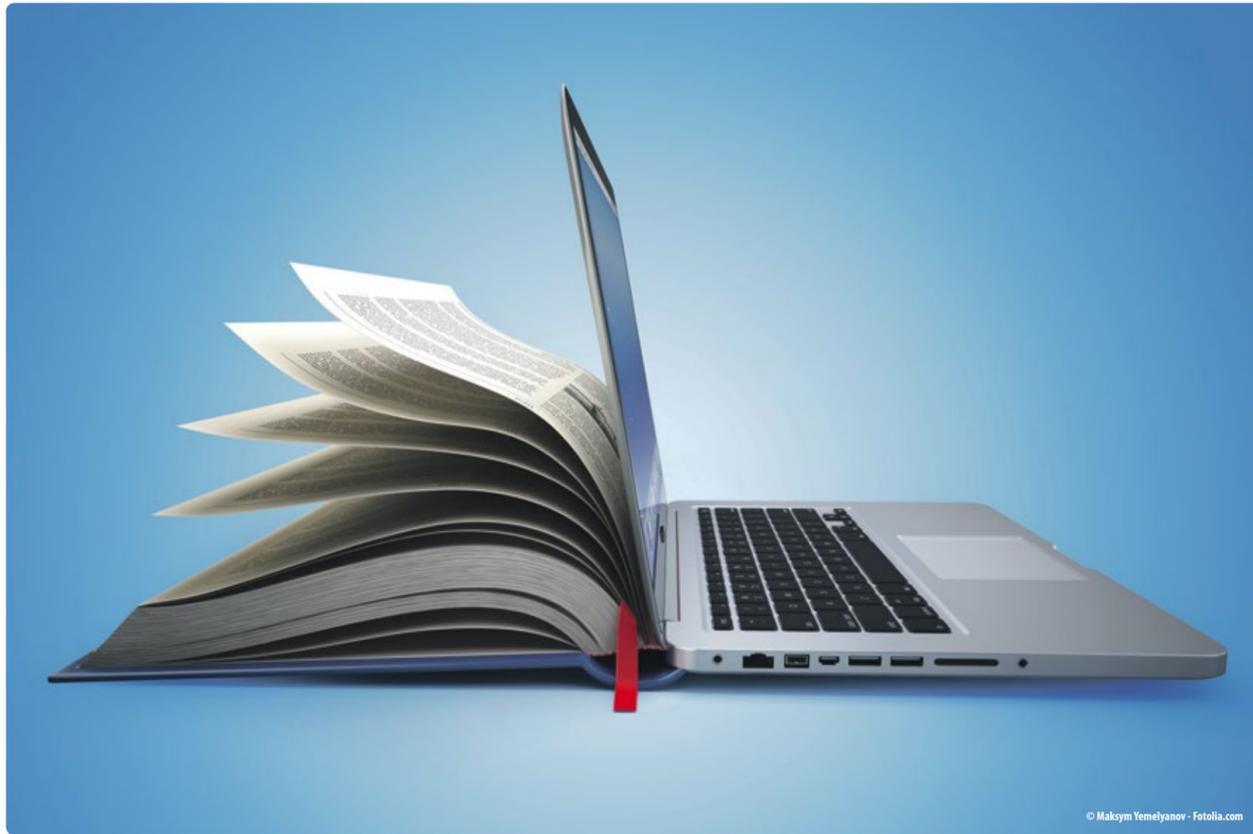


Wenn jedes Körnchen Wissen zählt

- Chemie- und Pharmafirmen nutzen Text-Mining, um ihre Forschungsprojekte zu beschleunigen -



Genau die Informationen zu finden, die man für die Entwicklung einer bestimmten Formulierung, Technologie oder Anwendung braucht, wird immer schwerer. Chemie- und Pharmakonzern, die das passende Wissen dazu schneller erwerben können, arbeiten nicht nur effizienter, sondern können auch eine schnellere Marktreife ihrer Produkte gewährleisten. Den notwendigen – zeitlichen und finanziellen – Vorsprung haben diejenigen, die Text-Mining gezielt einsetzen.

1,5 Millionen: so viele wissenschaftliche Artikel werden einem Bericht des Joint Information Systems Committee (JISC) zufolge pro Jahr veröffentlicht. Kein Wunder, dass wertvolle Informationen hier leicht untergehen. Schließlich ist es für viele Wissenschaftler gar nicht möglich, solch eine Menge an Informationen zu lesen und zu analysieren. Daher heißt das neue Zauberwort für die Forschung: Text-Mining.

Unter Text-Mining versteht man, in Analogie zu Data-Mining, die Extraktion von Fakten aus einem Textkörper, wie etwa einer Artikelsammlung. Und, was mindestens ebenso wichtig ist, das Erkennen von Zusammenhängen zwischen diesen Fakten. Dank Text-Mining können Forscher bekannte Hypothesen untermauern und bislang verdeckte oder unbekannte Verbindungen oder Kausalitäten zwischen bestimmten Fakten aufdecken.

Bahnbrechende Entdeckungen

In der Chemie etwa können dank Text-Mining chemische Strukturformeln in Dokumenten gefunden werden, bei denen eine Darstellung per Hand nicht oder nur für einige Strukturen vorgenommen wurde. Oder bei der eine Auszeichnung unwirtschaftlich ist, wie etwa bei den internen Berichten eines Unternehmens. Entscheidend ist, dass diese Strukturen an bestimmten Punkten innerhalb eines Dokuments auffindbar sind, weil man Fragen stellen kann wie „welche Chemikalien werden als Hemmer eines bestimmten Zielmoleküls erwähnt“ oder „welche Rolle spielt die Chemikalie in diesem Dokument“.



Kim Zwollo,
RightsDirect

So ist es etwa denkbar, dass ein Forschungsteam in der petrochemischen Industrie, das sich mit der Herstellung von biobasiertem Butadien beschäftigt, mithilfe von Text-Mining schneller auf die Idee kommt, ein spezielles Fermentationsverfahren zu entwickeln, das eine kosteneffiziente Umwandlung nachwachsender Rohstoffe in Butadien erlaubt.

Entscheidend dabei ist, dass solche Entwicklungen auf Erkenntnissen basieren, die im Laufe der Jahre nur durch Text-Mining zu Tage gefördert werden konnten. Weil die relevanten Informationen dazu in scheinbar grundverschiedenen Aufsätzen stehen, die ohne Text-Mining nicht miteinander in Verbindung gebracht worden wären.

Spätestens jetzt wird klar, dass Text-Mining größere Chancen für Entdeckungen und potentielle neue Einnahmequellen für die forschende Industrie bietet.

Data-Mining und Text-Mining: Der Unterschied

Data-Mining wird in der Regel mit stark strukturierten Daten betrieben. Text-Mining dagegen arbeitet mit wesentlich schwächer strukturierten Textdaten. Daraus ergeben sich zwei Folgen für die Prozesse im Text-Mining: Erstens müssen die Textdaten strukturiert werden und zweitens sind Mehrfachklassifikationen (anders als beim Data-Mining) geradezu erwünscht.

Die Unterscheidung zwischen Data-Mining und Text-Mining soll verdeutlichen, dass Text-Mining nicht einfach mittels Suchmaschinen erfolgen kann. Denn herkömmliche Suchmethoden über Stichwörter bspw. würden keine Ergebnisse für einen Text bringen, bei dem die wichtigsten Informationen in den Fußnoten und Quellenangaben stecken. Auch PDF-Formate könnten nicht durchsucht werden, und genauso wenig ganze Absätze und Hypothesen extrahiert oder gar die Zusammenhänge zwischen diesen hergestellt werden.

Die Lösung: Inhalte sammeln und normieren

Aus diesem Grund hat RightsDirect zusammen mit seiner Muttergesellschaft, dem Copyright Clearance Center (CCC), RightFind XML for Mining entwickelt. Die Lösung sammelt nicht nur Texte und Inhalte, sondern normiert sie auch in ein XML-Format. XML kann nicht nur Daten für die maschinelle Bearbeitung transportieren, sondern eignet sich auch für plattformunabhängige interaktive Anwendungen. Damit ist das Durchsuchen von Artikeln kein Problem mehr. Gleichzeitig deckt RightFind für alle verfügbar gemachten Inhalte die nötigen Copyright-Lizenzen ab. Und werden einmal mehr Informationen zum Suchbegriff gebraucht, hat der Benutzer Zugriff auf den Volltext. Mit einer einzigen Quelle. Für den Fall, dass die Rechte des Unternehmens dafür nicht ausreichen, kann der dementsprechende Text einfach zusätzlich über RightFind XML for Mining erworben werden. Zeitaufwändige Verhandlungen zu den Benutzungsrechten mit Verlagen und Autoren gehören der Vergangenheit an.

Mehr Effizienz durch Text-Mining

Wie die Beispiele zeigen, lassen sich mithilfe von Text-Mining die Gewinnung von Forschungsergebnissen in der Chemie- und Pharmaindustrie erheblich beschleunigen. Wollen Wissenschaftler, Chemiker oder Pharmaunternehmen etwa wissen, ob ihre Entwicklungen dem europäischen Chemikalienrecht REACH entsprechen, müssen sie bisher die dazu notwendigen Informationen nicht nur mühsam suchen, sondern jeden Verlag einzeln kontaktieren, einen Preis für die Rechte aushandeln, die Zeitschriftenartikel beziehen und dann noch in ein einheitliches, leicht durchsuch- und bearbeitbares Format umwandeln. Angenommen, dies würden 20 Unternehmen mit 20 Verlagen machen, bräuchte man allein hierfür 400 Vereinbarungen, 400 Artikelsammlungen und 400 Formatierungen.

RightFind XML for Mining nimmt diese Arbeit ab und erstellt einheitlich in XML formatierte Zeitschriftensammlungen zum jeweiligen Spezialgebiet. Und sogar die Weiterverarbeitung der Artikel mit Text-Mining-Programmen von Drittanbietern wie I2E von Linguamatics lässt sich selbstverständlich problemlos bewerkstelligen. So bleibt den Unternehmen mehr Zeit für die eigentliche Forschung und das Entwickeln von Innovationen. Allein der Aufwand, um die Genehmigungen für jede einzelne Publikation einzuholen, bedeutet für viele Unternehmen eine große Hürde. Was dies angeht, sind sich viele Unternehmen – oder besser gesagt deren Mitarbeiter – oftmals gar nicht darüber bewusst, wie weit ihre Rechte gehen. Hier nimmt ihnen RightFind for Mining das Wichtigste ab: Denn da sämtliche Publikationen mit RightFind für das kommerzielle Text-Mining vorab autorisiert sind, können sich die Benutzer darauf verlassen, dass bei ihren rechnerintensiven Projekten das Urheberrecht gewahrt wird. Hundertprozentig. (mr)

Kim Zwollo, Geschäftsführer,
RightsDirect, Amsterdam,
Niederlande

■ kzwollo@rightsdirect.com
www.rightsdirect.com

5 Gründe für Text-Mining

Text-Mining ist nicht ohne Grund ein aktuelles Thema, denn es bietet enorme Erleichterungen für die Wissenschaft. Finanziell und zeitlich.

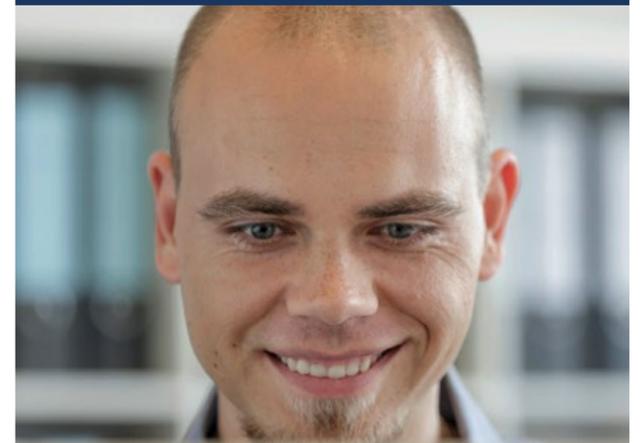
- Suchanfragen mit herkömmlichen Suchmaschinen waren gestern: von nun an werden nicht nur Abstracts, sondern komplette Volltextartikel durchsucht. Alle wichtigen Informationen werden somit schnell gefunden.
- Alle Texte und Inhalte werden in einem XML-Format normiert und lassen sich in den Workflow der unternehmenseigenen Plattform integrieren.
- Damit befinden Sie sich auf einer einzigen Plattform: das Zusammensuchen von Dokumenten wird überflüssig.
- Sämtliche verfügbaren Publikationen sind mit RightFind für das kommerzielle Text-Mining vorab autorisiert – zeitaufwändige Verhandlungen mit Verlagen und Autoren entfallen.
- Fazit: Unternehmen steigern die Effizienz im Bereich Forschung und Entwicklung.

Weitere Informationen: <http://www.rightsdirect.com/rightfind-xml-for-mining/>

»MEINE ABTEILUNG ARBEITET RECHTS-SICHER. GEORG GIBT JEDEM EINZELNEN VON UNS EINDEUTIGE AUFGABEN.«

Andreas Paschke

Leiter Organisationsentwicklung,
Röchling Engineering Plastics SE & Co. KG



»Nie mehr Gesetze lesen«

Die GEORG Compliance Experten ordnen nur die tatsächlich relevanten rechtlichen Pflichten Ihrem Unternehmen zu und unterstützen Sie mit ihrer umfangreichen Erfahrung. Das Selberlesen der komplizierten Rechtstexte entfällt. Sie senken den Personalaufwand und die Risiken.

Das GEORG Compliance Management System® bereitet gesetzliche Vorschriften individuell, einfach und zuverlässig auf.

Informieren Sie sich jetzt:

Martin Mantz GmbH
Hansaring 8
63843 Niedernberg
GERMANY

T +49-(0)6028 97919-0
F +49-(0)6028 97919-33

www.martin-mantz.de

GEORG
COMPLIANCE MANAGEMENT SYSTEM®