



wdilab

Web Data Integration

Workflowbasierte Datenintegration

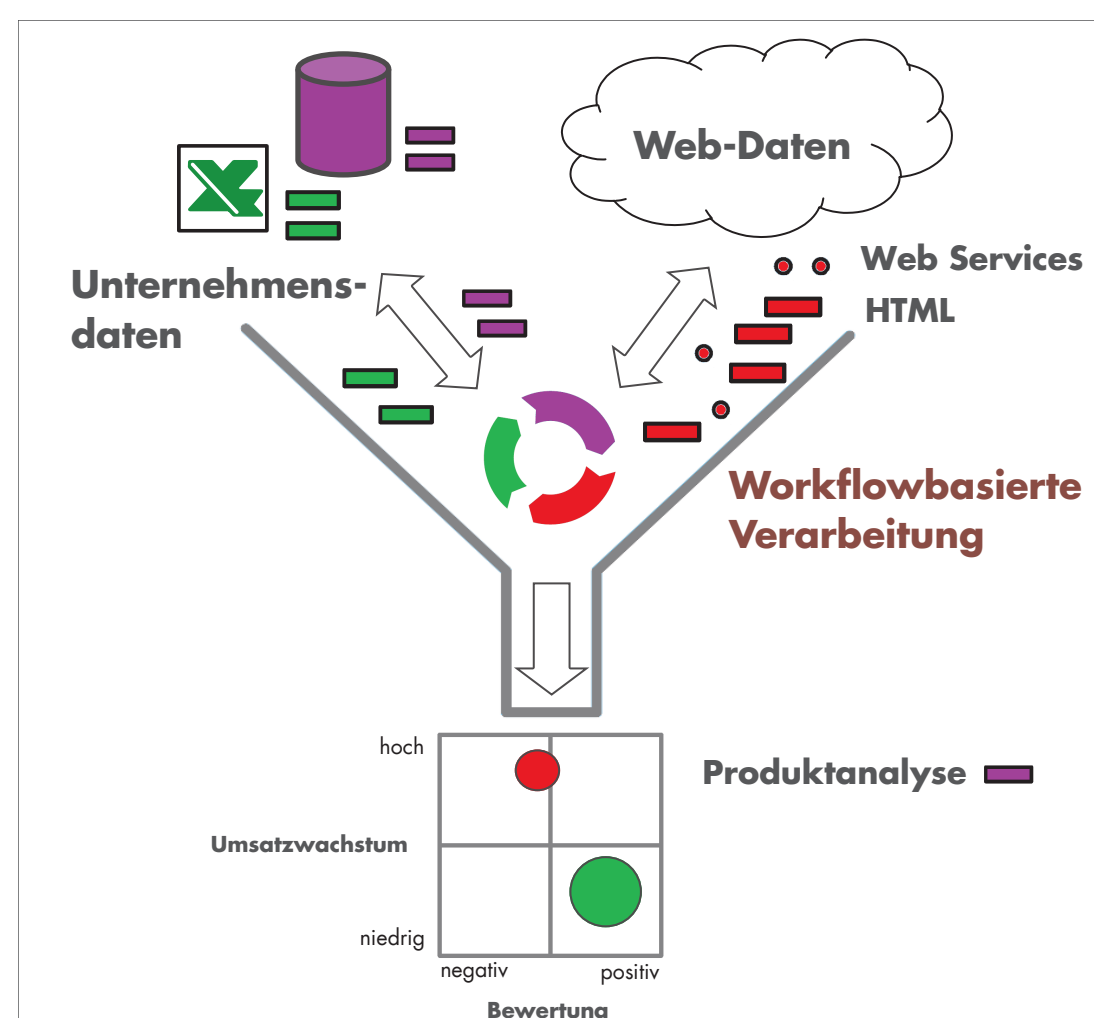
Motivation

Unternehmen integrieren ihre eigenen Daten mit Hilfe aufwändiger Datenbanktechnologien, z.B. für effiziente Analysen innerhalb von Business-Intelligence-Projekten.

- Hoher Aufwand für die Definition und Ausführung der Datenintegration
- Schwierig ist die Integration von Informationen, die nur über Webdatenquellen zugänglich sind
- Einhaltung der Prämissen: Datenqualität und Effizienz

Ziele

- Einfache, flexible Anwendung
- Generierung von optimierten Anfragen
- Workflowbasierte Extraktion, Transformation und Integration relevanter Web-Daten



Anwendungsbeispiel Produktanalyse

Anwendung

- Automatisches Web-Monitoring
- Integration relevanter Web-Daten in unternehmenseigene Datenbestände (z.B. Produktdaten, Produktbewertungen)

Innovation

- Zielgerichtetes Auffinden relevanter Daten sowie deren flexible Verknüpfung
- Kombination intelligenter Anfragen und anschließender Ermittlung von Übereinstimmungen (Matching)
- Neuartiger workflowbasierter Ansatz zur dynamischen Integration von Web-Daten und Services

Matching von Schemas und Ontologien

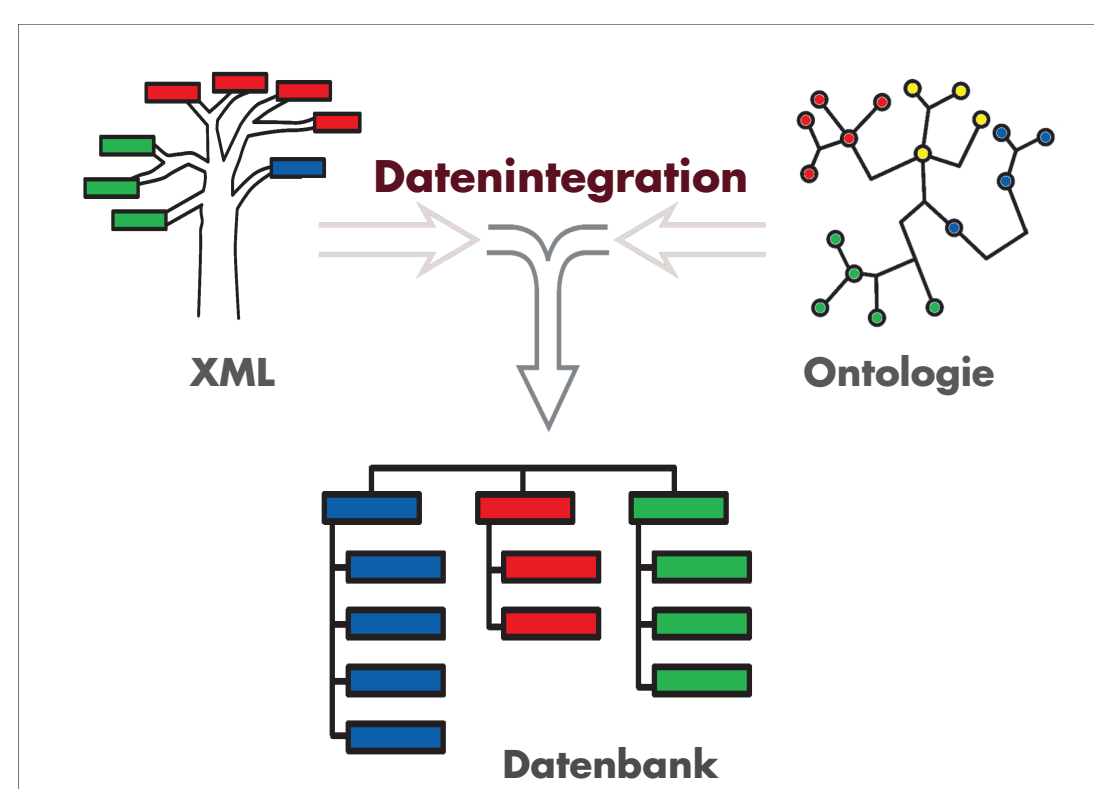
Motivation

Unternehmen strukturieren und beschreiben ihre Daten, z.B. Kunden und Bestellvorgänge, mit Hilfe von Schemas und Ontologien (z.B. Produktkataloge).

- Datenaustausch und -abgleich führt durch Heterogenität der Darstellungen zu einem aufwändigen und auch fehleranfälligen Prozess der Datenintegration
- Jede Weiterentwicklung von Schemas und Ontologien erfordert eine erneute Metadatenintegration

Ziele

- Generierung von Datentransformations-Queries
- Geringer Verwaltungsaufwand
- Optimierung für große Schemas
- Unterstützung für Schema-Evolution



Anwendungsbeispiel Datenintegration

Anwendung

- Datentransformation von Kommunikations-, Produkt- und Bestelldaten entsprechend der Unternehmensformate
- Integration von Daten bei Unternehmensfusionen
- Datentransformation für Business-Intelligence

Innovation

- Flexible Rahmenarchitektur zur optimierten und (semi-)automatischen Metadatenintegration
- Generischer Ansatz—erreicht qualitativ hochwertige Ergebnisse

Objekt-Matching, Dublettenbehandlung

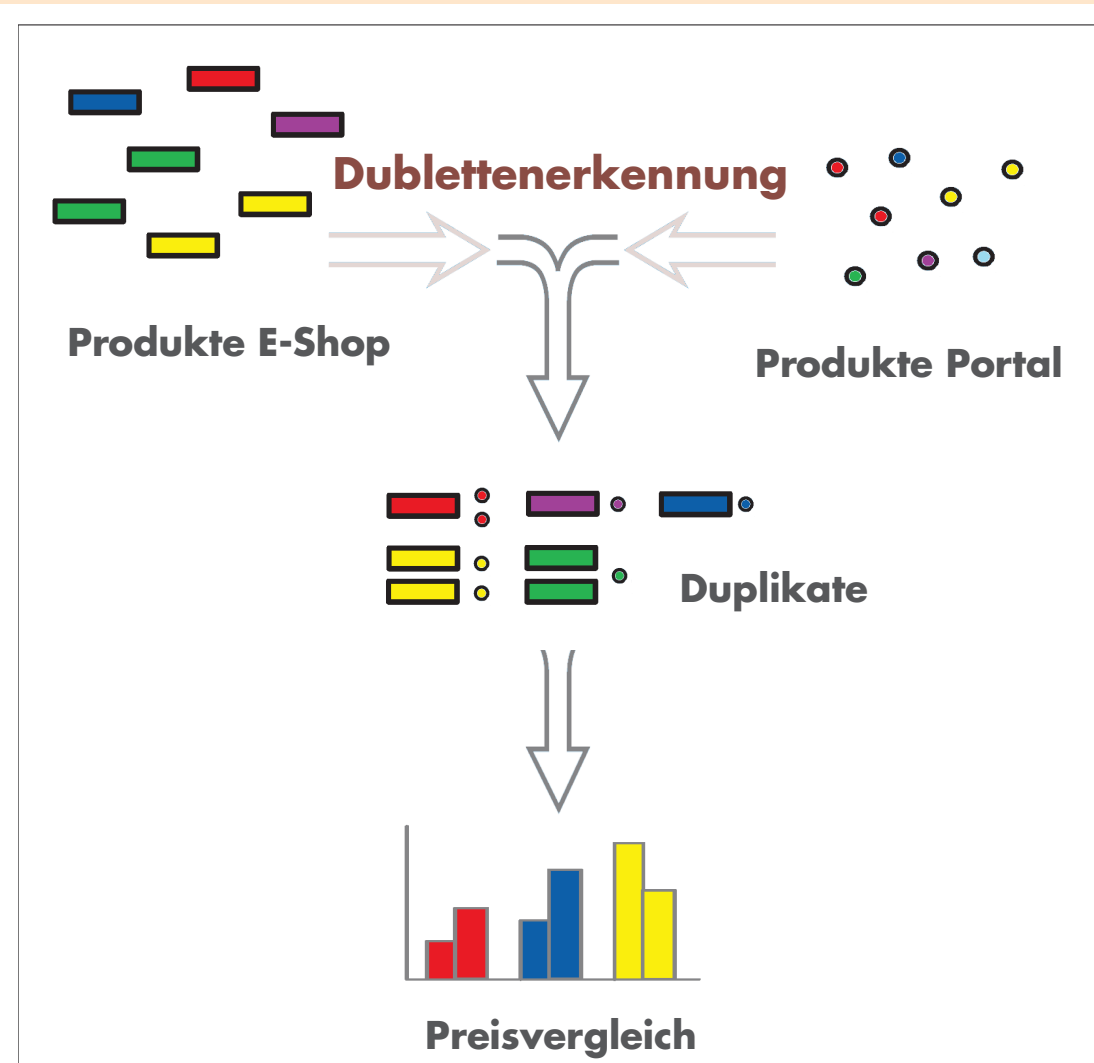
Motivation

Unternehmen führen häufig Daten aus verschiedenen Datenquellen zusammen, z.B. in Business-Intelligence-Projekten, bei Unternehmensfusionen oder Kundentransaktionen.

- Erhöhung der Datenqualität
- Aktualisierung von Datenbeständen und Datenanalyse

Ziele

- Kombination verschiedener Verfahren zur Sicherung einer sehr guten Ergebnisqualität
- Wiederverwendung von Ergebnissen zur Performance- und Qualitätssteigerung
- Automatisches Tuning



Anwendungsbeispiel Preisvergleich

Anwendung

- Integration und Aufbereitung unternehmensinterner bzw. unternehmensexterner Daten (Web-Daten) z.B. zur
- Bereinigung der Kundenbestände
- Erstellung von Konkurrenzanalysen (Produkt, Preis, Anbieter, Zielgruppen)

Innovation

- Flexible Rahmenarchitektur zur optimierten und (semi-)automatischen Dublettenerkennung
- Generischer Ansatz—erreicht qualitativ hochwertige Ergebnisse