# THOMSON REUTERS ENTITY IDENTIFICATION TOOL IMPROVES PLATFORM EFFICIENCY AT DOCUMENTCLOUD

*"Having a well-documented and easy to use API has made a big difference. The accuracy of the metadata and the consistent tagging across the entire documents repository is what enabled us to - effectively and very quickly - browse relevant documents on a particular entity or topic."*

*TED HAN*
*TECHNICAL DIRECTOR,*
*DOCUMENTCLOUD*

## THE CHALLENGE

If you're in need of information, a quick online search will usually garner positive results. Yet for journalists and researchers in particular, digging for information within documents oftentimes yields richer returns. Web-based DocumentCloud helps with that pursuit through its document research and publishing platform built to make document searches easier and faster.

Before the platform went live in 2010, to actually make it easier and faster the company knew that its platform needed an entity identification tool.

*"Search tools are mostly helpful when searching for particular needles in a haystack,"* explains Ted Han, Technical Director at DocumentCloud. *"Being able to get an aggregate sense of a document or set of documents for the purpose of browsing wasn't easily achievable without an entity identification tool. We couldn't go live without one which is why we looked for a vendor that could help us accomplish that goal."*

## DELIVERING THE SOLUTION

DocumentCloud wanted a solution that was either easy to install and maintain in a server environment, or a robust external API they could reliably query. Already aware of the breadth of solutions Thomson Reuters offered, DocumentCloud selected the Thomson Reuters Intelligent Tagging solution (an API) to boost the organization's news search capabilities.

Using Natural Language Processing, text analytics, and data-mining technologies, Intelligent Tagging quickly analyzes and tags documents, research reports, news articles, blog posts, and more. It's also the very same tagging solution that powers Thomson Reuters Eikon and Reuters News.

*"Having a well-documented and easy to use API has made a big difference. The accuracy of the metadata, and the consistent tagging across the entire documents repository is what enabled us to - effectively and very quickly - browse relevant documents on a particular entity or topic,"* says Mr. Han.

## CUSTOMER BENEFITS

The Intelligent Tagging solution has enabled DocumentCloud to provide its users with a search platform that effectively captures, collates, and makes sense of massive amounts information. The solution also returns information in order of relevance to the document, which is so vital for someone who is in need of quickly acquiring pertinent information from myriad documents simultaneously.

*See reverse side*

**THOMSON REUTERS™**

*"Intelligent Tagging helps our users, be it journalists or researchers, find leads into documents,"* says Mr. Han. *"Leads don't have to be 100 percent accurate, so long as they can save someone time in sifting through collections of documents. However, by using the relevance scores and broad range of entities and events, Intelligent Tagging significantly improves the search ranking and makes this task quicker and in some ways surfaces insights we didn't know that we should be aware of."*

## FUTURE IMPLEMENTATION

*"Early in its life DocumentCloud prototyped ways to let journalists use entities to cross reference documents,"* explains Mr. Han. This enables journalists to retrieve multiple highly relevant documents on a particular entity in one search. Mr. Han adds, *"We plan on rolling this feature out in the next year."*

**ABOUT DOCUMENTCLOUD**

DocumentCloud is a catalog of primary source documents and a tool for annotating, organizing and publishing them on the web. Documents are contributed by journalists, researchers and archivists. We're helping reporters get more out of documents and helping newsrooms make their online presence more engaging. DocumentCloud was founded in 2009 with a grant from the Knight News Challenge. After two years as an independent nonprofit organization, DocumentCloud became a project of Investigative Reporters and Editors in June 2011.

For more information on DocumentCloud, please visit www.documentcloud.org.

THOMSON REUTERS™