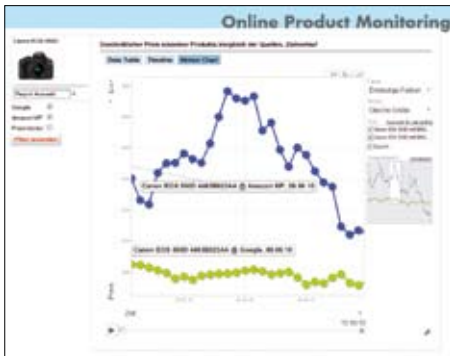


## ■ Integration von Webdaten



Beobachtung und Erfassung von Preisentwicklungen im Internet mit der Online Product Monitoring – Anwendung PROOF, z. B. für Produkte (Elektronik, Markenartikel), Flugpreise, Energiekosten.



**Unternehmen** nutzen das Internet immer stärker als Vertriebskanal sowie zur Beschaffung von Informationen für Business Intelligence-Auswertungen. Vielfach ist hierbei die Integration von Daten unterschiedlicher Herkunft erforderlich, etwa um in Preisvergleichsportalen Produktangebote vieler Anbieter zu bündeln. Die Daten sind dabei oft sehr unterschiedlich strukturiert, unvollständig oder gar fehlerhaft. Daten verschiedener Websites sind ferner oft nur durch aufwändige Such- und Extraktionsschritte zu gewinnen und unterliegen häufigen Änderungen. Diese Randbedingungen führen dazu, dass die korrekte Integration großer Mengen an Webdaten sehr schwierig ist und mit herkömmlichen Integrationsansätzen wie Data-Warehouse-Technologien nicht ausreichend gelöst werden kann.

Im WDI-Lab (Web Data Integration Lab) der Universität Leipzig werden daher unter Leitung von Prof. Dr. Erhard Rahm sehr leistungsfähige und innovative Integrationstechnologien und -lösungen für den Markteinsatz entwickelt. Das WDI-Lab ist ein Anfang 2010 gegründetes Innovationslabor, das vom BMBF im Rahmen des ForMaT-Programms (»Forschung für den Markt im Team«) gefördert wird. Im Rahmen mehrerer Arbeitsgruppen wurden bereits mächtige Werkzeuge entwickelt und erprobt, u. a. zur Definition von Workflows, zur Laufzeitintegration von Web-

daten, zum semantischem Abgleich (Matching) großer Ontologien (z. B. Produktkataloge) sowie zum Kategorisieren und Matching großer Mengen von Datenobjekten (z. B. Produktangebote). Auf Basis dieser Werkzeuge konnten bereits mehrere Anwendungslösungen realisiert werden, von denen im Folgenden einige beispielhaft vorgestellt werden.

Die im WDI-Lab entwickelte Lösung »Online Product Manager« bietet für Web-Shops und Internet-Portale eine automatische Kategorisierung und Matching von Produktangeboten in einer Anwendung und wird für große Datenmengen optimiert. Sie unterstützt die Integration von Produktangeboten aus unterschiedlichen Quellen

arbeitung und Anreicherung der Eingangsdaten. Diese Lösung wurde speziell für Portalbetreiber und E-Commerce Unternehmen konzipiert, eine Anpassung für andere Unternehmensbereiche und Branchen kann je nach Bedarf erfolgen.

Die Online Product Monitoring-Anwendung PROOF ermöglicht es Unternehmen, Preise und Produkte von Mitbewerbern im Internet automatisiert zu beobachten. Informiert zu sein über aktuelle Preisentwicklungen sichert Herstellern und Händlern gerade im Online-Handel die eigene Marktposition. PROOF lässt sich leicht um neue Datenquellen erweitern und auf verschiedene Branchen und Bereiche übertragen. Durch den Einsatz mächtiger Query-Technologien können

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung



in den eigenen Produktkatalog (Hierarchie von Produktkategorien). Zur Sicherstellung einer hohen Datenqualität dient eine umfassende Vorver-

*Hanna Köpcke (rechts) und Prof. Dr. Erhard Rahm (z. von rechts) präsentierten die im WDI-Lab entwickelten Anwendungslösungen auf der GFFT-Jahrestagung*



unterschiedliche Anfrageschnittstellen genutzt sowie die Qualität und die Effizienz der Datenabfragen optimiert werden. Die Steuerung der Monitoring-Jobs erfolgt über eine übersichtliche Browser-Oberfläche.

Für die weitere Forschungsarbeit ist das WDI-Lab sehr an Kooperationen mit Unternehmen unterschiedlicher Branchen interessiert. Ausführliche Informationen zu den hier vorgestellten und weiteren Anwendungen finden Sie unter <http://wdilab.uni-leipzig.de>.

Prof. Dr. Erhard Rahm  
Universität Leipzig  
Institut für Informatik  
WDI-Lab  
Dittrichring 5-7  
04109 Leipzig

Telefon (03 41) 97-3 22 41  
Fax (03 41) 97-3 22 09  
E-Mail [rahm@informatik.uni-leipzig.de](mailto:rahm@informatik.uni-leipzig.de)  
[www.wdilab.uni-leipzig.de](http://www.wdilab.uni-leipzig.de)