

Mining the knowledge locked in ECM

Dated: 18 December 2015

By Mitch DeFelice

Recent announcements from Google and Microsoft regarding machine-learning capabilities will provide the ability to transform corporate Enterprise Content Management (ECM) system from a simple unstructured data repository into an oracle of corporate data.

In their book *Smart Customer Stupid Companies...Why Only Intelligent Companies Will Thrive, and How to Be One of Them* – the authors Michael Hinshaw and Bruce Kasanoff articulate how customers are becoming “smarter” with technology advancements. The book presents a sound case that companies that do not evolve with their customers will become irrelevant.

There have been two recent announcements that have occurred (November 9th, 2015 and November 12th, 2015 respectively) that have the potential to turn the metaphorical phase “Stupid Companies” to mean literally that.

The first announcement was that Google open sourced TensorFlow, a type of machine learning system that uses unsupervised learning, i.e. “Deep Learning.” TensorFlow powers Google Photos, Google Translator and backbone features such as search and Smart Reply. Not to be outdone, Microsoft announced that it is open sourcing its “Deep Learning” system called Distributed Machine Learning Toolkit (DMTK).

Why would Google and Microsoft open their “secret sauces” to the world? There are a number of reasons one can speculate, but anytime you open up your secret sauce, it’s to win over programmer’s minds. In fact, machine learning and specifically Deep Learning subjects are not for the average corporate web application developer. You will need people who have strong mathematics and computer science skills along with machine learning background.

The impact of having access to these Deep Learning system capabilities will be truly disruptive, especially in the area of unstructured data. It is true Hadoop has all the underpinnings of a great ECM system with its distributed file system, map/reduce for large-scale data processing. Generating indexes associated with documents is a natural progression since Hadoop abundantly provides these capabilities.

However, ECM is much more than just large volumes of documents that is in need of indexing. ECM involves the whole life cycle of document management that includes: create, capture, indexing, approval (workflow/case management processing), publishing (version management), collaboration (share), archiving & defensible disposal (Records Management)

Having Deep Learning capabilities will transform ECM into a more advanced type of product. A product that can determine the content regardless of its content type (image, text, audio, and video). This will shift the technology from a simple content management solution to a knowledge management system.

Having the ability to automatically determine corporate enterprise records from ROT (Redundant, Obsolete, Trivial) will not only streamline document lifecycle management, but will allow you to build a knowledge management system that will transform the business.

Today, the best ECM systems can do is to classify your content by looking at metadata tags and keywords in documents. As an example, it will not be enough to look at a document and classify it as a legal contract. Deep Learning will take ECM to the next level, by not only classifying the document as a contract but also evaluating it to make sure it is an iron clad contract that has the necessary clauses to assure your company is protected!

Deep Learning will also provide Natural Language Processing (NLP) capabilities. You now have turned your corporate Enterprise Content Management system from a simple unstructured data repository into an oracle of corporate data.

Imagine how these new capabilities will change IT's ability to service the business. You can now tie your knowledge management solution to your business process to provide invaluable insights. For example, your medical claims processing can use a machine learning system for fraud detection, clinical treatment abuse, etc. You have now shifted your IT environment from simple "processing" transactions to "understanding" transactions from a level of service that has been delivered to your member.

Deep Learning systems will be able to "simulate" any new possible business opportunities under different market conditions. Having the ability to know upfront if a business opportunity will be profitable before spending millions of dollars to roll out an implementation solution is a very powerful position from which to leap frog your competition.

The reality is that technology is shifting faster than leadership can understand what threats it poses. The ability to adjust to any new threats will not be measured in months or years, rather in weeks. Large bureaucrat companies will struggle to move at the pace that will be required to keep up with demanding customers who will become ever savvier in a connected world. They will reward those companies that can provide the most value.

Mitch DeFelice is a TOGAF 9 Certified Sr. Solution Architect serving as a member of Health Care Service Corporation (HCSC) Enterprise Architect Services team with key focus on delivering unstructured & cognitive data solutions. Mitch blogs on his LinkedIn account at: <https://www.linkedin.com/in/mitchdefelice>

Published in IDM at <https://idm.net.au/article/0010851-mining-knowledge-locked-ecm>

Submitted by Alan Kong - PROV