

NEWS RELEASE



Media Contact:
Bob Campbell, Publishing Research Consortium
Tel: +44 (0)1865 476118
bcampbel@wiley.com

June 2011

JAM today: the PRC Study on Journal Article Mining [1]

If you have too much to read, or too much information to digest, could a machine do it for you? That is the essence of the motivation behind content mining, here including both text and data mining, examined in the latest Publishing Research Consortium (PRC) [2] study.

Headline findings in the report, which draw upon expert interviews and a survey of opinion are:

- Content mining is about to accelerate, will expand into new areas and develop further into automated information extraction and relationship analysis
- The focus is shifting from the traditional life sciences (especially drug discovery) to the social sciences, humanities, business, marketing and even law
- A majority of respondents to the survey supported three common solutions for facilitating content mining
 - More content standardization for mining-friendly formats
 - A shared content mining platform across publishers
 - Commonly agreed rules for the granting of mining permissions
- Third-party mining requests are received by most publishers (77% of all, 88% of large ones) but at a very low level (less than 10 per annum); most mining requests come from abstracting and indexing services followed by corporate R&D organisations.
- Over 90 % of publisher respondents grant research-focused mining requests, nearly 60 % of these in all or the majority of cases. The request will be granted by 60% of publisher respondents in most or all cases if it creates traffic drivers to their sites but just over half of these publishers (51%) will refuse in all or most cases if the results of the mining would compete with their own services
- A majority of publishers do not see Open Access as a prerequisite for content mining

Eefke Smit, who carried out the research with Maurits van der Graaf, said “ We found a lot of optimism for new opportunities in mining scholarly content among all stakeholder groups. Publishers expressed a clear intent to invest more in mining and new services that will reveal deeper levels of information. We can expect many more exciting developments in this area in the near future.”

Bob Campbell (Chairman of the PRC Steering Group) added: “This comprehensive study shows that publishers understand the potential of text and data mining. It demonstrates that many publishers grant permission for mining for research purposes. It is also understandable that many publishers are reluctant to allow mining if the outcome could replace or compete with their own services which can involve a considerable investment.”

The report focuses on the state of content mining in the arena of academic and professional publications, journal articles in particular. Academic and professional publishers frequently receive requests from parties wishing to mine their content and face uncontrolled downloads

or crawling. More and more publishers undertake content mining on their own journal content. This PRC study aims to provide more insight into practices, policies for permission requests, publishers' plans and possibilities to facilitate better content mining. In early 2011 the authors conducted 29 interviews with people involved in content mining projects and permission handling. During March and April of 2011 a survey was mailed to all publishers on the mailing lists of CrossRef and the International Association of STM Publishers. The report analysis is based on 190 responses.

[1] Journal Article Mining, PRC Study. Freely available on the PRC site (<http://www.publishingresearch.net/documents/PRCSmitJAMreport20June2011VersionofRecord.pdf>)

[2]About The Publishing Research Consortium (PRC):

The PRC is a group representing publishers and associations supporting global research into scholarly communication in order to enable evidence-based discussion and objective analysis (<http://www.publishingresearch.net>). PRC's objective is to support work that is scientific and pro-scholarship, in order to promote an understanding of the role of publishing and its impact on research and teaching.

June 2011