

IMPROVING UPTAKE OF TEXT AND DATA MINING IN THE EU

Facts

Project No: 665940
Program: H2020 | CSA | GARRI
Duration: 09/2015 - 08/2017

TDM Spotlight

The University Perspective

Who?

A research team at the University of Edinburgh – the sixth oldest university in the English-speaking world. They're using TDM to keep researchers updated on developments in their domain, by quickly identifying high-quality, relevant results from the millions of publications produced each year.

What kind of numbers are we talking about exactly?

In the healthcare sector, 1.3 million new pieces of research related to biomedical science alone are published each year. A typical database search returns about 80,000 hits, and only 4,000 of those are likely to be very relevant to a researcher's work.

How long would it take to scan those papers manually?

The team estimates that it takes 3-5 minutes to scan a paper, and every paper has to be double screened. If you didn't mind forgoing sleep, you could work around the clock to review one paper every 7 minutes. At that rate it would take a year to review all 80,000 hits, but since everyone needs to sleep a more realistic timeframe is 2-3 years. By the time you finished the first 80,000 hits, more results would have been added and your work would be out of date.

How does TDM help?

TDM techniques can already be used to zero in on the top 25% of papers which are most relevant to any given search query. Researchers believe that, with a little more work, it will be possible to use TDM to identify the top 10% of search results.

What else can TDM do?

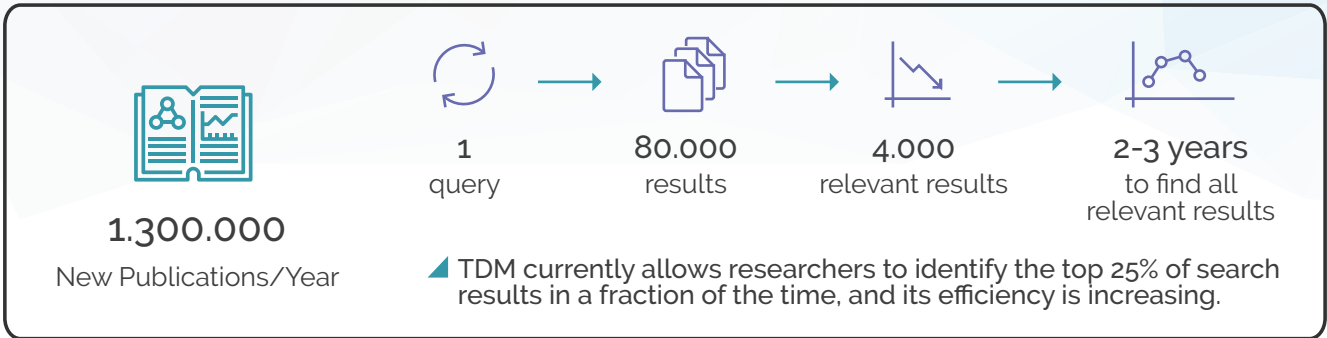
The research group, led by a Professor of Neurology and Translational Neuroscience at Edinburgh University has looked at the risk of bias in research. Bias is a problem because it can lead to effects and results being overstated.

As a result, future work may be based on a false premise and this could lead to numerous consequences, including higher failure rates. For laboratory research, money, time and animal lives could be wasted. In clinical trials, human subjects may be put at risk.

TDM is one potential way to create unbiased summaries of research, according to the Edinburgh team. It could help flag up studies in which effects have been overstated, and lead to improved efficiency in areas such as healthcare (eg. by guiding better design of clinical trials).



BIOMEDICAL LITERATURE AND THE IMPACT OF TDM



Why aren't we grasping this opportunity for scientific progress?

Like many others who want to use TDM, these Edinburgh University researchers face multiple barriers.

The research group is covered by the UK copyright exception for non-commercial TDM, but they can't share the full results with anyone outside of the institution who doesn't have the same access to subscriptions. For the same reason, it's difficult for the team to collaborate with the commercial sector in their search for tools and services which might help to reach their goals.

Technical barriers stand in the way of extracting the results of experiments, particularly where information is presented as tables and images instead of text. Often access is limited to the publication abstract. More content availability would greatly improve results.



Finally, not all publications are available through Edinburgh's university subscriptions, so the research team rely on interlibrary loans. This costs about £4 per publication, and delivery times are long. For smaller institutions, this is a major barrier standing in the way of the democratization of science.

"As we develop tools for TDM, it is absolutely essential that we have access to all relevant publications – otherwise we get a distorted view, seeing only the research published in open access. If, for instance, all the publications saying a drug didn't work were hidden away behind a paywall, we might end up doing a clinical trial when a complete understanding of the literature would have shown us that that was not justified."

Malcolm Macleod,
Professor of Neurology and Translational Neuroscience, University of Edinburgh

TDM BARRIERS

- ▲ Limited copyright exception hinders collaboration
- ▲ Technical barriers to content availability
- ▲ Access to publications and associated costs

Discover more
VISIT OUR COLLECTION

STORIES 

PROJECTS 

ORGANISATIONS 

TOOLS 

STUDIES 

