

Data discovery opportunities: Lessons learned from CORE

September 16, 2019 – OS FAIR, Porto

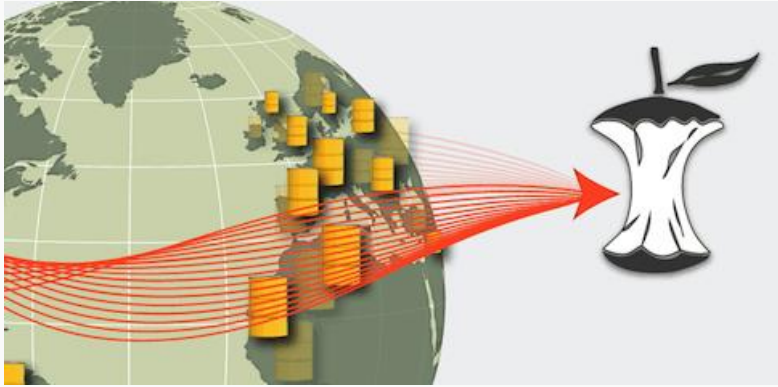
Dr Petr Knoth



Big Scientific Data and Text Analytics Group
Knowledge Media Institute, The Open University



CORE in a nutshell



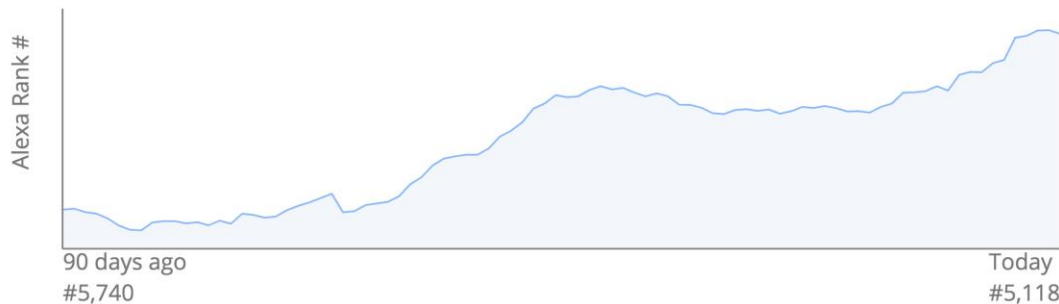
Mission: Aggregate all open access research articles worldwide, enrich this content and provide **seamless access** to it through a set of data services ...

Alexa Rank

This site ranks:
5,100 In global internet traffic and engagement
over the past 90 days

Estimate?

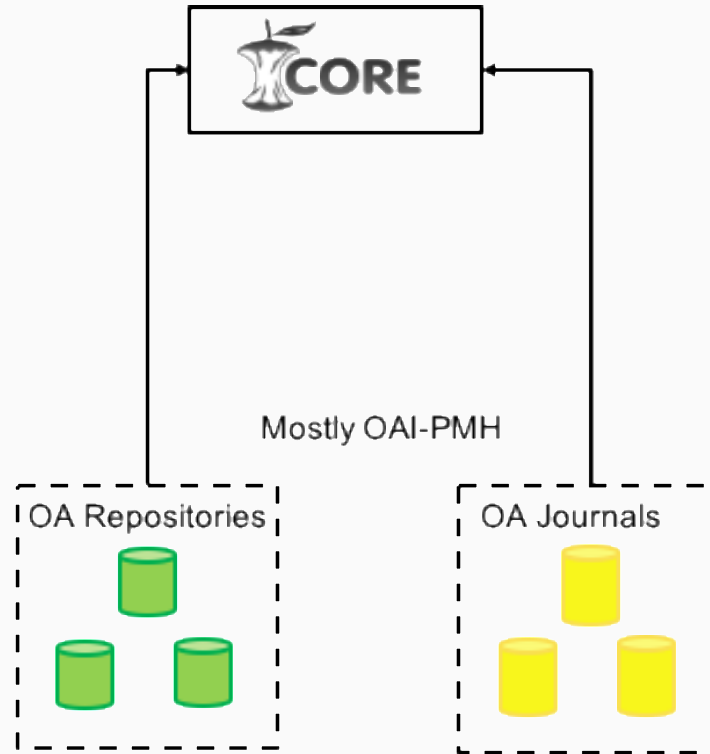
652



- core.ac.uk by usage in the **top 0.0009% of global websites**
- CORE provides access to 24,936,921 free to read full text research papers with **13,380,086 full texts hosted** directly by CORE

Harvesting is a prerequisite for discovery

- Data Registries
- Protocols
- Data providers and consumers willing to work together
- Centralisation \neq interoperability



Global network of repositories

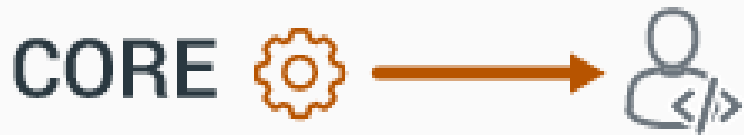
- “A single scientific repository is of limited value, real benefits come from the ability to **exchange data within a network** ...
- ... interoperability allows us to exploit today's computational power so that we can **aggregate, data mine, create new tools and services, and generate new knowledge** from repository content.”
 - Confederation of Open Access Repositories (COAR)

Lesson 1

Argument: The best idea about how to use your data might not be invented by you.

Action: We need programmable access to data.

What has CORE done to help. A widely popular raw data services. CORE API, CORE Dataset and CORE FastSync enabling others to build new applications on top of CORE.



Lesson 2

Argument: Discovery is about adding value where the users are.

Action: Need to bring features to users instead of users to features.

What has CORE done in this space: Integration with third party systems and discovery services:


- Microsoft Academic
- PMC
- institutional repositories
- CORE Discovery browser extension
- Lean Library
- ...

CORE integration in Microsoft Academic

Microsoft Academic

Differential effects of nitrogen and sulfur deprivation on growth and biodiesel feedstock production of Chlamydomonas reinhardtii

✕ 🔍 “ Sign up / Sign in




Differential effects of nitrogen and sulfur deprivation on growth and biodiesel feedstock production of Chlamydomonas reinhardtii

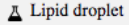
2012 BIOTECHNOLOGY AND BIOENGINEERING Volume: 109, Issue: 8, pp 1947-1957 DOI: 10.1002/bit.24474

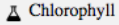
Turgay Cakmak^{1, 2}, Pinar Angun², Yunus Emre Demiray², Alper Devrim Ozkan² +2

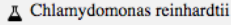
¹ Istanbul Medeniyet University, ² Bilkent University, ³ Kırıkkale University

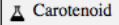
Biodiesel production from microalgae is a promising approach for energy production; however, high cost of its process limits the use of microalgal biodiesel. Increasing the levels of triacylglycerol (TAG) levels, which is used as a biodiesel feedstock, in microalgae has been achieved mainly by nitrogen starvation. In this study, we compared effects of sulfur (S) and nitrogen (N) starvation on TAG accumulation and related parameters in wild-type Chlamydomonas reinhardtii CC-124 mt(+) and CC-125 mt(-) strains. Cell division was interrupted, protein and chlorophyll levels rapidly declined while cell volume, total neutral lipid, carotenoid, and carbohydrate content increased in response to nutrient starvation. Cytosolic lipid droplets in microalgae under nutrient starvation were monitored by three-dimensional confocal laser imaging of live cells. Infrared spectroscopy results showed that relative TAG, oligosaccharide and polysaccharide levels increased rapidly in response to nutrient starvation, especially S starvation. Both strains exhibited similar levels of regulation responses under mineral deficiency, however, the degree of their responses were significantly different, which emphasizes the importance of mating type on the physiological response of algae. Neutral lipid, TAG, and carbohydrate levels reached their peak values following 4 days of N or S starvation. Therefore, 4 days of N or S starvation provides an excellent way of increasing TAG content. Although increase in these parameters was followed by a subsequent decline in N-starved strains after 4 days, this decline was not observed in S-starved ones, which shows that S starvation is a better way of increasing TAG production of C. reinhardtii than N starvation. Biotechnol. Bioeng. 2012;109: 1947-1957.

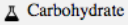
 Nutrient


 Lipid droplet

 Chlorophyll

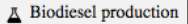
 Chlamydomonas reinhardtii

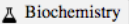
 Carotenoid


 Carbohydrate

 Botany

 Biology


 Biodiesel production

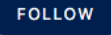
 Biochemistry


 Algae


View PDF: yoksis.bilkent.edu.tr


Website(s): core.ac.uk, onlinelibrary.wiley.com, cabdirect.org, ncbi.nlm.nih.gov


 GET FROM CORE


 FOLLOW


 CLAIM

 “





 41
References

 173
Citations*

Integrated full text discovery

Lesson 3

Argument: There are different types of discovery use cases.
Discovery \neq just search.

Action: Develop functionality for a range of discovery use cases

What has CORE done in this space:

A suite of services:

- CORE Search
- CORE Recommender
- CORE Discovery
- Supporting others

CORE Recommender

The screenshot shows the Apollo Open Content Mining interface. At the top, the University of Cambridge logo is on the left, and navigation links (Apollo Home, School of the Physical Sciences, Department of Chemistry, Unilever Centre for Molecular Informatics, Panton Discussions, View Item) are on the right. Below the navigation bar is the Apollo logo and a search box labeled 'Search Apollo' with an 'Advanced search' link. The main content area is titled 'Open Content Mining'. On the left, a sidebar contains a 'Browse' section with links to 'All of Apollo' (Communities & Collections, Authors, Titles, Keywords, Type) and 'This Collection' (Authors, Titles, Keywords, Type). Below this is a 'Statistics' section with a link to 'View Usage Statistics'. The main content area displays the record details for 'Open Content Mining'. It includes a thumbnail image of the document cover, a 'Citation' section with the text 'Murray-Rust, P. (2012). Open Content Mining.', a 'Description' section with the text 'Conference for the Fellows of OpenForum Academy - 24th September 2012 Brussels', an 'Abstract' section with the text 'Abstract— We present evidence that content-mining of scholarly articles is now technically feasible and highly valuable both. However researchers and information technologists are blocked by legal and contractual barriers from using it and developing the methodologies. We review the problems and propose changes in legal policy which we have already submitted to the UK's Hargreaves report on intellectual property reform. We put forward the fundamental rights of scholars and embed them in a manifesto: "The right to read is the right to mine", "Users and providers should encourage machine processing, and "Facts don't belong to anyone".', a 'Keywords' section with the text 'Open Content Mining, Index Terms—Open Knowledge, Content mining, Hargreaves process, Text mining, publishers, legal barriers', and an 'Identifiers' section with the text 'This record's URL: http://www.dspace.cam.ac.uk/handle/1810/243749'. Other fields include 'View / Open Files' (article text (PDF, 9Mb)), 'Authors' (Murray-Rust, Peter), 'Publication Date' (2012-09-24), 'ISBN' (to be assigned), and 'Language' (English).

UNIVERSITY OF CAMBRIDGE

Apollo Home / School of the Physical Sciences / Department of Chemistry / Unilever Centre for Molecular Informatics / Panton Discussions / View Item

Search Apollo
Advanced search

Browse

All of Apollo

- > Communities & Collections
- > Authors
- > Titles
- > Keywords
- > Type

This Collection

- > Authors
- > Titles
- > Keywords
- > Type

Statistics

- > View Usage Statistics

Open Content Mining

Citation
Murray-Rust, P. (2012). Open Content Mining.

Description
Conference for the Fellows of OpenForum Academy - 24th September 2012 Brussels

Abstract
Abstract— We present evidence that content-mining of scholarly articles is now technically feasible and highly valuable both. However researchers and information technologists are blocked by legal and contractual barriers from using it and developing the methodologies. We review the problems and propose changes in legal policy which we have already submitted to the UK's Hargreaves report on intellectual property reform. We put forward the fundamental rights of scholars and embed them in a manifesto: "The right to read is the right to mine", "Users and providers should encourage machine processing, and "Facts don't belong to anyone".

Keywords
Open Content Mining, Index Terms—Open Knowledge, Content mining, Hargreaves process, Text mining, publishers, legal barriers

Identifiers
This record's URL:
<http://www.dspace.cam.ac.uk/handle/1810/243749>

View / Open Files
[article text \(PDF, 9Mb\)](#)

Authors
Murray-Rust, Peter

Publication Date
2012-09-24

ISBN
to be assigned

Language
English

- Recommending relevant content to users from across all free content
- Recommender plugin for repositories
- <https://core.ac.uk/services/recommender/>

CORE Recommender


Type
Conference Object


Metadata
[Show full item record](#)


Rights
Attribution 2.0 UK: England & Wales
Licence URL:
<http://creativecommons.org/licenses/by/2.0/uk/>


Recommended or similar items


Suggested articles Suggested articles in Apollo


 **Effectively and Efficiently Mining Frequent Patterns from Dense Graph Streams on Disk**
Provided by: Elsevier - Publisher Connector | **Publisher:** The Authors. Published by Elsevier B.V. | **Year:** 2014
By Braun Peter, Cameron Juan J., Cuzzocrea Alfredo, Jiang Fan, Leung Carson K.

 **The right to read is the right to mine: Text and data mining copyright exceptions introduced in the UK.**
Provided by: LSE Research Online | **Publisher:** London School of Economics and Political Science | **Year:** 2014
By Mounce Ross

 **Global boom, local impacts: Mining revenues and subnational outcomes in Peru 2007-2011**
Provided by: EconStor | **Publisher:** Washington, DC: Inter-American Development Bank (IDB) | **Year:** 2014
By Zambrano Omar, Robles Marcos, Laos Denisse

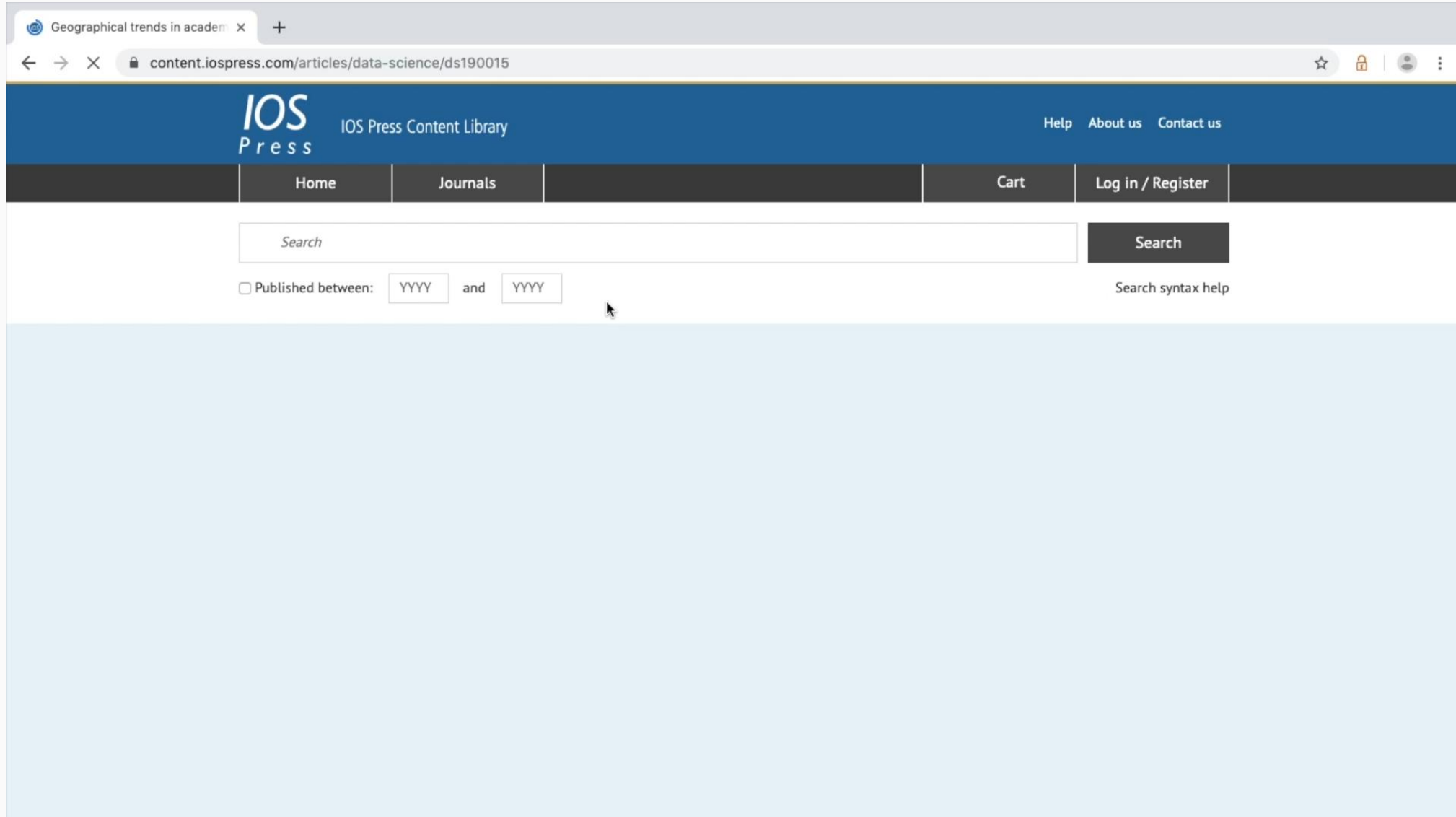
 **Environmental security, mining and good governance : mining regulation in the Kyrgyz region. A review**
Provided by: UEF Electronic Publications | **Publisher:** University of Eastern Finland
By Honkonen T

 **A Case Study of Data Analysis Process and Tools for a Consulting Company**
Provided by: Aaltodoc Publication Archive | **Year:** 2012
By Gong Peng

Powered by 

- Recommending relevant content to users from across all free content
- Recommender plugin for repositories
- <https://core.ac.uk/services/recommender/>

CORE Discovery demonstration



The screenshot displays the IOS Press Content Library website. The browser's address bar shows the URL `content.iospress.com/articles/data-science/ds190015`. The website's header features the IOS Press logo and navigation links for Help, About us, and Contact us. A dark navigation bar contains links for Home, Journals, Cart, and Log in / Register. Below this, a search bar with the placeholder text "Search" is accompanied by a "Search" button. A filter section includes a checkbox for "Published between:" followed by two "YYYY" input fields and an "and" connector. A "Search syntax help" link is also present. The main content area is a large, empty light blue rectangle.

Geographical trends in academi x +

content.iospress.com/articles/data-science/ds190015

IOS Press IOS Press Content Library

Help About us Contact us

Home Journals Cart Log in / Register

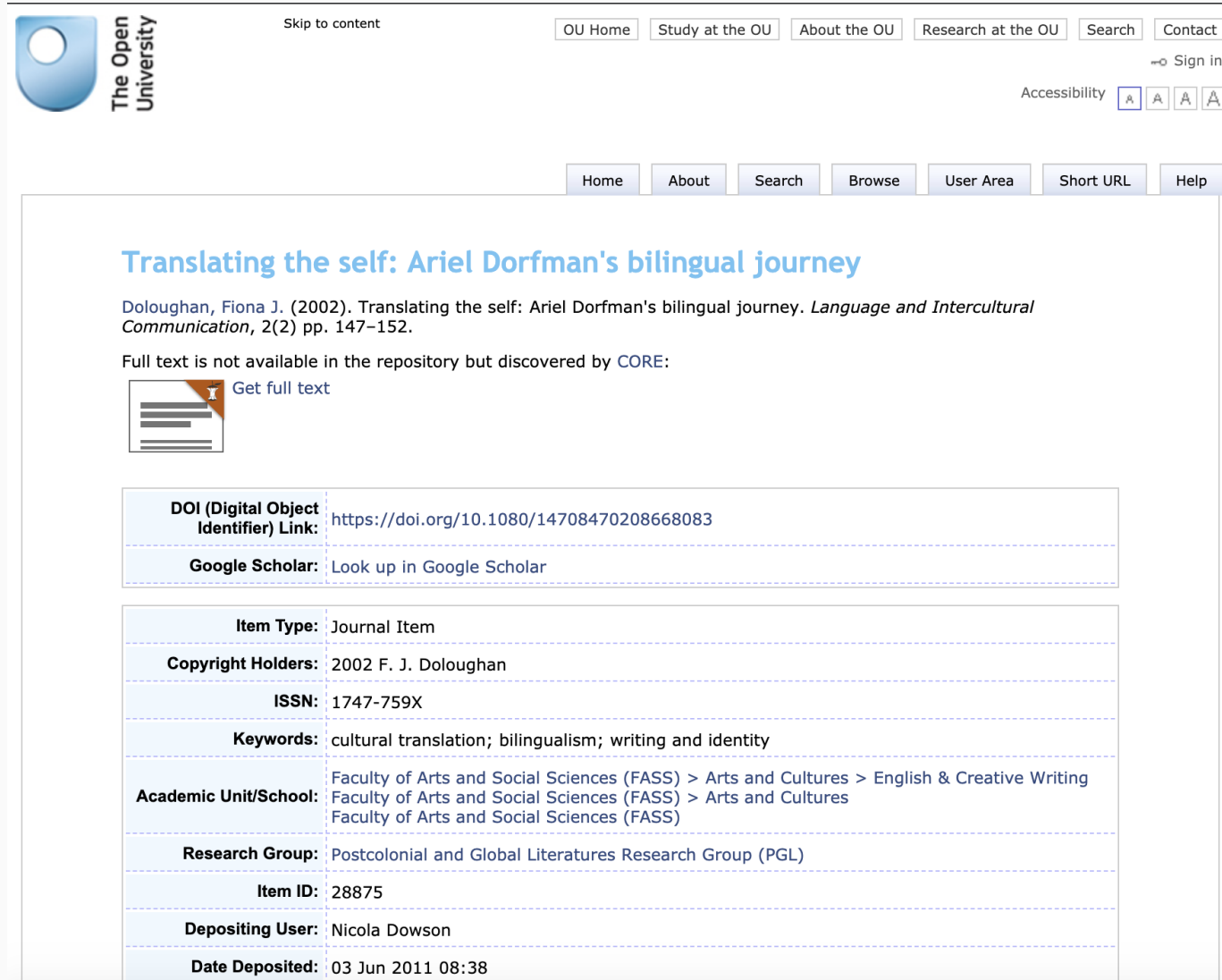
Search

☐ Published between: YYYY and YYYY

Search syntax help

CORE Discovery Repository integration

- Majority of articles in repositories metadata only.
- CORE Discovery repository plugin:
 - turns dead ends of user journeys into journeys fulfilling users' information needs
 - makes repository content more discoverable.



The screenshot displays the Open University's digital repository interface. At the top, there is a navigation bar with links for 'OU Home', 'Study at the OU', 'About the OU', 'Research at the OU', 'Search', and 'Contact'. A 'Skip to content' link is also present. Below the navigation bar, a secondary set of links includes 'Home', 'About', 'Search', 'Browse', 'User Area', 'Short URL', and 'Help'. The main content area features the title 'Translating the self: Ariel Dorfman's bilingual journey' in blue. Below the title, the citation is provided: 'Doloughan, Fiona J. (2002). Translating the self: Ariel Dorfman's bilingual journey. *Language and Intercultural Communication*, 2(2) pp. 147–152.' A note states 'Full text is not available in the repository but discovered by CORE:' followed by a 'Get full text' button and a document icon. Below this, there are two tables of metadata. The first table contains links for the DOI and Google Scholar. The second table provides detailed metadata about the item, including its type, copyright, ISSN, keywords, academic unit, research group, item ID, depositing user, and date deposited.

Skip to content

OU Home Study at the OU About the OU Research at the OU Search Contact

Sign in


Accessibility

Home About Search Browse User Area Short URL Help

Translating the self: Ariel Dorfman's bilingual journey

Doloughan, Fiona J. (2002). Translating the self: Ariel Dorfman's bilingual journey. *Language and Intercultural Communication*, 2(2) pp. 147–152.

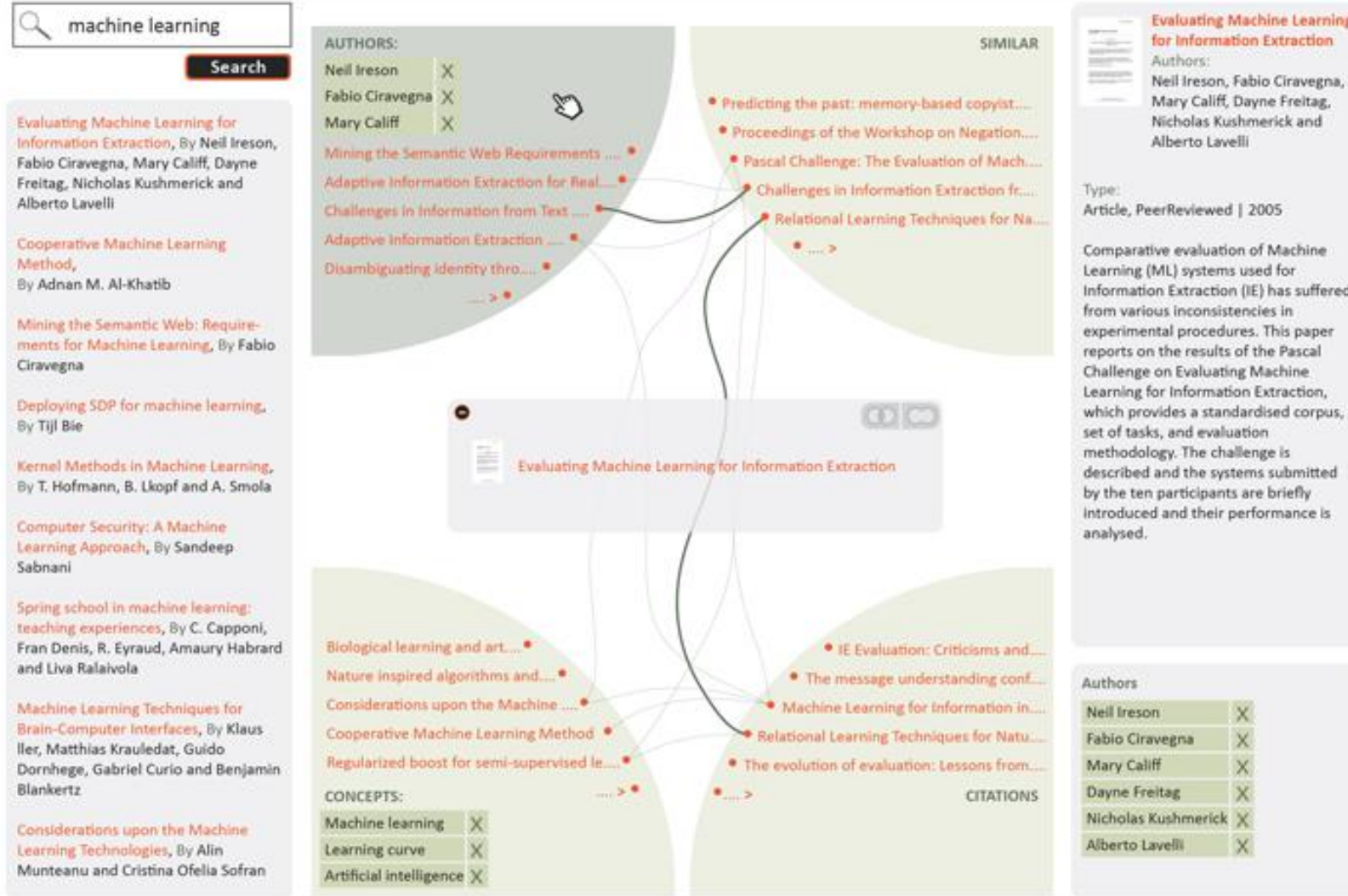
Full text is not available in the repository but discovered by CORE:

 [Get full text](#)

DOI (Digital Object Identifier) Link:	https://doi.org/10.1080/14708470208668083
Google Scholar:	Look up in Google Scholar

Item Type:	Journal Item
Copyright Holders:	2002 F. J. Doloughan
ISSN:	1747-759X
Keywords:	cultural translation; bilingualism; writing and identity
Academic Unit/School:	Faculty of Arts and Social Sciences (FASS) > Arts and Cultures > English & Creative Writing Faculty of Arts and Social Sciences (FASS) > Arts and Cultures Faculty of Arts and Social Sciences (FASS)
Research Group:	Postcolonial and Global Literatures Research Group (PGL)
Item ID:	28875
Depositing User:	Nicola Dowson
Date Deposited:	03 Jun 2011 08:38

Visual discovery interfaces



Herrmannova, D. and Knoth, P. (2012) [Visual Search for Supporting Content Exploration in Large Document Collections](#), D-Lib Magazine, 18, 7/8, Corporation for National Research Initiatives

Thank you!

<https://core.ac.uk>