



eScholarship
University of California



Enhancing eScholarship Discovery with Data for Research

Lisa Schiff, Ph.D.

Technical Lead

eScholarship Publishing Group

California Digital Library

lisa.schiff@ucop.edu

www.escholarship.org



eScholarship Services

Publishing the Work of the University of California

- eScholarship supports:
 - Publication of original journals and books
 - Deposit/Dissemination of:
 - Postprints
 - Preprints
 - Working papers
 - Manuscripts
 - Seminar series
- Currently supported by a vendor (bepress)
- ***Oct 2009: New in-house discovery interface***





Problem: Limited Intellectual Access to eScholarship Content

- 28,000+ scholarly objects from across the range of academic fields with no structured subject access
- Author-supplied metadata only
 - No consistency across objects, within domains or even for a given author
- Browsing only by genre or campus/unit affiliation





Effects on the User

- No ability to winnow large result sets
- No ability to browse across disciplines
- Inconsistency of results across types of searches



Full-Text="achievement gap"

Showing 1 through 25 of 628 matches for: Full Text is achievement gap:

Result Pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [23](#) [24](#) [25](#) [26](#)

↑ Author	Year	Title	Publication
	2001	Sociocultural Factors in Social Relationships: Examining Latino Teachers' and Paraeducators' Interactions with Latino Students	Center for Research on Education, Diversity & Excellence
	2002	Talking Leaves, Vol. 6, No. 2	Center for Research on Education, Diversity & Excellence
	2008	Engaging Contradictions: Theory, Politics, and Methods of Activist Scholarship	GAIA Books
	2008	Reinventing Britain: Constitutional Change under New Labour	GAIA Books
	2002	Educating Hispanic Students: Obstacles and Avenues to Improved Academic Achievement	Center for Research on Education, Diversity & Excellence
	2005	Breathless: Schools, Air Toxics, and Environmental Justice in California	Center for Justice, Tolerance, and Community
SW	2006	A Guide to the Atkinson Selected Papers	The Selected/Works of Richard Atkinson
David Abraham	2002	Citizenship Solidarity and Rights Individualism: On the Decline of National Citizenship in the U.S.	Center for Comparative Immigration Studies



Keywords="achievement gap"

Your query did not match any documents. Please try again.

Simple Search

Advanced Search

Advanced Search

Searching **eScholarship Repository**:

Keyword



includes



"achievement gap"



eScholarship
University of California



Abstract=“achievement gap”

Showing 1 through 3 of 3 matches for: Abstract includes achievement gap:

 <u>Author</u>	<u>Year</u>	<u>Title</u>	<u>Publication</u>
Margot I. Jackson	2007	<u>Understanding Links between Children's Health and Education</u>	<u>California Center for Population Research</u>
Alan Krueger, Jesse M. Rothstein, and Sarah Turner	2006	<u>Race, Income, and College in 25 Years: Evaluating Justice O'Connor's Conjecture</u>	<u>Center for Studies in Higher Education</u>
Hugh Mehan and Lea Hubbard	1999	<u>Tracking Untracking: Evaluating the Effectiveness of an Educational Innovation</u>	<u>Center for Research on Education, Diversity & Excellence</u>



Full Text="achievement gap" AND Subject=education

Showing 1 through 8 of 8 matches for: Full Text is "achievement gap" and Subject Area includes education:

<u>Author</u>	 <u>Year</u>	<u>Title</u>	<u>Publication</u>
Linnea Van Eman, Jerilyn Thorman Ph.D., Diane Montgomery Ph.D., and Stacy Otto Ph.D.	2009	The Balancing Act: Arts Integration and High-Stakes Testing	Journal for Learning through the Arts: A Research Journal on Arts Integration in Schools and Communities
Dolores Vazquez Donet	2008	Review: The Learning Leader: How to Focus School Improvement for Better Results by Douglas B. Reeves	InterActions: UCLA Journal of Education and Information Studies
 Nicki King, Ella R. Madsen, Marc Braverman, Carole Paterson, and Antronette K. Yancey	2008	Career Decisionmaking: Perspectives of Low-Income Urban Youth	Spaces for Difference: An Interdisciplinary Journal
Divya Sunder and Layli Uddin	2007	A Comparative Analysis of Bangladeshi and Pakistani Educational Attainment in London Secondary Schools	InterActions: UCLA Journal of Education and Information Studies





Solution: A Controlled Vocabulary

- Add a layer of intellectual structure to the content by...
 - Applying terms from a controlled vocabulary spanning the entire academy
- Ensure new content has terms assigned from this vocabulary
- Expose the vocabulary to users wherever possible





Challenges

- Which vocabulary?
 - *The National Academies Taxonomy of Fields and their Subfields*
(http://sites.nationalacademies.org/pga/Resdoc/PGA_044522)
- How to quickly apply to 28,000+ documents consistently?
 - *Automatic classification using Kea*





Classification Using Kea

- Kea: a tool for extracting keywords and/or assigning terms from a controlled vocabulary, with or without training data.

<http://www.nzdl.org/Kea/>

- Strategy
 - Manually assigned terms from the vocabulary to a cross-section of eScholarship content to create training and test sets (~800 docs =)
 - Trained Kea and then applied to remaining 27,000+ docs



Manual Term Assignment

eRep Discipline Tagging Tool - Mozilla Firefox

File Edit View History Delicious Bookmarks Yahoo! Tools Help

http://dev.escholarship.org/disciplines/

Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options

Last updated: November 17, 2008

eRep Academic Disciplines Taxonomy Manual Assignment Tool

Step 1: Avoid duplicate tagging! Enter the URL for the file you want to tag:

#Now this form collects the tags for a new file

Step 2: If the file hasn't been tagged, re-enter the URL here:

Step 3: Check all terms from the Academic Disciplines taxonomy below that apply to this file, choosing more precise terms over general terms.

Step 4: After choosing all appropriate terms, click this button:

Academic Disciplines

- Life Sciences
- Medicine and Health Sciences
- Physical Sciences and Mathematics
- Engineering
- Social and Behavioral Sciences
- Arts and Humanities
- Law
- Business
- Architecture
- Education

Done zotero



Huge Time Investment

- ~ 5 min/article to tag
- Training/Test Set
 - ~800 articles
 - ~ 70 hours
- Full Collection would be
 - ~ 28,000 articles
 - ~ 2335 hours or
 - ~ 292 FTE days



Results: Effective Browsing



eScholarship
University of California

[Home](#) [About](#) [Browse](#) [Publish](#) [Help](#) [My Items \(0\)](#)

[Advanced Search](#)

Digital Scholarship and Publishing at the University of California

Narrow results:

Peer-reviewed only

► Affiliation

▼ Discipline

► Arts and Humanities

► Life Sciences

► Law

► Social and Behavioral Sciences

► Architecture

► Engineering

► Medicine and Health Sciences

► Education

► Business

► Physical Sciences and Mathematics

You searched for: achievement gap , []

Displaying Results 1 - 10 of 36

Sort by:

Show:

◀◀ 1 | 2 | 3 | 4 ▶▶

1 **Title:** [Educating Hispanic Students: Obstacles and Avenues to Improved Academic Achievement](#)

Author: none

Published: Center for Research on Education, Diversity and Excellence, 2002

1 Match: ...has been to close the achievement gap between Hispanic and...

2 **Title:** [Review: Educating African American Males: Voices from the Field edited by Olatokunbo S. Fashola](#)

Author: Johnson, Jr., Stanley L.

Published: InterActions: UCLA Journal of Education and Information Studies (1548-3320) 4:1, 2008

6 Matches: ...its relationship to the achievement gap phenomenon (Perry,...
...toward closing the achievement gap of African American male...
...to widening the achievement gap between African American...

3 **Title:** [Understanding Links between Children's Health and Education](#)



eScholarship
University of California

But Mixed Results

Narrow results:

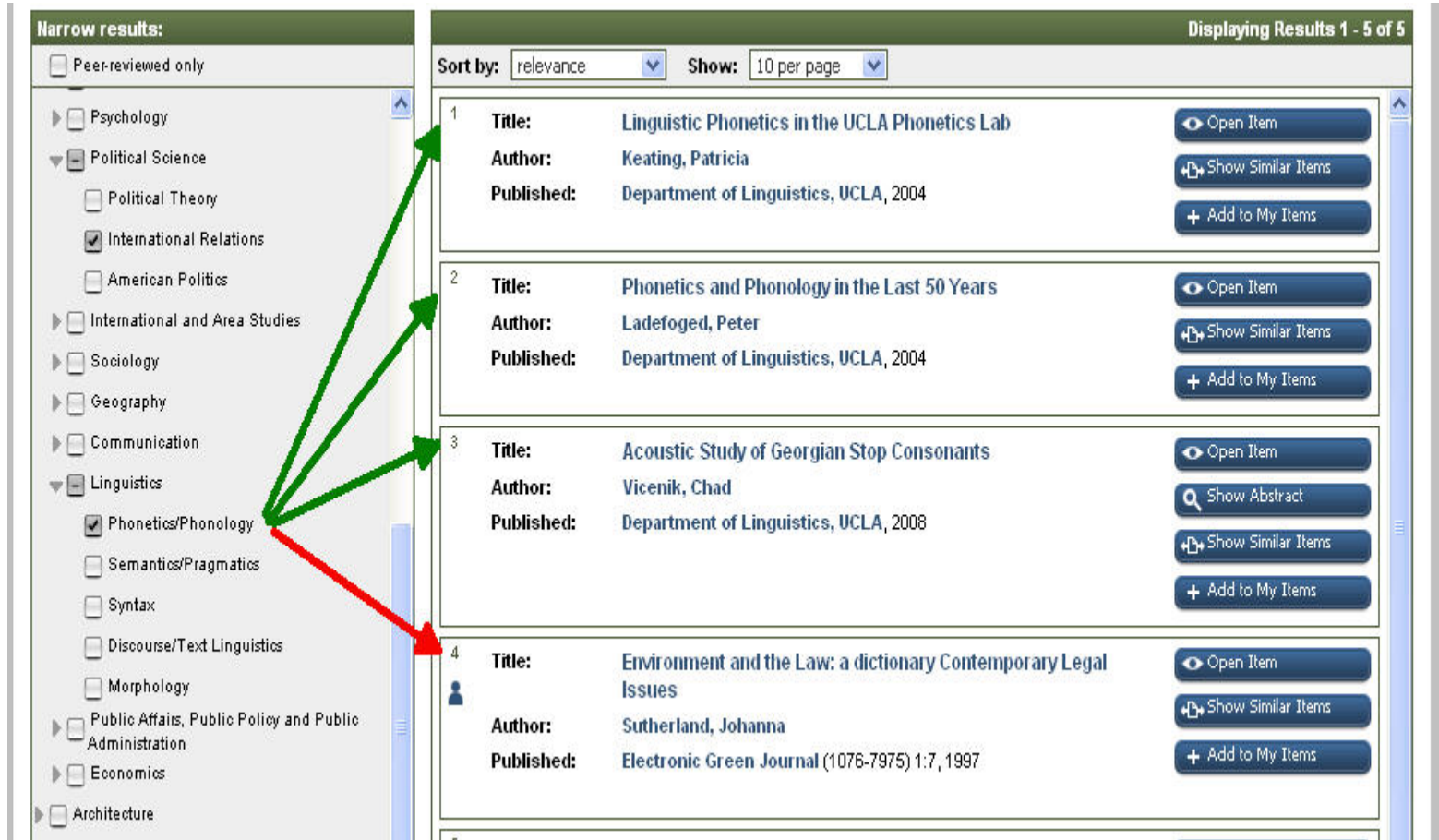
Peer-reviewed only

- Psychology
- Political Science
 - Political Theory
 - International Relations
 - American Politics
- International and Area Studies
- Sociology
- Geography
- Communication
- Linguistics
 - Phonetics/Phonology
 - Semantics/Pragmatics
 - Syntax
 - Discourse/Text Linguistics
 - Morphology
- Public Affairs, Public Policy and Public Administration
- Economics
- Architecture

Sort by: relevance **Show:** 10 per page

Displaying Results 1 - 5 of 5

- Title:** Linguistic Phonetics in the UCLA Phonetics Lab
Author: Keating, Patricia
Published: Department of Linguistics, UCLA, 2004
- Title:** Phonetics and Phonology in the Last 50 Years
Author: Ladefoged, Peter
Published: Department of Linguistics, UCLA, 2004
- Title:** Acoustic Study of Georgian Stop Consonants
Author: Vicenik, Chad
Published: Department of Linguistics, UCLA, 2008
- Title:** Environment and the Law: a dictionary Contemporary Legal Issues
Author: Sutherland, Johanna
Published: Electronic Green Journal (1076-7975) 1:7, 1997





Solution: DfR enriched Training Set

- DfR
 - Academic articles from JSTOR
 - Full-text
 - High quality subject metadata
 - Feasible cross-walk between DfR & eScholarship discipline vocabularies
 - Content exposed through an API





New Classification Strategy

- Identify DfR articles corresponding to eScholarship Disciplines vocabulary
- Programmatically request 10 most recent articles per term to train and test, e.g.
 - discipline=education AND (title or abstract or keywords)="educational psychology" AND type=research-article and lang=eng
 - Total articles required = $240 * 10 = 2,400$
- Represents the full scope of the vocabulary
- No discipline is under/over-trained





Expected Results

- Dramatically improved research experience
 - Consistency quality across disciplines
 - Fewer inaccurate results
 - Consistency within disciplines
 - Better precision for narrower terms





Stay Tuned

Visit www.escholarship.org during
Open Access week (Oct 19th-23rd) to
see results.

Questions?