

Unternehmensrelevante Informationen, aus dem Internet/Intranet effektiv strukturieren und auffinden

Unternehmensrelevante Informationen aus dem Internet/Intranet effektiv strukturieren und auffinden

Text-Mining und Suche: Technologien, Anbieter und Geschäftsmodelle

Seminar
16.11.2011

Berlin

Leitung
Dr. Alexander Löser

Deutsche Informatik-Akademie 2011

Motivation:

Zu jeder Zeit an relevante Informationen zu gelangen ist in der modernen Wissensgesellschaft ein unverzichtbares Gut – für Mitarbeiter gleichwohl wie für Unternehmen.

Doch genau hier liegt die aktuelle Herausforderung, ebenfalls für Mitarbeiter wie für Unternehmen.

Es existiert im Internet kein Informationsmonopol mehr, Konsumenten von Informationen erstellen und bearbeiten Inhalte in quantitativ und qualitativ entscheidendem Maße selbst. Dadurch werden zahlreiche Akteure zu Produzenten von Webinhalten unterschiedlichster Qualität in Wikis, Blogs, Foren sowie Videoportalen. Der Informationsdschungel wird somit immer dichter und „die Spreu vom Weizen zu trennen“ bei der Analyse von Daten immer schwieriger.

Anhand von drei Fallstudien zu den jeweiligen Anwendungsszenarien werden den Teilnehmern in dieser Veranstaltung Wege hierfür aufgezeigt:

- Business-News (Extraktion und Aggregation von Fakten aus Nachrichten-Datenquellen),
- Social Media Monitoring (Extraktion von Sentiments, also Meinungen und Gefühlsäußerungen zu Produkten und Unternehmen),

sowie Enterprise-Search (am Beispiel der IBM Intranet Suche bzw. Foren der SAP AG).

Zielsetzung

Analysemöglichkeiten aufzuzeigen die die Suchresultate der weltweit verfügbaren Daten und Informationen mit hoher Genauigkeit und Vollständigkeit effizient und wirtschaftlich erzielen, ist das Ziel dieses Seminars.

Das Seminar vermittelt Grundkenntnisse in der Extraktion, Suche und Analyse von Textdaten. Ebenfalls werden Marktplätze für unternehmensrelevante Daten vorgestellt.

Im Anschluss an das Seminar besteht die Möglichkeit einer individuellen Beratung durch den Veranstalter.

Gliederung

Vorstellung von Basistechnologien für die drei Phasen der Analyse: die Auswahl von Extraktoren, die Ableiten eines Analyse-Schemas und die Bewertung bzw. Optimierung der extrahierten Daten.

Die Extraktion von Textdaten ist oft mit hohem Berechnungsaufwand und damit auch Kosten verbunden. Zur Lösung dieses Problems wird eine parallele Abarbeitung des Extraktionsprozesses mittels der Open-Source-Plattformen Hadoop, UIMA und Lucene vorgestellt.

Ebenfalls werden Geschäftsmodelle für Daten als Dienstleistung vorgestellt und Kostentreiber benannt.

Schließlich werden einige Trends der zukünftigen Entwicklung im Text Mining und Web Mining gezeigt.

Einführung

- „Business Intelligence“ und Suche auf Text-Daten
- Typische Textmining-Szenarien

Text Mining: Projektstruktur

- Phasen, Akteure und Komponenten
- Kostentreiber
- Frei verfügbare Datenquellen

Text Mining: Out-of-the-Box Technologien

- Marktübersicht und Player
- Industriestandards, z.B. UIMA und INXIGHT
- Web Services, z.B. OpenCalais, bzw. Extractiv

Text-Mining: Design ‚eigener‘ Extraktoren

- Einführung in das Natural Language Processing
- Regel- und listenbasierte Extraktion, Eigennamen
- Offene und überwachte Faktenextraktion
 - Sprachen zur Beschreibung des Extraktionsflusses
 - Datenqualität vs. Ausführungszeit
 - Technologien für die englische und deutsche Sprache
 - Modelle zur Messung der Datenqualität

Massiv paralleles Text Mining mit Hadoop

- Einführung in Map-Reduce und Hadoop
- Integration von UIMA in Hadoop
- Map-Reduce-basierte Extraktion mit Hadoop
- Anfragesprachen für Hadoop (z.B. JAQL, HIVE)

Ausgewählte Anwendungsszenarien

- Hochgenaue Intranet-Suche auf 50 Millionen Seiten
- Social Media Monitoring im Web
- Extraktion von Millionen von Fakten aus News-Seiten

Geschäftsmodelle für Daten-als-Dienstleistung

- Internationale Marktplätze für Daten
- Revenue Sharing, Pay-Per Use, Abonnement
- Exkurs: MIA, ein Marktplatz für Informationen und Analysen auf dem deutschsprachige Web

Aktuelle Entwicklungen und Trends

- Parallele Anfrageprozessoren für Textdaten
- Verifikation von Daten aus dem Web

Verweise zu aktueller Forschung

Lehrmethode / Begleitmaterial

Vorträge, Diskussionen und Werkzeugdemonstrationen. Die Teilnehmenden erhalten Kopien der Vortragsunterlagen, Literaturhinweise und einschlägige Web-Adressen.

Adressaten

Führungskräfte, Entscheider, Analysten und Projektleiter aus Unternehmen aller Branchen und Größenordnungen, die mit der Informationsbeschaffung im Unternehmen oder der Informationsbereitstellung im Intranet betraut bzw. für diese verantwortlich sind.

Voraussetzungen

Grundverständnis von Prozessen für die Erzeugung textueller Inhalte, sowie IT-Grundkenntnisse und Interesse an einer Vertiefung selbiger.

Maximale Teilnehmerzahl

12

Referent

Dr. Alexander Löser ist leitender wissenschaftlicher Mitarbeiter am Fachgebiet Datenbanken und Informationssysteme an der TU Berlin. Mit seinem Team erforscht er Text- und Web-Mining Verfahren, insbesondere für die Verarbeitung von Webseiten in englischer und deutscher Sprache. Zuvor arbeitete er als Projektleiter in der industriellen Forschung, zunächst am IBM Almaden Research Center in San Jose und in später für die SAP AG im Forschungsprogramm "Data Management and Analytics". Seine Arbeit am Avatar Semantic Search System der IBM wurde in den Computer World Horizon Awards 2006 gewürdigt. Im Jahr 2011 erhielt er einen Preis vom BMWi für seine Projektidee MIA: Ein Marktplatz für Informationen und Analysen auf dem Datenbestand des deutschsprachigen Webs.

Anmeldung

TM1_2011

Hiermit melde ich mich zum Seminar

Unternehmensrelevante Informationen, aus dem Internet/Intranet effektiv strukturieren und auffinden

am 16.11.2011 in Berlin in verbindlich an.

Name: _____

Vorname: _____

Firma/
Institution: _____

Firmenanschrift: _____

Telefon: _____

E-mail: _____

[] GI-Mitglied, Mitglieds-Nr.: _____

[] VDE-Mitglied, Mitglieds-Nr.: _____

[] Beschäftigte(r) eines Mitgliedsunternehmens der GI, des BITKOM, BITMI bzw. des IT-Forum Rhein-Neckar

Teilnahmegebühr: EUR _____

Rechnungsanschrift: _____

Ort, Datum: _____

Unterschrift: _____

Veranstaltungsort

Spreepalais Berlin
Seminarraum Fraunhofer-Verbund IUK-Technologie
Anna-Louisa-Karch-Str. 2

10178 Berlin

Tel.: 030 726 15 66 0

Fax: 030 726 15 66 19

Dauer des Seminars

16.11.2011

09:30 bis 17:00 Uhr

Gebühren	Bei Eingang der Anmeldung	
	bis 12.09.2011	ab 13.09.2011
Normalgebühr	875,00 EUR	960,00 EUR
Gebühr für GI- bzw. VDE-Mitglieder sowie für Beschäftigte in Mitgliedsunternehmen der GI bzw. des BITKOM e.V.	790,00 EUR	875,00 EUR

* Beachten Sie die folgenden Preisvorteile

- 15% Rabatt bei Anmeldung von mindestens einem weiteren Kollegen/in derselben Firma

In der Gebühr sind die Begleitunterlagen sowie eine Tagungspauschale in Höhe von 55,00 EUR (inkl. USt.) je Tag enthalten, die das Mittagessen und die Kaffeepausen jeweils vormittags und nachmittags umfasst. Nach § 4 Nr. 22a UStG sind die Leistungen der Deutschen Informatik-Akademie GmbH (DIA) von der Umsatzsteuer befreit.

Allgemeine Teilnahmebedingungen

Anmeldung

Bitte per Internet, E-Mail, Fax oder Brief an:

Initiative D21 e.V.

Reinhardtstr. 38

10117 Berlin

Fon 030 / 5 26 87 22 50

Fax 030 / 5 26 87 22 60

E-Mail: kontakt@initiated21.de

<http://www.initiated21.de/seminar-dia>

Die Anmeldungen werden in der Reihenfolge ihres Eingangs berücksichtigt und schriftlich bestätigt. Mit der Anmeldungsbestätigung wird die Rechnung über die Teilnahmegebühr **durch die Deutsche Informatik-Akademie GmbH** übersandt. Telefonische Vorabreservierungen sind möglich.

Den Teilnehmenden der Veranstaltung wird eine Liste mit ihren Namen und Firmen / Institutionen ausgehändigt. Falls eine Aufnahme in diese Liste nicht erwünscht ist, so bitten wir bei der Anmeldung um einen entsprechenden Hinweis.

Rücktritt

Erfolgt ein Rücktritt von der Anmeldung bis zu zwei Wochen vor Veranstaltungsbeginn, so sind 50,00 EUR Bearbeitungsgebühr zu entrichten; bei einem Rücktritt später als zwei Wochen vor Veranstaltungsbeginn ist die volle Teilnahmegebühr zu entrichten. Eine Ersatzperson kann gestellt werden.

Veranstaltungsausfall

Sollte die Veranstaltung von uns aus wichtigen Gründen abgesagt werden müssen, so besteht Anspruch auf volle Rückerstattung der Teilnahmegebühr; Ansprüche darüber hinaus bestehen nicht.