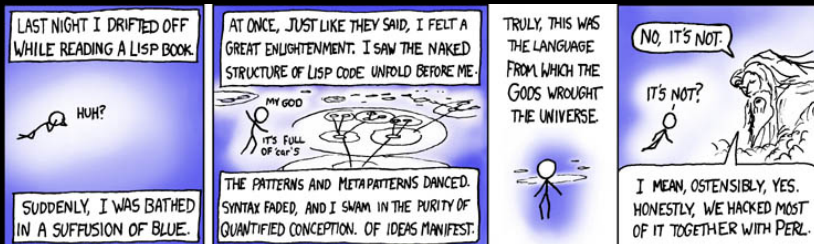


Literature Management with Perl and Emacs

Stefan Washietl

EMBL-European Bioinformatics Institute, Hinxton/Cambridge, UK



Part 1. A PubMed Interface for Emacs

Part 2. Automatic downloading of PDFs

Part 1.

A PubMed Interface for Emacs

EMACS

The Extensible, Customizable Self-Documenting Display Editor

by

Richard M. Stallman

Abstract: EMACS is a display editor which is implemented in an interpreted high level language. This allows users to extend the editor by replacing parts of it, to experiment with alternative command languages, and to share extensions which are generally useful. The ease of extension has contributed to the growth of a large set of useful features. This paper describes the organization of the EMACS system, emphasizing the way in which extensibility is achieved and used.

Keywords: Display, Editor, Extensible, Interactive, Self-documenting

Motivation

- ▶ Apparently no direct way to get PubMed references as BibTeX into Emacs.

	Min. no. of clicks	No. of applications
Shell script	5	3
Hubmed	5	2
JabRef	6	2
PubMode	0	1

Motivation

- ▶ Apparently no direct way to get PubMed references as BibTeX into Emacs.

	Min. no. of clicks	No. of applications
Shell script	5	3
Hubmed	5	2
JabRef	6	2
PubMode	0	1

- ▶ A few clicks can change user-experience and individual workflows.

PubMode

- ▶ PubMode does not impose any particular workflow. It works with ever workflow you like and plays well with BibTeX mode, AucTeX, RefTeX.
- ▶ 100% Emacs Lisp without any dependencies
- ▶ Can be installed with one line in `.emacs`

```
(add-to-list 'load-path " /.emacs.d/pubmode/lisp")
```

PubMode

- ▶ Implements the whole public API provided by NCBI including
 - ▶ LinkOut
 - ▶ Related articles
 - ▶ Spell correction
- ▶ Generates robust and customizable BibTeX
- ▶ Does not neglect corner cases. Tries also to get all your
“von Sørensen jr.” or “Bompfünewerer Consortium”
right.
- ▶ Tested and ready for prime-time.

Part 2.

Automatic downloading of PDFs

Firefox browser window titled "Pfreundschuh M [Author] - PubMed Results - Mozilla Firefox". The address bar shows the URL: <http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&Cmd=Search&Term=Pfreundschu>. The search results list several articles:

- [MabThera International Trial \(MINT\) Group. Lancet Oncol. 2008 May;9\(5\):435-44. Epub 2008 Apr 8. PMID: 18400558 \[PubMed - indexed for MEDLINE\] \[Related Articles\]\(#\)](#)
- 17: [First clinical use of ofatumumab, a novel fully human anti-CD20 monoclonal antibody in relapsed or refractory follicular lymphoma: results of a phase 1/2 trial.](#)**
Hagenbeek A, Gadeberg O, Johnson P, Pedersen LM, Walewski J, Hellmann A, Link BK, Robak T, Wojtkiewicz M, Pfreundschuh M, Kneba M, Engert A, Sonneveld P, Flensburg M, Petersen J, Losic N, Radford J.
Blood. 2008 Jun 15;111(12):5486-95. Epub 2008 Apr 4.
PMID: 18390837 [PubMed - indexed for MEDLINE] [Related Articles](#)
- 18: [Aspects of chemotherapy schedules in young and elderly patients with aggressive lymphoma.](#)**
Zwick C, Gleissner B, Pfreundschuh M.
Clin Lymphoma Myeloma. 2007 Dec;8 Suppl 2:S43-9. Review.
PMID: 18284715 [PubMed - indexed for MEDLINE] [Related Articles](#)
- 19: [Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: a randomised controlled trial \(RICOVER-60\).](#)**
Pfreundschuh M, Schubert J, Ziepert M, Schmits R, Mohren M, Lengfelder E, Reiser M, Nickenig C, Clemens M, Peter N, Bokemeyer C, Eimermacher H, Ho A, Hoffmann M, Mertelsmann R, Trümper L, Balleisen L, Liersch R, Metzner B, Hartmann F, Glass B, Poeschel V, Schmitz N, Ruebe C, Feller AC, Loeffler M; German High-Grade Non-Hodgkin Lymphoma Study Group (DSHNHL).
Lancet Oncol. 2008 Feb;9(2):105-16. Epub 2008 Jan 15.
PMID: 18226581 [PubMed - indexed for MEDLINE] [Related Articles](#)
- 20: [Dose-escalated CHOEP for the treatment of young patients with aggressive non-Hodgkin's lymphoma: I. A randomized dose escalation and feasibility study with bi- and tri-weekly regimens.](#)**
Trümper L, Zwick C, Ziepert M, Hohloch K, Schmits R, Mohren M, Liersch R, Bentz M, Graeven U, Wruck U, Hoffmann M, Metzner B, Hasenclever D, Loeffler M, Pfreundschuh M; German High-Grade Non-Hodgkin's Lymphoma Study Group (DSHNHL).
Ann Oncol. 2008 Mar;19(3):538-44. Epub 2008 Jan 22.
PMID: 18212092 [PubMed - indexed for MEDLINE] [Related Articles](#)

Click #1

Six versus eight cycles of bi-weekly CHOP-14 with ... [Lancet Oncol. 2008] - PubMed Result - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.ncbi.nlm.nih.gov/pubmed/18226581?ordinalpos=19&itool=EntrezSystem2.P...

me screenshot current window

Six versus eight cycles of bi...

NCBI PubMed A service of the U.S. National Library of Medicine and the National Institutes of Health www.pubmed.gov My NCBI Sign In Registered

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for Go Clear Advanced Search

Limits Preview/Index History Clipboard Details

Display AbstractPlus Show 20 Sort By Send to

All: 1 Review: 0

1: Lancet Oncol, 2008 Feb;9(2):105-16. Epub 2008 Jan 15. THE LANCET Oncology

Comment in:
[Lancet Oncol. 2008 Feb;9\(2\):84-6.](#)

Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: a randomised controlled trial (RICOVER-60).

[Pfreundschuh M, Schubert J, Ziegert M, Schmits R, Mohren M, Lengfelder E, Reiser M, Nickenig C, Clemens M, Peter N, Bokemeyer C, Einermacher H, Ho A, Hoffmann M, Mertelsmann R, Trümper L, Balleisen L, Liersch R, Metzner B, Hartmann F, Glass B, Poeschel V, Schmitz N, Ruebe C, Feller AC, Loeffler M: German High-Grade Non-Hodgkin Lymphoma Study Group \(DSHNHL\).](#)

Internal Medicine, Saarland University Medical School, Saarland University, Homburg, Germany. inmpfr@uniklinikum-saarland.de

BACKGROUND: Cyclophosphamide, doxorubicin, vincristine, and prednisolone (CHOP) is used to treat patients with non-Hodgkin lymphoma. Interval decrease from 3 weeks of treatment (CHOP-21) to 2 weeks (CHOP-14), and addition of rituximab to CHOP-21 (R-CHOP-21) has been shown to improve outcome in elderly patients with diffuse large B-cell lymphoma (DLBCL). This randomised trial assessed whether six or eight cycles of R-CHOP-14 can improve outcome of these patients compared with six or eight cycles of CHOP-14. **METHODS:** 1222 elderly patients (aged 61-80 years) were randomly assigned to six or eight cycles of CHOP-14 with or without rituximab. Radiotherapy was planned to sites of initial bulky disease with or without extranodal involvement. The primary endpoint was event-free survival; secondary endpoints were response, progression during treatment, progression-free survival, overall survival, and frequency of toxic effects. Analyses were done by intention to treat. The trial is registered on National Cancer Institute website, number NCT00052936 and as EU-20243. **FINDINGS:** 3-year event-free survival was 47.2% after six cycles of CHOP-14 (95% CI 41.2-53.3), 53.0% (47.0-59.1) after eight cycles of CHOP-14, 66.5% (60.9-72.0) after six cycles of R-CHOP-14, and 63.1% (57.4-68.8) after eight cycles of

Related Articles

- Review** Rituximab: a review of its use in non-Hodgkin's lymphoma and chronic lymphocytic leukaemia. [Drugs. 2003]
- CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell [Lancet Oncol. 2006]
- Prognostic significance of maximum tumour (bulk) diameter in young patients with good-prognosis diffuse large-B-cell lymphoma treat [Lancet Oncol. 2008]
- Rituximab plus CHOP (R-CHOP) overcomes bcl-2-associated resistance to chemotherapy in elderly patients with diffuse large B-cell lymph [Blood. 2003]
- Review** The best treatment for diffuse large B-cell lymphoma: a German perspective. [Oncology (Williston Park). 2005]

» See Reviews... » See All...

Patient Drug Information

- Cyclophosphamide (Cytoxan®, Neosar®) Your doctor has ordered the drug cyclophosphamide to help treat your illness. The drug can be taken by mouth in tablet form or be given by injection into a vein. » read more...
- Daunorubicin (Cerubidine®, DaunoXome®) Your doctor has ordered the drug daunorubicin to help treat your illness. The drug is given by injection into a vein. » read more...

Click #2



Select a website below to get this article.

[Help](#)

Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: a randomised controlled trial (RICOVER-60)

The Lancet Oncology, Volume 9, Issue 2, Pages 105-116

M. Pfreundschuh, J. Schubert, M. Ziepert, R. Schmits, M. Mohren, E. Lengfelder, M. Reiser, C. Nicklenig, M. Clemens, N. Peter

You can get the full-text article here...

...if you are:



- visitor or subscriber to the website
- interested in pay-per-view article purchase
- healthcare practitioner
- student or faculty member
- researcher or librarian



- visitor or subscriber to the website
- healthcare practitioner
- student or faculty member
- researcher or librarian



- based in and accessing the site from Norway



- visitor or subscriber to the journal website
- interested in pay-per-view article purchase
- healthcare practitioner

Click #3

ScienceDirect - The Lancet Oncology : Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive...
File Edit View History Bookmarks Tools Help
http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6W85-4RKMCR5-1&_user=...
Six versus eight cycles of bi-... ScienceDirect - The Lancet O...
ScienceDirect
Login Register
Home Browse Search My Settings Alerts Help
Quick Search All fields Author
search tips Journal/book title Volume Issue Page Clear Go Advanced Search
THE LANCET Oncology
Volume 9, Issue 2, February 2008, Pages 105-116
Font Size: - +
Article Figures/Tables References PDF (283 k) Thumbnails | Full-Size Images
doi:10.1016/S1470-2045(08)70002-0
Cite or Link Using DOI
Copyright © 2008 Elsevier Ltd. All rights reserved.
Fast track — Articles
Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: a randomised controlled trial (RICOVER-60)
Prof Michael Pfreundschuh MD^a, Joerg Schubert MD^a, Marita Ziepert Dipl-Math^a, Rudolf Schmits MD^a, Martin Mohren MD^a, Eva Lengfelder MD^a, Marcel Reiser MD^a, Christina Nickenig MD^a, Michael Clemens MD^a, Norma Peter MD^a, Prof Carsten Bokemeyer MD^a, Hartmut Eimermacher MD^b, Prof Anthony Ho MD^b, Martin Hoffmann MD^b, Prof Roland Mertelsmann MD^b, Prof Lorenz Trümper MD^b, Leopold Balleisen MD^b, Ruediger Liersch MD^b, Bernd Metzner MD^b, Frank Hartmann MD^b, Bertram Glass MD^b, Viola Poeschel MD^b, Norbert Schmitz MD^b, Prof Christian Ruebe MD^{b,1}, Prof Alfred C Feller MD^b, Prof Markus Loeffler MD^b and for the German High-Grade Non-Hodgkin Lymphoma Study Group (DSHNHL)
^aInternal Medicine, Saarland University Medical School, Saarland University, Homburg, Germany
^bRadiotherapy Hospital, Saarland University, Homburg, Germany
Article Toolbox
E-mail Article
Cited By
Save as Citation Alert
Citation Feed
Export Citation
Add to my Quick Links
Add to Scollab
Permissions & Reprints
Cited By in Scopus (24)
Supplementary Content within this Article
Webpanel: Drugs used in this trial
View Within Article
Related Articles in ScienceDirect

Click #4

Bookmarks

Options ▾

- Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients
 - Introduction
 - Methods
 - Patients
 - Treatment
 - Statistical analysis
 - Role of the funding source
 - Results
 - Discussion
 - Acknowledgment
 - References

Articles

Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: a randomised controlled trial (RICOVER-60)

Michael Pfeundscher, Joerg Schubert, Marit Zepert, Rudolf Schmits, Martin Mohren, Ewa Lengfelder, Marcel Belzer, Christina Nicklisch, Michael Clemens, Norval Felix, Carsten Bokemuehl, Hartmut Eimermacher, Anthony Ho, Martin Hoffmann, Roland Mertenmann, Lorenz Trunper, Leopold Ballmann, Rüdiger Liesch, Bernd Metzner, Frank Hartmann, Bertram Glass, Viola Poeschl, Norbert Schmitz, Christian Ruebe, Alfred C Feller, Markus Loeffler, for the German High-Grade/Non-Hodgkin Lymphoma Study Group (DSHNHL)

Summary

Background Cyclophosphamide, doxorubicin, vincristine, and prednisolone (CHOP) is used to treat patients with non-Hodgkin lymphoma. Interval decrease from 3 weeks of treatment (CHOP-21) to 2 weeks (CHOP-14), and addition of rituximab to CHOP-21 (R-CHOP-21) has been shown to improve outcome in elderly patients with diffuse large B-cell lymphoma (DLBCL). This randomised trial assessed whether six or eight cycles of R-CHOP-14 can improve outcome of these patients compared with six or eight cycles of CHOP-14.

Methods 1222 elderly patients (aged 61–80 years) were randomly assigned to six or eight cycles of CHOP-14 with or without rituximab. Radiotherapy was planned to sites of initial bulky disease with or without extranodal involvement. The primary endpoint was event-free survival; secondary endpoints were response, progression during treatment, progression-free survival, overall survival, and frequency of toxic effects. Analyses were done by intention to treat. The trial is registered on National Cancer Institute website, number NCT00052936 and as EU-20243.

Findings 3-year event-free survival was 47.2% after six cycles of CHOP-14 (95% CI 41.2–53.3), 53.0% (47.0–59.1) after eight cycles of CHOP-14, 66.5% (60.9–72.0) after six cycles of R-CHOP-14, and 63.1% (57.4–68.8) after eight cycles of R-CHOP-14. Compared with six cycles of CHOP-14, the improvement in 3-year event-free survival was 5.8% (–2.8–14.4) for eight cycles of CHOP-14, 19.3% (11.1–27.5) for six cycles of R-CHOP-14, and 15.9% (7.6–24.2) for eight cycles of R-CHOP-14. 3-year overall survival was 67.7% (62.0–73.5) for six cycles of CHOP-14, 66.0% (60.1–71.9) for eight cycles of CHOP-14, 78.1% (73.2–83.0) for six cycles of R-CHOP-14, and 72.5% (67.1–77.9) for eight cycles of R-CHOP-14. Compared with treatment with six cycles of CHOP-14, overall survival improved by –1.7% (–10.0–6.6) after eight cycles of CHOP-14, 10.4% (2.8–18.0) after six cycles of R-CHOP-14, and 4.8% (–3.1–12.7) after eight cycles of R-CHOP-14. In a multivariate analysis that used six cycles of CHOP-14 without



 Latest (first 2008, 9, 105-16)

Published Online
 January 15, 2008
 DOI:10.1016/S1473-0142(08)70002-0

See Reflection and Reaction page 14

Internal Medicine, Saarland University Medical School (Prof M Pfeundscher MD, J Schubert MD, R Schmits MD, F Hartmann MD, V Poeschl MD) and Radiotherapy Hospital (Prof C Ruebe MD), Saarland University, Homburg, Germany; Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University, Leipzig, Germany (M Zepert Dipl Math, Prof M Loeffler MD); University Hospital M Aglberg, Magdeburg, Germany (M Mohren MD); Klinikum Mannheim, University of

Click #5 (3 more to actually save it)

Difficulties for automatic PDF retrieval

- ▶ No direct PDF link considered in LinkOut specification
- ▶ No other efforts by publishers to provide a standardized interface to PDFs
- ▶ Use of meta-tags not widely adopted; Most publishers ignore them, some use them incorrectly.
- ▶ Many publishers present PDFs without any care for web-standards, semantics and accessibility.

Idiotically linked PDFs: a business



An open source solution

- ▶ Fully automatic crawler not feasible:
 - ▶ Would need to crawl up to three levels which is too slow
 - ▶ Publisher sites don't like crawlers
- ▶ We propose a rule-based semi-automatic agent to find PDFs

Basic principle of the system

URL



URL



Basic principle of the system

URL



...

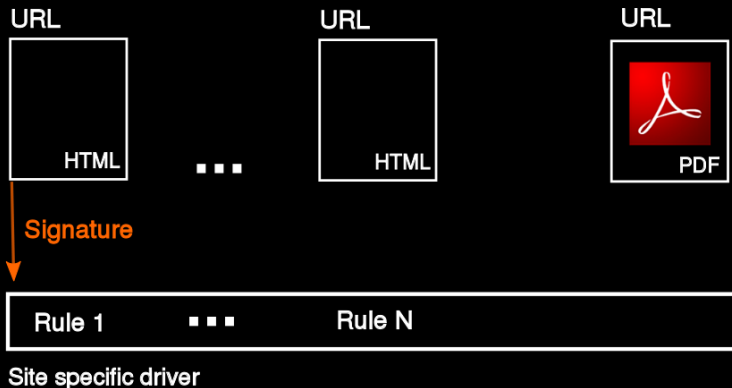
URL



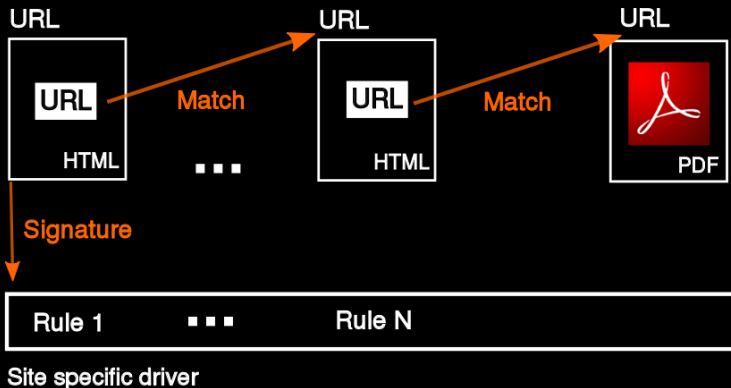
URL



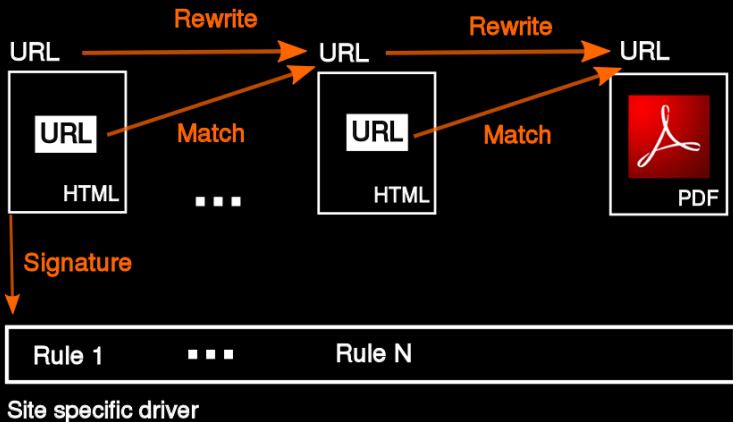
Basic principle of the system



Basic principle of the system



Basic principle of the system





A service of the U.S. National Library of Medicine
and the National Institutes of Health

www.pubmed.gov

My NCBI

[Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for Go Clear [Advanced Search](#) [Save Search](#)

Limits Preview/Index History Clipboard Details

Display AbstractPlus Show 20 Sort By Send to

All: 1 Review: 0

We found 1 article:

1: [Indian J Cancer](#). 2007 Jan-Mar;44(1):36-7.

[LinkOut to related resource](#) Links

Frontal convexity primary lymphoma masquerading meningioma: a case report and review of literature.

[Agrawal A](#), [Makannavar JH](#), [Shetty JP](#), [Shetty RK](#), [Shetty L](#).

Department of Neurosurgery, K. S. Hegde Medical Academy, Mangalore, India. dramit_in@yahoo.com

Primary non-Hodgkin's lymphoma of the skull with extra- and intracranial extension without systemic or skeletal manifestation in a nonimmunocompromised patient is extremely rare. These lesions often cause difficulty in diagnosis because they mimic other conditions. We report a case of primary lymphoma involving scalp, skull vault, meninges and invading the brain parenchyma and masquerading clinically and radiologically as a meningioma.

PMID: 17401223 [PubMed - indexed for MEDLINE]

Related Articles

- ▶ **Review** Primary malignant non-Hodgkin's lymphoma of cranial vault: a case report. [Surg Neurol. 1993]
- ▶ Intradiploic meningioma of the skull. Case report and review of literature. [J Neurosurg Sci. 1999]
- ▶ Extracranial meningioma and pregnancy: a rare diagnosis. [J Plast Reconstr Aesthet Surg. 2007]
- ▶ Primary cranial vault lymphoma mimicking a meningioma. [Neurochirurgia (Stuttg). 1983]
- ▶ **Review** Intraosseous calvarial meningioma of the skull presenting as a solitary osteolytic skull lesion: case report. [Acta Neurochir (Wien). 1994]

▶ See Reviews... | ▶ See All...

Recent Activity

Turn Off Clear

NAVIGATE here

Search

- » Next Article
- » Previous Article
- » Table of Contents

RESOURCE Inks

- » Similar in PUBMED
- » Search Pubmed for

- Agrawal A
- Makannavar JH
- Shetty JP
- Shetty RK
- Shetty L

- » Article in PDF (97 KB)
- » Citation Manager
- » Access Statistics
- » Reader Comments
- » Email Alert *
- » Add to My List *

* Registration required (free)

Meningioma Brain Tumor

Key research articles at your fingertips...free sign up
www.touchoncologicaldisease.co

Medical Negligence Claims

Suffered as a result of negligence? Make your Compensation Claim Now!
National-Accident-Helpline.co.uk

Lymphoma

Find info on lymphoma facts, symptom & treatment.
AllCancer-Info.com

Higher Earnings

Many opportunities to increase your earnings working in France!
www.professionmedecin.fr

Ads by Google

CASE REPORT

Year : 2007 | Volume : 44 | Issue : 1 | Page : 36-37

Frontal convexity primary lymphoma masquerading meningioma: A case report and review of literature

A Agrawal¹, JH Makannavar², JP Shetty², RK Shetty³, L Shetty³

¹ Department of Neurosurgery, K. S. Hegde Medical Academy, Mangalore, India

² Department of Pathology, K. S. Hegde Medical Academy, Mangalore, India

³ Department of Radiology, K. S. Hegde Medical Academy, Mangalore, India

Correspondence Address:

A Agrawal
Department of Neurosurgery, K. S. Hegde Medical Academy, Mangalore
India

Login to access the email ID

Indian Journal of Cancer: PDF - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.indianjcancer.com/downloadpdf.asp?issn=0019-509X;year=2007;v...

Frontal convexity primary ly... Indian Journal of Cancer: PDF

Home Feedback Subscribe Top cited articles Login

Users Online :3

Indian Journal of CANCER

About Current Issue Archive Search Instructions Advertise Contact

NAVIGATE here

- My Preferences
- Manuscript submission
- Advanced Search


Join Us
Be a reviewer for IJC

See Statistics
Most popular articles

Email Alerts
Get latest news from us


INDIAN
CANCER
SOCIETY

» Download PDF

To download PDF version of selected article [click here](#) 

This is a temporary file and hence do not link it from a website. Instead link the URL of this page if you wish to link the PDF file.

To read this article you will require Adobe® Acrobat Reader, if you do not have this installed you can download it from [here](#)



The PDF provided from this link is may be protected to disallow full editing and/or printing. The HTML version provides free full text articles. By no means restricting the PDF printing is contradicting to the definition or philosophy of open access. By restricting PDF printing we wish to avoid commercial exploitation of our content.

You could get a non-protected PDF from Infotrieve, Copyright Clearance Centre or write to subscriptions [AT] medknow [DOT] com.

IndianJournalofCancer44136-3404883_092728.pdf (application/pdf Object) - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.indianjancer.com/temp/IndianJournalofCancer44136-3404883_09272

Frontal convexity primary ly... Indian Journal of Cancer: PDF IndianJournalofCancer44136-...

1 / 2 120% Find

[Downloaded free from <http://www.indianjancer.com> on Monday, February 16, 2009]

Case Report

Frontal convexity primary lymphoma masquerading meningioma: A case report and review of literature

Agrawal A, Makannavar JH*, Shetty JP*, Shetty RK**, Shetty L**
Departments of Neurosurgery, *Pathology and **Radiology, K. S. Hegde Medical Academy, Mangalore, India

Correspondence to: Dr. Amit Agrawal, E-mail: dramit_in@yahoo.com

Abstract

Primary non-Hodgkin's lymphoma of the skull with extra- and intracranial extension without systemic or skeletal manifestation in a nonimmunocompromised patient is extremely rare. These lesions often cause difficulty in diagnosis because they mimic other conditions. We report a case of primary lymphoma involving scalp, skull vault, meninges and invading the brain parenchyma and masquerading clinically and radiologically as a meningioma.

Example of a site driver

```
<site name="Medknown">  
  <signature>  
    <url>!article.asp\?issn=\S*?aulast=\S*!</url>  
  </signature>
```

Example of a site driver

```
<site name="Medknown">
  <signature>
    <url>!article.asp\?issn=\S*?aulast=\S*!</url>
  </signature>
  <rule>
    <pattern>
      <rewrite>!(.*)!$1;type=2!</rewrite>
    </pattern>
```

Example of a site driver

```
<site name="Medknown">
  <signature>
    <url>!article.asp\?issn=\S*?aulast=\S*!</url>
  </signature>
  <rule>
    <pattern>
      <rewrite>!(.*)!$1;type=2!</rewrite>
    </pattern>
    <pattern>
      <match>!(temp/\S*?.pdf)&quot;!</match>
    </pattern>
  </rule>
```

Example of a site driver

```
<site name="Medknown">
  <signature>
    <url>!article.asp\?issn=\S*?aulast=\S*!</url>
  </signature>
  <rule>
    <pattern>
      <rewrite>! (.*)!$1;type=2!</rewrite>
    </pattern>
    <pattern>
      <match>! (temp/\S*?.pdf)&quot;!</match>
    </pattern>
  </rule>
  <test>
    <url>http://www.indianjcancer.com/article.asp?issn=...</url>
  </test>
</site>
```

Driver in YAML format

Medknown:

signature:

url:

- '!article.asp\?issn=\S*?aulast=\S*!'

rule:

- pattern:

- rewrite: '!(.*)!\$1;type=2!'
- match: '!(temp/\S*?.pdf)!"!'

test:

url:

- http://www.indianjcancer.com/article.asp?issn=...

Publisher and journals in PubMed

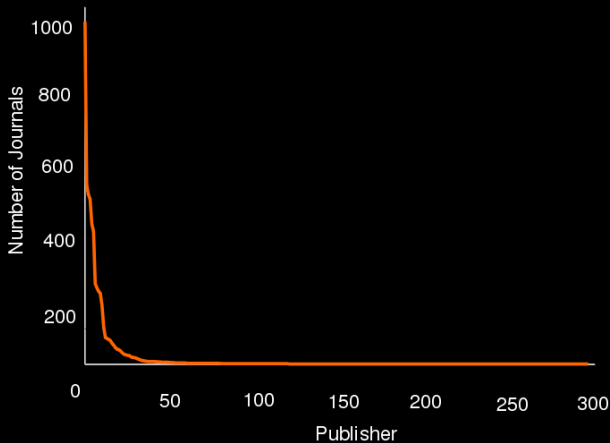
- ▶ 6574 journals (Nov 2008)

Publisher and journals in PubMed

- ▶ 6574 journals (Nov 2008)
- ▶ 302 publisher

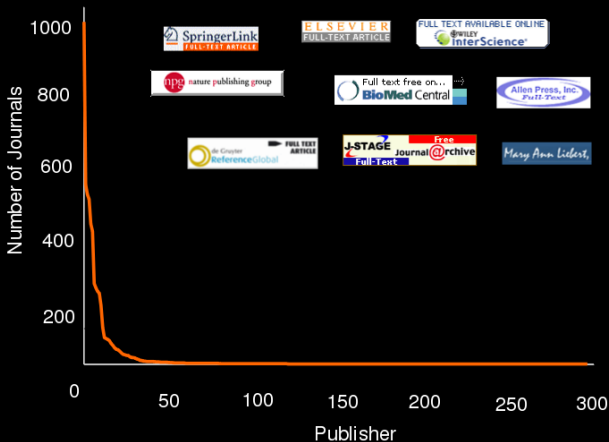
Publisher and journals in PubMed

- ▶ 6574 journals (Nov 2008)
- ▶ 302 publisher



Publisher and journals in PubMed

- ▶ 6574 journals (Nov 2008)
- ▶ 302 publisher



Some sites are easy

Springer:

signature:

url:

- '!springerlink!'

rule:

- pattern:

- match: '!(/content/\S*fulltext\.pdf)!'

Some sites are easy

Springer:

signature:

url:

- '!springerlink!'

rule:

- pattern:

- match: '!(/content/\S*fulltext\.pdf)!'

NPG:

signature:

url:

- '!nature\.com!'

rule:

- pattern:

- match: '!(/\S*\.pdf)!'

Implementation

- ▶ Perl module available that implements an intelligent agent that
 - ▶ interprets site driver files and searches for PDFs
 - ▶ supports all necessary browser technologies (cookies, proxy...)
 - ▶ validates file type identity
 - ▶ downloads PDF
- ▶ Currently 17 proof-of-concept drivers available covering vast majority of papers in PubMed.
- ▶ No site encountered yet that could not be crawled using simple signature/match/rewrite rules.
- ▶ Automatic test suite to routinely validate set of drivers

Conclusions

- ▶ The lack of automatic access to PDFs hindered the development of any sane and complete literature management system for years.
- ▶ I have shown a simple yet robust open source solution for the problem which can be part of any larger system in any language.
- ▶ Note for Pubget and QUOSA: I wish you good luck with a business essentially built on proprietary software that can be replaced by 220 lines of Perl.

github.com/wash