Copyright Clearance Center’s Text Mining Solution
CCC: The Rights Licensing Experts

• Not-for-profit founded in 1978
• Global licensing and content solutions
• Manages 600+ million individual rights
• Subsidiaries in Europe and Japan
• Outsell 10 to Watch
• EContent 100 Companies that Matter Most in the digital content industry for five years
Making Copyright Work

Rightsholders

600+ million rights from:
- Publishers
- Authors
- Agents
- Creators

Content Users

- 35,000 companies
- Employees worldwide
- Users in 180 countries

Licensing Solutions
Rights Management
Content Delivery
Copyright Education
Text Mining Today

- Mining limited to abstracts
- High cost to obtain formatted full-text content and permission from multiple publishers
- Multiple formats
- Researchers can’t mine content to which they are not subscribed
Introducing CCC’s Text Mining Solution

Build a collection of full-text articles in XML format for mining

- Identify full-text articles.
- Download full-text, XML versions of articles.
- Import into your preferred text mining software.

CCC’s Text Mining Solution

Text Mining Software
A New Opportunity in the Corporate Market

• Life sciences expanding use of text and data mining (TDM)
• Rightsholder opportunity to monetize content being used in new text mining workflow
• New solution offering licenses and delivers XML content for mining purposes
• Built in collaboration with six major publishers.
Macro-level: TDM Opportunity

“$100 Billion in value annually across the US health-care system, by optimizing innovation…”
- McKinsey.com April 2013

$35B market growing 10% IDC. Software and service text-analytics revenues now total $835 million globally. 2010 market study

Big Data
+27% CAGR

TDM

Industrial R&D spending
+3-5% CAGR

Business Intelligence and Text Analytics
+9.7% CAGR

More than $500B spent on R&D key sectors. Moderate growth expected in 2014.
Battelle.org
Rights-Aware Text Mining Article Access

Publishers provide content and rights.

Publishers

Full-Text Search and Rights Engine

CCC Solution

STM
The global voice of scholarly publishing
Rights-Aware Text Mining Article Access

1. Publishers provide content and rights.

2. Companies create article sets using keywords or other criteria. (Including unsubscribed content)
Rights-Aware Text Mining Article Access

1. Publishers provide content and rights.

2. Companies create article sets using keywords or other criteria. (Including unsubscribed content)

3. Companies preview results including subscription status. Optional content purchase.

4. Final results are imported into text mining tools.
## Value Provided by CCC

<table>
<thead>
<tr>
<th>Publisher Provides</th>
<th>CCC Provides</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Feed of XML</td>
<td>• Normalization of XML metadata</td>
</tr>
<tr>
<td>• Agreement w/ CCC</td>
<td>• Hosting and storage of content</td>
</tr>
<tr>
<td></td>
<td>• User base of customers</td>
</tr>
<tr>
<td></td>
<td>• Reduction in engineering costs for multiple feeds</td>
</tr>
<tr>
<td></td>
<td>• Search and API software</td>
</tr>
<tr>
<td></td>
<td>• Article set delivery</td>
</tr>
</tbody>
</table>

CCC provides many valuable services with this offering. The effort and expense involved in the above are what affect the service fees.
Generate **New Value** from Subscription Business

- Increase renewals
- Boost content visible in the article discovery process
- Increase royalty payments
- Reduce cost of providing content for mining to multiple users
Reporting for Rightsholder Analysis

• 1st royalty payments to Rightsholders 12-24 months after launch.
• Reports include # of downloads per period, by title
• Revenue reporting attributed to title
• Reports will be available in RightsCentral, CCC’s reporting platform
## Scope of License
Limited to TDM Use

<table>
<thead>
<tr>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global service</td>
</tr>
<tr>
<td>Metadata and abstracts download in XML</td>
</tr>
<tr>
<td>Full text for subscribed-to articles</td>
</tr>
<tr>
<td>XML storage plus two additional copies</td>
</tr>
<tr>
<td>Transform and import text into customer’s text mining tools</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>User cannot create, license, sell, distribute or make available a database using the content.</td>
</tr>
<tr>
<td>Cannot convert to HTML, PDF or other reading format</td>
</tr>
<tr>
<td>Use of content limited to Internal use</td>
</tr>
<tr>
<td>Projects are limited to 10,000 articles</td>
</tr>
</tbody>
</table>
# Participating Publishers

<table>
<thead>
<tr>
<th>Large STM</th>
<th>Mid-sized STM</th>
<th>Society Publications</th>
</tr>
</thead>
</table>
| - Cambridge University Press  
- Hindawi Publishing  
- Informa Healthcare  
- Karger AG  
- PLOS  
- Springer Science & Business Media  
- Taylor & Francis  
- Thieme  
- Walter de Gruyter  
- CoAction Publishing  
- Future Medicine  
- Future Science  
- IOS Press B.V.  
- Landes Bioscience  
- MA Healthcare  
- Maney Publishing  
- R C N Publishing Co.  
- World Scientific Publishing | - American Diabetes Association  
- American Society for Nutrition  
- American Society for Microbiology  
- Inderscience  
- Royal Society of Chemistry |
Demo: Set Criteria for Your Project

Users request articles using keyword search or other criteria.
Demo: Refine and Filter Content Sets

CCC returns article results with subscription status into a customized index for each organization.
Demo: Download Result Sets in Standard XML Format

Request download archive

If you want to create the download archive with the current result set, please select an archive format and click on the Download button. Otherwise, click on the Cancel button.

- ZIP
- TAR

Users can then download the XML versions of the articles for loading into their text mining tools.
Content needed ASAP to support sales of the license.
Time’s Up!

About your speaker:

Name: Matthew Pedersen
Company: Copyright Clearance Center
Tel: +1-978-646-2426
Email: mpedersen@copyright.com
Social Media: @copyrightclear