

Dokument- und Kontaktsynchronisation mit mobilen Datenbanken: Anforderungen und Lösungsansätze aus Sicht von Unternehmensportalen

Florian Matthes, Vanda Lehel
Arbeitsbereich Softwaresysteme, TU Hamburg-Harburg
f.matthes@tuhh.de, lehel@tuhh.de

Abstract

Dieser Workshopbeitrag befasst sich mit der Synchronisation von Informationsobjekten zwischen personalisierbaren Unternehmensportalen und mobilen Endgeräten aufbauend auf den Erfahrungen des Arbeitsbereichs beim Entwurf und der Realisierung von Unternehmensportalen und Informationsbrokern. Entgegen der gängigen Praxis plädieren wir für eine weitgehende softwaretechnische und funktionale Gleichbehandlung des "persönlichen (mobilen) Portals" und des Unternehmensportals. Diese eher symmetrischen Peer-to-Peer-Architektur führt zu neuen Anforderungen und Lösungsansätzen für das klassische Problem der Dokument- und Kontaktsynchronisation mit mobilen Endgeräten.

1 Problemstellung

Unternehmensportale bieten einen uniformen, personalisierten Zugang über Internet-Technologien zu Informationen für verschiedene Benutzerkreise an [Wegn02]. Diese Informationen setzen sich aus der Integration der Verwaltung der Inhalte von Dokument-Management-Systemen (DMS), Content-Management-Systemen (CMS) sowie Datenbank-Management-Systemen (DBMS) zusammen.

Da das gesamte Management des persistenten Informationsbestandes zentral geschieht, weisen Unternehmensportale gewöhnlich eine klassische Client/Server-Architektur auf.

Ferner ist eine Benutzerverwaltung generischer Bestandteil von Unternehmensportalen. Hierbei werden neben Benutzerinformationen auch Nutzerprofile verwaltet, um eine personalisierte Präsentation der vorhandenen Inhalte zu ermöglichen. Während der Kommunikation der Client-Systeme mit dem Unternehmensportal werden die entsprechenden Daten unidirektional den Clients übermittelt.

Die nachfolgende Abbildung 1.1 stellt das beschriebene Modell eines Unternehmensportals dar.

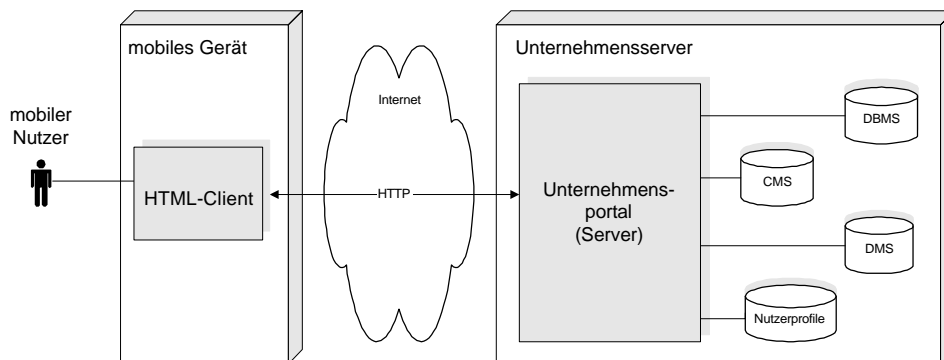


Abbildung 1.1: Unternehmensportal mit klassischer Client/Server – Architektur

Ziel der Arbeit ist es, ausgehend von diesem Modell auch die Client-Systeme als persönliche Informationsportale zu betrachten. Diese sind autonom und können in ihrem Informationsbestand einerseits eigene, lokale Daten verwalten. Andererseits ist auch ein Teil des Informationsbestandes dem des Unternehmensportals äquivalent. Infolgedessen entsteht eine dezentrale Verwaltung des Informationsbestandes, wobei logisch zusammengehörige Daten sich unabhängig voneinander ändern können, da auf den Clients auch offline gearbeitet werden kann.

In diesem Fall ist ein bidirektionaler Datenaustausch zwischen Client und Server erforderlich, um die Unterschiede durch einen Synchronisationsmechanismus auszugleichen.

Das Modell für diese Architektur ist in Abbildung 1.2 dargestellt.

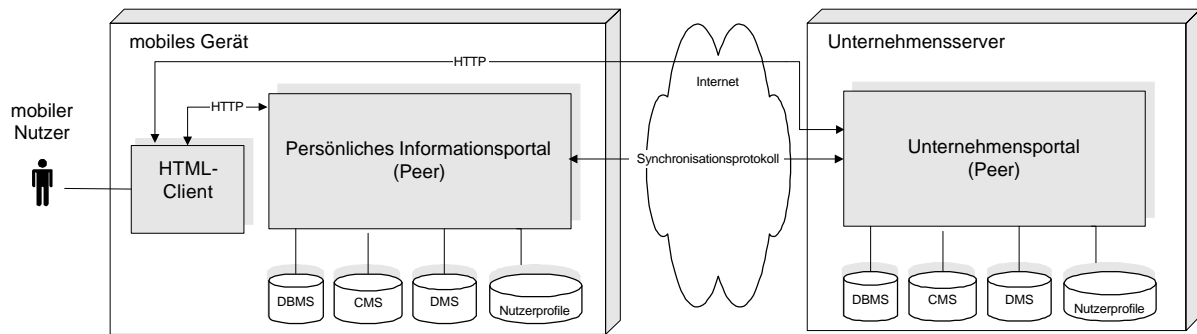


Abbildung 1.2: Kooperation zwischen Informationsportalen

In dieser Arbeit werden Konzepte für die Synchronisation von Informationsobjekten in Unternehmensportalen vorgestellt, wobei Objekte zur Darstellung von Kontakt- und Dokumentinformationen aufgrund ihrer hohen praktischen Relevanz im Vordergrund des Interesses stehen. Als Beispiel wird an dieser Stelle das konzeptuelle Modell des infoAsset Brokers [Wegn02] betrachtet, der eine Basissoftware für Unternehmensportale darstellt (vgl. Abbildung 1.3).

Bei diesem Modell wird jedes Informationsobjekt, das persistent gemacht werden kann, als Asset bezeichnet. In Abbildung 1.3 sind alle Asset-Objekte durch eine schwarz gefärbte linke obere Ecke gekennzeichnet. Es gibt verschiedene Asset-Typen, beispielsweise Personen, Gruppen, Verzeichnisse, Dokumente oder auch Begriffe und zahlreiche assoziative und hierarchische Beziehungen zwischen ihnen.

Für eine Synchronisation von Dokument- und Kontaktinformationen wesentliche Informationsobjekte dieses Unternehmensportals sind einerseits Person-Objekte, die Gruppenmitgliedschaften innehaben, sowie als Autoren und Eigner von Dokumenten fungieren können. Ferner spielen die Dokument-Objekte selbst eine wichtige Rolle, die in Verzeichnissen strukturiert sind und Begriffen zugeordnet werden können. Begriff-Objekte ihrerseits verfügen über eine spezielle Taxonomie durch eine eindeutige Unterordnung sowie eine beliebige Spezialisierung durch andere Begriffe.

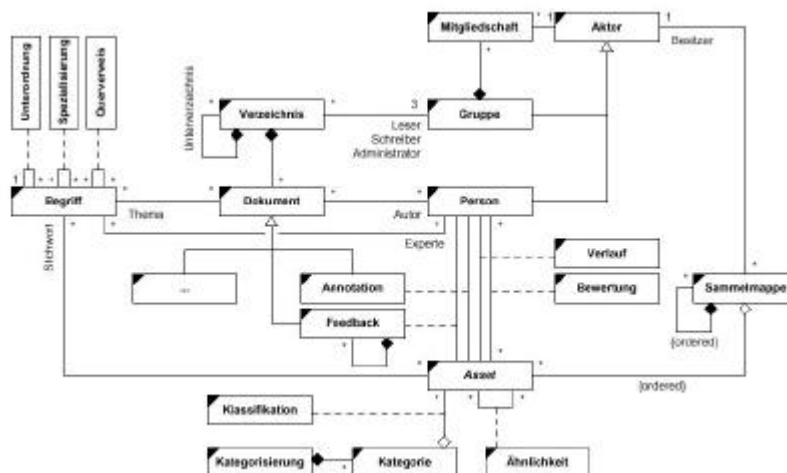


Abbildung 1.3: Konzeptuelles Modell eines Unternehmensportals nach [Wegn02]

In diesem Modell werden also auf der Ebene des Unternehmensportals die Inhalte von darunter liegenden Ebenen des Dokument-Management-Systems, Content-Management-Systems sowie Datenbank-Management-Systems integriert. Eine solche Integrations-schicht ist bei anderen Portalen ebenfalls zu finden wie dem Microsoft Sharepoint Portal Server, dem Hyperwave Information Server und Portal oder dem Lotus Discovery Server und K-Station, siehe [Wegn02].

Im folgenden werden die Anforderungen an einen Synchronisationsmechanismus für Unternehmensportale vorgestellt und die auftretenden Komplexitäten analysiert.

Abschnitt 3 ist einer Übersicht von bereits vorhandenen Synchronisationsmodellen und -werkzeugen gewidmet, die auf unterschiedlichen Ebenen eines Systems eingesetzt werden.

In Abschnitt 4 werden Konzepte für das Design eines Synchronisationsmodells für das Unternehmensportal infoAsset Broker diskutiert.

2 Anforderungen an das Synchronisationsmodell aus der Sicht des CSCW

Bei der Synchronisation von Informationsobjekten in Unternehmensportalen muss eine Reihe von speziellen Anforderungen berücksichtigt werden. Das typische Szenario für die Notwendigkeit einer Synchronisation in dem in Abschnitt 1 vorgestellten dezentralen Modell stellt das Offline-Arbeiten an Informationsobjekten auf autonomen Client-Systemen dar. Die Aufgabe eines Synchronisationsmodells ist demnach die effektive Unterstützung solcher Szenarien, die die im folgenden aufgeführten Anforderungen beinhaltet.

Zum einen ist in diesem Fall ein konsistenter Zustand auch dann gegeben, wenn der gesamte Informationsbestand der Portale auf dem Client- und dem Server-System nicht identisch ist. Es ist somit erforderlich, dem Benutzer eine Funktionalität zur flexiblen, prädikativen Auswahl einer Teilmenge der Objekte anzubieten und nur diese selektive zu synchronisieren. Dabei ist aber auch die benutzerfreundliche Präsentation der möglichen Filterfunktionen zu beachten, um die Semantik der sich daraus ergebenden Synchronisationsoperationen zu verdeutlichen.

Die Informationsobjekte im Unternehmensportal sind semistrukturierte Daten und setzen sich einerseits aus Metadaten, andererseits aus Inhalten (Texten, Bildern, ...) zusammen. Außerdem können Attribute eines Objektes Referenzen auf andere Informationsobjekte enthalten. Daraus ergibt sich, dass die Synchronisation eines Informationsobjektes auch die Synchronisation weiterer referenzierter Objekte nach sich ziehen kann, also eine Erweiterung der Menge der zu synchronisierenden Objekte.

Dem gegenüber steht eine Einschränkung der Ausführung von Synchronisationsoperationen bei fehlenden Rechten auf Informationsobjekten. Eine feingranulare Rechtevergabe für die Art des Zugriffs eines Benutzers auf die Objekte ist im Unternehmensportal festgelegt und darf durch die Synchronisationsoperationen nicht umgangen werden.

Das Zurückschreiben von Änderungen auf den Informationsobjekten muss durch den Benutzer manuell kontrollierbar sein, damit er die Möglichkeit hat, auf die Ausführung von Synchronisationsoperationen mit komplizierter Semantik Einfluss zu nehmen. Ferner kann er in diesem Fall auch die Strategie für die Behebung von Konfliktsituationen, die bei mehreren unterschiedlichen Änderungsoperationen auf dem gleichen Objekt entstehen, bestimmen.

3 Vorhandene Synchronisationsmodelle und -werkzeuge

In diesem Abschnitt werden bereits existierende Synchronisationsmodelle und -werkzeuge betrachtet und analysiert, ob sie den im Falle der Synchronisation von Informationsobjekten eines Unternehmensportals relevanten Anforderungen gerecht werden können.

Eine Synchronisation von Daten kann auf verschiedenen Ebenen eines Systems stattfinden, von der Ebene des Betriebssystems (DMS) über die Datenbankebene (DBMS) bis hin zur Applikationsschicht.

Bei einem Betriebssystem ist eine Synchronisation auf dem Dateisystem möglich, wobei Verzeichnisse und ihre Inhalte, also die Dateien, miteinander verglichen und auf einen identischen Stand gebracht werden können. Hierfür gibt es spezielle Werkzeuge wie Rsync für UNIX [Rsyn01], das möglichst effizient mit der Übermittlung von Änderungsdifferenzen von Dateien arbeitet, Backer für Windows-Systeme [Back02], das eine Vielzahl von Optionen zur selektiven Synchronisation anbietet, Unison [Pier01] oder Reconcile [Reco99]. Die zugrundeliegenden Modelle sind jedoch nur für die Synchronisation von individuellen Dateien aber nicht vernetzten Metadaten geeignet.

Ein anderes Synchronisationsmodell für Dokumente bieten Versionskontrollsysteme. Diese bauen auf einem Workspace-Modell auf, wobei Kopien der Dokumente von einem zentralen Server ausgecheckt und nach einer lokalen Modifikation wieder eingechekkt werden. Konkrete Systeme sind Perforce

[Perf01], CVS [Shar01] oder PRCS [MDon01]. Es können in diesem Fall nur die vom Server verwalteten Dokumente bei der Synchronisation erfasst werden.

Auf der Datenbankebene kann ebenfalls eine Synchronisation der Daten erfolgen, und zwar durch eine Datenbankreplikation und Snapshots, siehe [Hoep01] für einen Kurzüberblick. Diese Funktionalität wird durch DBMS unterstützt.

Schließlich bezieht sich die Synchronisation auf Applikationsebene auf Objekte, bzw. auf einen durch Referenzen zwischen Objekten entstehenden Objektgraphen.

Im vorliegenden Fall handelt es sich um die Ebene des generischen Objektmodells des Unternehmensportals. Die vorgegebenen Anforderungen erfüllen sich durch die Nutzung bereits vorhandener Integrationsdienste für das Content-Management sowie die Benutzerverwaltung mit Zuweisung von Rechten auf Objekte. Ferner können die Suchfunktionen für Informationsobjekte im Unternehmensportal für eine selektive Synchronisation (Filterfunktionen) genutzt werden.

Ein offener Standard für ein Synchronisationsmodell für mobile Geräte wird durch das SyncML Synchronisationsprotokoll in [SyML01a] und [SyML01b] spezifiziert. Hier werden auch standardisierte Austauschformate für praxisrelevante Informationstypen (z.B. vCard-Standard für Kontaktinformationen, vCalendar für Termine) angeboten.

4 Design eines Synchronisationsdienstes

Ebene der Synchronisation

Aufbauend auf dem generischen Objektmodell des infoAsset Brokers (vgl. Abbildung 4.1) soll ein Synchronisationsmodell entworfen werden.

Wie in Abschnitt 1 erläutert, werden Informationsobjekte im Unternehmensportal infoAsset Broker als Assets bezeichnet.

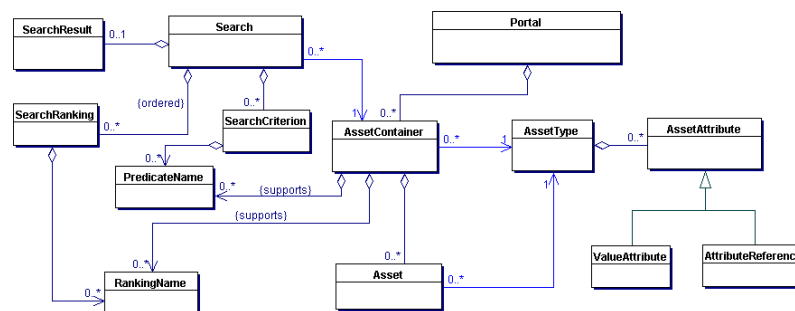


Abbildung 4.1: Generisches Objektmodell eines Unternehmensportals

Assets können zwei Arten von Attributen haben, nämlich Attribute mit einem bestimmten Wert oder mit einer Referenz auf ein anderes Asset. Alle Assets eines bestimmten Typs werden von einem AssetContainer verwaltet. Dies bedeutet zum einen die Verwaltung des Lebenszyklus von Assets, also das Erstellen und Löschen der Assets durch ihren zugehörigen AssetContainer. Zum anderen wird über den AssetContainer auch die Suche (Search) nach Assets, deren Attribute bestimmte Kriterien erfüllen (SearchCriterion), ermöglicht. Die Qualität der Suchergebnisse wird wiederum nach vordefinierten Merkmalen beurteilt (SearchRanking) und entsprechend sortiert.

Wahl des Synchronisationsprotokolls

Für die Wahl des Synchronisationsprotokolls gibt es verschiedene Möglichkeiten. Einerseits kann man sich an Protokollen und offenen Standards orientieren, die sich in der Praxis für vergleichbare Zwecke bereits bewährt haben. Andererseits ist es vorstellbar, ein proprietäres, speziell angepasstes Protokoll für die Synchronisation von Informationsobjekten zu entwickeln.

Das Synchronisationsprotokoll muss einerseits das Transportprotokoll zwischen den Systemen festlegen, andererseits aber auch die Repräsentation der Informationsobjekte und der Synchronisationsoperationen während der Übermittlung.

Für das Transportprotokoll bietet sich beispielsweise HTTP an, um auch hierbei auf Standardportaldienste bei den Request/Response-Zyklen zurückgreifen zu können. Außerdem kann auch das Sessionmanagement den Anforderungen gerecht erweitert werden. Weiterhin können Sicherheitsdienste unmittelbar genutzt werden, wie die Überprüfung von Berechtigungen oder Verschlüsselungsmechanismen. Weiterhin sind durch die Nutzung eines Standardprotokolls auch andere mobile Clients synchronisierbar. Bei einem speziellen Transportprotokoll müssen alle erforderlichen Dienste neu entwickelt werden; allerdings kann jeder einzelne optimal den Bedürfnissen des Synchronisationsmechanismus angepasst werden.

Die Repräsentation der Informationen kann an offene Standards wie SyncML, siehe [SyML01a] und [SyML01b], angelehnt werden. Es können jedoch bei der Synchronisation von Informationsobjekten Erweiterungen notwendig sein, um die in Abschnitt 2 erläuterten Anforderungen zu erfüllen. Ein wesentlicher Fall ist auch das frühe Ausschließen von Synchronisationsoperationen, also vor der Übermittlung der entsprechenden Informationsobjekte. Dies tritt ein, falls der Benutzer nicht berechtigt ist, die gewünschten Operationen ausführen zu lassen. Hierfür ist eine eigens definierte Funktionalität des Protokolls erforderlich.

Synchronisationsprofile und –aufgaben

Die Synchronisation von Informationsobjekten erfolgt zwischen zwei Portalen (vgl. Abbildung 4.2), wobei eines der Systeme die Rolle des Client Portals, das andere die Rolle des Server Portals übernimmt. Dabei initiiert immer das Client Portal aus einer SynchronisationSession die Synchronisation. Dies kann sowohl manuell als auch ereignisgesteuert und damit automatisch erfolgen.

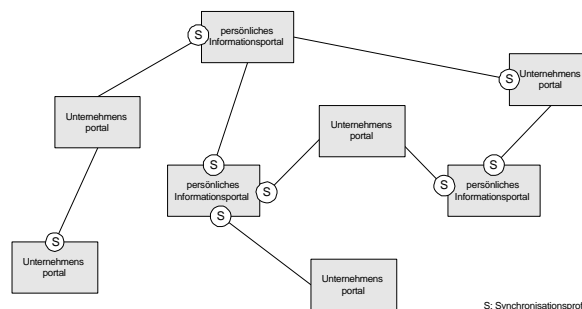


Abbildung 4.2: Synchronisation zwischen Portalen

Das Client-Portal verwaltet außerdem das SynchronisationsProfile, das die entsprechenden Servereinstellungen enthält. Ein Synchronisationsprofil besteht aus mehreren SynchronisationTasks, die jeweils über die Suchfunktionalität eine gefilterte Auswahl der Menge der zu synchronisierenden Objekte pro Assettyp beinhalten sowie die Richtung der Synchronisation festlegen. Danach richtet sich auch die Semantik der beiden Informationsbestände auf Client und Server, nämlich des Source Container und des Target Container.

Bei der Push-Methode einer One-Way-Synchronisation publiziert der Client dem Server seine Änderungen, somit ist der Source Container beim Client und der Target Container beim Server. Für die Pull-Methode, bei der der Client Änderungen des Servers übernimmt, gilt die entgegengesetzte Semantik. Bei der Two-Way-Synchronisation werden Informationsobjekte in beide Richtungen ausgetauscht. Die Informationsobjekte, die synchronisierbar sind, können in diesem Modell Personen, Dokumente oder Verzeichnisse sein.

Für die Synchronisation wesentliche Aspekte sind einerseits die Identifikation der äquivalenten Objekte in den beiden Portalen, die voneinander unabhängig ObjektIDs vergeben können. Bei unterschiedlichen ObjektIDs ist eine IDMapping der entsprechenden Informationsobjekte notwendig, die auch dem Synchronisationsprofil für den jeweiligen Server zugeordnet wird. Die Abfolge der Modifi-

kationen an Informationsobjekten kann durch eine Verwaltung von ChangeLogs realisiert werden, die bei der zusätzlichen Anwendung von Merging-Tools eine wichtige Rolle spielt. Andernfalls ist es möglich, auch über das Attribut timeStamp von Assets Änderungen an einem Objekt festzustellen. Wenn nur Zeitstempel benutzt werden, sind keine komplizierten Merging-Strategien auf die Inhalte anwendbar. Strukturierte Metadaten können jedoch auf diese Weise miteinander verglichen werden.

Die letztendlich durch Benutzerinteraktion beeinflussten ausgeführten Synchronisationsoperationen (SynchronisationAction) werden ebenfalls dem entsprechenden Synchronisationsprofil zugeordnet.

Das integrierte Synchronisationsmodell in das Objektmodell des infoAsset Brokers ist in Abbildung 4.3 dargestellt.

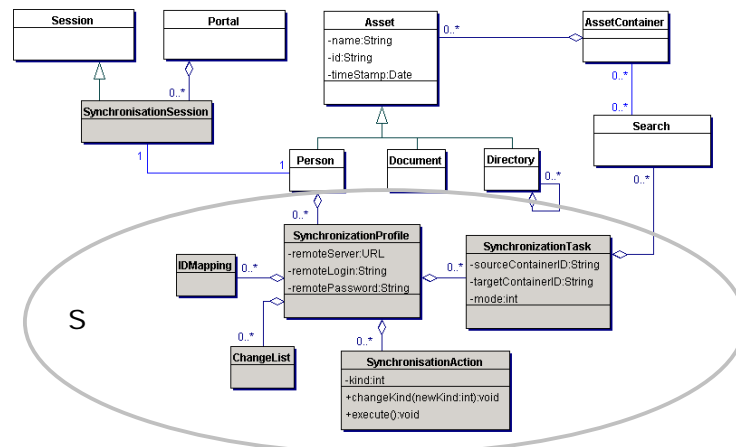


Abbildung 4.3: Synchronisationsmodell für den infoAsset Broker

Weitere Optionen umfassen eine besonders fein definierbare Filtersemantik durch Angabe der Art eines Filters. Es ist möglich, die prädikative Selektion der zu synchronisierenden Objekte zusätzlich zum Einschluss-Filter durch Filter mit Ausschlusssemantik sowie durch Filter zum Erzwingen der Synchronisation für bestimmte Objekte zu ergänzen wie in [Back02].

Als Erweiterung können für die Informationsobjekte zusätzlich zu den eigentlichen Attributen auch spezifische Informationen verwaltet werden, die während des Synchronisationsprozesses als Argumente für Vergleichs- oder Mergetools dienen.

Literatur

- [Back02] BACKER 5 Produktinformation, 2002
<http://www.leanware.com/deutsch/produktinformation.htm>
- [BaPi98] BALASUBRAMANIAM S.; PIERCE, Benjamin C. What is a File Synchronizer? , Mobile Computing and Networking S. 98-108, 1998
- [Goll01] GOLLMICK, Christoph. Unterstützung mobiler Clients durch erweiterte Client/Server-Datenbanksysteme, Grundlagen von Datenbanken 2001, S. 43-47
- [Hoep01] HOEPFNER, Hagen. Grundlagen von Replikationstechniken für mobile Datenbanken, BTW Studierenden-Programm 2001, S. 7-9
- [Lubi00] LUBINSKI, Astrid. Replizieren und Reduzieren von Daten für ressourcenbegrenzte Umgebungen, Grundlagen von Datenbanken 2000, S. 66-70
- [MDon01] MACDONALD, Joshua. PRCS Project Revision Control System , <http://prcs.sourceforge.net/> , 2001
- [Perf01] PERFORCE P4 Produktinformation <http://www.perforce.com/perforce/products.html>, 2001
- [PhBa99] PHATAK, Shirish Hemant; BADRINATH, B. R.: Conflict Resolution and Reconciliation in Disconnected Databases, DEXA Workshop 1999: S. 76-81
- [Pier01] UNISON File Synchronizer, <http://www.cis.upenn.edu/~bcpierce/unison/manual.html>, 2001
- [Pitou95] PITOURA, Evaggelia; Bhargava, Bharat K. Maintaining Consistency of Data in Mo-

- bile Distributed Environments, ICDCS 1995: S. 404-413
- [Reco99] RECONCILE User's Guide, <http://www.merl.com/reports/TR99-14/>, 1999
- [Rsyn02] RSYNC 2.5.2 Dokumentation, <http://samba.anu.edu.au/rsync/>, 2002
- [Shar01] SHARMA, Kamil; CVS Client-Server Version Control, <http://www.linuxgazette.com/issue66/sharma.html>, 2001
- [SyML01a] SYNCML Sync Protocol 1.0.1, http://www.syncml.org/docs/syncml_protocol_v101_20010615.pdf, 2001
- [SyML01b] SYNCML Representation Protocol 1.0.1, http://www.syncml.org/docs/syncml_represent_v101_20010615.pdf, 2001
- [Wegn02] WEGNER, Holm. Analyse und objektorientierter Entwurf eines integrierten Portalsystems für das Wissensmanagement, Dissertation, 2002
- [YODN01] YEE, Wai Gen; OMIECINSKI, Edward; DONAHOO, Michael J.; NAVATHE, Shankant B. Scaling Replica Maintenance In Intermittently Synchronized Mobile Databases, CIKM 2001: S.450-457