

Das Web als Firmendatenbank: Big Data für Marketing, Vertrieb und Pressearbeit

Wednesday, October 11, 2017 3:00 PM (30 minutes)

Informationen zu Bestandskunden und potentiellen Kunden sind von großer Bedeutung für erfolgreiche PR-, Marketing- und Vertriebsprozesse. Die aktuellste und größte Informationsquelle für solche Informationen ist das Web. Der größte Teil der Informationen liegt dort allerdings unstrukturiert in Textform vor ist daher für gezielte Abfragen und quantitative Analysen nicht verfügbar –und genau dies zu erreichen ist das Ziel der Echobot Media Technologies GmbH aus Karlsruhe. Echobot betreibt Webcrawler, die täglich ca. eine halbe Mio. Nachrichten-Artikel sowie 2-3 Mio. Social-Media-Postings aus dem Web erfassen und eine Vielzahl von Unternehmenswebsites in regelmäßigen Abständen indexieren. Um diese Informationen verwertbar zu machen, d.h. Wissen aus den Daten zu generieren und es den betroffenen Firmen zuzuordnen, forscht Echobot an Analysemethoden, die entsprechend skalieren. Im Vortrag wird ein Überblick über die damit verbundenen Problemstellungen gegeben und ein Projekt, das in Zusammenarbeit mit dem KIT durchgeführt wurde, genauer vorgestellt. Ziel des Projektes war die Prädiktion der Branchenzugehörigkeit einer Firma auf Basis der Firmen-Website. In einem Zwischenschritt wurden die Website-Inhalte zunächst in Themen-Vektoren überführt, auf deren Basis dann verschiedene Klassifikatoren trainiert und evaluiert wurden. Die hohe Feature-Dimensionalität und große Anzahl verschiedener Klassen stellten dabei besonderen Herausforderungen dar, auf die im Vortrag näher eingegangen wird.

Track

SDIC

Authors: Dr KÖNIG, Christoph (KIT/SCC); Dr KELLER, Matthias (Echobot)

Presenters: Dr KELLER, Matthias (Echobot); Mr SCHLITTER, Nico (KIT/SCC)

Session Classification: Use-Cases