

HighQ: Informationsintegration mit dem IBM Information Server

Hintergrund

Die erfolgreiche Abwicklung inner- und zwischenbetrieblicher Geschäftsprozesse hängt in einem hohen Maße von einer effektiven Integration der benötigten Informationssysteme ab. Diese Informationssysteme sind verteilt und autonom; die darin gespeicherten Daten sind **heterogen** und **verschmutzt**. Integrationssysteme ermöglichen den einheitlichen Zugriff auf diese Daten und bieten Nutzern und Anwendungen eine einheitliche und integrierende Schnittstelle. Oft müssen dabei neben klassischen Datenbanken auch Systeme integriert werden, die innovative Technologien wie Service-orientierte Architekturen (SOA), Web Services, XML oder Radio Frequency Identification (RFID) nutzen.

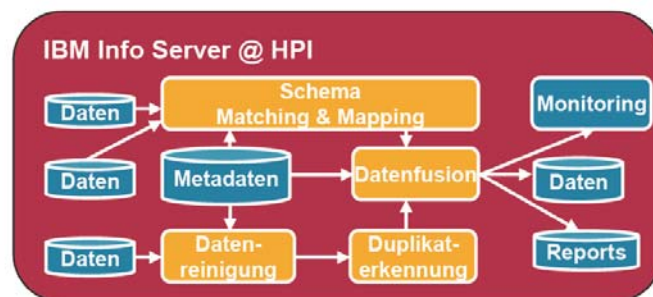
Die Problematik der Integration bestehender Datenbestände, z.B. Kundendatenbestände, ist der Hintergrund dieses Bachelorprojekts. Hinter dem Begriff der Informationsintegration verbergen sich Techniken der **Datenreinigung** (data cleansing) und **Duplikaterkennung**, der Informationsfusion und Datenkonsolidierung und des Data Warehousing.

Rahmenbedingungen

Das Projekt wird vom Hasso Plattner Institut, vertreten durch das Fachgebiet Informationssysteme, in **Zusammenarbeit mit IBM Deutschland** durchgeführt. Dafür stellt IBM u.a. den IBM Information Server, eine Software Plattform zur Informationsintegration, zur Verfügung. Zudem wird mittels IBM-Fördergeldern ein „**Information Server Lab**“ am HPI (2 Server und 10 hochwertige Desktops) eingerichtet.

Der IBM Information Server wird im Fachgebiet Informationssysteme bereits erfolgreich in Lehre und Forschung eingesetzt. Begleitend zu diesem Projekt werden zusätzliche Veranstaltungen, Workshops und Diskussionen stattfinden, zu denen die Teilnehmer des Bachelorprojekts herzlich eingeladen sind.

Die Teilnehmerzahl ist auf 8 Mitglieder beschränkt. Die technische Umsetzung erfolgt mit Java.



Projektbeschreibung

Der IBM Information Server umfasst bereits umfangreiche Funktionalitäten zum Verständnis, Bereinigen, Umwandeln und Bereitstellen unabhängiger, heterogener Daten aus verschiedenen Quellen. Ziel dieses Projektes ist es, den IBM Information Server um **neue Methoden aus der Forschung**, z.B. zur Duplikaterkennung, zu erweitern, die eine optimale Integration heterogener und autonomer Informationsquellen in verschiedensten Szenarien ermöglichen.

Zusätzlich unterstützt der IBM Information Server den Aufbau einer **Service-orientierten Architektur** (SOA), um implementierte Funktionalitäten als Integrationsservices in verschiedene Prozessen und Szenarien wiederverwenden zu können.

Es werden vom Fachgebiet Informationssysteme und IBM mehrere **Integrationsszenarien (use cases) aus dem SAP Umfeld** zur Verfügung gestellt, für die unter Verwendung des erweiterten IBM Information Servers ein konkretes Integrationssystem implementiert werden soll.

Projektvorbereitung

In der Vorbereitungsphase werden Grundlagen der Informationsintegration vorgestellt. Die Teilnehmer lernen Techniken der Informationsintegration und deren Optimierung kennen. Darüber hinaus werden Spezialthemen, etwa Methoden der **Informationsqualität** und Datenreinigung, der Informationssuche und des **Metadatenmanagements** behandelt. Diese Themen werden im Rahmen regelmäßiger Treffen sowohl theoretisch als auch praktisch in regelmäßigen Treffen im Wintersemester 2007/08 durch die Teilnehmer bearbeitet und vorgestellt. Detaillierte Kenntnisse zum Arbeiten mit dem IBM Information Server werden im Vorbereitungsseminar zusätzlich vermittelt und erste Methoden werden umgesetzt.

Projektbeginn: 15.10.2007

Kontakt

Für weiterführende Informationen stehen Prof. Dr. Felix Naumann und Alexander Albrecht zur Verfügung. Eine Terminabsprache ist über das Sekretariat von Prof. Naumann möglich: office-naumann@hpi.uni-potsdam.de

http://www.hpi.uni-potsdam.de/naumann/lehre/bachelorprojekte/ibm_information_server.html

<http://www.ibm.com/de/entwicklung/>

<http://www.ibm.com/software/de/db2/integration/>