

# Das Zusammenspiel von Agenten- und Föderationstechniken innerhalb von BUSINESS-MV

*Gunnar Weber    Andreas Heuer*  
Lehrstuhl Datenbank- und Informationssysteme  
Fachbereich Informatik, Universität Rostock

*Claudia Herzig*  
Fraunhofer Institut für  
Graphische Datenverarbeitung

## 1 Das Projekt BUSINESS-MV

Die Kooperationsplattform BUSINESS-MV<sup>1</sup> [3, 2] wird kleinen und mittelständischen Unternehmen sowie den in die Zusammenarbeit einbezogenen Institutionen und Behörden die Möglichkeit bieten, sich in Projekten, Gemeinschaften und Konsortien bis hin zu "Virtuellen Unternehmen" einzubringen und innerhalb dieser in einfacher Weise zu kommunizieren und zu kooperieren. Im Rahmen des Projektes werden speziell über das Internet zugängliche Anwendungsdienste für die Zusammenarbeit zwischen kooperierenden Unternehmen, Unternehmen und Behörden sowie zwischen Unternehmen und Kunden bereitgestellt. Um vorhandene bzw. entstehende Informationsdienste und Kommunikationsmedien in Mecklenburg-Vorpommern effektiv nutzen zu können, müssen diese in das System integriert werden. Ziel ist zum einen die Einbindung verschiedenartigster Informationsquellen (Datenbanken unterschiedlichen Typs, Suchmaschinen, WWW-Seiten, Textdokumente, ...) und zum anderen die Gewährleistung einer hohen Qualität der Integration. Aus diesem Grund wurde der in Abbildung 1 skizzierte Ansatz gewählt, der

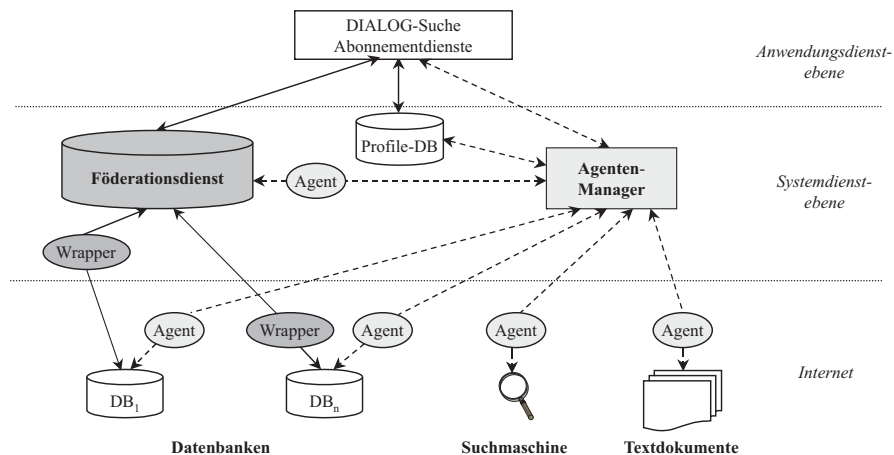


Abbildung 1: Agenten- und Föderationstechniken im Zusammenspiel

sowohl Agenten- als auch Föderationstechniken nutzt. Suchmaschinen wie SWING [5] und Textdokumente werden über Agenten in das System eingebunden. Die Integration von Datenbanken,

<sup>1</sup>BUSINESS-MV ist ein vom Land Mecklenburg-Vorpommern gefördertes Verbundprojekt. Partner sind das Fraunhofer IGD Rostock, die Medeocom GmbH, die Universität Rostock und das ZGDV Rostock.

die einen Vollzugriff<sup>2</sup> bieten, erfolgt über den Föderationsdienst und einen stationären Agenten. Datenbanken, die nur über eine Web-Schnittstelle angesprochen werden können, nehmen eine Zwitterstellung ein: hier muß man in Abhängigkeit von der Qualität der Schnittstelle entscheiden, ob die Datenquelle in den Föderationsdienst eingebunden wird. Im folgenden werden beide Integrationsmethoden detaillierter vorgestellt sowie Vor- und Nachteile diskutiert. Ein Vergleich der beiden Techniken erfolgt in Abschnitt 4.

## 2 Der Agentenansatz

Der Einsatz autonomer mobiler Agenten zur Informationsgewinnung erfolgt nach folgenden Prinzip. Der Agent bewegt direkt sich zur Datenquelle (z.B. DB-Server). Auf dieser befindet sich ein speziell an diese Datenquelle angepaßter stationärer Agent, welchen der mobile Suchagent kontaktiert und die Anfrage übermittelt. Der stationäre Agent sucht daraufhin die entsprechenden Informationen und teilt sie dem Suchagenten mit, welcher sie mit bereits gesammelten Ergebnissen vergleicht und bewertet. Nach Eliminierung eventueller Duplikate und weniger relevanter Ergebnisse bewegt sich der Suchagent dann weiter zur nächsten Datenquelle oder zurück zum Rechner, auf dem er vom Agenten-Manager gestartet wurde. Von dort aus werden die Ergebnisse an den Nutzer übertragen [4].

Die Art der Erzeugung und Aktivierung der Suchagenten ist abhängig vom Anwendungskontext. Zeitkritische Anfragen, bei denen der Anwender online seine Anfrage stellt und auf die Anzeige des Ergebnisses im Browserfenster wartet, werden anders behandelt als Abonnements oder ad-hoc-Anfragen, deren Ergebnis per Fax oder E-Mail zugestellt werden soll. Im ersten Fall werden mehrere Agenten pro Anfrage aktiviert, welche parallel zu jeweils einer Datenquelle reisen, im zweiten Fall besucht ein einzelner Agent sequentiell alle Datenquellen (Abbildung 2).

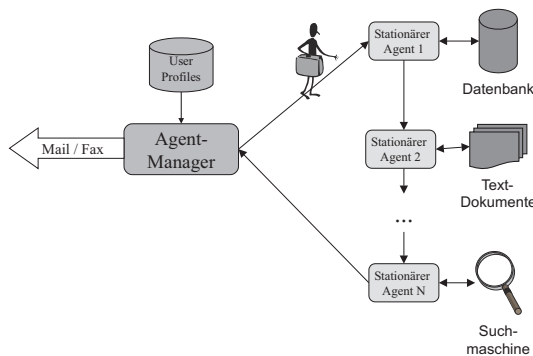


Abbildung 2: Agentenansatz

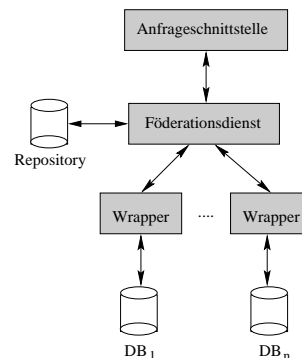


Abbildung 3: Föderationsdienst

### 2.1 Vorteile

Der Einsatz von Agenten ermöglicht die einfache Einbindung der verschiedenartigsten Datenquellen. Der Zugriff auf die jeweiligen Datenquellen wird von einem stationären Agenten realisiert, mit dem der mobile Suchagent auf immer dieselbe Art kommuniziert. Dieser stationäre Agent kann vom Anbieter der Datenquelle selbst auf relativ einfache Weise implementiert werden, wodurch der Anbieter die volle Kontrolle über den Zugriff auf seine Daten behält.

<sup>2</sup>In diesem Fall können beliebige Anfragen an die Datenbank gestellt werden und das Ergebnis wird in einem strukturierten Ausgabestrom geliefert.

Durch die Verwendung mobiler Agenten ergibt sich insbesondere bei sequentiellm Besuch der Datenquellen ein zusätzlicher Vorteil [6]. Da sich die Agenten direkt an der Datenquelle befinden, dort lokal kommunizieren und nur die für den Anwender relevantesten Daten mit sich nehmen, reduziert sich das insgesamt übertragene Datenvolumen.

Der Einsatz von Agenten erlaubt eine Personalisierung der angebotenen Dienste. Zur Bewertung der Relevanz kann der Agent bei Bedarf selbständig Informationen aus dem Nutzerprofil heranziehen. Ein Agent ist in der Lage, auf der Basis bereits gesammelter Erfahrungen eine Optimierung der Suchanfragen vornehmen - beispielsweise könnten Datenquellen, die für den entsprechenden Anwender in der Vergangenheit gute Ergebnisse geliefert haben, bevorzugt angesteuert werden.

## 2.2 Nachteile

Die Verwendung eines stationären Agenten zum Zugriff auf die Datenquellen erfordert die Implementation eines solchen Agenten für jede Datenquelle, wobei diese je nach Art der Datenquelle recht aufwendig werden kann. Um den Empfang mobiler Agenten und eine Kommunikation mit dem stationären Agenten zu gewährleisten, muß ein Agenten-Server (in diesem Fall der Voyager ORB) installiert werden. Die damit verbundenen Kosten und Risiken könnten Datenbankanbieter von einer Unterstützung der Agenten Abstand nehmen lassen.

Aufgrund der beabsichtigten Einbindung verschiedenartigster Datenquellen können zur Kombination und Bewertung der Ergebnisse nicht die in Datenbanksystemen bereits vorhandenen Mechanismen genutzt werden, sondern die Algorithmen (Merging, Sortierung, Duplikateliminierung, ...) müssen "von Hand" implementiert werden.

## 3 Der Föderationsansatz

Der Föderationsdienst bietet eine weitere Zugriffsmöglichkeit auf die in das System eingebundenen Informationsquellen. Da diese thematisch stark voneinander abweichen, werden inhaltlich verwandte Quellen in einer Anwendungsdomäne geclustert. Für jede Domäne wird dann eine einheitliche Sicht auf die zugeordneten Datenquellen definiert. Diese Sicht ist vergleichbar mit dem globalen Schema in föderierten Datenbanksystemen (FDBMS), die Datenquellen werden im folgenden wie in FDBMS üblich als Komponentendatenbanken bezeichnet [7, 1].

Da die Unterschiede zwischen den einzelnen Komponentendatenbanken hinsichtlich des Datenmodells, des Schemas, des zugrundeliegenden Datenbanksystems, der Möglichkeit des Zugriffs<sup>3</sup> bzw. der Abfragemöglichkeiten sehr groß sein können, wurde eine Architektur (Abbildung 3) gewählt, die sehr viel Spielraum in alle Richtungen läßt.

Im Repository werden die globalen Schemata für die Anwendungsdomänen mit verschiedenen Metainformationen sowie die relevanten Schemabestandteile der Komponentendatenbanken gespeichert. Den Kern des Repositories bilden die Abbildungsvorschriften, die für die Übersetzung einer globalen Anfrage an den Föderationsdienst in lokale Anfragen an die Komponentendatenbanken benötigt werden. Diese Regeln umfassen zum einen Vorschriften zum Ermitteln der in Frage kommenden Anwendungsobjekte (Selektionsregeln) und zum anderen Vorschriften zum Bestimmen der Attributwerte für diese Objekte (Projektionsregeln).

Der Föderationsdienst analysiert zunächst die globale Anfrage und generiert dann für die ermittelten Komponentendatenbanken auf Grundlage der Informationen aus dem Repository die loka-

---

<sup>3</sup>Der Zugriff auf die Datenquelle kann direkt über JDBC/ODBC (entweder lokal auf Datenbankseite oder remote auf Föderationsdienst-Seite) oder nur über eine Web-Schnittstelle möglich sein.

len Anfragen. Diese werden dann an den für die Komponentendatenbank zuständigen Wrapper weitergeleitet, der die Verbindung zur Datenquelle aufbaut und die Anfragen abarbeitet. Jeder Wrapper liefert eine Ergebnisrelation, die entsprechend dem globalen Schema aufgebaut ist und eine Sicht auf die jeweilige Komponentendatenbank darstellt. Somit werden die Schemakonflikte bereits auf der Wrapperebene beseitigt. Die einzelnen Ergebnisrelationen werden dann vom Föderationsdienst zum Gesamtergebnis zusammengefügt, wobei gleiche Realweltobjekte erkannt und eventuell vorkommene Datenheterogenitäten beseitigt werden müssen.

### 3.1 Vorteile

Der Föderationsdienst bietet eine hohe Integrationsqualität, da sowohl Konflikte auf Schema- als auch auf Datenebene bereinigt werden. Die Auflösung der Schemakonflikte erfolgt größtenteils durch die Abbildungsregeln<sup>4</sup>. Somit ist bei Schemaänderungen in den meisten Fällen nur eine Redefinition der Abbildungsregeln notwendig.

Die Anfragesprache des Föderationsdienstes kann die volle Mächtigkeit von Datenbanksystemen bieten - zum einen werden die Anfragemöglichkeiten der Komponentendatenbanken bei der dynamischen Generierung der Anfragen ausgenutzt und zum anderen stehen auf Föderationsdienst-Seite globale Anfragefunktionalitäten zur Verfügung. Diese Mächtigkeit kann z.B. bei der Verknüpfung von Datenquellen, die in einer Fremdschlüssel-Beziehung stehen, ausgenutzt werden.

### 3.2 Nachteile

Die hohe Integrationsqualität beschränkt die Anzahl der Datenquellen, die in die Föderationsdienst eingebunden werden können - die Datenquelle muß strukturierte Anfragen verarbeiten und ein Ergebnis liefern, das in das globale Schema der in Frage kommenden Anwendungsdomäne integrierbar ist.

Das Konzept des Föderationsdienstes sieht keine Nachbearbeitung von Anfrageergebnissen vor. Somit ist keine Ergebnisverdichtung möglich, d.h. es können keine Informationen aus dem Nutzerprofil zum Reduzieren der Ergebnismenge herangezogen werden. Die Verarbeitung der Profilinformationen in der Anfrage stellt auch keine befriedigende Lösung dar, da es sich bei diesen Informationen eher um vage als um exakte Angaben handelt.

## 4 Vergleichende Betrachtung

Die beiden vorangegangenen Abschnitte haben gezeigt, daß in Abhängigkeit vom Anwendungsfall entschieden werden muß, welcher Ansatz besser geeignet bzw. in welchem Fall eine Kombination beider Techniken vorteilhaft ist. Die Tabelle 1 zeigt einen Vergleich der beiden Ansätze anhand ausgewählter Kriterien.

Die Agententechnik erlaubt die Einbindung auch solcher Informationsquellen, die nicht oder nur mit sehr hohem Aufwand in den Föderationsdienst integriert werden können. Somit ist der Einsatz der Agententechnik vor allem für allgemeine Suchanfragen geeignet, bei denen keine hohe Qualität, aber eine große Abdeckung der vorhandenen Quellen gefordert ist. Die Verwendung von Agenten ist z.B. auch notwendig, wenn Informationen aus dem Nutzerprofil zur Verdichtung des Ergebnisses herangezogen werden. Sollen exakte, scharfe Anfragen mit weniger Informationsquellen, aber hoher Integrationsqualität verarbeitet werden, dann kommt der

---

<sup>4</sup>In einigen Fällen werden Nachbearbeitungsmethoden in den Wrappern benutzt, um Formatanpassungen vorzunehmen.

Kriterium	Agententechnik	Föderationstechnik
Ausführungszeitpunkt	zeitunabhängig	zeitnah
Kopplung	lose	eng
Autonomie	sehr autonom	eingeschränkt autonom (Abbildungsregeln)
Qualität der Integration	niedrig	hoch
Anfragebearbeitung	flexibel (parallel, sequentiell)	parallel
Mächtigkeit der Anfragesprache	gering	voll
Verknüpfung von Datenquellen (über gleiche Attribute)	möglich, aber hoher Programmieraufwand	über Joins
Einbeziehung von Profilen	in der Ergebnisanalyse Fuzzy-Profil	in der Anfrage Exaktes Profil
Lastverteilung	immer (mobile Agenten)	möglich (Wrapper auf DB-Seite)

Tabelle 1: Vergleich beider Techniken

Föderationsdienst zum Einsatz. Ein weiterer Anwendungsfall für Föderationstechniken ergibt sich, wenn z.B. Firmen- und Produktdatenbanken mit überschneidenden Inhalten über gleiche Attribute miteinander verknüpft werden sollen.

Agenten können auf den Föderationsdienst aufsetzen. Dies ist beispielsweise dann sinnvoll, wenn eine Anpassung eines vom Föderationsdienst gelieferten Ergebnisses an Nutzeranforderungen notwendig ist.

## 5 Zusammenfassung

Der in BUSINESS-MV gewählte Ansatz, sowohl Agenten- als auch Föderationstechniken zur Integration von Informationsquellen zu nutzen, zeichnet sich durch seine Flexibilität aus. In Abhängigkeit vom Anwendungsszenario kann entweder eine Vielzahl von Informationsquellen durchsucht oder eine hohe Qualität des Ergebnisses in Bezug auf den Suchauftrag erreicht werden. In einigen Fällen werden die Vorteile beider Techniken kombiniert. Agenten können die vom Föderationsdienst gelieferten hochwertigen Ergebnisse nutzen und mit den von anderen Quellen gesammelten Ergebnissen kombinieren bzw. eine Ergebnisverdichtung vornehmen.

## Literatur

- [1] S. Conrad. *Föderierte Datenbanksysteme - Konzepte der Datenintegration*. Springer, Berlin, 1997.
- [2] N. Günther and G. Flach. Business-MV - Die Kooperationsplattform für Mecklenburg-Vorpommern. In *2. IuK-Tage Mecklenburg-Vorpommern*. Wirtschaftsministerium Mecklenburg-Vorpommern, 1999.
- [3] N. Günther, G. Flach, and T. Courvoisier. Business-MV -Internet-based Services for Database Federation, Communication and Cooperation. *Computer Graphic topics*, 2000(2):7,8, 2000.
- [4] C. Herzig, B. Hockauf, and B. Urban. Agent based Services in Business-MV. *Computer Graphic topics*, 2000(2):5,6, 2000.
- [5] A. Heuer, H. Meyer, and G. Weber. SWING: Die Suchmaschine des Landesinformationssystems MV-Info. In *2. IuK-Tage Mecklenburg-Vorpommern*. Wirtschaftsministerium Mecklenburg-Vorp., 1999.
- [6] D. B. Lange and M. Oshima. *Programming an Deploying Java Mobile Agents with Aglets*. Addison-Wesley, Reading, 1998.
- [7] A. Sheth and J. Larson. Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. *ACM Computing Surveys*, 22(3):183–236, 1990.