

Integrated repositories: usage, citations and other metrics

Tim Brody
University of Southampton, UK

Contents

- Making best use of existing services
- Integration of ISI
- Integration of Google Scholar
- Use of citation data in the repository
- Repository usage statistics

ISI Web of Science Web Services

- An API for the science-citation index (and related databases)
- Supports: search, record retrieval, citing/cited article lists
- Search over most ISI fields e.g. title, author-name, author-affiliation
- Retrieve full records: bibliographic data + abstract + citings
- SOAP (XML query/response)

What we can do with ISI

- Add citation counts for our records
- (Re)publish most bibliographic fields for our publications e.g. filling in blanks
- *Required to add a note that the data comes from Web of Science*
- **Can't** republish list of citing/cited articles
- **Can't** republish abstracts (publisher copyright)
- No technical means to link to ISI record (unstable URLs) ???

Integration with ISI

- Built a wrapper for SOAP API that:
 - Provides a search() method
 - Extracts result record XML
- Script that searches repository for relevant records to update:
 - is journal article OR conference item
 - Queries ISI ...
 - Sets cited count for any matches found

Querying ISI

- Find the matching record in ISI's database
- First-named author:
 - *Kiang, K. M., Frampton, K., Monro, T. M., ...*
 - au = Kiang K*
- Title:
 - *Extruded single-mode non-silica glass holey optical fibres*
 - ti = Extruded single mode ...
- (au = Kiang K*) AND (ti = Extruded single mode ...)

Performance of Matching

- Problems with inconsistent (or incorrect) titles
- (TODO: Check false-positives)
- Easy to refine search query

Google Scholar

- Web site with a boolean query interface
- Can search for citing articles
- Blocks anything that queries too often – no service agreement
- No structured data, except for cite count



Integration with Scholar

- A fake Web user that:
- Queries for title + name:
 - “Extruded single-mode ... author:Kiang K”
- Searches the first 100 matches:
 - a record that links to the repository **OR**
 - where the title link is identical
- Updates record with “Cited by X” count

Adding Citations to the Repository

- Citation counts added as new metadata fields:
 - ISI Total Citations
 - Google Scholar Total Citations
- Added to search results
- Added as a sort-by value

Date is between 01 January 2008 and 01 January 2009 inclusive



Displaying results 1 to 20 of 686.

[Refine search](#) | [New search](#) | [1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [6](#) | [7](#) | [8](#) | [9](#) | [10](#) | [11](#) | [Next](#)

Order the results:

Export 686 results as

 [RSS 1.0](#)  [Atom](#)  [RSS 2.0](#)

1. Feng, X., Loh, W. H., Flanagan, J. C., Camerlingo, A., Dasgupta, S., Petropoulos, P., Horak, P., Frampton, K., White, N. M., Price, J. H. V., Rutt, H. and Richardson, D. J. (2008) Single-Mode Tellurite Glass Holey Fiber with Extremely Large Mode Area for Infrared Nonlinear Applications. *Optics Express*, 16 (18). pp. 13651-13655. (Cited 6 times in ISI and 7 times in Google Scholar.) 
2. Shadbolt, N. and Berners-Lee, T. (2008) Web Science Emerges. *Scientific American* . (Cited 5 times in ISI and 5 times in Google Scholar.) 

Not forgetting Elsevier's Scopus



THE UNIVERSITY OF LIVERPOOL RESEARCH ARCHIVE

[Login](#)

[About](#)
[Deposit an Item](#)
[Deposit Guide](#)
[Policies](#)
[Copyright Awareness](#)
[Depositors Helpline](#)

**Scholarly
Communication**

[Overview](#)

[News](#)

[Wellcome Trust Open
Access Policy](#)

[Contacts](#)

[Archive Admin](#)

CpG island methylation phenotype (CIMP) in oral squamous cell carcinoma: associated with a marked inflammatory response and less aggressive tumour biology

Shaw, Richard J.; Hall, Gillian L.; Lowe, Derek; Bowers, Naomi L; Liloglou, Triantafillos; Field, John K.; Woolgar, Julia A. and Risk, Janet M. (2007) *CpG island methylation phenotype (CIMP) in oral squamous cell carcinoma: associated with a marked inflammatory response and less aggressive tumour biology*. *Oral Oncology*, 43 (9). pp. 878-886. ISSN 1368-8375



[PDF - Accepted Version](#)
496Kb

Official URL: <http://www.elsevier.com/wps/find/journaldescriptio...>

Cited 14 times in **SCOPUS**



Abstract

Studies in several tumour sites highlight the significance of the CpG island methylation phenotype (CIMP), with distinct features of histology, biological aggression and outcome. We utilise pyrosequencing techniques of quantitative methylation analysis to investigate the presence of CIMP in oral squamous cell carcinoma (OSCC) for the first time, and evaluate its correlation with allelic imbalance, pathology and clinical behaviour. Tumour tissue, control tissue and PBLs were obtained from 74 patients with oral squamous cell carcinoma. Pyrosequencing was used to analyse methylation patterns in 75–200 bp regions of the CpG rich gene promoters of 10 genes with a broad range of cellular functions. Allelic imbalance was investigated using a multiplexed panel

Reporting to research managers

- A tool that summarises the repository content
- Not being used for comparing (competing) *people, groups*
- Are faculty depositing their research?
- Proportion of deposits that are primary research outputs (vs. grey literature)?
- What are the notable papers (by cites)?

Depositor Activity Report

Date Deposited

2008-01-01-2009-12-31

Item Type

- Article
- Conference or Workshop Item
- Book
- Book Section
- Technical Report
- Thesis
- Other

Research Group

- Intelligence, Agents, Multimedia
-
- BIO@ECS Research Group
- Communications Research Group
- Dependable Systems and Software Engineering Research Group
- Electrical Power Engineering
- ...High Voltage Laboratory
- Electronic Systems and Devices Group
- Electronics Construction, Maintenance and Logistics
- Grid and Pervasive Computing Research Group
- High Performance Computing Research Group
- IT Innovation Centre

All of these ▾

Report Format

Select an output format. The CSV format can be saved and loaded into Microsoft Excel.

HTML ▾

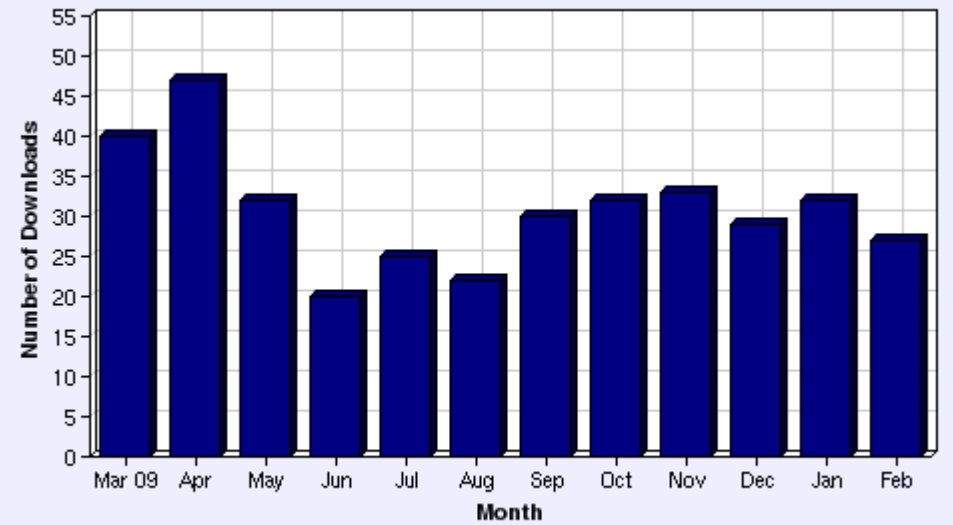
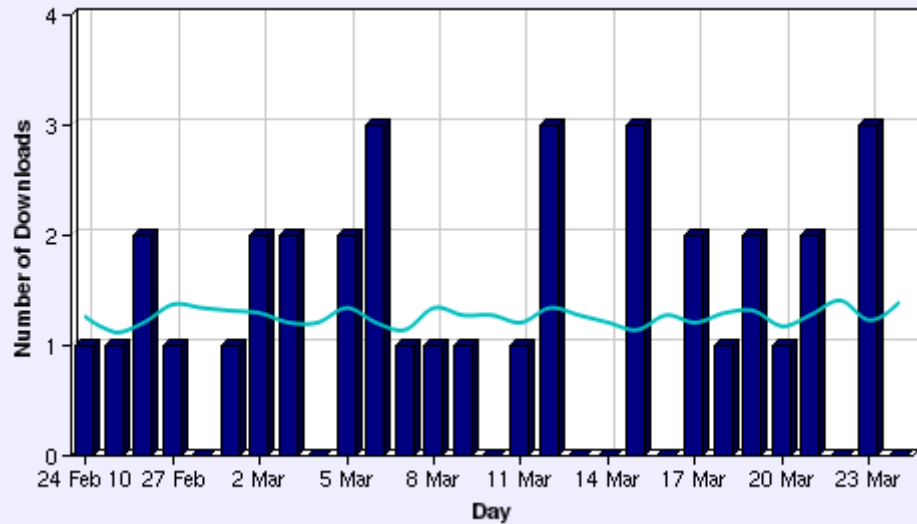
Search Reset the form

Name	Username	Authored	Journal	Conference	In-ISI	ISI-Cites	In-GScholar	GScholar-Cites
Alahmari, Saad	saa08r	1	0	1	0	0	0	0
Alani, Harith	ha	13	3	10	4	7	7	52
Aldeco Perez, Rocio	raap06r	1	0	1	0	0	0	0
Andre, Paul	pa2	13	1	12	3	0	8	52
Au Yeung, Ching Man	cmay06r	11	0	11	1	0	6	23
Bandara, H	locked_2404	1	0	1	1	1	0	0
Berners-Lee, Tim	timbl	2	2	0	2	11	2	42
Braines, Dave	db06v	13	0	13	1	0	12	40
Brody, Tim	tdb2	3	1	2	0	0	1	1
Brody, Timothy	locked_1403	1	1	0	1	3	1	198
Byde, Andrew	ab06v	2	1	1	1	0	2	3
Carr, Les	lac	23	6	17	3	3	7	206
Chalkiadakis, Georgios	gc2	8	0	8	1	0	4	30
Chapman, Archie	locked_5438	5	1	4	0	0	0	0
Chen, Zheng	zc05r	3	0	3	1	1	3	218
Correndo, Gianluca	gc3	13	2	11	2	0	6	35
Croitoru, Madalina	locked_5463	6	0	6	5	0	5	15
Crowder, Richard	rmc	13	1	12	5	0	5	7
Dash, Rajdeep	rkd	3	3	0	2	0	3	6
De Roure, David	dder	34	12	22	11	4	11	75
Diamantas, Sotirios	sd04r	1	0	1	1	0	0	0
Dunnlaw David	dnd	8	2	6	7	2	5	16

Usage Data

- Analyse usage of **full-texts** in the repository
 - Feedback to authors
 - Compare with cites
- Future: integrate repository usage with other venues
 - Merge journal and repository usage

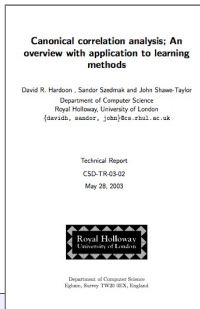
IRStats for Brody's PhD Thesis



Every Eprint Tells a Story

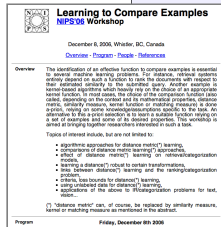
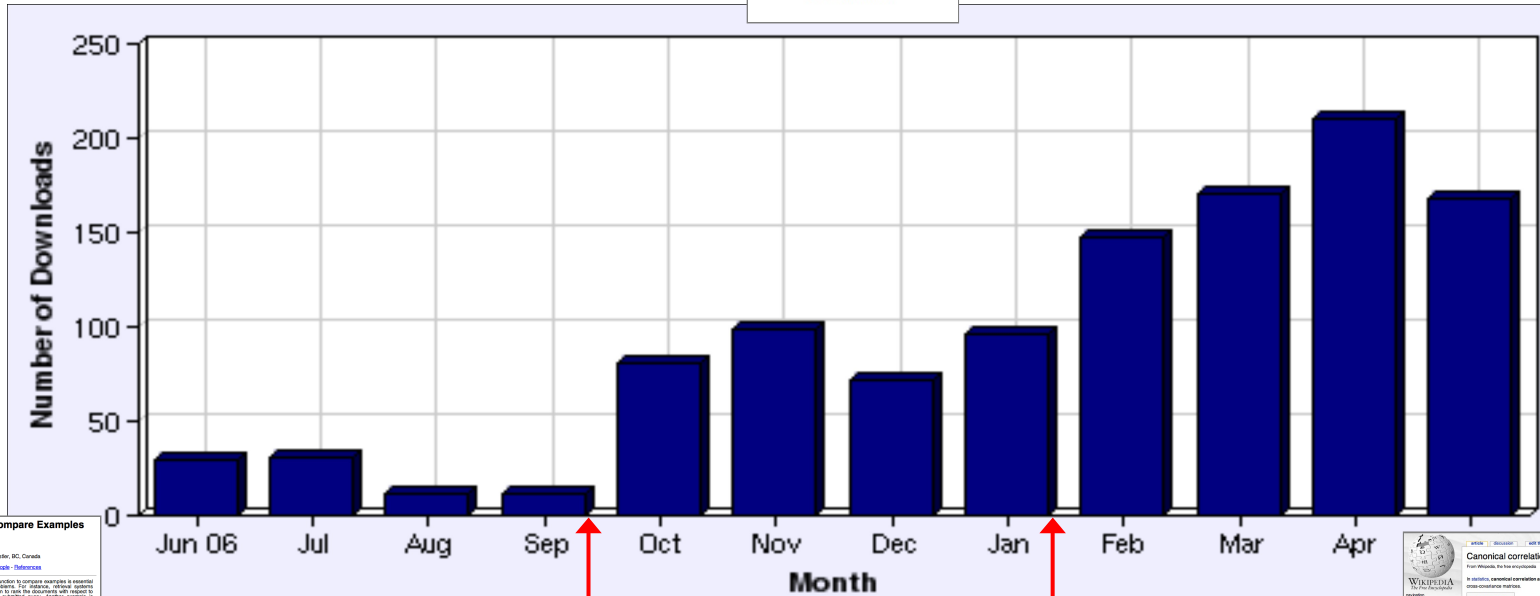
What explains its download profile?

Why makes it popular ... or not?



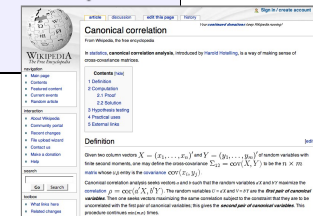
Hardoon, D. R., Szedmak, S. and Shawe-Taylor, J. (2003) Canonical correlation analysis; An overview with application to learning methods. Technical Report CSD-TR-03-02, Computer Science Department, Royal Holloway, University of London.

<http://eprints.ecs.soton.ac.uk/9225/>



End Sept 2006: NIPS 2006 workshop linked to this eprint from its webpage
<http://www.idiap.ch/lce/>

End Jan 2006: external reference on the Wikipedia "Canonical correlation" page at
http://en.wikipedia.org/w/index.php?title=Canonical_correlation



Both external links bring new visitors directly to this unpublished technical report, but another effect is to increase the Google search traffic as a result of increased PageRank.

Thanks

- ISI tool developed as part of the Readiness4REF project:
 - <http://www.kcl.ac.uk/iss/cerch/projects/portfolio/r4r.html>
- IRStats developed as part of the Interoperable Repository Statistics project:
 - <http://irs.eprints.org/>

Process

