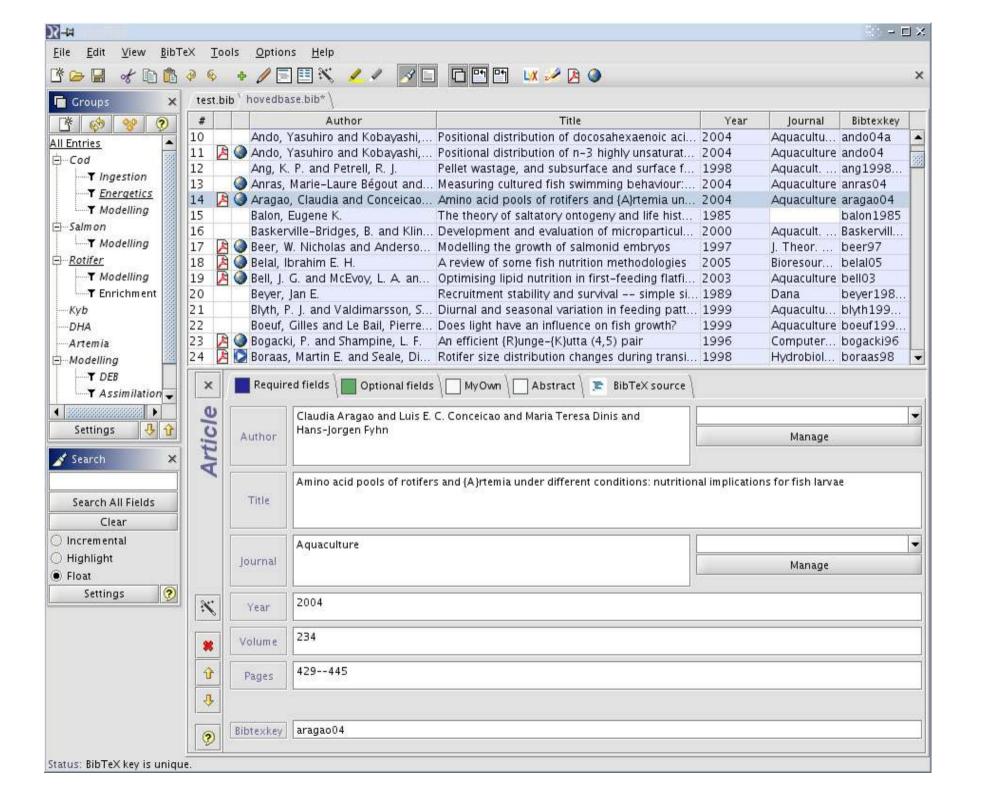
Three Reference Management Systems

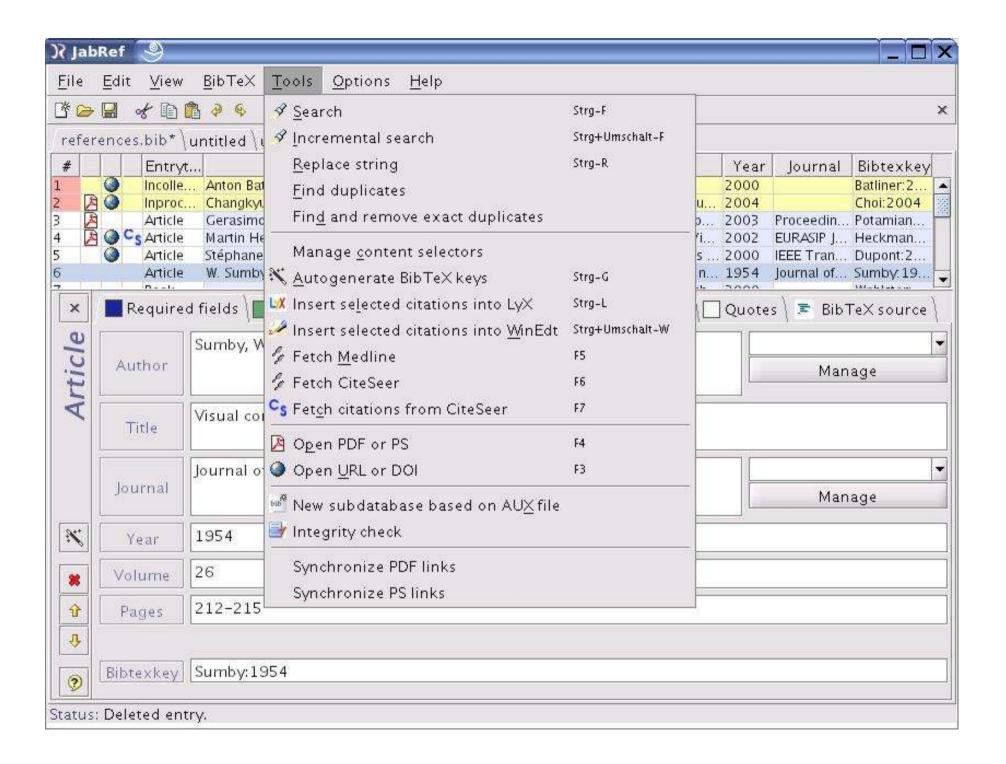
Three RM Systems

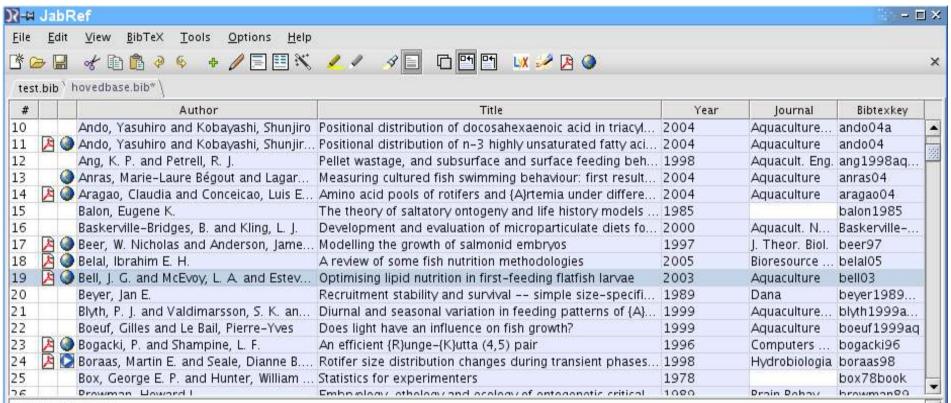
- (1) JabRef: primarily designed for single users, interface to BibTeX DB
- (2) CiteULike: an open reference management system on the web, can import/export .bib
- (3) RefDB: multi-user system with underlying SQL DB in RIS format, can import/export .bib

(1) JabRef

- Open-source Java GUI to BibTeX entries
 - Since Oct '03, dev. status: 5 Production/Stable
- Some management of keywords, authors, journal names etc.
- Import/export of other formats:
 - In: BibTeXML, CSA, Refer/Endnote, ISI Web of Science, SilverPlatter, Medline/Pubmed (xml), Scifinder, OVID, INSPEC, Biblioscape, Sixpack, JStor and RIS.
 - Out: HTML, Docbook, BibTeXML, MODS, RTF, Refer/Endnote and OpenOffice.org.
 - Automatic download of fields from Citeseer (and Medline)
- Customisable interface







Article (bell03)

Status:

Bell, J.G.; McEvoy, L.A.; Estevez, A.; Shields, R.J. & Sargent, J.R. Optimising lipid nutrition in first-feeding flatfish larvae

Aquaculture, 2003, 227, 211-220

Abstract: Although global production of flatfish has increased in recent years, both in terms of numbers of fish and diversification into new species, problems still remain with low survival rates and difficulties with metamorphosis. This short review highlights some advances made in optimising lipid nutrition in an attempt to overcome some of these problems. Copepod nauplii are the best live prey for first-feeding flatfish larvae. Rotifers can provide a useful method of essential fatty acid delivery but Artemia are poorer in this regard, especially at first feeding. Copepods are nutritionally beneficial due to their naturally high levels of the essential highly unsaturated fatty acids (HUFA), 20:5n-3 (eicosapentaenoic acid; EPA) and 22:6n-3 (docosahexaenoic acid; DHA), which are predominantly in the form of phospholipids. Rotifers can be enriched with fish oil emulsions to provide compositions similar to copepods, while enriched Artemia are difficult to enrich with high levels of

(2) CiteULike

- Based on the same principle as flickr.com
- community talks according to a "folksonomy"
- BibTeX/Endnote import/export
- Manage multiple sets of references
- Can find users with similar interests

- Main feature:NO NEED TO TYPE
- CiteULike "Bookmarklet" can automatically import entries from amazon.com, SpringerLink, ...
- Fixed list of supported sites, other references are added manually/imported
- Free, Web-based (U Manchester is hosting it)

Article title



watch

Search

Navigation

- Home
- News and Status
- Discussion list

Your Library

- View articles
 - Recent
 - To read
- Authors and tags
- Search
- Post article
- Manually post article
- Export to BibTeX/EndNote
- Change password

Your Watchlist

- View watchlist
- Manage watchlist

Iournals

leidner's library ordered by **importance**

Recent papers added to leidner's library. You can also search leidner's library or view all of leidner's authors and tags.

 Ontological Semantics (Language, Speech, and Communication) by Nirenburg S, Raskin V posted to ont by leidner as *** and 1 other ... on 2005-09-16 13:43:00



Information Theory, Inference & Learning Algorithms by Mackay DIC posted to mI by leidner as *** and 5 others ... on 2005-11-01 16:57:35



Readings in Information Retrieval

by Jones K5, Willett P posted to ir by leidner as *** and 1 other ... on 2005-10-11 11:40:46



Elements of Information Theory

by Cover TM, Thomas JA posted to ml by leidner as *** and 1 other ... on 2005-10-13 01:57:10



Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW (With CD-ROM)

by Belew RK posted to ir by leidner as ★★★★ on 2005-07-25 19:58:13





leidner's tags

All tags in leidner's library

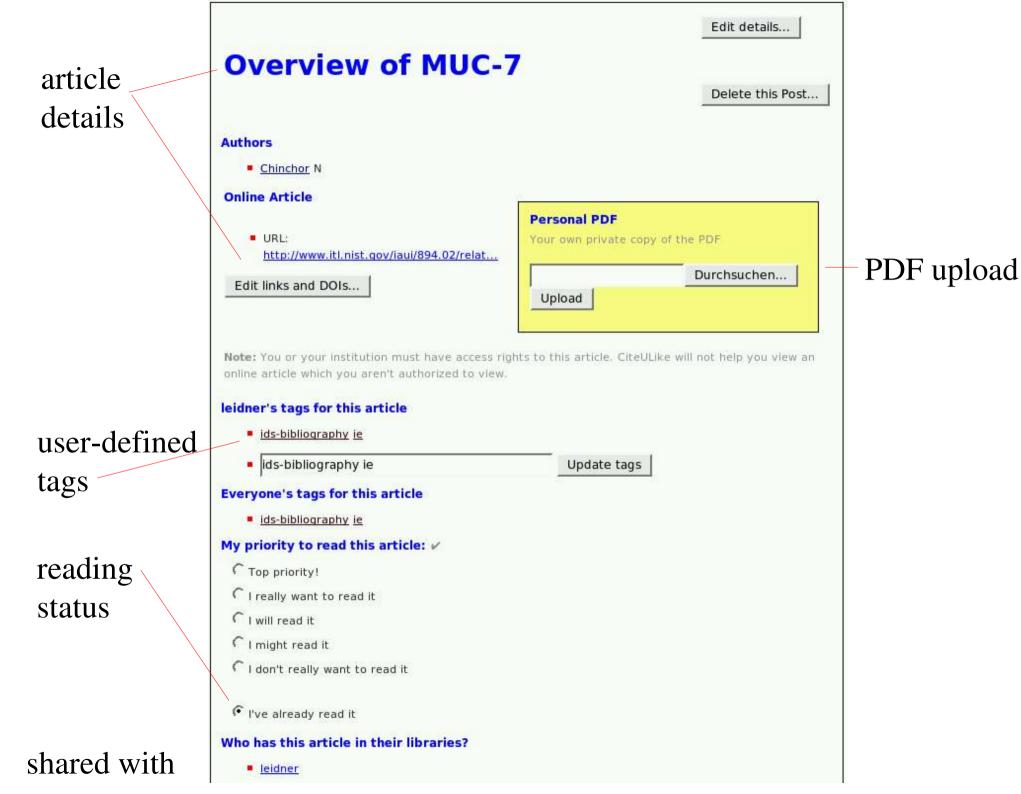
Filter:

ai asr bioinf c compilers cs epist eval geb geoinf gir gis grid hci

ids-bibliography ie ir lbs leidner ml nlp

nlu ont prg ga semweb tdt topres tr

"tag soup"



(3) RefDB

- Multi-user reference management system
 - Open Source, status: 4 Beta (initial release May 2001)
 - Operating Systems: POSIX OSes; OSX;/Darwin Windows+Cygwin
 - Implemented in C, some Perl and shell scripts
 - Client/server architecture
 - Good documentation and other support
 - All character encodings supported by your platform
- Underlying format: RIS
 - common data exchange format, originally defined by Research
 Information System, creators of Reference Manager (c)
 - can be generated and imported by all major reference databases on Windows

RefDB Functionality

- Underlying SQL database
 - powerful search (booleans, reg expressions, on any combination of any fields)
 - tables for journals, authors etc.
 - two types of entries: references and extended notes
- Automatic bibliography creation
- Explicitly supports sharing:
 - user accounts, concurrent access
 - personal vs. full reference lists
 - user-specific vs. generic data
 - personal vs. public extended notes

Personal (User-associated) Information

- Can restrict search to those references or extended notes associated with your username.
- Can also add references from others' personal lists.
- Personal part of reference information (reprint status, availability, notes) is saved for each user individually.

Import & Export in RefDB

• In:

RIS, RISX, Medline (tagged and XML), BibTeX, MARC, and DocBook. More input filters can be created.

• Out:

RIS, RISX, plain text, HTML, BibTeX reference list, DocBook bibliography element (SGML or XML), TEI listBibl element (XML). API provided for creation of more formats.

The RIS Format

- A tagged text file format.
- Fixed set (standard!), but some room for personal use.
- Following kinds of tags exist:
 - Title and Reference Type
 - Authors
 - Year and Free Text Field (personal notes & abstract)
 - Keywords and Reprint Status
 - Periodical Tags
 - Periodical and Publisher Tags
 - Misc. Tags

An Example of a RIS Record

TY - JOUR

A1 - Baldwin, S.A.

A1 - Fugaccia,I.

A1 - Brown, D.R.

A1 - Brown, L.V.

A1 - Scheff, S.W.

T1 - Blood-brain barrier breach

following

cortical contusion in the rat

JO - J.Neurosurg.

Y1 - 1996

VL - 85

SP - 476

EP - 481

RP - Not In File

KW - cortical contusion

KW - blood-brain barrier

KW - horseradish peroxidase

KW - head trauma

KW - hippocampus

KW - rat

N2 - Adult Fisher 344 rats were

subjected to a unilateral [...]

ER -

Extended notes

- 'Extended' as opposed to normal notes fields (N1, AB) in RIS
- An xnote is a text item which can be linked to zero or more references, authors, keywords, journal names or journal abbreviations.
- It can optionally have a title, keywords, and URLs associated with it.
- Can be private or public.

An example

• A private note linked to a reference and an author:

• The content can be plain text or contain markup.

Bibliography Creation

- Different layout styles exist, more can be created.
- Citation and reference styles are defined in XML.
- For XML/SGML:
 - RefDB can format the document for you, like bibtex/latex does.
- Remember for Latex, we'd export reference list in bibtex format.

Existing RefDB interfaces

- Command line
- Web interface:
 very basic, at present incomplete PHP interfaces, work in progress.
- Emacs mode: RIS mode and REFDB mode

RefDB Command Line Interface

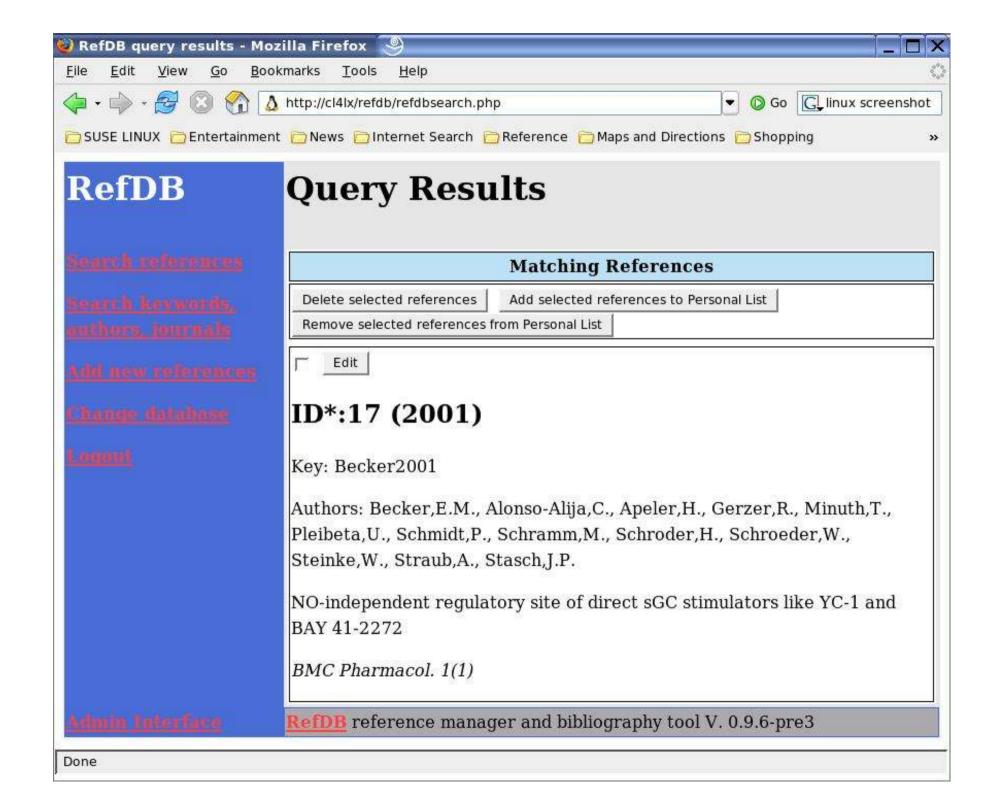
Several clients:

- refdba for administrator
- refdbc to add, edit, delete, and search reference entries, as well as extended notes. Note that we can only add references from a file, in RIS or RISX format (i.e. no built-in editor):

```
addref -U doe -g .default.ris -E ISO-8859-1 foo.ris
```

- refdbib to create bibliographies
- data format conversion tools

🖏 RefDB query form - Mozil	la Firefox
<u>File Edit View Go Boo</u>	kmarks <u>T</u> ools <u>H</u> elp
	http://cl4lx/refdb/refdbquery.html
SUSE LINUX Entertainment News Internet Search Reference Maps and Directions Shopping »	
RefDB	Reference Query Form
Search references	Display also the following fields:
outhors, journals	□ Notes □ Abstract □ Address □ Reprint □ ISSN □ URL □ Publisher □ City □ U1 □ U2 □ U3 □ U4 □ U5 □ M1 □ M2 □ M3
Add now references Change database	Scope of your query: full reference list
Longoni	Simple query
	Is LIKE ▼ Becker in author ▼ AND ▼
	Is LIKE ▼ in author ▼ AND ▼
	Is LIKE ▼ In author ▼ AND ▼
	IS LIKE ▼ in author ▼
	Send Simple query Reset form
	Advanced query
	Query string: ":ID:>0"
	Send Advanced query Reset form
Admin Interface	RefDB reference manager and bibliography tool V. 0.9.6-pre3
Done	



RefDB Emacs Modes: ris-mode

• Colour coding to spot syntax errors, too much data for a field

• Short-cuts:

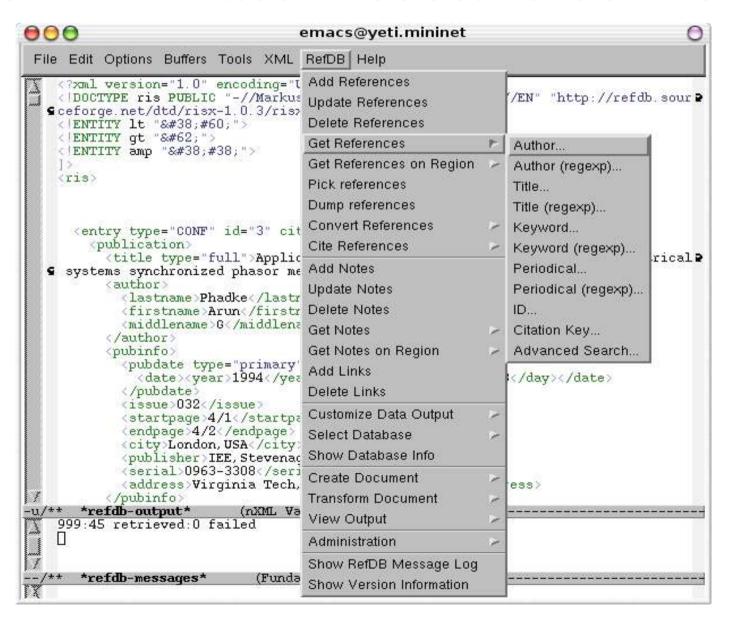
- · Insert a new reference skeleton,
- · insert a tag,
- · duplicate tag,
- move between references,
- narrow buffer to current reference

```
emacs@tipi.mininet
File Edit Options Buffers Tools RefDB Help
     - Inhibition of superoxide anion generation by YC-1 in rat neutrophils throu
 sqh cyclic GMP-dependent and -independent mechanisms.
  N2 - 3-(5'-Hydroxymethyl-2'-furyl)-1-benzyl indazole (YC-1), a soluble quanylyl≥
 ¶ cyclase (sGC) activator, inhibited formyl-methionyl-leucyl-phenylalanine (fMLP)
 ¶-induced superoxide anion (0(2)*(-)) generation and 0(2) consumption in rat neut₽
 Grophils (IC(50) values of 12.7+/-3.1 and 17.7+/-6.9 microM, respectively)
      - Department of Education and Research, Taichung Veterans General Hospital,
 $160 Chung Kang Road, Sec. 3, 407, ROC, Taichung, Taiwan. w1994@vghtc.gov.tw
       - Wang, Karl Heinz
       - Chang, Ling C
       - Raung, S. L.
       - Hsu. M. F
       - Huang, L. J.
       - Kuo, S. C.
       - 2002///
       - 2002
      - Biochem Pharmacol
       - Biochemical Pharmacology
       - 63
      - EC 4, 6.1.2 (Guanylate Cyclase)

    ON REQUEST (01/01/99)

      sample.ris
```

RefDB Emacs Modes: refdb-mode



A Quick Overview

- JabRef: + fully BibTeX compatible, + Java GUI,
 sharing only possible with restrictions
- **CiteULike**: + web-based, + convenient data entry from certain sources, no tight control over data format (for export)
- **RefDB**: + neutral data format, + adaptable export/import, + good support for sharing,
 - (currently) incomplete GUI but emacs modes,
 - more work to setup/maintain.

LSV BibTeX Collector

- Simple, low-maintencance solution to support the sharing of BibTeX entries in a group
- Crawls user directories for references every night
- Uses Bibclean to sort, pretty-print and merge entries,
 - eliminating duplicates
- Interoperable with e.g. JabRef
- for UNIX and Linux
- Free

```
jochen@ws13lx:~ <3>
jochen@ws13lx:"> /vol/references/bin/collect-bib-all.pl
LSV BibTeX Collector 1.0
|Written by Jochen L. Leidner <jochen.leidner@lsv.uni-saarland.de>
Gathering BibTeX data from user dietmar...
Gathering BibTeX data from user marcela...
                                                [failt]
Gathering BibTeX data from user stefan...
                                                failt]
Gathering BibTeX data from user jochen...
Gathering BibTeX data from user nobody...
                                                failt
Gathering BibTeX data from user diana...
Gathering BibTeX data from user abeschorner...
Gathering BibTeX data from user dan ...
Gathering BibTeX data from user irene...
                                                failt
Gathering BibTeX data from user kgaroufi...
                                                [failt]
Gathering BibTeX data from user dietrich...
Gathering BibTeX data from user christoph...
                                                [failt]
Gathering BibTeX data from user smartweb...
                                                [failt]
Gathering BibTeX data from user barbara...
Gathering BibTeX data from user munawar...
                                                [failt]
Gathering BibTeX data from user andrea...
                                                [failt]
Gathering BibTeX data from user testuser...
                                                [failt]
Gathering BibTeX data from user yasir...
                                                [failt]
Gathering BibTeX data from user amerkel...
jochen@ws13lx:~>
```

Take-Home Message

- Reference Management do it! :-)
- Can use: BibTeX, JabRef, CiteULike, RefDB...
- Many other options, especially if you're a Windows or Mac user
- Sharing references has some advantages
- See References for further information

References

- Overview of reference management software: http://bibliographic.openoffice.org/biblio-sw.html
- BibTeX: many guides and tutorials, e.g. http://newton.ex.ac.uk/tex/pack/bibtex/btxdoc/btxdoc.html
- JabRef: http://jabref.sourceforge.net/
- CiteULike: http://www.citeulike.org/
- RefDB: http://refdb.sourceforge.net/
- RIS format:
 http://www.refman.com/support/risformat_intro.asp