Citation Services for Institutional Repositories: Citebase Search

Tim Brody
Intelligence, Agents, Multimedia Group
University of Southampton
Content

• The Open Access Literature
• Why Open Access?
• Citation Services for IRs
  – Distributed Archives
  – Citebase Search
• Effect of Open Access
Impact cycle begins:
Research is done

Researchers write pre-refereeing “Pre-Print”

Submitted to Journal

Pre-Print reviewed by Peer Experts – “Peer-Review”

Pre-Print revised by article’s Authors

Refereed “Post-Print” Accepted, Certified, Published by Journal

Researchers can access the Post-Print if their university has a subscription to the Journal

New impact cycles: New research builds on existing research
Open Access Literature

• Research Archives ("author self-archiving")
  – 340,000 arXiv.org
  – 700,000 citeseer
  – 1,000s in institutional & other repositories

• Open Access Journals
  – BioMed Central/PLOS/ [DOAJ]

• Time-delayed access
  – PubMed Central
  – HighWire Press

• Personal Web pages
"Skywriting": All research, accessible to all potential users, anywhere, anytime

New impact cycles:
Self-archived research impact is greater (and faster) because access is maximized (and accelerated)

New impact cycles:
New research builds on existing research

Impact cycle begins:
Research is done

Researchers write pre-refereeing "Pre-Print"

Submitted to Journal

Pre-Print reviewed by Peer Experts – "Peer-Review"

Pré-Print revised by article’s Authors

Refereed “Post-Print” Accepted, Certified, Published by Journal

Post-Print self-archived to University’s Eprint Website

Researchers can access the Post-Print if their university has a subscription to the Journal
Why Open Access?

• Maximise research impact through maximised access [evidence of 50-250% more citations]
• Efficiency
  – ADS Est. to provide $250 million benefit to astronomy
• Continuous and comprehensive assessment
• Periphery benefits
  – Institutional management
  – Publicly funded research publicly accessible
  – Developing World Access
  – Easier to identify plagiarism (do a Google search!)
OA Benefit: Services

- General Web-based full-text search engines
- Research literature specific: A&I, publisher, library, aggregator, journal contents, society …
Citebase Search

• “Citebase Search is a semi-autonomous citation index for the free, online research literature. It harvests pre- and post-prints (most author self-archived) from OAI-PMH compliant archives, parses and links their references and indexes the metadata in a search engine.”
New Dimensions at a Millimeter to a Fermi and Superstrings at a TeV

... is proposed which does not rely on low energy supersymmetry or technicolor. The gravit... strong gravity such as the production of black holes. The new dimensions can be probed by events with large missing energy carried off by gravitons escaping into ...

The Black Hole in Three Dimensional Space Time

... negative cosmological constant, admit a black hole solution. The 2+1 black hole - characterized by mass, angular momentum ... de Sitter space appears as a negative energy state separated by a mass gap from the continuous black hole spectrum. Evaluation of the partition ...

The World as a Hologram

... Planck scale. The consequences for high energy particle collisions are described. The ... s related to information spreading near black hole horizons. The considerations of this pa ...

String Solitons

...
New Dimensions at a Millimeter to a Fermi and Superstrings at a TeV

Authors: Antoniadis, I.; Arkani-Hamed, N.; Dimopoulos, S.; Dvali, G.

Recently, a new framework for solving the hierarchy problem has been proposed which does not rely on low energy supersymmetry or technicolor. The gravitational and gauge interactions unite at the electroweak scale, and the observed weakness of gravity at long distances is due to the existence of large new spatial dimensions. In this letter, we show that this framework can be embedded in string theory. These models have a perturbative description in the context of type I string theory. The gravitational sector consists of closed strings propagating in the higher-dimensional bulk, while ordinary matter consists of open strings living on D3-branes. This scenario raises the exciting possibility that the LHC and NLC will experimentally study both ordinary aspects of string physics such as the production of narrow Regge-excitations of all standard model particles, as well more exotic phenomena involving strong gravity such as the production of black holes. The new dimensions can be probed by events with large missing energy carried off by gravitons escaping into the bulk. We finally discuss some important issues of model building, such as proton stability, gauge coupling unification and supersymmetry breaking.

Comment: 12 pages, latex

Full-text available from: Cached PDF
Linked PDF (experimental)

- This Article's Citation/Hit History
- All Articles Cited by this Article (Reference List)
- Top 5 Articles Citing this Article
- All Articles Citing this Article
- Top 5 Most Co-cited Articles with this Article
- All Articles Co-Cited with this Article

Citebase is currently only an experimental demonstration. Users are cautioned not to use it for academic evaluation yet. Citation coverage and analysis is incomplete and hit coverage and analysis is both incomplete and noisy.

**Summary**

<table>
<thead>
<tr>
<th>Caution!</th>
<th>Citations</th>
<th>Full-text Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>To this Article</td>
<td>1868</td>
<td>114</td>
</tr>
<tr>
<td>To authors (mean)</td>
<td>25,000</td>
<td>25,000</td>
</tr>
</tbody>
</table>

**This Record's Reference List (explain?)**


Top 10 Countries

- United Kingdom: 70.13%
- Belgium: 4.39%
- Italy: 2.63%
- France: 1.75%
- Switzerland: 1.75%
- Ireland: 1.75%
- Mexico: 0.88%
- Poland: 0.88%
- Unknown (IP): 7.02%
- Unknown (net): 5.00%

All Downloads over Time

- 2005-10
- 2005-06
- 2005-05
- 2005-04
- 2005-03
- 2005-02
- 2005-01
- 2004-10
- 2004-08
- 2004-06

- United Kingdom: 80
- Unknown (IP): 8
- Belgium: 5
- Unknown (net): 5
- Italy: 3
- France: 2
- Unknown (com): 2
- Sweden: 2
- Switzerland: 2
- Ireland: 2
Citebase Search: Navigation by Citation Links

Future

Current Article

Related

Co-cited

Past

Article with reference list

Reference link
<table>
<thead>
<tr>
<th>Order</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Large Mass Hierarchy from a Small Extra Dimension</td>
</tr>
<tr>
<td>2</td>
<td>An Alternative to Compactification</td>
</tr>
<tr>
<td>3</td>
<td>This Record's Reference List</td>
</tr>
<tr>
<td>4</td>
<td>Top 5 Records Citing this Record</td>
</tr>
</tbody>
</table>

**A Large Mass Hierarchy from a Small Extra Dimension**

We propose a new higher-dimensional mechanism for solving the Hierarchy Problem. The Weak scale is generated from a large scale of order the Planck scale through an exponential hierarchy. However, this exponential arises not from gauge interactions but from the background metric (which is a slice ... Comment: 9 pages, LaTeX)

**An Alternative to Compactification**

(Randall, Lisa; Sundrum, Raman (1999-05-03) in Physical Review Letters)
Co-cited

The Hierarchy Problem and New Dimensions at a Millimeter
Arkani-Hamed, Nima; Dimopoulos, Savas; Dvali, Gia (1998-03-11) in Physics Letters
We propose a new framework for solving the hierarchy problem which does not rely on either
a large number of extra dimensions or a large number of particles. In our framework, the
gravitational and gauge interactions become unified at the weak scale, while the
distance scale in nature. The observed w... Comment: 16 pages, latex, no figures

A Large Mass Hierarchy from a Small Extra Dimension
Randall, Lisa; Sundrum, Raman (1999-05-03) in Physical Review Letters
1055
We propose a new higher-dimensional mechanism for solving the Hierarchy Problem. The
Weak scale is generated from a large scale of order the Planck scale through an exponential
hierarchy. However, this exponential arises not from gauge interactions but from the
background metric (which is a slice... Comment: 9 pages, LaTeX

Phenomenology, Astrophysics and Cosmology of Theories with Sub-Millimeter Dimensions and TeV Scale Quantum Gravity
Arkani-Hamed, Nima; Dimopoulos, Savas; Dvali, Gia (1998-07-11) in Physical Review D
835
We recently proposed a solution to the hierarchy problem not relying on low-energy
supersymmetry or technicolor. Instead, the problem is nullified by bringing quantum
gravity down to the TeV scale. This is accomplished by the presence of n ≥ 2 new
dimensions of sub-millimeter size, with the SM field... Comment: 51 pages, LaTeX

An Alternative to Compactification
Randall, Lisa; Sundrum, Raman (1999-06-06) in Physical Review Letters
735
Conventional wisdom states that Newton's force law implies only four non-compact
dimensions. We demonstrate that this is not necessarily true in the presence of a
non-factorizable background geometry. The specific example we study is a single 3-brane
embedded in five dimensions. We show that even... Comment: LaTeX, 9 pages

Heterotic and Type I String Dynamics from Eleven Dimensions
Horava, Petr; Witten, Edward (1995-10-29) in Nuclear Physics B
402
A Large Mass Hierarchy from a Small Extra Dimension


We propose a new higher-dimensional mechanism for solving the Hierarchy Problem. The Weak scale is generated from a large scale of order the Planck scale through an exponential hierarchy. However, this exponential arises not from gauge interactions but from the background metric (which is a slice ... Comment: 9 pages, LaTex

An Alternative to Compactification


Conventional wisdom states that Newton's force law implies only four non-compact dimensions. We demonstrate that this is not necessarily true in the presence of a non-factorizable background geometry. The specific example we study is a single 3-brane embedded in five dimensions. We show that even ... Comment: LaTex, 9 pages

Heterotic and Type I String Dynamics from Eleven Dimensions

Horava, Petr; Witten, Edward (1995-10-29) In Nuclear Physics B 460 506 (1996)

We propose that the ten-dimensional $E_8 \times E_8$ heterotic string is related to an eleven-dimensional theory on the orbifold $\mathbb{R}^6 / \mathbb{Z}_2 \times \mathbb{Z}_2$. In the same way that the Type IIA string in ten dimensions is related to $\mathbb{R}^{10} / \mathbb{Z}_2$. This in particular determines the strong coupling behavi ... Comment: 27 pages, 2 postscript figures, harvmac, epsf, a reference added

Large N Field Theories, String Theory and Gravity

Citebase Search: Usage

• 8000 users per-day ("visits")
  – 60,000 hits
• 350,000 full-text records
• 12 million references
  – Of which 2.7 million linked to full-text
• 5 million Web download hits (uk.arXiv.org)
OAI-based Infrastructure

Data Provider → Service Provider (“Harvester”)

Data Provider

XML Records (Dublin Core etc.)
Citebase Search

Repositories → Metadata Harvest (OAI-PMH) → Meta Database

Repositories → Full-text Harvest → References Database

References Database → Citation Database

Citation Database → Web Interface

Emailed Web Logs → Meta Database

Citebase

28/04/2009
Future Citation Ranking

• Citation- and Web- usage by:
  – Article
  – Authors
  – Journal

• Fan-in/Fan-out type rank
  – Hub/Authority (Google PageRank)

• Co-cites
  – Latent semantic analysis

• [All by time-trends]
Citation Indices for OA

• Free to access:
  – Google Scholar
  – Citeseer
  – Astrophysics Data Service (NASA ADS)

• Subscription:
  – Scopus
  – ISI WCI
Analysis: Effect of Open Access

1. Correlation of Citation Impact with Web Impact (downloads)
   • (Based on arXiv.org)
2. Effect of Open Access on citation behaviour
3. OA Impact Advantage
Citation Latency

Paper B cites A

Citation Latency = 10 Months

2002-04-05 2003-02-14
OA Advantage

• (University of Quebec)
• ISI on CD-ROM, records linked to arXiv.org equivalents
  – In arXiv.org = OA
  – Not in arXiv.org = Not OA
Same-Journal Control, No Sample-Size Equalizing, incl. Self-Citations

OA vs. Non-OA Citation Impact Advantage (Physics)

On average 134.0 journals/year (incl. self-citations) - Fri Oct 7 11:30:32 2005
Correlations: $OAA \cdot OAP = 0.157$, $OAP \cdot Year = 0.921$, $OAA \cdot Year = 0.182$
OA Advantage Analysis

• Open Access (OA)
  + Competitive advantage vs. subs-only

• arXiv Advantage (AA)
  + Primary resource for physicists

• Quality Bias (QB)
  + Author self-selection (publish anything, self-archive the best …)

• Early Advantage
  + Cites to pre-print + (later) cites to post-print
Conclusions

• High impact papers are read more (and this can be measured online)
  – Web downloads may be an pre-indicator of impact

• Faster access leads to reducing Citation Latencies
  – Hence faster research cycles, higher impact, and more productivity
Summary

• The Web makes Open Access research literature possible, and hence more effective scholarship
• Services compete without holding the literature hostage
• OAI allows repositories to concentrate on getting and storing the literature
• Citebase Search provides citation navigation for OAI archive(s)
  – Or anyone else who wants to provide a similar service
The Last Slide

• Tim Brody tdb01r@ecs.soton.ac.uk
  – http://opcit.eprints.org/ (papers & presentations)
• Citebase Search
  – http://www.citebase.org/
• EPrints.org
  – (advocacy, answers & software)
  – http://www.eprints.org/

I am a doctoral student in the Intelligence, Agents, Multimedia Group at the University of Southampton working with digital library systems: Citebase Search, E-Prints UK, TARDIS & OAI.
Prof. Stevan Harnad <harnad@ecs.soton.ac.uk>
The Research Literature

- The grey literature
  - Technical reports
  - Monographs
  - Presentations
- Royalty literature
  - Books
- Refereed journal corpus
The Refereed Journal Literature

- Written without the expectation of royalties
- Akin to ‘Advertising’ for authors and their work (=maximise use & uptake)
- Reviewed for free by peers
- Est. 20,000 Peer-reviewed Journals
  - B.L. archives 60,000 serials
- Est. 2,000,000 Articles Annually
Open Access: A Definition

- Immediate, free access via the Web to the author give-away peer-reviewed literature
  1. Author posted pre-prints, post-prints ("e-prints") linked to journal version
  2. Open Access publishing