

Subject Re: [Sigmetrics] 1findr: research discovery & analytics platform
From David Wojick <dwojick@craigellachie.us>
Sender SIGMETRICS <sigmetrics-bounces@asist.org>
To Éric Archambault <eric.archambault@science-metrix.com>
Cc sigmetrics@mail.asis.org <sigmetrics@mail.asis.org>
Date 25-04-2018 13:14

Yes, Eric, I think that language analysis is potentially far superior to citation analysis, now that full text is available. Citations were great when they were all we had. In addition to no lag, language is more objective than citations. Authors can often choose who to cite or not, but specific language is necessary in order to state their results.

By the same token, language use is probably a better indicator of the extent and evolution of research areas than subjectively defined and applied discipline categories. New ideas almost always require new language, new phrases at least, sometimes even new words. As I like to put it, the science frontier is a language frontier. The words we use have the meanings they do because we are trying to say what is true. Much follows from this.

David

On Apr 25, 2018, at 6:49 AM, Éric Archambault<eric.archambault@science-metrix.com> wrote:

Good points David.

For recommenders, using bibliographic coupling is among the most powerful tools and articles need not use the same vocabulary. There has been much emphasis placed on co-citation in the bibliometric/scientometric community but bibliographic coupling is extremely powerful and way more convenient (stable and no lag contrary to co-citation analysis which needs to way for citations to materialize and then is always evolving as the citation graph builds up).

Again, I don't mean to say you're not right to emphasize the need for full-text searching, but in many cases there are workaround. One case where full-text is more particularly useful is corpus building for e.g. text/data mining and literature related discovery (LRD) studies.

Éric

Eric Archambault, PhD

CEO | Chef de la direction

C. 1.514.518.0823

eric.archambault@science-metrix.com

science-metrix.com & 1science.com

From: David Wojick <dwojick@craigellachie.us>
Sent: April-24-18 5:41 PM
To: Éric Archambault <eric.archambault@science-metrix.com>
Cc: sigmetrics@mail.asis.org
Subject: Re: [Sigmetrics] 1findr: research discovery & analytics platform

I agree up to a point, Eric. Metadata (especially including an abstract) is usually sufficient for what we might call a standard search. This is one where what we are looking for is the central topic of the article.

But there are many other sorts of search, where the thing sought is relatively secondary to the article and here only full text search works. Examples might include those climate change articles that rely on a specific model, or nuclear physics that uses the Monte Carlo method. A great many questions of this form can arise, in research and in science metrics.

Then too there is the powerful "more like this" (MLT) function which requires full text. This finds closely related research that does not use the same language. An example is author disambiguation versus name identity. Google Scholar's version of MLT is very useful.

In fact I developed an algorithm for DOE OSTI that uses "more like this" technology to find all and only those articles closely related to a given topic, ranked by closeness. When you get full text will be happy to show it to you.

But the fact that your system does not do everything now is not a criticism, merely a direction for possible progress.

Best of luck,

David

On Apr 24, 2018, at 4:18 PM, Éric Archambault <eric.archambault@science-metrix.com> wrote:

David,

Thanks for your encouraging comments. You are right, we don't do full text indexing search – yet. We want to get there though as a bibliometrician I have always been a tad skeptical about the need to go much beyond high quality metadata. When you can't find a paper and you have title, journal, abstract, references/citations, chances are the paper won't be all that sharp for most of the mainstream applications. Probably the term is not that key if it can't be found anywhere in the metadata. I'm not saying there are absolutely no cases for searching in the metadata but most of the people want sharp results, and though we are all impressed by zillions of results, we rarely if ever use the long tail. This only became stronger with Google that made us lazy, naïve, and not curious enough. 1findr is not perfect as it is, but it presents a nice compromise being sharp and being extensive enough. But duly noted we may miss a few diamonds, and have a shorter tail in our results.

Over time, we hope to have more publishers helping us built a high quality full-text. We've started with Karger who likes to think outside the box. We've started experimenting with the Frontiers corpus as well. This is still small scale but we are careful and reflective about our development. Once we'll have determined the investments in technical complexity and index size is worth our while to improve the user experience, we'll start deploying full-text indexing on a larger but progressive scale, at least for those publishers who want their material to be discoverable to the maximum extent.

Éric

Eric Archambault, PhD

CEO | Chef de la direction

1335, Mont-Royal E

Montréal QC Canada H2J 1Y6

T. 1.514.495.6505 x.111

C. 1.514.518.0823

eric.archambault@science-metrix.com

science-metrix.com & 1science.com

<image003.png> <image004.png>

From: SIGMETRICS <sigmatrics-bounces@asist.org> **On Behalf Of** David Wojick

Sent: April 24, 2018 2:16 PM

To: sigmatrics@mail.asis.org

Subject: Re: [Sigmetrics] 1findr: research discovery & analytics platform

It appears not to be doing full text search, which is a significant limitation. I did a search on "chaotic" for 2018 and got 527 hits. Almost all had the term in the title and almost all of the remainder had it in the abstract. Normally with full text, those with the term only in the text are many times more than those with it in title, often orders of magnitude more.

But the scope is impressive, as is the ability to filter for OA.

David

David Wojick, Ph.D.

Formerly Senior Consultant for Innovation

DOE OSTI <https://www.osti.gov/>

At 08:00 AM 4/24/2018, you wrote:

```
Content-Language: en-US
Content-Type: multipart/related;
    type="multipart/alternative";
    boundary="-----=_NextPart_001_00EE_01D3DBBD.BC977220"
```

Greetings everyone,

Today, 1science announced the official launch of 1findr, its platform for research discovery and analytics. Indexing 90 million articles of which 27 million are available in OA it represents the largest curated collection worldwide of scholarly research. The platform aims to include all articles published in peer-reviewed journals, in all fields of research, in all languages and from every country.

Here are a few resources if you're interested in learning more:

- Access 1findr platform: www.1findr.com
- Visit the 1findr website: www.1science.com/1findr
- Send in your questions: 1findr@1science.com
- See the press release: www.1science.com/1findr-public-launch

Sincerely,

Grégoire

Grégoire Côté

President | Président

Science-Metrix

1335, Mont-Royal E
Montréal, QC H2J 1Y6
Canada

<image001.png><image002.png>

T. 1.514.495.6505 x115

T. 1.800.994.4761

F. 1.514.495.6523

gregoire.cote@science-metrix.com

www.science-metrix.com

Content-Type: image/png;
name="image001.png"
Content-Description: image001.png
Content-Disposition: inline;
creation-date=Tue, 24 Apr 2018 12:00:30 GMT;
modification-date=Tue, 24 Apr 2018 12:00:30 GMT;
filename="image001.png";
size=1068
Content-ID: <image001.png@01D3DB57.02A76980>

Content-Type: image/png;
name="image002.png"
Content-Description: image002.png
Content-Disposition: inline;
creation-date=Tue, 24 Apr 2018 12:00:30 GMT;
modification-date=Tue, 24 Apr 2018 12:00:30 GMT;
filename="image002.png";
size=1109
Content-ID: <image002.png@01D3DB57.02A76980>

SIGMETRICS mailing list
SIGMETRICS@mail.asis.org
<http://mail.asis.org/mailman/listinfo/sigmetrics>

SIGMETRICS mailing list
SIGMETRICS@mail.asis.org
<http://mail.asis.org/mailman/listinfo/sigmetrics>