Deni Auclair
VP & Lead Analyst
T. +1 508-785-8384
dauclair@outsellinc.com

Insights
Copyright Clearance Center's Newest Service: RightFind XML for Mining

CCC announces the launch of RightFind XML for Mining, providing a solution to researchers looking to ease the burden of finding hidden knowledge in vast oceans of text.

Important Details

Adhering to the company's philosophy of providing solutions to pain points around licensing and content access, Copyright Clearance Center (CCC) announced the launch of its text mining service, RightFind XML for Mining (RXM). The product has been in development since 2013, when issues around mining massive amounts of text had long reached critical mass within the STM and commercial pharma arenas.

To clarify, text mining is the extraction of facts from a body of text, such as a corpus of articles, and the identification of relationships between these facts. Text mining helps researchers confirm known hypotheses and reveal hidden, obscure, or previously unknown connections or causalities. Examples of medical literature-based discoveries stem from the seminal work of _Don R. Swanson_, an information scientist who, in 1986, took a set of articles about Raynaud's disease and another set about dietary fish oil and found a previously undiscovered relationship between the condition and fish oil, confirmed three years later in a clinical trial. He also discovered the previously unknown association between migraine and magnesium deficiency using the same methodology.

Specific issues in text mining revolve around the research process and hurdles that have to be overcome along the way. CCC's RXM provides a solution for the first key steps of that process and its pain points, starting with the aggregation of content in XML, a structured format for mining, the normalization of that content (since conversion from PDF isn't always a smooth process), and the covering of all articles in the corpus through a single license. A researcher using RXM can create projects containing a list of articles from the entire corpus, then further explore by journal, publisher,

and year, as well as by additional search terms. The context of the search term within the full-text article is also visible to the researcher, which would not be the case without purchase of the corpus.

The researcher then determines accessibility to full-text articles, since RXM analyzes what is available through open access or a company's holdings and what is behind a paywall. Whatever is not available through the researcher's institutional subscription can be purchased through the RXM module. Absent RXM, this would entail contacting a list of publishers to negotiate permissions. Even if only five or six publishers, it is a time-consuming process. With RXM, determining access and mining permissions is automated and CCC pre-negotiates a suite of blanket or transactional licenses with publishers allowing for text mining.

Once the researcher finalizes the list of desired articles, they can be downloaded (in XML format) so they can be processed by a third-party text mining software tool or custom software. At this point, the text mining process moves from RXM to various text mining software providers such as Elsevier, SAS, or TEMIS, once an integration is implemented. An integration is already in place for the I2E software module of Linguamatics, CCC's development partner. Similar integrations are in progress with other text mining vendors and software providers.

Implications

Without automation, the entire process can take weeks or months for a single query. Multiply that by the number of questions a lab generates each day and the effort, the hours, and the cost of access required become astronomical. This is, of course, one reason why text (and data) mining is such a hot topic – because the pain involves both time and money, and RXM can do the job in a matter of hours or even minutes. The bigger reason is that it supports innovation through simplifying and speeding a process that can be so daunting researchers often give up on attempting to mine large bodies of text.

RightFind XML for Mining is an example of an offering primed for success for several reasons:

- It reaffirms CCC's role in the information arena as an independent intermediary sitting between content users and providers, helping to find efficiencies and create bridges between the two groups.
- The organization is leveraging its relationship with existing customers to sell additional services to that customer base.
- It has identified and addressed pain points, creating a tool for research organizations looking to save time and money on the front end of the text mining process.
- Benefits are clearly visible: publishers do not have to negotiate with each research organization for text mining rights, and those organizations do not have to negotiate with however many publishers have text they wish to mine.
- Finally, the product is on an existing platform, thereby functioning as an additional module rather than a separate offering, and it is scalable to other industries, so there is opportunity for growth.

With the future of text and data mining as promising as it is in support of population health, to name just one application, RightFind XML for Mining promises to have an equally bright future with the potential to save research organizations time and money. Even more important, successfully addressing these pain points supports innovation in scientific research.

About Insights

Redistribution: Insights is an annual subscription service. Workgroup subscribers may distribute content freely only within their workgroup as defined in their agreement with Outsell. Company subscribers may distribute content freely within their companies, agencies, and institutions, but not to persons outside their organizations without permission. Individuals subscribe for their own use and may not distribute, disseminate, disclose, or otherwise make use of the information herein without permission.

For more information about subscriptions to Insights, visit:
http://www.outsellinc.com/store/subscription/insights
For contact details, e-mail: insights@outsellinc.com

About Outsell

Outsell, Inc. is the world's only research and advisory firm focused solely on media, information, and technology. We use a time-tested, quality-proven, top-secret blend of data, brains, community, and pizzazz to produce extraordinary value for our clients. Elite information industry executives from all over the world look to Outsell for trusted advice, bold insights, and confidential access to exclusive intelligence and decision support. Outsell has your back. You'll stay more focused, save time, grow revenue in a fast-changing global digital environment, and sleep better at night. Guaranteed.

www.outsellinc.com	Burlingame, CA USA	London, United Kingdom
info@outsellinc.com	+1 650-342-6060	+44 (0)20 8090 6590