

RightFind™ XML for Mining

Text and Data Mining of Scholarly Publications

Presented to: IFRRO Business Models Forum, November 2015

Babis Marmanis, CTO & Vice President, Engineering & Product Development



What Is Text and Data Mining?

- Automate the extraction of “Entities” from full text
- Find relationships between entities
- Find patterns between relationships
- Produce hypotheses of interest (exploration)
- Provide supporting evidence for hypotheses (validation)



Text Mining Today

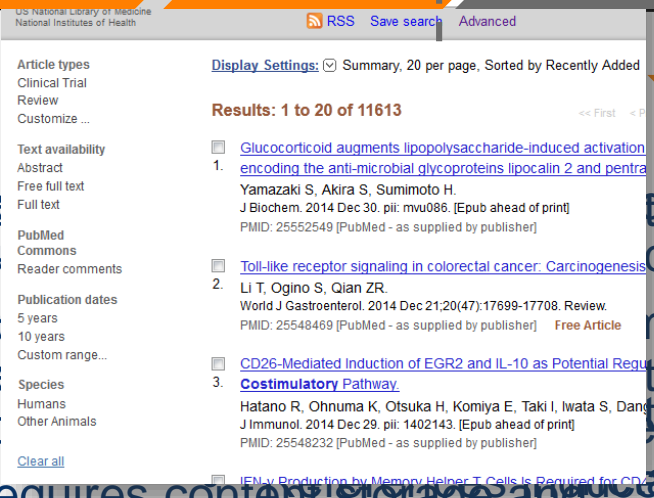
- Mining limited to abstracts
- High cost to obtain full text content and permission from multiple publishers
- Multiple formats
- Researchers can't mine content to which they are not subscribed



Text Mining Today – Example Workflow



- Perform search
- Obtain permission from publishers to mine for commercial use
- Requires content storage and feed management
- Requires structuring text into XML
- "Text does not have words"
- Combining content from multiple sources takes time to normalize the data



RightFind™ XML for Mining

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.

XML for Mining

Text Mining Software

XML for Mining Benefits



Better results from text mining

- Many key facts, data, and findings don't exist in the abstract
- Updated daily with the latest published research from publishers



Reduced labor and effort

- Simplification of licensing, content aggregation and storage
- Rapidly get full text indexed in TDM tool
- 4-8 person-weeks of labor savings per project



Simplifies copyright compliance