

## 10.2 Intelligente Informationsagenten für Wissensentdeckung und Data Mining im Internet

**Matthias Klusch**

**DFKI GmbH Saarbrücken, Deutschland**

[klusch@dfki.de](mailto:klusch@dfki.de), <http://www.dfki.de/~klusch/>

### 1 Das Internet, Web und Agenten

Inzwischen ist die Verfügbarkeit einer schier unerschöpflichen Informationslandschaft, elektronischen Dienstleistungen und die Entwicklung von entsprechenden Märkten im Internet real; die 'Infosphere' ist längst nicht mehr nur eine Vision. Dazu gehören neben populären Anwendungen wie Web-TV, Online Banking oder Shopping auch ortstransparenter Zugriff auf integrierte, multimedial unterstützte Dienstleistungen im Internet mit mobiler Telekommunikation (WAP-fähige Webhandies und Palms).

Die wesentlichen, damit einhergehenden Probleme sind vor allem die Bewältigung von Masse, Verteilung und Vielfalt der im Netz verfügbaren Daten sowie deren Auswahl und Verarbeitung nach Bedarf. Dabei handelt es sich nicht nur um die bis dato schätzungsweise 400 Millionen HTML-Seiten im hypertextbasierten Informationsdienst Worldwide Web sondern gerade auch um jegliche Arten von Altlastendatenbanken, Dateisysteme, Multimediasysteme und andere Softwareapplikationen. Im Jahr 2001 werden mehr als 220 Millionen Menschen Zugang zum Internet besitzen; sie alle werden mit dem Problem einer Suche nach für sie relevanten Informationen in dann etwa einer Milliarde indizierten Webseiten konfrontiert sein. Ohne jegliche Hilfe sind hierbei besonders problematisch

- die Verwendung von verschiedenen Anfragesprachen und Tools für Anfragen und Verarbeitung von Daten, wie Text, Sprache oder Bildinhalte, die in verschiedenen Datenmodellen, semi-strukturiert oder gar unstrukturiert vorliegen, sowie
- Auswahl und Einsatz von geeigneten Ontologien und Vokabularen bei der semantischen Korrelation von verschiedenen Informationen.

Eine weitere Herausforderung besteht insbesondere in der effizienten Handhabung von kulturellen, sprachlichen und sozialen Differenzen zwischen Nutzer(gruppen) im globalen Informationsraum des Internet. Zudem ist noch nicht abzusehen, welche Folgen die laufenden Bemühungen um eine drastische Anhebung der Übertragungsraten im Internet der nächsten Generation (Internet-2/Next Generation Internet) auf 2.4 Gbit/s Ende 2000 und weiter bis zu geplanten 9.6 Gbit/s auf die Situation der Nutzer im öffentlichen Internet haben wird.

#### 1.1 Informationssuche im Web: Indexe und Suchmaschinen

Ein grundlegendes Problem ist die Suche nach relevanten Informationen im Internet. Die bisher populärste Lösung besteht in der Verwendung von sog. Webverzeichnissen und Suchmaschinen (Lawrence & Giles 1998).

- *Web Directories*. Verzeichnisse im Web, wie beispielsweise *Yahoo!*, *looksmart*, *Magellan* oder *InfoMine*, bieten dem Nutzer Anfragen über annotierten und indizierten Einträgen zu verschiedenen, hierarchisch strukturierten Bereichen oder Kategorien von im Web verfügbaren Informationen an. Meist werden auch bewertende Zusammenfassungen der zugrundeliegenden Informationsquellen, sog. ratings und reviews, angeboten. Webverzeichnisse werden fast ausschließlich manuell erstellt und auf den neuesten Stand

gebracht. Ältere, monolithische und ebenso hierarchisch strukturierte Verzeichnisse samt Anfragedienst im Internet sind die bekannten Systeme *Harvest*, *Gopher* und *WAIS*.

- *Suchmaschinen*. Eine weitgehend automatisierte Suche und Indizierung von Informationen im Web erfolgt durch sog. Suchmaschinen, wie *Lycos*, *AltaVista*, *HotBot*, *InfoSeek*, *Excite* oder *WebCrawler*. Jede Suchmaschine (SearchBots 1999) bedient sich eines eigenen Softwareroboters, einem sog. SearchBot, der ausgehend von einem initialen Satz an Verweisen auf Webseiten nach weiteren, indizierten und nach dem Robot Exclusion Protocol frei verfügbaren HTML-Seiten im Web rekursiv nach relevanten Texten und Verweisen durchforstet. Dabei kommen grundlegende Techniken des Information Retrieval zum Einsatz (Gudivada 1997). Aus den gelieferten Daten erzeugt die Suchmaschine dann einen eigenen Web-Index, den sie periodisch durch ihren Search Robot auf den neuesten Stand bringen läßt. Wie bei den Webverzeichnissen werden dem Benutzer einer Suchmaschine nur eine auf einfachen regulären Ausdrücken beruhende Anfragesprache geboten, die er zur Formulierung seiner Suchanfragen nutzen kann. Die Suchergebnisse werden in einer nach verschiedenen Kriterien bestimmten Rangfolge angezeigt.
- *Meta-Suchmaschinen* bieten ihren Nutzern eine Erweiterung des Suchraums und damit auch eine eventuelle, nicht nur quantitative Verbesserung der Ergebnisse (hit rate) durch eine nebenläufige Bearbeitung einer Anfrage über mehreren, verschiedenen Suchmaschinen.

Eine aktive und intelligente Unterstützung wird von allen diesen Lösungen jedoch nicht geboten; das sollen spezielle Softwareagenten, sog. intelligente Informationsagenten (Klusch 1999), leisten. In diesem Kontext ist in Analogie zur Natur der Begriff einer sog. Informations-Nahrungskette (Etzioni 1996) sehr populär geworden. Sie veranschaulicht den Anspruch und Intention der Informationsagententechnologie als eine der Schlüsseltechnologien für das Internet und Web.



Schaubild 1: Die sog. „Information Food Chain“ im Internet

## 1.2 Intelligente Softwareagenten

Die Begriffe 'Softwareagent', 'Agentenbasiertes System' und 'Multiagentensystem' werden in verschiedenen Forschungsgebieten, wie Verteilte Systeme, Software Engineering und der Verteilten Künstlichen Intelligenz unterschiedlich gebraucht. Eine umfassende Einführung in die Thematik von intelligenten Softwareagenten ist u.a. in den Arbeiten (Wooldridge & Jennings 1995; Bradshaw 1997; Klusch 1998, S. 5 – 19; Weiss 1999) zu finden. Im folgenden wird der Begriff 'Agent' als eine Metapher für eine Art von (Software-)Programmen verwendet, deren Funktion und Art der Ausführung in einer offenen und verteilten Umgebung mit bestimmten menschlichen Eigenschaften wie *autonomes, zielgerichtetes Verhalten, soziale Kooperation, Intelligenz und Lernfähigkeit* verbunden wird. Art und Umfang von Aufgaben, Wissen und Fähigkeiten eines Agenten hängen zum Teil stark von der jeweiligen Anwendung ab, für die er entwickelt wurde.

Aktuelle Forschungsprojekte untersuchen den Einsatz von intelligenten Softwareagenten und Multiagentensystemen u.a. in den Bereichen Logistik und Produktionsplanung (Fischer 1999), Finanzplanung und Börsenmanagement sowie dezentrale Organisation von Stromnetzen (Contreras et al. 1999). Der wohl attraktivste Einsatz von sog. Informationsagententechnologie ist jedoch eine proaktive und wissensbasierte Suche nach relevanten Informationen im Internet.

Wir werden im folgenden Abschnitt einen Überblick über diese Technologie geben und daran anschließend in Abschnitt 3 kurz einige Möglichkeiten für den Einsatz von Informationsagenten im Bereich der Wissensentdeckung und des Data Mining aufzeigen. Dabei ist anzumerken, daß die on-line Erfassung, Verarbeitung und Präsentation von privaten Daten an Dritte durch autonome, pro-aktiv handelnde Agenten im Internet aus Gründen der Ethik und vor allem des Datenschutzes problematisch sein können.

## 2 Informationsagenten für das Internet: Ein Überblick

Bislang gibt es keine exakte, einheitliche Vorstellung davon, was unter einem 'Informationsagenten' zu verstehen ist. Erste Schritte zur Konvergenz der entsprechenden Begrifflichkeiten zu diesem Thema werden jedoch sichtbar (Klusch 1999). Ein *Informationsagent* ist ein Softwareagent, der in der Lage ist, für seine Benutzer proaktiv und autonom nach relevanten Informationen in einer Menge von verschiedenen, eventuell geographisch verteilten Informationsquellen zu suchen, die für ihn verfügbar sind. Es wird allgemein verlangt, daß solche Agenten eine oder mehrere der folgenden Anforderungen erfüllen: (a) *Erwerb und Management von Informationen*, (b) *Synthese und Präsentation von Informationen*, (c) *Intelligente, individuelle Benutzerassistenz*. Diese Eigenschaften zeichnen ihn insbesondere gegenüber den Suchmaschinen im Web und sog. intelligenten Front-Ends von Datenbanken aus, die seit Mitte der 80er Jahre im kommerziellen Einsatz sind.

### 2.1 Klassifikation von Informationsagenten

Ein Informationsagent kann nach den folgenden Eigenschaften kategorisiert werden.

- *Nicht-kooperativ oder kooperativ*, entsprechend der Fähigkeiten des Agenten, für die Ausführung seiner Aufgaben mit anderen Agenten zu kooperieren.

- Ein *adaptiver* Agent ist in der Lage, sich an bestimmte Arten von dynamischen Veränderungen in der Umgebung anzupassen. Dies betrifft Veränderungen sowohl im Netzwerk als auch Ort und Inhalt von Informationsquellen sowie Benutzerverhalten.
- *Rationale* Informationsagenten entscheiden sich ökonomisch nutzenorientiert für die Ausführung von Aktionen. Sie agieren und kooperieren rational untereinander, um im wesentlichen ihren eigenen, individuellen Nutzen und damit auch den ihrer Benutzer oder Auftraggeber zu erhöhen. Ein solches Verhalten ist insbesondere für einen Einsatz im Bereich des elektronischen Handels und Geschäftsverkehrs geeignet, frei nach dem Motto "Information gegen Geld".
- Ein *mobiler* Informationsagent ist in der Lage, selbständig im Internet von einer Site zur anderen zu migrieren, um dort jeweils lokal seine (Teil-)Aufgaben auszuführen und mit relevanten Daten zu seinen Benutzern zurückzukehren.

Die folgenden Abschnitte behandeln jede dieser Kategorien von Informationsagenten. Es werden dabei kurz die relevanten, zugrundeliegenden Technologien und entsprechende Beispiele von (Systemen von) Informationsagenten benannt.

## 2.2 Kooperative Informationsagenten und Systeme

Der Begriff eines *kooperativen Informationsagenten* wurde 1992 von Papazoglou, Selis und Laufmann im Zusammenhang mit dem Paradigma kooperativer Informationssysteme eingeführt (Papazoglou et al. 1992). Ausgangspunkt ist eine Menge von verschiedenen, autonomen Datenbanken, die untereinander vernetzt sind. Jeder dieser Datenbanken ist ein Informationsagent zugeordnet, der sie nach außen gegen unberechtigte Zugriffe abschirmt und für seine Benutzer aktiv nach relevanten Informationen sucht. Dabei muß ein Agent eine Vielzahl von Problemen lösen, die durch die Verteilung und Autonomie der Datenbanksysteme sowie die Heterogenität der Daten in den Datenbanken entstehen (Sheth et al. 1999). Es wird vorausgesetzt, daß er hierfür genügend Wissen sowie ausreichende Inferenz- und Kooperationsfähigkeiten besitzt.

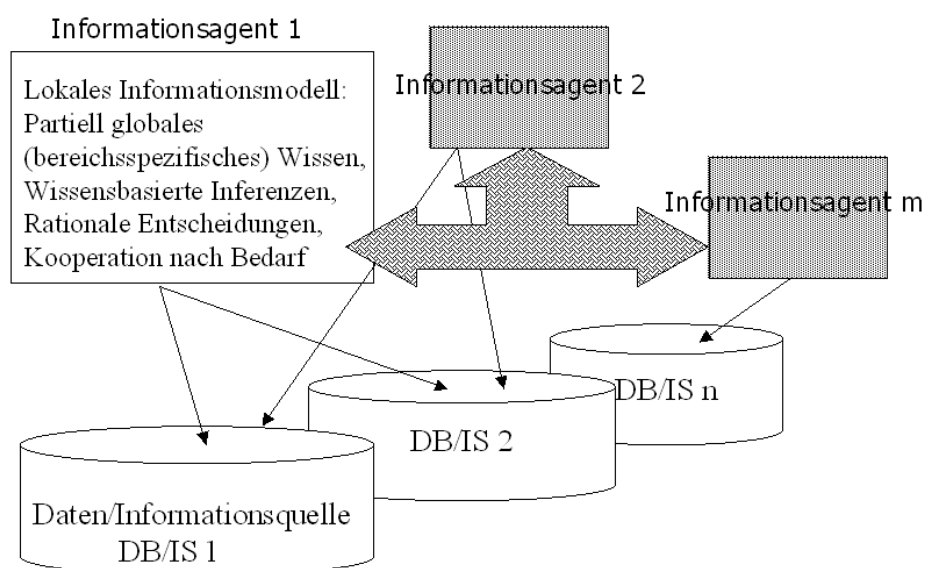


Schaubild 2: Makrostruktur eines Systems von kooperativen Informationsagenten.

Die gesamte anwendungsabhängige Kooperation zwischen den einzelnen Datenbanken im Netz erfolgt ausschließlich über ihre Agenten. Diese 1:1-Beziehung zwischen Informationssystem und Agent wurde in der Folgezeit als zu restriktiv angesehen und ist mittlerweile aufgehoben worden (s. Schaubild 2; Papazoglou & Schlageter 1998). Prominente Beispiele für Architekturen und Systeme von kooperativen Informationsagenten sind *RETSINA* (Sycara & Zeng 1996), *InfoSleuth* (Nodine & Fowler 1999), *IMPACT* (Arisha et al. 1999) und *SCOPES* (Ouksel 1999).

Hinsichtlich der Aufgabe kooperativer Informationsagenten eine intelligente Interoperabilität zwischen heterogenen Systemen zu ermöglichen und einen qualitativen Mehrwert an für den Benutzer verfügbaren Diensten zu erzielen, werden sie oft als ein Bestandteil von sog. intelligenter Middleware für verteilte Informationssysteme betrachtet. In diesem Zusammenhang sind auch die sog. Mediatorbasierten Informationssysteme (Wiederhold 1992; Huhns & Singh 1999) zu nennen, wie beispielsweise *SIMS/ARIADNE* (Arens et al. 1996), *Infomaster* (Genesereth et al. 1997), *TSIMMIS* (Garcia-Molina et al. 1995), *OBSERVER* (Sheth et al. 1996) und UCSD's Mediator *MIX* (Baru 1999). In Schaubild 3 ist die Grundstruktur und Funktionsweise eines solchen Systems beispielhaft veranschaulicht.

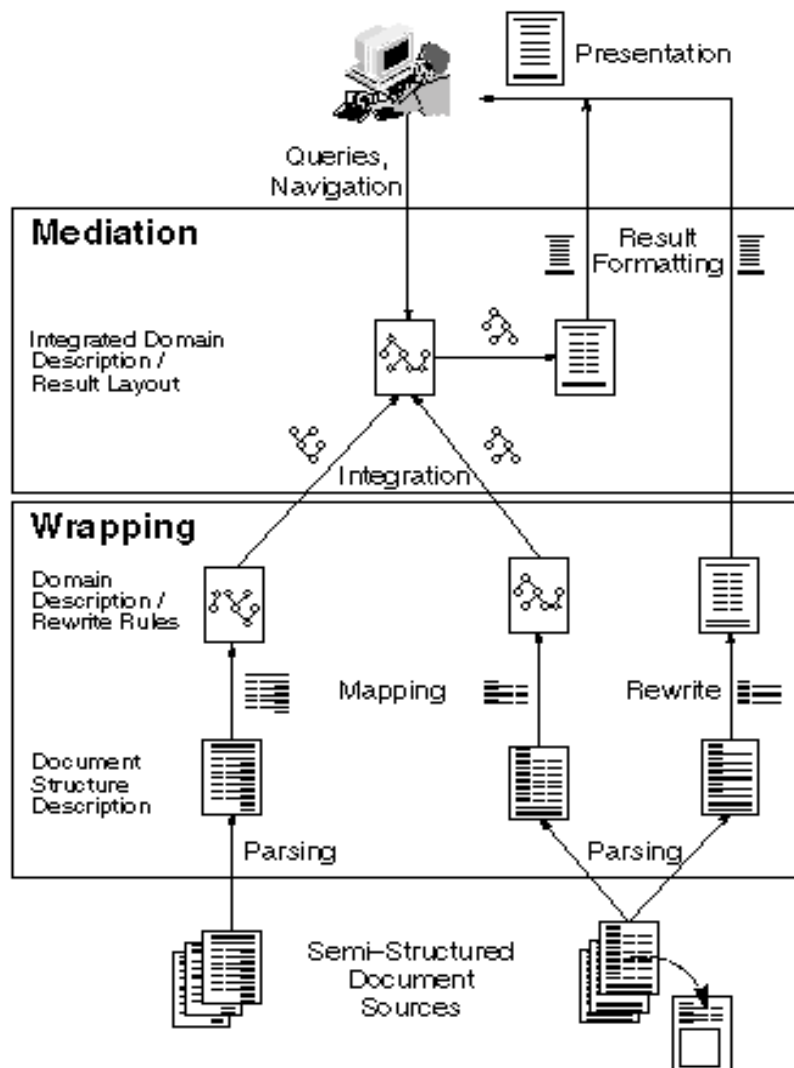


Schaubild 3: Beispiel für ein Mediatorbasiertes Informationssystem

Für jedes betrachtete Informationssystem wird jeweils ein sog. Wrapper-Agent verwendet, der die (semi-) strukturierten Inhalte des lokalen Systems syntaktisch und semantisch auf ein

gemeinsames Datenmodell abbildet und bei Bedarf einem oder mehreren Mediatoren zur Verfügung stellt. Ein Mediator(-agent) integriert anfragerrelevante Daten aus verschiedenen Quellen zu Informationen in ein einheitliches Format und Modell; der Benutzer kann diese Informationen auf verschiedene Art und Weise individuell nutzbringend vom Mediator abrufen. Konzeptuelle Heterogenitäten in den lokalen Informationen werden vom Mediator mit Hilfe von entsprechend geeigneten, ontologischen Abbildungen bearbeitet. Eine Kooperation von bereichsspezifischen Mediatoren untereinander würde dem Konzept eines kooperativen Informationssystems entsprechen; sie ist in der Referenzarchitektur (*DARPA-I3*) vorgesehen, aber bis heute in keinem Mediatorbasierten System prototypisch realisiert.

### **2.2.1 Basistechniken und Methoden zur Kommunikation und Kooperation**

Für die Entwicklung von kooperativen Informationsagenten sind insbesondere hinsichtlich eines transparenten Zugriffs auf heterogene Systeme die Verwendung von Standards zur Middleware wie *OMG's CORBA* (Common Object Request Broker Architecture), *SUN's JDBC* (Java Database Connectivity) und Microsoft's *ODBC* (Open Database Connectivity) geeignet. Jede dieser Frameworks bietet eine abstrakte Schnittstellenbeschreibungssprache und Dienste an, die es einem Entwickler erlauben, verteilte Objekte verschiedener Funktionalität anwendungsorientiert zu definieren, lokalisieren und aufzurufen. Die Dienste zur Interkonnektivität von Systemen folgen im wesentlichen dem Client-Server Koordinationsmodell. Der primäre Vorteil solche Frameworks zu verwenden besteht darin, die Heterogenität von Altlastensystemen und Anwendungen transparent in standardisierten, interoperablen Einheiten, den Wrappern, einzukapseln zu können. Einen aktiven und intelligenten Informationserwerb können jedoch weder solche Frameworks noch die sog. föderierten oder Multidatenbanksysteme (Bukhres & Elmagarmid 1996) leisten.

Eine wesentliche Eigenschaft kooperativer Informationsagenten liegt darin, bei Bedarf mit anderen Agenten zielgerichtet und ohne globale Kontrollinstanz zusammenzuarbeiten. Das beinhaltet insbesondere, daß potentielle Kooperationspartner verstanden und mit ihnen gemeinsam über eine Zusammenarbeit verhandelt werden können. Relevante Techniken und Standards betreffen

- die Entwicklung und Wiederverwendung von standardisierten Bereichs- oder Globalontologien (Steels 1998; Jacobs & Shea 1997, Luke et al. 1997),
- sprechaktbasierte Agentenkommunikationssprachen und Protokolle wie *KQML* und *FIPA ACL* (Labrou et al. 1999; Finin et al. 1994),
- Formate und Sprachen für die Repräsentation und den Austausch von semantischen Metadaten und Wissen, wie *ANSI KIF* (Knowledge Interchange Format), *XML*, *RDF* oder die XML-basierte Modellbeschreibungssprache *PMML* (DML.org 1999).

Zudem kommen Methoden zur Koordinierung von kollaborativer Arbeit zwischen heterogenen Agenten wie Techniken der Dienstvermittlung via Matchmaking und Brokering (Jha et al. 1998; Kuokka & Harrada 1996) hinzu. Diese Techniken basieren auf geeigneten Beschreibungen der Fähigkeiten von Agenten, d.h., den von ihnen angebotenen und angefragten Diensten. Standardisierte Inhaltsbeschreibungssprachen wie *SGML* bzw. *XML* und spezielle Agentenbeschreibungssprachen wie *LARKS* (Sycara et al. 1999) sind dabei ebenso notwendig wie entsprechend geeignete, effiziente und flexible Vermittlungsprozesse und integrierte Methoden zur Bewältigung von Informations- und Systemheterogenität (Sheth et al. 1999).

Wesentliche Aspekte der Kollaboration zwischen intelligenten Informationsagenten betreffen vor allem die dezentrale Planung und Durchführung der Verteilung von komplexen Aufgaben, Teambildung (Kinny et al. 1994; Tambe 1998; Sullivan et al. 1999), adaptive Wahl einer

Koordinationsstrategie (Prasad & Lesser 1999; Sen & Sekaran 1998) und Verhandlungen in offenen, verteilten Umgebungen.

Relevante Fragen hierzu behandeln u.a. die Modellierung von und Analyse der Beziehungen zwischen Rollen von Agenten in verschiedenen Anwendungsbereichen und –szenarien sowie geeignete Typen von Konversation und Konventionen, die zur Erreichung eines Konsenses für die Art und Weise einer Erfüllung von gegebenen oder selbständig von den Agenten emergent ermittelten Zielen in einer offenen, verteilten Umgebung notwendig sind (Shoham & Tenenholz 1997; Cavedon & Sonenberg 1998). Die Forschung wird dabei insbesondere von Arbeiten aus den Gebieten des CSCW (Computer Supported Collaborative Work) sowie den kognitiven und sozialen Wissenschaften inspiriert (Dautenhahn & Numaoka 1998). Die Arbeiten (Huhns & Stephens 1999; Durfee 1999; Ossowski 1999; Doran et al. 1997) bieten einen guten Überblick über mögliche Koordinations- und Planungsmechanismen in heterogenen Agentengesellschaften.

## 2.2.2 Beispiele: InfoSleuth und RETSINA

### a. InfoSleuth

*InfoSleuth* ist ein System, welches ganz speziell zur Suche und Verarbeitung von Information in föderierten Systemen (z.B. im Internet) von der Firma MCC (USA) seit 1995 entwickelt wurde. Das System arbeitet auf der Basis einer ontologischen Darstellung von sog. Ressourcen, die im wesentlichen Datenbanken sind und setzt dabei verschiedene Typen von Agenten ein. Diese Agenten kommunizieren über eine gemeinsame Sprache *KQML* und die einzelnen Aufträge werden im Wissensaustauschformat *ANSI KIF* formuliert. *InfoSleuth* enthält Agenten mit unterschiedlichen Aufgaben: *User*-, *Task Execution*-, *Broker*- und *Resource*-Agenten.

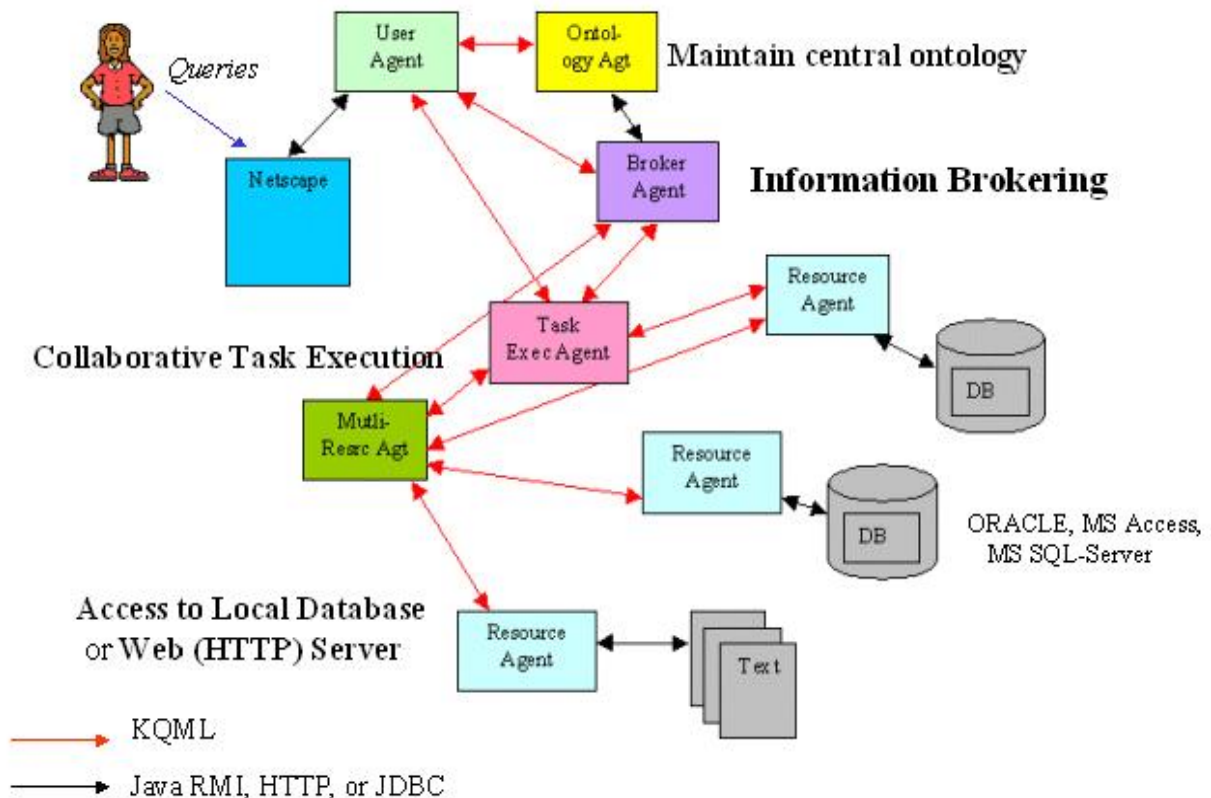


Schaubild 4: Architektur des Multi-Agenten Systems von *InfoSleuth*

Ein *User-Agent* stellt ein intelligentes Interface für den Benutzer zu InfoSleuth dar. Er nimmt Anfragen des Benutzers entgegen und gibt die Anfrageergebnisse wieder an den Benutzer zurück. Jeder *Task Execution Agent* erstellt einen Ausführungsplan für Benutzeranfragen und koordiniert die Sammlung von Information auf einem hohen Abstraktionsniveau (auf niedrigem, d.h. technischem Niveau ist das Aufgabe der Resource Agenten). Es werden deklarative Ausführungspläne verwendet und die Planbearbeitung verläuft datengetrieben. Ein *Broker-Agent* sucht für Aufträge von Benutzern die entsprechenden Dienstleister (-agenten) und vermittelt zwischen beiden. Die gesamte Kommunikation und Datenübertragung erfolgt über diesen Agenten. Der Broker-Agent ist im weitesten Sinne vergleichbar mit einem *CORBA*-Broker mit dem Unterschied, daß hier die Dienstvermittlung über einen semantischen Abgleich von Angebot und Nachfrage geschieht (in *CORBA* geschieht dies rein syntaxorientiert, d.h. der Anfragende muß den benötigten Dienst bereits kennen).

Jeder der *Resource-Agents* stellt ein Interface zu einer Informationsquelle (z.B. einer Datenbank) dar und macht diese Resource für Retrieval- und Update-Dienste im Netz verfügbar. Dieser Agententyp übersetzt die in *KQML/KIF* formulierten, eingehenden Anfragen in Anfragen der Sprache des jeweiligen Informationssystems.

Die gemeinsame Wissensrepräsentation aller Agenten (bzw. des gesamten Systems) geschieht - wie üblich - über Ontologien (Gruber 1993). Jeder Agent enthält seine eigene anwendungs- und bereichsspezifische Begriffswelt in Form einer lokalen Ontologie und veröffentlicht diese in einem Anmeldevorgang zu Beginn seiner Aktivitäten in einer zentralen Ontologie des Systems. Die Verwaltung und Kontrolle dieser zentralen Ontologie untersteht einem speziellen Resource-Agenten (einem sog. *Ontology Agent*); sie ist in drei Ebenen strukturiert, in denen jeweils die übergeordnete Ebene die Begriffe oder Konzepte zur Darstellung der untergeordneten Ebene enthält. Die lokalen Ontologien können in verschiedenen Formaten vorliegen; sie sind selbstbeschreibend und die Agenten können gegenseitig ihre Konzepte erfragen, bzw. je nach Agententyp auch abgleichen. Als besondere Eigenschaften der semantisch gesteuerten Dienstvermittlung werden genannt:

- *Intelligente Weitergabe von Anfragen*: Analog zur Auswertung eines Verteilungsschemas in verteilten Datenbanken, werden Anfragen nur an diejenigen Resource-Agenten weitergegeben, aus deren Ontologien der Broker-Agent erkennen kann, daß sie für die Anfrage relevante Information liefern können.
- *Dynamische Anbindung von Ressourcen*: Durch die ständige Bereitschaft des Broker-Agenten können jederzeit Ressourcen an- bzw. abgemeldet werden. Das System ist damit stets auf dem aktuellen Stand und kein Benutzer muß über die Anzahl und die Art der vorhandenen Ressourcen Wissen besitzen.
- *Skalierbarkeit*: Wird nicht mehr als ein Problem angesehen, da das System fast vollständig dynamisiert ist. Es paßt sich selbst an die Menge der jeweils verfügbaren Ressourcen an.

#### b. *RETSINA*

In einem *RETSINA*-Multiagentensystem (Sycara & Zeng 1996) kooperieren drei verschiedene Arten von Agenten miteinander, um relevante Informationen im Internet zu finden: *Interface*-, *Aufgaben*- und *Informationsagenten*.





Schaubild 5: *RETSINA*-Architektur eines kooperativen Agentensystems

Die *Interface-Agenten* stellen für den Benutzer eines *RETSINA*-Agentensystems den einzigen Zugang zum System dar. Sie helfen ihm interaktiv bei der Spezifizierung von Aufgaben und geben die Ergebnisse ihrer Bearbeitung aus. Alle *Aufgabenagenten* besitzen jeweils ein umfangreiches bereichsspezifisches Wissen; jeder Agent plant und kontrolliert die Ausführung der Aufgaben, die er von einem oder mehreren Interface-Agenten erhält. Eventuell zerlegt er eine Aufgabe in mehrere Teilaufgaben und delegiert ihre Ausführung entweder an andere geeignete Aufgabenagenten oder direkt an die entsprechenden Informationsagenten. Ein *Informationsagent* hat einen direkten Zugriff auf eine oder mehrere Informationsquellen. Er kann die Anfragen eines Aufgabenagenten je nach Anforderung einmal oder wiederholt (periodisch) bearbeiten sowie eine bedingte Überwachung von Datenbeständen in den Informationsquellen durchführen. Eine besondere Art der Informationsagenten in einem *RETSINA*-Agentensystem sind die sog. *Matchmaker-Agenten*. Im Gegensatz zu den Broker-Agenten bei *InfoSLeuth* beschränken sich diese Agenten nur auf das Vermitteln von relevanten Agenten für die Ausführung einer Aufgabe, überlassen aber alles weitere den davon betroffenen Agenten. Eine Kooperation findet je nach Bedarf zwischen allen Arten von Agenten statt. Die Kommunikation zwischen den Agenten erfolgt in der Sprache *KQML*.

Die Grundidee bei der Entwicklung von Agenten eines *RETSINA*-Multiagentensystems ist die Wiederverwendbarkeit von (Java-)Programmcode: Die bereichsabhängigen und bereichsunabhängigen Kontrollstrukturen für Kommunikation, Kooperation, Scheduling und Monitoring sowie Planfragmente und Wissensbasis eines generischen Agenten werden je nach Anwendung für die jeweiligen Agenten geeignet instantiiert. Beispiele für *RETSINA*-Agentensysteme sind für verschiedene Anwendungen bekannt:

- *WARREN* (Handel und Informationen über Aktien an verschiedenen Börsen),
- *PLEIADES* (Terminplanung und Büroautomation) und
- *THALES* (Vorhersage der Position/Sichtbarkeit eines Satelliten).

## **2.3 Adaptive Informationsagenten**

Ein adaptiver Informationsagent besitzt die Fähigkeit, sich selbständig auf stetige Veränderungen in offenen Umgebungen und Anforderungen in seinem Aufgabenbereich anzupassen. Das betrifft zum Beispiel wechselnde Präferenzen und Bewertungen der Relevanz von Informationen durch seine Benutzer, die dynamische Klassifikation von nicht statischen, vagen oder unsicheren Inhalten aus eventuell ortsveränderlichen Informationsquellen und flexible Anpassung an wechselhafte Verfügbarkeit von Ressourcen im Netz wie Bandbreite und Auslastungen von aktuell besuchten Servern. Durch den Einsatz von (semi-)automatischen Lernverfahren für Informationsagenten wird vor allem eine flexiblere und benutzerfreundlichere, wenn auch nicht immer schnellere Informationssuche erwartet. Es ist dabei generell zwischen dem Lernen eines einzelnen Agenten in einem Multiagentensystem und der Adaption eines Multiagentensystems als Ganzes zu unterscheiden (Weiß & Sen 1995).

Anwendungen von adaptiven Informationsagenten umfassen insbesondere intelligente Benutzerschnittstellen für das Web (André 1999; Lieberman 1999), digitale Bibliotheken (Durfee et al. 1999) und Manufacturing (Brauer & Weiß 1998).

### **2.3.1 Methoden und Techniken für die Entwicklung von adaptiven Agenten**

Die bei der Entwicklung von adaptiven Informationsagenten angewendeten Verfahren stammen zumeist aus den Gebieten des maschinellen Lernens (Mitchell 1998; Ballard 1999). Relevante Aspekte betreffen (Sen & Weiß 1999)

- angewandte Strategien für Lernen durch Analogie, Beispiel oder Entdeckung,
- Art des Feedbacks für Agenten bei Methoden des Reinforcement, überwachten oder selbstorganisierenden Lernens,
- Typen der (multimedialen) Interaktion zwischen Agenten, Benutzern und dem Multiagentensystem mit der Umgebung,
- den Zweck des Lernens zur Verbesserung der individuellen Fähigkeiten eines einzelnen Agenten oder eines Agentensystems, den dabei zu erwartende Nutzen sowie
- die Verteilung von Daten und nebenläufigen Berechnungen für eine Adaption im System.

Derzeit werden Lernverfahren zumeist nur bei nicht kooperativen Informationsagenten eingesetzt, wie beispielsweise eine heuristische Breitensuche im Web (Lieberman 1995), neuronale Netze (Pannu & Sycara 1996) und genetische Algorithmen (Sheth & Maes 1993). Andere grundlegende Methoden für die Entwicklung von adaptiven Informationsagenten sind das

- *inhaltsbasierte* Filtern von Informationen (Losee 1998; Salton 1989) und Erstellen von approximativen, gelernten Benutzerprofilen durch Beobachtung von on-line Aktivitäten der betreffenden Nutzer sowie
- *kollaboratives* Filtern von Informationen (Balabonovic & Shoham 1997) anhand von Vergleichen individueller Präferenzen mit denen von anderen Benutzern nach bestimmten Gemeinsamkeiten.

Zudem sind effektive Techniken der multimedialen Mensch-Maschine/Agent-Interaktion (Lewis 1998) von immenser Bedeutung, da sie einen wesentlichen Beitrag zur komfortablen, zeitgemäßen Visualisierung von relevanten Informationen leisten und damit helfen, den Grad der Akzeptanz von Informationsagenten durch seine Benutzer zu verbessern. Dies kann

insbesondere mit Hilfe von glaubwürdigen, sog. lebensechten Charakteren und Avataren erfolgen (Elliott & Brzezinski 1998); dabei werden in zunehmendem Maße eine integrierte Verarbeitung von affektiven, audiovisuellen Eingaben wie Gestik, Mimik (Picard 1997) und natürliche Sprache mit einbezogen. Die Entwicklung solcher Techniken geht einher mit signifikanten Fortschritten in Forschung und Standardisierung in relevanten Bereichen wie beispielsweise der Repräsentation und Integration von Multimedia, 3D-Grafikbeschleunigung, Massenspeichern und Übertragungsraten im Next Generation Internet.

### 2.3.2 Beispiele

Persönliche, adaptive Benutzerassistenten und Schnittstellenagenten für das Web und Anwendungen wie Web-TV sind beispielsweise *TrIAs* (Bauer & Dengler 1998), *WebPersona* (André et al. 1998) und *Let's Browse* (Lieberman 1999). In Schaubild 6 sind als Beispiele für animierte Charaktere für interaktive Benutzerschnittstellen und Präsentationen im Web die „Personae“ Jane, Merlin und Max von Extempo Inc. sowie der Cartoon-Agent Cyberella der DFKI GmbH zu sehen.



Schaubild 6: Beispiele von Charakteren für intelligente Benutzerschnittstellen im Internet.

Gelegentlich sind in Internet Chats auch sog. Chatterbots zu finden, die ähnlich der durch *ELIZA* bekannten, einfachen aber wirkungsvollen Art von Konversation zwischen Softwareroboter und Mensch arbeiten; sie sind vornehmlich für den Einsatz in sog. Call Centern zur Beantwortung der Masse von stereotypischen Fragen gedacht.

Es gibt derzeit nur sehr wenige Implementierungen von Systemen kooperativer *und* adaptiver Informationsagenten. Prominente Beispiele betreffen die Anwendung von kollaborativem Filtern von Informationen für den elektronischen Handel bzw. im Nachrichtenbereich wie *Firefly* bzw. *GroupLens* (Konstan et al. 1997). Transitive Recommender-Systeme wie *Histos* (Moukas et al. 1999) bewerten dagegen die Historie der Reputationen von und Präferenzen in Benutzergruppen für induktive Folgerungen. Andere Beispiele sind die Systeme *Amalthea* (Moukas 1998) und *Arachnid/InfoSpiders* (Menczer & Monge 1999); sie wenden genetische Algorithmen (Mitchell 1996) für eine adaptive Informationssuche an: diejenigen Agenten, die ihren Nutzern 'relevante' Informationen liefern, können sich nach evolutionären Prinzipien vermehren und liefern graduell immer bessere Informationen, die anderen 'sterben aus'.

## 2.4 Rationale Informationsagenten für den elektronischen Handel

### 2.4.1 Elektronischer Handel

Elektronischer Handel (E-Commerce) kann definiert werden als die Menge aller Aktivitäten, die direkt den on-line Handel von Waren und Dienstleistungen im Internet betreffen; damit ist E-Commerce Teil des allgemeineren elektronischen Geschäftsverkehrs, des sog. E-Business. Je nach Ausrichtung des Handelsflusses zwischen Geschäftsbereichen bzw. Warenanbietern

und Konsumenten kann nach verschiedenen Marktsegmenten des E-Commerce unterschieden werden. Diese Segmente sind

- Business-to-Business (B2B) traditionell über EDI/EDIFACT und WebEDI Transaktionen,
- Business-to-Consumer (B2C) zumeist durch on-line Retailer und proprietäre Auktionen,
- Consumer-to-Consumer (C2C) wie z.B. durch freie, virtuelle Marktplätze und Auktionen
- sowie Consumer-to-Business (C2B) wie inverse Auktionen, in denen Unternehmen sich um die Durchführung oder Lieferung von speziellen Dienstleistungen bzw. Produkten bewerben, die zuvor von Kunden allgemein angefordert und ausgeschrieben wurden.

#### **2.4.2 Potential von und Basistechnologien für E-Commerce**

E-Commerce ist ein immenser Wachstumsmarkt seit den letzten fünf Jahren, der mittlerweile fast alle Branchen erfaßt hat; Berichten von führenden Marktforschungsunternehmen wie Meta Group und Forrester Research zu Folge, steigen die in Europa bzw. USA durch E-Commerce erzielten Firmenumsätze im Jahr 2000 auf einen Schätzwert von 200 bzw. 350 Milliarden US\$. Bemerkenswert ist, daß weltweit 2003 das B2B-Segment das des B2C voraussichtlich mit 1,3 bzw. 0,1 Billionen US\$ um eine Größenordnung (noch immer) übersteigt. Trotz oder gerade wegen des enormen Potentials von E-Commerce ist der Einsatz von Agenten auch in diesem Bereich aus Sicht der Unternehmen, Kunden und Forschung hoch attraktiv (Merz, 1999; McKnight & Bailey 1997). Die Entwicklung von E-Commerce Lösungen greift im wesentlichen auf folgende Basistechnologien zurück:

- Standardisierte Repräsentation, Erwerb und Austausch von Informationen, beispielsweise durch *XML*, *UML*, *EDIFACT* und wiederverwendbare Bereichsontologien sowie Methoden des Information Retrieval für das Web (Baezo-Yates & Ribeiro-Neto 1999; Gudivada 1997) und Data Mining,
- sichere Benutzerprofile und Datensicherheit (Damgard 1999), wie durch (a)symmetrische Kodierung, digitale Signaturen, W3C's *P3* (Privacy Preferences Project),
- sicherer, elektronischer Zahlungsverkehr (Schuster et al. 1997), z.B. via *VISA/MC's SET* für Kreditkartenzahlungen, digitales Geld oder periodische Abbuchung von Kundenkonten zentral koordinierter, virtuelle Märkte, sowie
- Standardprotokolle, die alle wesentlichen Prozesse des elektronischen Handels beschreiben, wie etwa *IETF's IOTP* (Internet Open Trading Protocol).

Weitere Aspekte des E-Commerce umfassen u.a. Methoden für einen graduellen Aufbau und nachhaltige Pflege von *Vertrauen* zwischen on-line Händler und Kunde, besser ihren Agenten (Castelfranchi et al. 1998; Marsh 1994), mit handhabbarem Risiko für alle Beteiligten. Das geht über Authentifizierung, Zertifizierung von dritter Seite oder kryptografische Methoden weit hinaus.

#### **2.4.3 Informationsagenten für E-Commerce**

Aufgrund der fortschreitenden Kommerzialisierung des Internet ist abzusehen, daß Informationsagenten für bestimmte Dienstleistungen bezahlen und von ihren Nutzern bezahlt werden. Damit sind Methoden ökonomisch rationaler Entscheidungsfindung für die Ausgestaltung des Verhaltens solcher Agenten wesentlich. Das betrifft nicht nur vereinzelte Verhandlungen zwischen ihren Nutzern als potentielle Kunden von Unternehmen, sondern auch dynamische Entscheidungen zur Kooperation mit anderen Agenten im Internet (Indermaur 1998). Die Begriffe 'Rationalität' und 'rationales Multiagentensystem' sind in der KI und der Ökonomie weitestgehend unabhängig voneinander betrachtet worden (Boutilier et

al. 1997); interdisziplinäre Arbeiten führten erst seit Beginn der 90er Jahre zu verwertbaren Ergebnissen auf dem Gebiet der Multiagentensysteme in der Verteilten KI (Rosenschein & Zlotkin 1994; Sandholm & Lesser 1997; Klusch & Shehory 1996; Shehory & Kraus 1998, Ketchpel 1994).

Virtuelle Institutionen des E-Commerce wie Marktplätze (Wellman 1995) und Auktionen (Wolfstetter 1996) bieten geeignete Szenarien für die Anwendung von Agententechnologie. Die für eine Entwicklung von Agentenbasierten E-Commerce Lösungen (Kerschberg & Banerjee 1999) wesentlichen Techniken und Methoden umfassen neben dezentralen, multilateralen Verhandlungen, Variationen des bekannten Kontraktnetzprotokolls wie *TRACONET* (Sandholm 1996), Arbitrierungsschemata für Matchmaking (Tesch & Fankhauser 1999) und den bereits erwähnten intelligenten Interaktionen und Benutzerprofilen sowie dem kollaborativen Filtern von Informationen auch die sog. Produktvergleiche und Koalitionsformierung zwischen Agenten (Shehory & Kraus 1998).

Die meisten der derzeit im Web verfügbaren Kaufagenten, sog. ShopBots, wie *evenbetter.com* oder *mySimon.com* und solchen die integraler Bestandteil von Webportalen sind, wie bei den Verzeichnissen *yahoo!* oder *lycos.com*, bieten ihren Nutzern eine Strategie des sog. „Best-Price-Buy“ an. D.h., sie führen einen Vergleich von Produkten der gewünschten Kategorie nur über den Preis von mehreren, dem einzelnen ShopBot verfügbaren Händlerdatenbanken durch; andere Attribute, wie beispielsweise Lieferzeiten, Kundenservice und Reputation der betreffenden Händler spielen keine Rolle. Ausnahmen sind hier die Multiagentensysteme für E-Commerce *Kasbah/MarketMaker* (Chavez et al. 1997), *UMDL* (Durfee 1999), *AuctionBot* (Wellman et al. 1998b) und *Tête-a-Tête* (Guttman et al. 1999). Ersteres bietet eine Wertung von Teilnehmern vorangegangener Geschäfte an, den sog. Better Business Bureau Service; letzteres basiert seine Produktkaufempfehlungen auf der sog. Mehrattribut-Theorie, welche im wesentlichen auf der verteilten Lösung einer Menge von Attributwertebedingungen beruht. Ein umfassender Überblick über neuere Entwicklungen im agentenbasierten elektronischen Handel (*AMET* Agent-Mediated Electronic Trading) wird beispielsweise in den Arbeiten (Noriega & Sierra 1998; Guttman et al. 1999) gegeben.

## **2.5 Mobile Informationsagenten**

Die Fähigkeit zur Mobilität wird von Informationsagenten nicht dringend gefordert, ist aber in Bezug auf ein Orts- und Zeitunabhängiges, intelligentes Informationsmanagement von hoher Attraktivität für den Benutzer. Ein weiterer, erwarteter Vorteil von mobilen Agenten (Pham & Karmouch 1998) ist eine unter bestimmten Bedingungen signifikant reduzierte Netzlast durch lokale Datenverarbeitung. Mobile Agenten folgen dem Paradigma des Remote Programming im Gegensatz zu dem des Remote Procedure/Function Call (*RPC/RFC*) der traditionellen Client-Server Systeme; sie sind in der Lage unter gegebenen Umständen für die Ausführung ihrer Aufgaben auf mehrere, geographisch verteilte Server zu migrieren (Shehory 1999).

Unterstützt wird die Verwendung solcher Agenten in geeigneten Softwareapplikationen auch durch die steigende Verbreitung von mobilen Telekommunikationsgeräten, die eine Verbindung zum Internet bieten (*WAP*-fähige Webhandies, Kommunikatoren und Palms) (Krause & Magedanz 1996, Pentland 1998), intelligentem Breitband (Venieris & Hussmann 1998) und drahtlosen, Satellitengestützten Datennetzwerken. Desweiteren werden die fortschreitende Verteilung von relevanten Daten und Informationen und die Anforderung, (eventuell a priori unbekannt) Server oder deren Proxies bei Bedarf dynamisch selektieren sowie auf mögliche Störungen von (Funk-) Netzwerkbereinigungen flexibel reagieren zu können, zum Einsatz von mobilen Agenten in der Zukunft führen. Es wird zudem an einer standardisierten Integration von mobilen Agentensystemen in existierende *RPC*-basierte

Middlewareplattformen wie *CORBA* (Magedanz et al. 1998) gearbeitet; *OMG's CORBA MASIF* (Mobile Agent Systems Interoperability Facility) ist ein Beispiel hierfür.

Derzeit populäre Entwicklungsumgebungen für Systeme von mobilen Agenten sind beispielsweise *IBM's Aglets* (Lange & Oshima 1998), *ObjectSpace's Voyager*, *Mitsubishi's Concordia* (Wong et al. 1997), *MOLE* der Universität Stuttgart sowie skriptsprachenbasierte Systeme wie *D'agents/AgentTcl* (Rus et al. 1997) und *ARA* (Peine & Stolpmann 1997). Eine vergleichende Übersicht mobiler Agentensysteme und ein praktisches Beispiel für eine Anwendung von mobilen Agenten für verteiltes Information Retrieval ist in (Brewington et al. 1999) zu finden. Ein anderes Beispiel bietet *ACORN* (Marsh 1997) mit dem Konzept von sog. Agenten-Cafés, in denen mobile Agenten untereinander Informationen präsentieren, verhandeln und austauschen können. Das aktuelle Hauptanwendungsgebiet für mobile Agenten betrifft dezentrale, effiziente Dienstleistungen in der Telekommunikation (Weihmayer & Velthuijsen 1998; Plu 1998).

Im Kontext mobiler Agenten werden relevante Sicherheitsaspekte immer wieder lebhaft und mit Hinweis auf Ähnlichkeiten zu bösartigen Viren im Internet diskutiert. Dabei sind Attacken aus dreierlei Perspektiven zu betrachten: Server-contra-Agent, Agent-contra-Server und Agent-contra-Agent (Greenberg 1998; Tschudin 1999; Vigna 1998). Aktuell diskutierte Lösungsansätze sind beispielsweise vertrauenswürdige Ausführungsumgebungen für mobile Agenten auf fremden Servern, Mechanismen zur Entdeckung oder Vorbeugung von Codemodifikationen, wie z.B. durch Verwendung sicherer Hardware, Proof-Carrying Code oder kryptografische, zeitbegrenzte Ausführung von Agenten.

Mögliche Schritte zur Verbesserung der Akzeptanz von mobilen Agenten sind eine weitere Verbreitung von (nur einmal vom Server zum Client bzw. umgekehrt migrierenden) mobilem Code (Picco et al. 1999) von Java-Applets bzw. Servlets und Proxy Server von ISPs, die einen von Nutzern gesendeten mobilen Code zur bezahlten Ausführung akzeptieren. Der nächste Schritt wäre der der Agententechnologie in relativ sicheren Intranet-Umgebungen hinaus ins öffentliche Internet, und hier bleibt die Frage nach der Sicherheit von mobilen Prozessen, d.h., unbeschränkt migrierfähigen Agenten in der laufenden Forschung zu klären.

### **3 Einsatz von Informationsagenten für Wissensentdeckung und Data Mining im Internet**

*Wissensentdeckung in Datenbanken* (KDD Knowledge Discovery) ist ein seit 1991 etabliertes, interdisziplinäres Forschungsgebiet (Piatetsky-Shapiro & Frawley 1991); es hat das Ziel, Verfahren und Techniken zu entwickeln, die für die Identifikation von neuartigen, gültigen, potentiell nützlichen und verständlichen Mustern in großen Mengen von Daten geeignet sind. *Data Mining* (Chen et al. 1996) ist dabei ein wesentlicher Bestandteil im Gesamtprozeß des KDD; es umfaßt den Einsatz von Methoden zur quantitativen und qualitativen Analyse der betreffenden Daten für eine Bestimmung von entsprechenden Arten von Datenmustern. Man steht hierbei u.a. vor dem Problem, die Vielzahl der in einem Suchprozeß konstruierten und anhand von Testdaten überprüften Hypothesen und Aussagen auf diejenigen zu beschränken, die für einen Benutzer am interessantesten sind (Müller 1999). In der Literatur werden die Begriffe KDD und Data Mining häufig synonym verwendet; ein KDD-Projekt durchläuft jedoch außer dem des Data Mining noch mehrere andere Phasen, wie die der Datenaufbereitung, Evaluierung und Dokumentation des Projekts im aktuellen Geschäftsumfeld (Nakhaeizadeh et al. 1997). Eine ausführliche Darstellung der einzelnen Phasen von Wissensentdeckung wird im ersten Teil dieses Buches gegeben.

Relevante Gebiete des KDD sind insbesondere die der Datenbanktechnologie, Statistik und des Maschinellen Lernens. Wir werden im folgenden sehr kurz auf einige, grundlegende Perspektiven und neuere Forschungsrichtungen im Bereich des Data Mining und KDD im

allgemeinen eingehen. Daraus ergeben sich insbesondere Einsatzmöglichkeiten für intelligente Informationsagenten, die den komplexen Prozeß des KDD und Data Mining auf vielfältige Weise effektiv und effizient unterstützen können.

### **3.1 Perspektiven von Data Mining und KDD**

Die bekanntesten Perspektiven des Data Mining sind die der Induktion, Kompression, Anfrage und Approximation von Datenmustern (Ramakrishnan & Grama 1999):

- *Induktion.* Verfahren zur Induktion kehren den deduktiven Prozeß prädikatenlogischer Inferenzen um, indem sie versuchen, ausgehend von einer speziellen Situation durch eine (fokussierte) Suche in einem konstruierten Hypothesenraum zu allgemeineren Aussagen zu kommen. Bekannte Verfahren zur Induktion sind z.B. ILP-Algorithmen (Inductive Logic Programming).
- *Kompression* im Data Mining bezeichnet das Prinzip des Suchens nach möglichst einfachen Datenmustern (das sog. Occam's Razor Prinzip), deren gelernte Beschreibungen minimal im Vergleich zur vollständigen Aufzählung der originalen Daten sind (theoretisch fundiert durch das sog. Minimum Description Length Prinzip).
- *Anfrage.* Data Mining kann auch als Menge geeigneter Anfrageprozesse über Datenbeständen in einer Datenbank gesehen werden; das erfordert ausdrucksstärkere Anfragesprachen als z.B. SQL und entsprechende Datenmodelle.
- *Approximation* schwächt iterativ ein gegebenes akurates, exaktes Modell der Daten durch approximative Transformationen ab, um mögliche versteckte Muster zu entdecken. Bekanntes Beispiel für ein solches Verfahren ist Latent Semantic Indexing (Furnas et al. 1988) aus dem klassischen Gebiet des Information Retrieval.

Die grundlegenden Methoden zu diesen Perspektiven des Data Mining umfassen Algorithmen u.a. für die Erstellung von Entscheidungsbäumen, Assoziations- und Klassifikationsregeln sowie Instanzbasiertes Lernen (Witten & Frank 1999). Dabei spielt die Skalierbarkeit der Algorithmen auf sehr große Datenmengen eine ebenso große Rolle wie ihre Fähigkeit, mit unpräzisen, falschen oder fehlenden Attributwerten, kontinuierlichen und diskreten Werteskalen kosteneffizient umgehen zu können. Die erkannten quantitativen oder qualitativen Datenmuster können hinsichtlich Genauigkeit, Ausdrucksstärke, Interpretierbarkeit, Präzision, Interessantheit oder operative Verwertbarkeit für den Geschäftsverkehr bewertet werden.

Einige klassische Anwendungsprobleme für Data Mining sind

1. die *Marktbasierte Analyse* von Kundentransaktionen mit sog. Apriori-Methoden der Partitionierung, sequentiellen Muster, des Hashing und Sampling zur effizienten Reduktion der Menge von potentiellen Kandidaten bei der Auswahl relevanter Datensätze (Ganti et al. 1999a),
2. das *Clustering*, d.h., die ähnlichkeitsbasierte Gruppierung von Daten entsprechend ihrer Eigenschaften (Ganti et al. 1999b) mit Systemen wie *Birch\** (Zhang et al. 1996), *CURE* (Guha et al. 1998) und *Chameleon* (Karypis et al. 1999) sowie
3. die *Klassifikation (Regression)* von Daten auf der Basis historischer Informationen; das Ziel ist ein Model, das ausgehend von gegebenen, symbolischen (oder numerischen) Werten für Bestimmungsattribute (predicator attributes) den eines Zielattributs (dependent attribute) bestimmt. Hierfür können z.B. Methoden der Statistik, neuronale Netze und Konstruktion von Entscheidungsbäumen wie *C4.5*, *CART/SPRINT* (Shafer et al. 1996) und *RainForest* (Gehrke et al. 1998) verwendet werden.

Einige besondere Herausforderungen für die Forschung im Bereich des Data Mining und KDD im allgemeinen sind die folgenden.

- Eine *dynamische, proaktive Unterstützung von Ad-hoc und Anfrageorientiertem Data Mining* im Internet, Intranet oder Extranet, beispielsweise für *OLAM-Systeme* (Online Analytical Mining) (Han et al. 1999) und Data Warehouses.
- *Anytime-Algorithmen für eine Anfrageverarbeitung*, die vom Nutzer jederzeit abgebrochen werden kann und bei der die Qualität der Antworten monoton steigt. Das ist besonders bei Beschränkungen von verfügbaren Systemressourcen und Zeit von Vorteil.
- *Interaktives Data Mining* für spezifische Anwendungen, bei denen beispielsweise eine aktive Begrenzung des Suchraums für Datenmuster durch einen Benutzer-System-Dialog sinnvoll ist. Der Einsatz von *ECA* (Event-Condition-Action)-Regelbasierten Komponenten von aktiven Datenbanksystemen (Widom & Ceri 1996) kann die Automatisierung einer solchen Dialoggesteuerten Suche unterstützen.
- *Verteiltes und paralleles Data Mining*. Die verteilte, nebenläufige oder parallele Suche in verschiedenen Datenmengen nach Datenmustern (auf gleiche oder verschiedene Weise) ist u.a. in heterogenen, asynchronen, dynamischen Umgebungen mit hoher Fehleranfälligkeit geeignet; es kann auf Kosten eines höheren, lokalen Berechnungsaufwandes die zu übertragenden Datenmengen signifikant reduzieren.

### **3.2 Anwendungspotentiale von Agenten für Data Mining**

Aus den erwähnten Perspektiven und Herausforderungen des Data Mining ergeben sich einige Anwendungspotentiale von intelligenten Agenten:

- *Interaktives Data Mining*,
- *Dynamische Selektion und Extraktion von Informationen*,
- *Multi-Strategieinsatz für komplexe Aufgaben der Wissensentdeckung*,
- *Skalierung von Data Mining auf die Analyse von großen, verteilten Datenmengen*.

Es zu beachten, daß es bislang an umfassenden Forschungsergebnissen in diesen Bereichen fehlt; daher bleiben die nachfolgenden Beschreibungen von möglichen Anwendungsgebieten von Agenten im Data Mining skizzenhaft und sind vor allem als Anreiz für weitere Forschung gedacht.

*Interaktives Data Mining* bedeutet vor allem eine wesentlich stärkere Benutzerorientierung während der Datenanalyse durch intelligente, lernfähige Agenten. Adaptive Agenten können beispielsweise bei der Eingrenzung des Suchraums, der Bewertung und Präsentation von Zwischenresultaten helfen, indem sie selbständig mit Vorschlägen zur weiteren Vorgehensweise bei der Suche und Interpretation von Datenmustern aufwarten. Dabei kann u.a. auf zuvor erlernte, individuelle Benutzerprofile und Ergebnisse von vorangegangenen Sitzungen zurückgegriffen werden.

Ein denkbares Szenario wäre ein asynchron verteiltes Data Mining in einem Netzwerk, das von mehreren Agenten kollaborativ für Mitglieder einer bestimmten Benutzergruppe durchgeführt wird. Dabei hat jedes Mitglied eine andere Sichtweise auf die Daten und wird entsprechend von einem persönlichen Agenten im Netzwerk repräsentiert. Zeitaufwendige Verhandlungen zwischen den einzelnen Mitgliedern über eine allgemein akzeptierte Dateninterpretation sowie globale Konsistenzüberprüfungen können von den Agenten initiiert und bei Bedarf selbst kooperativ oder interaktiv mit den jeweiligen Mitgliedern durchgeführt



werden. Individuelle Präferenzen für mögliche oder vorgeschlagene Interpretationen werden von den jeweiligen Agenten in die Verhandlung eingebracht.

Eine andere, einfachere Perspektive ist die statische Ermittlung von Parameterwerten für verwendete, repetitive Data Mining Algorithmen. Hierbei beschränkt sich die Interaktion zwischen Agent und Benutzer lediglich auf eine ansprechende, leicht bedienbare und redundanzfreie Eingabe von notwendigen Werten.

Bei komplexen Aufgaben der Wissensentdeckung wie beispielsweise umfangreichen Anfragen zu Aussagen über Datenmengen in verschiedenen Geschäftsbereichen ist die Anwendung von nur einem Data Mining Verfahren nicht immer hinreichend. Eine geeignete Kombination von verschiedenen Strategien zur Wissensentdeckung und Data Mining könnte in solchen Fällen von Vorteil sein. Derartige *Multi-Strategien* sind von intelligenten Agenten erlernbar; dies kann ähnlich dem Erlernen einer Auswahl einer geeigneten Koordinationsstrategie in einem Multiagentensystem erfolgen (Prasad & Lesser 1999). Agenten in einem System lernen, welches Data Mining Verfahren unter welchen Umständen anzuwenden ist, um welche Art von Ergebnissen zu erzielen. Die effiziente Anwendung von Techniken des maschinellen Lernens auf das Lernen von Multi-Strategien im Data Mining bleibt Gegenstand zukünftiger Forschung.

Die *Skalierung* von Data Mining auf große Datenmengen in einer Größenordnung von Giga, Terra oder gar Peta kann unter Einsatz von paralleler Hardware und entsprechenden, parallelen Verfahren angegangen werden. Mobile Agenten (s.a. Abschnitt 2.5) können hier eine Alternative sein. Die Kerneigenschaften von mobilen Agentensysteme umfassen in dieser Hinsicht vor allem die

- *Reduktion der Last des originären Servers* (d.i. der Server, der den Client besitzt, der den mobilen Agenten initiiert hat) und des betreffenden Netzwerks (Intranet oder Internet),
- *Flexibilität in der Ausführung von Operationen auf Gastservern* sofern dies den mobilen Agenten in einer kontrollierten Umgebung erlaubt wird, wie beispielsweise das Anpassen der Anwendungslogik vom Agenten bei unvorgesehenen Veränderungen.

Im Kontext des Data Mining im Web kann der Einsatz von mobilen Agenten durch ihre Migration und lokale Verarbeitung auf fremden Datenservern eine signifikante Reduktion der Datenvolumen vor der Übertragung auf einen zentralen Data Mining Server bedeuten.

Ein *dynamischer Informationserwerb* im Web ist mit Hilfe von nicht notwendigerweise mobilen Agenten zur Unterstützung von *OLAP*-Anwendungen (Online Analytical Processing) und Data Warehousing in Pull- und Push-Szenarien möglich. Proaktiv handelnde und lernfähige Agenten können Methoden des Data Mining dynamisch entweder lokal oder entfernt einsetzen. Ihre Aufgaben umfassen u.a. eine

- adaptive Auswahl von geeigneten Informationsquellen im Internet (Sen et al. 1999) anhand aktueller Kriterien wie zu erwartender Informationsgehalt, Netzwerk- und Serverlasten,
- geeignete Extraktion von Inhalten und ihre Aufbereitung (z.B. zum Aufbau und zur dynamischen Pflege von Inhalten in einem Data Warehouse) sowie
- kontinuierliche (periodisch oder Ereignisbasierte) Überwachung, sog. Monitoring von vorgegebenen Informationsquellen, um Änderungen im Datenbestand für relevante Informationen schnellstmöglich zu erkennen.

In diesem Zusammenhang könnten mobile Agenten auch selbständig nach bisher unbekanntem, geeigneten Quellen im Internet suchen; ein solches sog. Resource Discovery (Chakrabarti et al. 1999) ist Gegenstand laufender Forschung zu mobilen Agenten. Planungen mobiler Agenten für Beobachtung und Informationsbeschaffung im Kontext von dynamischem Data Mining im Web umfassen also insbesondere die folgenden Aufgaben.

1. Verwendung *zeitoptimaler Migrationspfade* durch das Internet (für gegebene Aufgaben und aktuell bekannte Netzwerkbedingungen).
2. Effiziente *Beobachtung und Erkennung von Änderungen* in der Netzwerkumgebung und einer gegebenen Menge von Webseiten (schnellstmögliche Erkennung von Änderungen bei minimalem Aufwand).

Das erste Planungsproblem ist in der Fachliteratur auch als sog. Traveling Agent Problem (TAP) bekannt: Ein mobiler Agent hat zu entscheiden, in welcher Reihenfolge er welche Rechner (Server, Websites) im Netzwerk besucht, um die Gesamtzeit für das Finden von gewünschten Informationen zu minimieren. Wie auch das verwandte Problem des Handlungsreisenden ist das TAP nicht mit vertretbarem Aufwand zu entscheiden (NP-vollständiges Problem). Unter bestimmten, einschränkenden Annahmen über die Netzwerkumgebung und zugrunde gelegte Architektur des Planungssystems, die für die Zwecke der Data Mining Anwendung annehmbar erscheinen, sind polynomiale Lösungen theoretisch und in der Praxis möglich (Brewington et al. 1999).

Auch zu dem zweiten Planungsproblem einer zeit- und kosteneffizienten Beobachtung und Erkennung von Änderungen in der Netzwerkumgebung (Ranganathan et al. 1997) und Dokumenten sind handhabbare Lösungsansätze vorhanden. Die Ergebnisse von Tests, die mit dem Einsatz von mobilen Filteragenten bei gängigen Suchmaschinen im Web gemacht wurden, sind positiv: es konnte eine nach gegebenen (Agenten- Daten, Resultatstransferkosten-) Metriken definierte Kostenersparnis bei der Durchführung von Anfragen von bis zu 90% erzielt werden (Theilmann & Rothermel 1999a,b).

Ein weiterer Vorteil von mobilen Informationsagenten für Wissensentdeckung liegt bei der zeit- und ortsunabhängigen Beschaffung und Verarbeitung von Daten; die Ergebnisse können dem Besitzer des Agenten dann persönlich auf mobilen Telekommunikationsendgeräten, wie WAP-fähige Webhandies, Palms oder Laptops, zusammenfassend präsentiert werden. Das ermöglicht dynamisches Data Mining auch aus der Perspektive der Unterstützung der eigenen operativen Geschäftsprozesse – eine ständige Erreichbarkeit und Versorgung mit aktuellen Informationen ist mit mobilen Agenten unabhängig von dem Status der eigenen Vernetzung möglich.

Ausführliche Forschungsberichte zu dieser Thematik sind derzeit allerdings nicht oder nur sehr beschränkt verfügbar; neuere Fallstudien geben jedoch Anlaß zu einer positiven Bewertung hinsichtlich einer zum Teil erheblichen Wertschöpfung durch den Einsatz mobiler Agenten (Theilmann & Rothermel 1999b, Brewington et al. 1999). Eine Verallgemeinerung dieser Studien steht derzeit noch aus; ihre Durchführung müßte in einem breit angelegten Feldversuch im Internet unter Beteiligung und im Konsens mehrerer Institutionen erfolgen.

Eine weite Verwendung des Standard XML für semantische Inhaltsbeschreibung von Quellen im Web unterstützt die Suche und Erwerb von relevanten Informationen durch Agenten. Im Schaubild 7 ist beispielhaft die Grundstruktur und Funktionsweise des XML-basierten Brokeragenten der *GMD* (Gesellschaft für Mathematik und Datenverarbeitung) skizziert (Huck et al. 1998). Der Agent beobachtet gegebene und sucht relevante Webseiten im HTML-Format. Für die Extraktion von Informationen werden die Seiten mit Hilfe einer gegebenen Extraktionsgrammatik des sog. *JEDI*-Parsers von HTML in XML übersetzt. Die XML-Dokumente werden effizient im *Document Object Model (DOM)* Standard (W3C-DOM) in

einem XML-Warehouse gespeichert und können vom Benutzer selektiv mit Hilfe eines XML-Browsers angefragt werden. Bei der Beantwortung einer Anfrage werden die relevanten XML-Dokumente und ihre Dokumenttypdefinitionen (DTD) kombiniert und angezeigt.

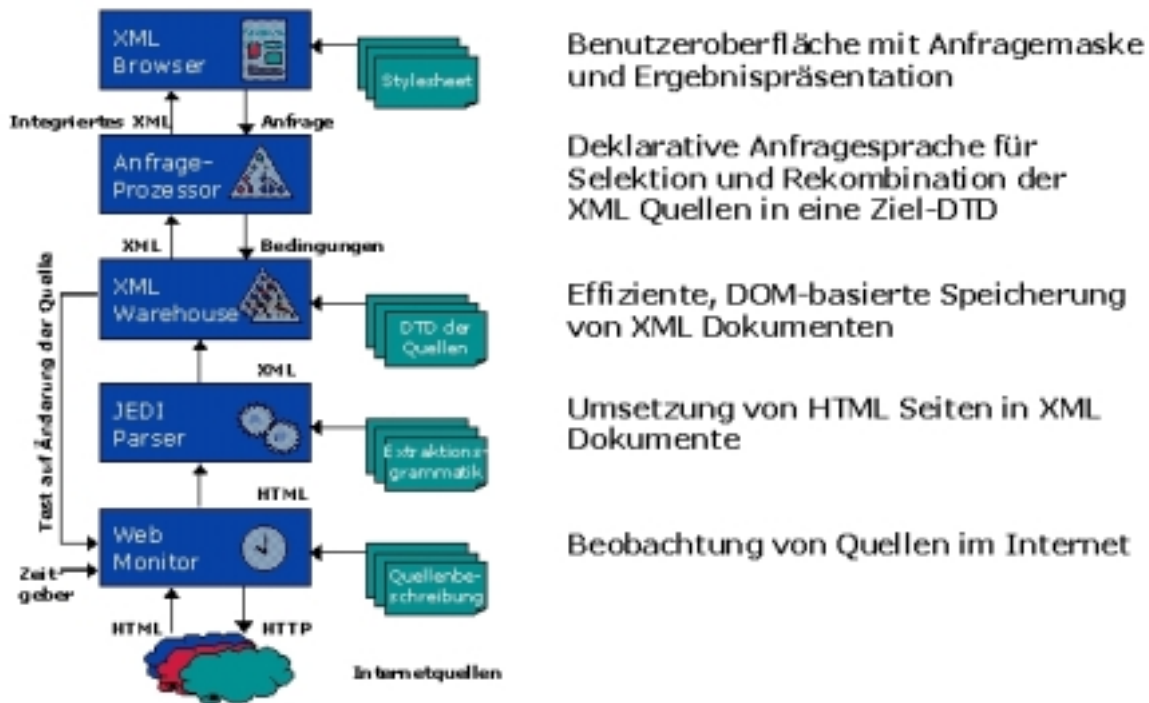


Schaubild 7: Beispiel - Grundstruktur und Funktionsweise des *GMD XML-Brokers* im Web

Einige Beispiele für einzelne Agenten, die Methoden des Data Mining im Web anwenden sind *SportsFinder* (Lu et al. 1997) für die Extraktion von Informationen über Sportergebnisse, und *SpamMiner* (Zighed et al. 1997) für das adaptive Filtern von unerwünschter E-Mail (sog. Spam-Mail).

In (Mladenic 1999) ist eine umfassende Übersicht über adaptive Agenten und agentenbasierte Systeme wie *Personal WebWatcher* und *CiteSeer* gegeben. Alle diese Systeme betreiben sog. Text Mining im Web mit Hilfe einer geeigneten Kombination von Methoden des Maschinellen Lernens und Information Retrieval. Der aktuelle Trend geht dabei in Richtung einer Integration von Sprach- und Bildverarbeitung in den semi-automatischen Prozess der Informationsextraktion aus dem Web.

Systeme von kollaborativen Agenten für verteiltes Data Mining in mehreren, verschiedenen Datenbanken im Internet, wie beispielsweise zur Erkennung von Kreditkartenbetrug (Chan et al. 1999), gibt es bislang nicht.

## 4 Literatur

1. André, E. (1999) Hrsg.: Applied Artificial Intelligence Journal, Special Double Issue on Animated Interface Agents. Vol. 13, No. 4-5.
2. André, E.; Rist, T.; Müller, J. (1998): WebPersona: A Life-Like Presentation Agent for the World-Wide Web. Knowledge-Based Systems, 11(1):25-36.
3. ANSI KIF. Knowledge Interchange Format draft proposed American National Standard (dpANS) NCITS.T2/98-004, <http://logic.stanford.edu/kif/dpans.html>
4. Arens, Y.; Knoblock C.A.; Hsu, C. (1996): Query Processing in the SIMS Information Mediator. Austin Tate (Ed.), Advanced Planning Technology, AAAI Press, CA.

5. Arisha, K.; Kraus, S.; Ozcan, F.; Ross, R.; Subrahmanian, V.S. (1999): IMPACT: Interactive Maryland Platform for Agents Collaborating Together, IEEE Intelligent Systems, March-April. Siehe auch: (Subrahmanian et al. 2000), <http://www.cs.umd.edu/users/vs/agent/impact.html>
6. Baeza-Yates, R.; Ribeiro-Neto, B. (1999): Modern Information Retrieval. Addison-Wesley/ACM Press New York.
7. Ballard, D. (1999): *An Introduction to Natural Computing*. MIT Press.
8. Bauer, M.; Dengler, D. (1998): TriAS-An Architecture for Trainable Information Assistants. In Working Notes of the Autonomous Agents'98 Workshop on Agents in Interaction-Acquiring Competence through Imitation, 1998. Also in Working Notes of the AAI'98 Workshop on AI and Information Integration.
9. Balabonovic, M.; Shoham, Y. (1997): Combining Content-Based and Collaborative Recommendation. Communications of the ACM, March.
10. Beugnard, J. et al. (1999): Making components contract aware. IEEE Computer, July/August.
11. Boutilier, C.; Shoham, Y.; Wellman, M. (1997): Economic principles of multi-agent systems. AI Journal. Vol 94.
12. Bradshaw, J. (1997) Hrsg.: Software Agents. MIT Press.
13. Brauer, W.; Weiß, G. (1998): Multi-Machine Scheduling - a multiagent learning approach. Proc. ICMAS-98, Paris (France), IEEE Press.
14. Brewington, B. et al. (1999): Mobile agents for distributed information retrieval. Chapter 15 in (Klusch 1999).
15. Baru, C. et al. (1999): XML-based information mediation with MIX. Proceedings ACM SIGMOD Conference.
16. Castelfranchi, C.; Falcone, R.; Firozabadi, B.S.; Tan, Y.-H. (1998) Hrsg.: Deception, fraud and trust in agent societies. Proc. Intl. Workshop, Autonomous Agents Conference (Agents-98), NRC Report, Rome, Italy.
17. Cavedon, L.; Sonenberg, L. (1997): On Social Commitment, Roles and Preferred Goals. In Proceedings of Intl Conference on Multiagent Systems ICMAS-98, Paris.
18. Chakrabarti, SW., Dom, B., van den Berg, M. (1999): Focussed crawling – A new approach for topic-specific resource discovery. Proc. 8<sup>th</sup> WWW Conference, Elsevier.
19. Chavez, A.; Dreilinger, D.; Guttman, R.; Maes, P. (1997): A Real-Life Experiment in Creating an Agent Marketplace. Proc. PAAM-97, London (UK), April.
20. Chen, M.S.; Han, J.; Yu, P.S. (1996): Data Mining: An overview from a database perspective. IEEE Transaction Knowledge and Data Engineering, December, pp. 866-883.
21. Contreras, J.; Klusch, M.; Vielhak, T.; Yen, J.; Wu, F. (1999): Multi-Agent Coalition Formation in Power Transmission Planning: Bilateral Shapley Value and Kernel Approaches. Proc. 13th International Power Systems Computation Conference (PSCC-99), Trondheim, Norway.
22. Damgard, I. (1999) Hrsg.: Lectures on data security. Springer, LNCS, Vol. 1561.
23. Dautenhahn, K.; Numaoka, C. (1998) Hrsg.: International Journal on Applied Artificial Intelligence, Special Issue on Socially Intelligent Agents, Vol. 12 (7-8).
24. DARPA I3 Initiative (Intelligent Integration of Information): <http://www.dc.isx.com/I3/>
25. DMG.org (1999): PMML 1.0 – Predictive Model Markup Language. Data Mining Group at University of Illinois at Chicago, USA. <http://www.dmg.org/public/techreports/>
26. Durfee, E. (1999): Distributed Problem Solving and Planning. Chapter 3 in (Weiß 1999).
27. Durfee, E. et al. (1999), Strategic reasoning and adaptation in an information economy. Chapter 8 in (Klusch 1999).
28. Elliott, C.; Brzezinski, J. (1998): Autonomous Agents as Synthetic Characters. AI Magazine, 19(2).
29. Etzioni, O. (1996): Moving up the information food chain: deploying softbots on the WWW. Proceedings of the AAI-96 Conference, AAI Press, 1996.
30. Finin, T.; Fritzson, R.; McKay, D.; McEntire, R. (1994): KQML as an Agent Communication Language. Proc. 3<sup>rd</sup> Intern. Conference on Information and Knowledge Management CIKM-94, ACM Press.
31. Fischer, K. (1999): Agent-based design of holonic manufacturing systems. Journal of Robotics and Autonomous Systems, 27:3-13.
32. Furnas, G.W.; Deerwester, S.; Dumais, S.T.; Landauer, T.K.; Harshman, R.A.; Streeter, L.A.; Lochbaum, K.E. (1988): Information retrieval using a singular value decomposition model of latent semantic structure. In Proceedings of 11<sup>th</sup> Annual Intl ACM SIGIR Conference on Research & Development in IR.
33. Ganti, V.; Gehrke, J.; Ramakrishnan, R. (1999a): Mining very large databases. IEEE Computer, August.
34. Ganti et al. (1999b): Clustering large datasets in arbitrary metric spaces. 15. Int. Conf Data Engineering, IEEE Press, Los Alamitos, pp. 502-511.
35. Garcia-Molina, H. et al. (1995): The TSIMMIS Approach to Mediation: Data Models and Languages. Proc. Workshop NGITS, <ftp://db.stanford.edu/pub/garcia/1995/tsimmis-models-languages.ps>
36. Gehrke, J., Ramakrishnan, R., Ganti, V. (1998): RainForest – A framework for fast decision tree construction of large datasets. Proc. 24th Int Conf Very Large Databases VLDB, Morgan Kaufmann.
37. Genesereth, M.R.; Keller, A.M.; Duschka, O. (1997): Infomaster: An Information Integration System. Proc. ACM SIGMOD Conference, May.

38. Good, N.; Schafer, J.B.; Konstan, J.; Borchers, A.; Sarwar, B.; Herlocker, J.; Riedl, J. (1999): Combining Collaborative Filtering with Personal Agents for Better Recommendations. Proceedings of the 1999 Conference of the American Association of Artificial Intelligence (AAAI-99).
39. Greenberg, M.S. et al. (1998): Mobile Agents and Security. IEEE Communications, Vol. 36(7), July.
40. Gruber, T.R. (1992): *Ontolingua: A Mechanism to Support Portable Ontologies*. Stanford University, Knowledge Systems Laboratory, Technical Report KSL-91-66, March 1992. Siehe auch: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, Vol. 5(2): 199-220, 1993.
41. Gudivada, V.N. (1997): Information retrieval on the world wide Web. IEEE Internet Computing, 1(5).
42. Guha, S.; Rastogi, R.; Shim, K. (1998): CURE: An efficient clustering algorithm for large databases. Proc. ACM SIGMOD Conference on Management of Data, ACM Press.
43. Guttman, R.; Moukas, A.; Maes, P. (1999): Agents as Mediators in Electronic Commerce. In (Klusch 1999), Part II, Chapter 6, Springer.
44. Han, J; Lakshmanan, L.V.S.; Ng R.T. (1999): Constraint-Based, multidimensional data mining. IEEE Computer, August.
45. Huck, G.; Fankhauser, P.; Aberer, K.; Neuhold, E.J. (1998): JEDI: Extracting and Synthesizing Information from the Web. Proceedings of Intl Conference on Cooperativ Information Systems COOPIS-98, New York, August, IEEE Computer Society Press.
46. Huhns, M.; Singh, M.P. (1999): Social abstractions for information agents. In (Klusch 1999).
47. Huhns, M.; Stephens, L.M. (1999): Multiagent systems and Societies of Agents. Chapter 2 in (Weiß 1999).
48. Indermaur, K. (1998): At your service? A state-of-the art report on electronic agents in e-commerce. Magazine Database Management Systems DBMS, Vol. 11(10), September.
49. Jacobs, N.; Shea, R. (1997): Carnot and InfoSleuth - Database Technology and the WWW. ACM SIGMOD Intern. Conf. on Management of Data, May 1995. See also: R.J. Bayardo et al.. InfoSleuth: Agent-Based semantic integration of information in open and dynamic environments. Proc. ACM SIGMOD Intern. Conf. on Management of Data, 1997.
50. Jha, S.; Chalasani, P.; Shehory, O.; Sycara, K. (1998): A Formal Treatment of Distributed Matchmaking. Proc. of 2<sup>nd</sup> Intern. Conference on Autonomous Agents (Agents-98), Minneapolis (MN, USA), May 1998.
51. Karypis, G.; Han, E.-H.; Kumar, V. (1999): Chameleon: Hierarchical clustering using dynamic modeling. IEEE Computer, August.
52. Kerschberg, L.; Banerjee, S. (1999): An Agency-Based Framework for Electronic Business. In (Klusch et al. 1999)
53. Ketchpel, S. (1994): Forming coalitions in face of uncertain reward. In Proc. of AAAI-94 Conference..
54. Kinny, D. et al. (1994): Planned team activity. Springer, LNAI, Vol. 830.
55. Klusch, M. (1999) Hrsg.: Intelligent Information Agents. Springer.
56. Klusch, M.; Shehory, O. (1996): Coalition Formation Among Rational Information Agents. Proceedings of Intl Workshop. MAAMAW-96, Springer, LNAI 1038.
57. Klusch, M.; Shehory, O. (1996): A Polynomial Kernel-Oriented Coalition Algorithm for Rational Information Agents. Proc. 2<sup>nd</sup> Intern. Conf. on Multi-Agent Systems ICMAS-96, Kyoto, AAAI Press.
58. Klusch, M.; Weiß, G. (1998) Hrsg.: Cooperative Information Agents II. Proc. 2<sup>nd</sup> Intern. Workshop CIA-98, Paris, France, Springer, LNAI, Vol. 1435.
59. Klusch, M.; Shehory, O.; Weiß, G. (1999) Hrsg.: Cooperative Information Agents III. Proc. 3<sup>rd</sup> Intern. Workshop CIA-99, Uppsala, Sweden, Springer, LNAI, Vol. 1652.
60. Konstan, J. et al. (1997): GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM 40,3.
61. Krause, S.; Magedanz, T. (1996): Mobile service agents enabling intelligence on demand in telecommunications. Proc. of IEEE GLOBECOM-96 Global Communications Conference. siehe auch <http://www.tinac.com>
62. Kuokka, D.; Harrada, L. (1995): On using KQML for Matchmaking. Proc. 3rd Intl. Conf. on Information and Knowledge Management CIKM-95, pp. 239-45, AAAI/MIT Press.
63. Labrou, Z.; Finin, T.; Peng, Y. (1999): Agent communication languages: The current landscape. IEEE Intelligent Systems, March/April.
64. Lange, D.B.; Oshima, M. (1998): Programming and deploying Java mobile agents with Aglets. Addison Wesley.
65. Lawrence, S.; Giles, C.L. (1998): Searching the world wide Web. Science, Vol. 280.
66. Lewis, M. (1998): Designing for Human-Agent Interaction. AI Magazine, 19(2).
67. Lieberman, H. (1995): Letizia: An Agent That Assists Web Browsing. Proc. 14th International Joint Conference on Artificial Intelligence}, Montreal, August.
68. Lieberman, H. (1999): Personal assistants for the Web: A MIT perspective. Chapter 12 in (Klusch 1999).
69. Losee, R.M. (1998): Text retrieval and filtering. Kluwer.
70. Lu, H.; Sterling, L.; Wyatt, A. (1997): Knowledge discovery in SportsFinder – An agent to extract sports results from the Web. Proc. Knowledge Discovery and Data Mining.

71. Luke, S.; Spector, L.; Rager, D.; Hendler, J. (1997): Ontology-based Web agents. Proc. 1<sup>st</sup> Intern. Conference on Autonomous Agents (Agents-97), Marina del Rey, USA.
72. Magedanz, T.; Breugst, M.; Busse, I.; Covaci, S. (1998): Integrating mobile agent technology and CORBA middleware. AgentLink Newsletter, Issue 1, November. <http://www.agentlink.org> (Newsletter)
73. Marsh, S. (1994): Formalising Trust as a Computational Concept. PhD Thesis, University of Stirling, UK
74. Marsh, S. (1997): A Community of Autonomous Agents for the Search and Distribution of Information in Networks. In Proceedings Information Retrieval Research, 19th BCS-IRSG Annual Colloquium, J. Furner, D. Harper (Hrsg.), Aberdeen, UK.
75. McKnight, L.W., Bailey, J.P. (1997): Internet economics: When constituencies collide in Cyberspace. Internet Computing, November/December.
76. Menczer, F.; Monge, A.E. (1999): Scalable Web search by adaptive online agents: An InfoSpiders case study. Chapter 14 in (Klusch 1999).
77. Merz, M.(1999): Electronic Commerce. dpunkt Verlag.
78. Mitchell, M. (1996): An Introduction to Genetic Algorithms. MIT Press.
79. Mitchell, T. (1997): Machine Learning. McGraw-Hill.
80. Mladenic, D. (1999): Text-Learning and related intelligent agents: A survey. IEEE Intelligent Systems, July/August.
81. Moukas, A. (1998): An Evolving Multi-Agent Information Filtering and Discovery Sytem for the WWW. Journal of Autonomous Agents and MultiAgent Systems, Vol. 1(1):59-88.
82. Moukas, A.; Zacharia, G.; Maes, P. (1999): Amalthea and Histos: Multiagent Systems for WWW sites and reputation recommendation. Chapter 13 in (Klusch 1999).
83. Müller, M. (1999): Interessantheit bei der Entdeckung von Wissen in Datenbanken. Künstliche Intelligenz, KI 3/99, Springer, S. 40 – 42.
84. Nakhaeizadeh, G.; Reinartz, T.; Wirth, R. (1997): Wissensentdeckung in Datenbanken und Data Mining: Ein Überblick. In: Nakhaeizaeh, G. (Hrsg.): Data Mining: Theoretische Aspekte und Anwendungen. Heidelberg, Physica Verlag, S. 1-33.
85. Nodine, M.; Fowler, J. (1999): An overview of active information gathering in Infosleuth. Proc. Intern. Conference on Autonomous Agents, USA, AAAI Press.
86. Noriega, P., Sierra C. (1998) Hrsg.: Agent Mediated Electronic Commerce. Proc. AMET-98 Intl. Workshop on Agent-Mediated Electronic Trading, Springer, LNAI Vol. 1571.
87. Ossowski, S. (1999): Co-ordination in artificial agent societies. Springer, LNAI, Vol. 1535.
88. Ouksel, A. (1999): A Framework for a scalable agent architecture of cooperating heterogeneous knowledge sources. Chapter 5 in (Klusch 1999).
89. Pannu, A.; Sycara, K. (1996): A learning personal agent for text filtering and notification. Proc. Intern. Conference on Cooperative Knowledge-based Systems CKBS-96.
90. Papazoglou, M.P.; Schlageter, G. (1998) Hrsg.: Cooperative Information Systems: Trends and Directions. Academic Press.
91. Papazoglou, M.P.; Laufmann, S.; Sellis, T.K. (1992): An organizational framework for cooperating intelligent information systems. Intern. Journal of Cooperative Information Systems, Vol. 1(1):169-202.
92. Peine, H.; Stolpmann, T. (1997): The architecture of the ARA platform for mobile agents. In K. Rothermel and R. Populescu-Zeletin (Eds.), *Mobile Agents*, Proc. 1<sup>st</sup> Intern. Workshop on Mobile Agents MA-97, Springer, LNCS, Vol. 1219.
93. Pentland, A.P. (1998): Wearable Intelligence. Scientific American, Vol. 9(4).
94. Pham, V.A., Karmouch, A. (1998): Mobile software agents: An overview. IEEE Communications, July.
95. Piatetsky-Shapiro, G.; Frawley, W.J. (1991) Hrsg.: Knowledge Discovery in Databases. Menlo Park, CA: AAAI Press/MIT Press.
96. Picard, R.W. (1997): Affective Computing. MIT Press, Cambridge. siehe auch: <http://www-white.media.mit.edu/vismod/demos/affect/>
97. Picco, G.; Mascolo, C.; Roman, G.-C. (1999): A Fine-Grained Model for Code Mobility. Proceedings of the European Software Engineering Conference held jointly with the 7th ACM SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE '99).
98. Plu, M. (1998): Software technologies for building agent based systems in telecommunication networks. In N.R. Jennings and M. Wooldridge (Hrsg.), *Agent Technology*, Springer.
99. Prasad, M.V.N.; Lesser, V. (1999): Learning situation-specific coordinating in cooperative multi-agent systems. Intern. Journal on Autonomous Agents and Multi-Agent Systems.
100. Ramakrishnan, N.; Grama, A.Y. (1999): Data Mining: From Serendipity to Science. IEEE Computer, August.
101. Ranganathan, M. et al. (1997): Network-aware mobile programs. Proc. USENIX-97 Annual Conference, Anaheim, USA.
102. Rosenschein, J.S.; Zlotkin, G. (1994): Rules of Encounter: Designing Conventions for Automated Negotiation among Computers. The MIT Press.

103. Rus, D.; Gray, R.; Kotz, D. (1997): Transportable Information Agents. In M. Huhns and M.P. Singh (Hrsg.), Readings in Agents, Morgan Kaufmann siehe auch: Journal of Intelligent Information Systems, Vol. 9:215-238.
104. Sandholm, T. (1996): TRACONET - An implementation of the contract net protocol based on marginal cost calculations. Journal on Group Decision and Negotiation.
105. Sandholm, T.; Lesser, V. (1997): Coalitions among computationally bounded agents. AI Journal, Vol. 94.
106. Salton, G. (1989): Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer, Addison Wesley.
107. Schuster, R.; Färber, J.; Eberl, M. (1997): Digital Cash. Springer.
108. SearchBots (Suchmaschinen): <http://ourworld.compuserve.com/homepages/aschick/SUCHMASC.HTM>, <http://www.botspot.com/>
109. Sen, S.; Sekaran, M. (1998): Individual learning of coordination knowledge. Journal of Experimental and Theoretical AI, Special Issue on Learning in DAI Systems.
110. Sen, S.; Weiß, G. (1999). Learning in Multiagent Systems. Chapter 6 in (Weiß 1999).
111. Sen, S.; Biswas, A.; Gosh, S. (1999): Adaptive choice of information sources. In (Klusch 1999).
112. Shafer, J., Agrawal, R., Mehta, M. (1996): SPRINT: A scalable parallel classifier for data mining. Proc. 22. Conf. Very Large Databases VLDB, ACM Press, New York.
113. Shehory, O. (1999): Spawning Information Agents on the Web. Chapter 17 in (Klusch 1999).
114. Shehory, O.; Kraus, S. (1998): Feasible Formation of Coalitions Among Autonomous Agents in Non-Super-additive Environments. Computational Intelligence, Vol. 15(3).
115. Sheth, B.; Maes, P. (1993): Evolving agents for personalized information filtering. Proc. 9th IEEE Conference on AI for Applications, 1993.
116. Sheth, A.; Mena, E.; Illaramendi, A.; Kashyap, V. (1996): OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. Proc. of Intl. Conf. on Cooperative Information Systems CoopIS-96, IEEE Computer Soc. Press.
117. Sheth, A.; Kashyap, V.; Lima, T. (1999): Semantic Information Brokering – How can a multi-agent approach help? In (Klusch et al. 1999).
118. Shoham, Y.; Tenenholz, M. (1997): On the emergence of social conventions: modeling, analysis, and simulations. AI Journal, Vol. 94.
119. Steels, L. (1998): The origins of ontologies and communication conventions in multi-agent systems. Journal on Autonomous Agents and Multi-Agent Systems, Vol. 1(2).
120. Subrahmanian, V.S.; Bonatti, P.; Dix, J.; Eiter, T.; Kraus, S.; Ozcan, F.; Ross, R. (2000): Heterogeneous Agent Systems: Theory and Implementation. MIT Press, Cambridge, USA.
121. Sullivan, D.G. et al. (1999): Intention reconciliation in the context of teamwork: an initial empirical investigation. In (Klusch et al. 1999)
122. Sycara, K.; Zeng, D. (1996): Coordination of multiple intelligent software agents. Intern. Journal of Cooperative Information Systems, Vol. 5(2/3), pp. 181 - 211, World Scientific.
123. Sycara, K.; Zeng, D. (1997): Benefits of learning in negotiation. Proc. AAAI-97, Providence (USA).
124. Sycara, K.; Klusch, M.; Lu, J.; Widoff, S. (1999): Dynamic service matchmaking among agents in open information environments. Journal ACM SIGMOD Record, Spec. Issue on Semantic Interoperability in Global Information Systems.
125. Tambe, M. (1998): Implementing Agent Teams in Dynamic Multiagent Environments. Applied AI Journal 12(2-3).
126. Tesch, T.; Fankhauser, P. (1999): Arbitration and matchmaking for agents with conflicting interests. In (Klusch et al. 1999).
127. Theilmann, W., Rothermel, K. (1999a): Maintaining specialized search engines through mobile filter agents. In (Klusch et al. 1999a).
128. Theilmann, W., Rothermel, K. (1999b): Disseminating Mobile Agents for Distributed Information Filtering. Proc. Joint Symposium ASA/MA'99 of 1st Int. Symp. on Agent Systems and Applications (ASA'99) and 3rd Int. Symp. on Mobile Agents (MA'99), IEEE Press, 1999, pp. 152-161
129. Tolksdorf, R. (1999) Hrsg.: Journal on Autonomous Agents and Multi-Agent Systems, Special Issue on Coordination Mechanisms and Patterns for Web Agents, 1999.
130. Tschudin, C. (1999): Mobile Agent Security. Chapter 18 in (Klusch 1999).
131. Venieris, I.; Hussmann, H. (1998) Hrsg.: Intelligent Broadband Networks, Wiley.
132. Vigna, G. (1998) Hrsg.: Mobile Agents and Security. Springer, LNCS Vol. 1419.
133. W3C-DOM Document Object Model. <http://www.w3.org/DOM/>
134. Weihmayer, R.; Velthuijsen, H. (1998): Intelligent agents in telecommunications. In N.R. Jennings and M. Wooldridge (Hrsg.), Agent Technology, Springer.
135. Weiß, G. (1999) Hrsg.: Multiagent Systems. MIT Press.
136. Weiß, G. (1997) Hrsg.: Distributed Artificial Intelligence meets Machine Learning. Springer, LNCS 1221.
137. Weiß, G.; Sen, S. (1995) Hrsg.: Adaptation and Learning in Multi-Agent Systems. Proceedings of IJCAI-95 Workshop, CA, Springer, LNCS Vol. 1042.

138. Wellman, M.P. (1995): Market-oriented Programming: Some Early Lessons. In S.H. Clearwater (Hrsg.), Market-Based Control: A Paradigm for Distributed Resource Allocation, World Scientific.
139. Wellman, M.P.; Wurman, P.R. (1998): Market-aware agents for a multiagent world. *Robotics and Autonomous Systems* 24:115-125.
140. Wellman, M.P.; Wurman, P.R.; Walsh, W.E. (1998b): The Michigan Internet AuctionBot: A configurable auction server for human and software agents. *Proc. 2<sup>nd</sup> Conference on Autonomous Agents (Agents-98)*.
141. Wiederhold, G. (1992): Mediators in the architecture of future information systems. *IEEE Computer*, 25.
142. Widow, J.; Cer, S. (1996) Hrsg.: *Active Database Systems*. Morgan Kaufmann, CA.
143. Witten, I.; Frank, E. (1999): *Data Mining*. Morgan Kaufmann.
144. Wolfstetter, E. (1996): Auctions: an introduction. *Journal of Economic Surveys*, Vol. 10(4), pp. 367-420.
145. Wong, D. et al. (1997): *Concordia: An infrastructure for collaborating mobile agents*. Springer, LNCS 1219.
146. Wooldridge, M; Jennings, N.R. (1995): *Intelligent Agents: Theory and Practice*. *Knowledge Engineering Review*, Vol. 10(2), 1995. Siehe auch: M. Wooldridge. *Intelligent Agents*. Chapter 1 in (Weiß 1999).
147. Zighed, A.; Cote, M.; Troudi, N. (1997): The data mining, and the technology of agents to fight the illicit electronic messages. *Proc. Knowledge Discovery and Data Mining*.
148. Zhang, T., Ramakrishnan, R., Livny, M. (1996): Birch: An efficient data clustering method for large databases. *Proc ACM SIGMOD Int. Conf. Mgmt of Data*, ACM Press, New York.