



Implementing the CORE Recommender in Strathprints: a “whitehat” improvement to promote user interaction

by George Macgregor, Institutional Repository Coordinator, University of Strathclyde

This guest blog post briefly reviews why the CORE Recommender was quickly adopted on Strathprints and how it has become a central part of our quest to improve the interactive qualities of repositories.

Back in October 2016 my colleagues at the CORE Team released their Recommender plugin. The CORE Recommender plugin can be installed on repositories and journal systems to recommend similar scholarly content. On this very blog, Nancy Pontika, Lucas Anastasiou and Petr Knoth, announced the release of the Recommender as a:

...great opportunity to improve the functionality of repositories by unleashing the power of recommendation over a huge collection of open-access documents, currently 37 million metadata records and more than 4 million full-text, available in CORE.*

(Note from CORE Team: the up-to-date numbers are 80,097,014 metadata and 8,586,179 full-text records.)*

When the CORE Recommender is deployed a repository user will find that as they are viewing an article or abstract page within the repository, they will be presented with recommendations for other related research outputs, all mined

from CORE. The Recommender sends data about the item the user is visiting to CORE. Such data include any identifiers and, where possible, accompanying metadata. The CORE response to the repository then delivers CORE’s content recommendations and a list of suggested related outputs are presented to the user in the repository user interface. The algorithm used to compute these recommendations is described in [the original CORE Recommender blog post](#) but is ultimately based on content-based filtering, citation graph analysis and analysis of the semantic relatedness between the articles in the CORE aggregation. It is therefore unlike most standard recommender engines and is an innovative application of open science in repositories.

Needless to say, we were among the first institutions to proudly implement the CORE Recommender on our EPrints repository. The implementation was on [Strathprints](#), the University of Strathclyde’s institutional repository, and was rolled out as part of some wider work to improve repository visibility and web impact. The detail of this other work can be found in [a poster presented at the 2017 Repository Fringe Conference](#) and [a recent blog post](#).

Make pages primarily for users, not for search engines

Enhancing the visibility and web impact of repository content necessitates numerous improvements to the technical configuration of repository pages and to the wider repository platform. Such changes are designed to optimise the repository for crawling / indexing by search engines or other academic search services and are often referred to as “whitehat” improvements. Whitehat changes are those that have the potential to improve search performance whilst simultaneously maintaining the integrity of the repository and keeping within the boundaries of what search services consider “good practice”. But not all whitehat changes are necessarily valued by these search services. For example, consider Google’s most important [indicator of relevance in search](#): *“Provide high-quality content on your pages [...] this is the single most important thing to do. If your pages contain useful information, their content will attract many visitors and entice webmasters to link to your site.”*

Google then goes on to recommend that if websites want to be easily discovered through their searches, they should ensure their websites are:

- More valuable and useful than other sites
- More useful and informative
- Of a high quality, and
- Provide engaging experiences for their users by interacting with them

Content and engagement is therefore king because these are the things that users ultimately care about. And look at the inclusion guidelines for most search services and you will discover that they provide very similar recommendations.

“Horizontal” information seeking behaviour is characterised by users’ need to skip across multiple, disparate sites in order to gather the information they need to complete their task. Scholarly research work, especially literature searching, not only encourages this horizontal information seeking but is itself an essential part of the scholarly research process; moving from site to site, from page to page, following links and citations wherever they may take you. By exposing their content to a wide variety of search services, and owing to the nature of repository content itself, repositories encourage – and are conducive to – “horizontal” information seeking strategies. They are, after all, created to serve scholarly content. The issue facing most repositories therefore is that this sort of information seeking also results in high “bounce rates” – the rate at which visitors to a website leave the site after viewing only one page.

But if bounces are an inevitable consequence of scholarly information seeking strategies, what if a repository could be demonstrated to be, say, more useful than other scholarly sites, or provide more in the way of user interaction and engagement, might it then be possible retain the user on the repository for longer? The answer to this question is, of course, “yes” – and one of the ways it can be achieved is through the CORE Recommender.

Developing methodologies for large scale wave and tidal stream marine renewable energy extraction and its environmental impact : an overview of the TeraWatt project

Side, J. and Gallego, A. and James, M. and Davies, I. and Heath, M. and Karunarathna, H. and Venugopal, V. and Vögler, A. and Burrows, M. (2017) *Developing methodologies for large scale wave and tidal stream marine renewable energy extraction and its environmental impact : an overview of the TeraWatt project* *Ocean and Coastal Management*, 147. pp. 1-5. ISSN 0964-5691



Text (Side-etal-OCM2017-large-scale-wave-and-tidal-stream-marine-renewable-energy-extraction) Side_etal_OCM2017_large_scale_wave_and_tidal_stream_marine_renewable_energy_extraction.pdf - Final Published Version
License: **OGL**
[Download \(771kB\)](#) | [Preview](#)

Official URL: <http://dx.doi.org/10.1016/j.ocecoaman.2016.11.015>

Abstract

This article presents an introduction and overview of a modelling project “TeraWatt: Large scale interactive coupled 3D modelling for wave and tidal energy resource and environmental impact”. The project was funded by the Sustainable Power GENERation (SUPERGEN) “Marine Hub”. SUPERGEN (<http://www.rcuk.ac.uk/research/xrcprogrammes/energy/energyresearch/supergen>) is an initiative of the Research Councils UK (RCUK) Energy Programme, led by the Engineering and Physical Sciences Research Council (EPSRC). It is focused on supporting resilient and sustainable generation, supply, transmission and storage of the UK’s energy. The Marine Challenge addresses the UK Centre for Marine Energy Research’s (UKCMER) mission to support the wave and tidal energy sector beyond their current state of development.

Item type: Article
ID code: 59521
Notes: (c) Crown Copyright [Open Government Licence - OGL3] renewable energy, tidal power, wave power, pentland firth, marine renewable energy, tidal stream.
Keywords: environmental impact, energy extraction, Environmental Sciences, Hydraulic engineering, Ocean engineering, Energy(all), Ocean Engineering
Subjects: Geography Anthropology Recreation > Environmental Sciences Technology > Hydraulic engineering Ocean engineering
Department: Faculty of Science > Mathematics and Statistics
Depositing user: Pure Administrator
Date deposited: 20 Jan 2017 14:25
Last modified: 18 Oct 2017 11:00
Related URLs: Journal or Publication
URI: https://strathprints.strath.ac.uk/id/eprint/59521
Export data: <input type="text" value="RDF+XML"/> <input type="button" value="Export"/>

Suggested articles Suggested articles in University of Strathclyde Institutional Repository

Developing methodologies for large scale wave and tidal stream marine renewable energy extraction and its environmental impact : an overview of the TeraWatt project
Provided by: University of Strathclyde Institutional Repository | Year: 2016
By Side J., Gallego A., James M., Davies I., Heath M., Karunarathna H., Venugopal V., Vögler A., Burrows M.

National Policy Framework for Marine Renewable Energy within the United Kingdom



Quick search

[Advanced search](#) -- [Help](#)

Browse research content

- By author or creator
- By year
- By subject
- By department or faculty
- By journal or other publication

Explore Strathprints

- [Strathprints - home](#)
- [Latest additions](#)
- [Atom](#) [RSS 1.0](#) [RSS 2.0](#)
- [About Strathprints](#)
- [Open Access @ Strathclyde](#)
- [Usage statistics](#)
- [Follow @ Tumblr -- Twitter](#)

See more details

- Tweeted by 2
- 12 readers on Mendeley
- 0 readers on Connotea
- 0 readers on CiteULike

Contact us

Open Access enquiries: openaccess@strath.ac.uk
Repository enquiries: strathprints@strath.ac.uk

[Follow @strathclydeos](#) 516 followers

Strathprints metadata page with the CORE Recommender

The CORE Recommender was configured to generate content recommendations at the foot of the Strathprints abstract pages (see screen above). Those familiar with recommendation engines will understand, in broad terms, what they attempt to deliver, but from the screen snippets below it can be observed clearly that some of the following signals are computed to provide meaningful

suggestions:

- The recommended article reports on a similar topic, perhaps even matching some of the prominent concepts within the recommendations;
- Users who read this article on the repository were also interested in reading a particular recommended article;
- The recommended article provides an influential citation to the article the user has discovered on the repository;
- The recommended article is often co-cited with this article.

The screenshot displays the Strathprints repository interface. At the top, it shows the last modified date (18 Oct 2017 11:00) and related URLs, including a specific URI: <https://strathprints.strath.ac.uk/id/eprint/59521>. Below this, there is an 'Export data' section with a dropdown menu set to 'RDF+XML' and an 'Export' button. The main content area is titled 'Suggested articles' and lists five recommended articles, each with a thumbnail, title, provider, publisher, year, and author information. The sidebar on the right includes a 'Contact us' button, open access enquiries email (openaccess@strath.ac.uk), repository enquiries email (strathprints@strath.ac.uk), and a Twitter follow button for @strathclydeoa with 516 followers. At the bottom, there is a 'Powered by CORE' logo and a navigation bar with 'Open Access' and 'Strathprints home' links.

Strathprints configuration of the CORE Recommender

Remember too that the CORE Recommender allows users to drill recommendations with greater specificity by allowing them to explore recommendations from within the very repository the user is visiting and driving additional traffic to Strathprints via CORE. For such recommendations the Recommender also attempts to understand whether:

- The recommended article is related to the reference article and comes from the same repository;
- The recommended article shares a common author with the reference article;

The screen snippet below illustrates how toggling between the Recommender tabs reveals Recommendations from within the same repository whilst using all other aspects of the CORE Recommender Algorithm.

The screenshot displays a web interface for suggested articles. At the top, there are two tabs: "Suggested articles" (selected) and "Suggested articles in University of Strathclyde Institutional Repository". Below the tabs, five article cards are listed, each with a thumbnail, title, source, year, and authors. The articles are:

- Developing methodologies for large scale wave and tidal stream marine renewable energy extraction and its environmental impact : an overview of the TeraWatt project**
Provided by: University of Strathclyde Institutional Repository | Year: 2016
By Side J., Gallego A., James M., Davies I., Heath M., Karunarathna H., Venugopal V., Vögler A., Burrows M.
- Large scale three-dimensional modelling for wave and tidal energy resource and environmental impact : methodologies for quantifying acceptable thresholds for sustainable exploitation**
Provided by: University of Strathclyde Institutional Repository | Year: 2016
By Gallego A., Side J., Baston S., Waldman S., Bell M., James M., Davies I., O'Hara Murray R., Heath M., Sabatino A., McKee D., McCaig C., Karunarathna H., Fairley I., Chatzirodou A., Venugopal V., Nemalidinne R., Yung T.Z., Vögler A., MacIver R., Burrows M.
- Economic and socio-economic assessment methods for ocean renewable energy : public and private perspectives**
Provided by: University of Strathclyde Institutional Repository | Year: 2015
By Dalton Gordon, Allan Grant, Beaumont Nicola, Georgakaki Alikei, Hacking Nick, Hooper Tara, Kerr Sandy, O'Hagan Anne Marie, Reilly Kieran, Ricci Pierpaolo, Sheng Wenan, Stallard Tim
- Modelling the sensitivity of suspended sediment profiles to tidal current and wave conditions**
Provided by: University of Strathclyde Institutional Repository | Year: 2016
By Heath Michael, Sabatino Alessandro, Serpetti Natalia, McCaig Chris, O'Hara Murray Rory
- Energy storage in the UK electrical network : estimation of the scale and review of technology options**
Provided by: University of Strathclyde Institutional Repository | Year: 2010
By Wilson Ian Allan Grant, McGregor Peter G., Hall Peter J.

At the bottom right of the interface, there is a logo that says "Powered by CORE".

Toggling between CORE Recommender tabs

Following the introduction of some other [technical improvements documented here](#), we performed a little before and after analyses of Strathprints which provided some encouraging results. But we also used the work as an opportunity to examine the potential influence of whitehat improvements on user engagement.

We discovered that the average time users spent on Strathprints upon arrival increased considerably. 01:29 up from 00:58 – so users typically spent 58% longer on Strathprints, strongly indicative that improvements to the user interface and especially the introduction of the CORE Recommender, was enough to persuade users to defer their bounce and instead read content, or explore alternative content within the CORE Recommender.

In other words, it was possible to improve the “dwell time” of Strathprints repository users and this, in a roundabout way, actually helps to improve search engine visibility. But why?

This is because focusing on bounce rates is not a reliable metric, especially given what we already know about horizontal information seeking and repositories. Indeed, we found that the Strathprints bounce rate only decreased slightly after our changes were implemented. But experts are more aware than ever that “dwell time” is perhaps a more important metric than bounce rate—and dwell time is certainly critical to understanding repository engagement. For example, a user might spend 25 mins reading content on your repository, taking notes and chaining references, but then they might leave. In bounce rate terms that user has “bounced” because they failed to navigate to another page on the repository. Yet, the user spent 25 mins consuming repository content and clearly found that content useful. It is for this reason that many search services, Google included, now factor in “dwell time”. Like PageRank more generally, we do not know quite how Google calculates dwell time or what weighting it is assigned in calculating PageRank, but we do know it is important.

So, the ability of a repository to increase dwell time is critical to repository discovery and can only be achieved by providing high quality content (which repositories do anyway because they serve scholarly content!) and creating interaction opportunities on the repository, something which the CORE

Recommender clearly promotes.

Concluding thoughts

The role of the CORE Recommender in promoting a 58% increase in user dwell time on Strathprints has been an important component of our wider repository visibility and web impact strategy. And from our more detailed analyses we can observe that the Recommender has also led to an increase in traffic from CORE to Strathprints too (via the Recommender). Of course, the CORE Recommender can only get better, bringing additional benefits to Strathprints. CORE intend to continually enhance the Recommender and its Algorithm. They are constantly re-evaluating it and refining it and are continually introducing additional signals and sources, such as readership, citation count, popularity and research area topics, thereby delivering even better, more successful recommendations.

□ 31st October 2017 □ nancyontika CORE, recommender, repositories □ open access, recommendation systems, recommender, statistics, strathclyde university, strathprints, tdm, text and data mining

PREVIOUS

CORE reaches a new milestone: 75 million metadata and 6 million full text

NEXT

CORE visits Ethiopia and participates in an Open Science training session

Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 4.0 International license .



Proudly powered by WordPress