

Tobias Kraft, Carmen Constantinescu, Nicola Hönle, Frank Wagner

Bericht zur BTW2003

1 Allgemeines

Vom 26. bis 28. Februar 2003 hatte sich die Datenbank-Community zum zehnten Mal zur zweijährig stattfindenden Fachtagung des GI-Fachbereichs Datenbanken und Informationssysteme in der Alma Mater zu Leipzig eingefunden. Passend zum zehnten Jubiläum wurde der etwas angestaubte Titel der Tagung an den Zeitgeist angepasst, indem die Bedeutung des Kürzels BTW entsprechend neu definiert wurde. BTW steht nun offiziell für *Business, Technologie und Web* und nicht mehr für *Büro, Technik und Wissenschaft*. Eine weitere Neuerung ist der unter <http://www.btw2003.de/proceedings/proceedings.shtml> elektronisch verfügbare Tagungsband. Neben den Tagungsbeiträgen der gedruckten Version kann sich dort jeder Interessierte kostenlos auch die zugehörigen Folien herunterladen und die Bilder der Vortragenden betrachten. Außerdem wurde in DBLP (Digital Bibliography & Library Project) eine Verlinkung auf die Volltexte der Beiträge erstellt. Die Fachtagung für Kommunikation in Verteilten Systemen KiVS fand diesmal zeitgleich in Leipzig statt. Auf der BTW konnten mit 272 Teilnehmern (davon ca. 50 aus der Industrie und 65 Studenten) deutlich mehr registriert werden als in den letzten Jahren; beide Fachtagungen zusammen brachten es sogar auf beinahe 500 Teilnehmer. Das BTW-Programmkomitee unter der Leitung von Prof. Dr. Gerhard Weikum wählte aus ca. 80 wissenschaftlichen Einreichungen 14 Lang- und 17 Kurzbeiträge aus. Das vom Industriekomitee unter dem Vorsitz von Dr. Harald Schöning zusammengestellte Industrieprogramm umfasste 11 Beiträge unter Beteiligung der führenden Datenbank-system-Hersteller.

Im Vorfeld der BTW vom 24. bis 26. Februar 2003 fanden in Leipzig die 9. Datenbank-Tutorientage (DBTT 2003) statt. Diese wurden vom Fachbereich Datenbanken und Informationssysteme (DBIS) der GI und der Deutschen Informatik-Akademie (DIA) gemeinsam organisiert. An den von Prof. Dr. Klaus Küspert koordinierten Tutorien nahmen insgesamt über 150 Teilnehmer teil. Außerdem wur-

den parallel dazu zwei Workshops veranstaltet, die mit je ca. 40 Teilnehmern gut besucht waren. Das Studierendenprogramm wurde ebenfalls schon im Vorfeld durchgeführt. Die von BTW und KiVS gemeinsam ausgerichtete Eröffnungsveranstaltung fand am Mittwochmorgen im Mendelssohn-Saal des Gewandhauses statt. Die Grußworte kamen von Prof. Dr. V. Bigl (Rektor), Dr. Dr. G. Maibaum (SMWK) und D. Schubert (Stadt Leipzig). Die musikalische Umrahmung übernahm das Quartett des Akademischen Orchesters Leipzig. Das anschließende Konferenzprogramm setzte sich aus einem wissenschaftlichen Programm, einem Industrieprogramm, mehreren eingeladenen Vorträgen und einem Rahmenprogramm zusammen. Die eingeladenen Vorträge bildeten hierbei den Auftakt zu jedem Konferenztag; das Industrieprogramm fand parallel zum wissenschaftlichen Programm statt. Zum Abschluss der BTW wurden die Dissertationspreise verliehen und nach einigen Dankesworten an das Organisationskomitee der Veranstaltungsort der nächsten BTW bekannt gegeben: 2005 wird die BTW wieder an ihre »Geburtsstätte« nach Karlsruhe zurückkehren, wo sie 1985 zum ersten Mal stattgefunden hat.

Nicht nur die Vorträge sondern auch das Rahmenprogramm fand großen Zuspruch bei den Teilnehmern. Am Mittwochabend bei der »Get-Together«-Veranstaltung konnten so neue Kontakte geknüpft oder alte wieder aufgefrischt werden. Die von BTW und KiVS gemeinsam organisierte Abendveranstaltung fand am Donnerstagabend in Auerbachs Keller, Leipzigs wohl bekanntestem Restaurant, statt. Unter dem Motto »auf Goethes Spuren« erschienen um Punkt neun Uhr Mephisto und Gretchen höchstpersönlich im historischen Gewölbekeller und bildeten mit ihrer theatralischen Einlage den Höhepunkt des Abends. Am Freitagmittag wurde für die, die noch Zeit und Lust hatten, ein Stadtrundgang angeboten, auf dem die historisch bedeutendsten und schönsten Gebäude der Stadt aufgesucht wurden.

Während der gesamten Tagung

herrschte eine angenehme Atmosphäre. Daher wollen wir auch von unserer Seite ein großes Lob an alle, die an der Organisation und Durchführung der Tagung beteiligt waren, insbesondere an das Organisationskomitee um Prof. Dr. Erhard Rahm, aussprechen. Sie haben ein wunderbares Tagungs- und Rahmenprogramm zusammengestellt, das wohl allen Teilnehmern noch lange in positiver Erinnerung bleiben wird. In diesem Zusammenhang sind natürlich auch die zahlreichen Helfer zu erwähnen, die einen reibungslosen Ablauf garantiert haben; nicht zu vergessen die vielen Sponsoren, die durch ihre finanzielle Unterstützung diese Tagung erst möglich gemacht haben und teilweise auch durch Informationsstände auf der Tagung vertreten waren.

Im Folgenden wird ein kurzer Überblick über die verschiedenen Schwerpunkte der einzelnen Programmteile gegeben.

2 Tutorien

Trends in der SQL-Norm – SQL:200x und Information Integration (SQL/MED)

Das erste Tutorium beschäftigte sich mit dem Thema Information Integration und der Ausprägung von SQL/MED im neuen SQL-Standard. Jan-Eike Michels von IBM aus San Jose, Kalifornien, beschäftigt sich mit der Standardisierung von SQL und sitzt sozusagen direkt »an der Quelle«, so dass er in diesem Tutorium den aktuellen Stand präsentieren konnte.

Der SQL/MED-Standard zum Management externer Daten bietet Datalinks und die Integration von Daten über Wrapper. Datalinks sind Datentypen für Verweise auf Dateien, wobei diese Dateien in einem Datalinker verwaltet werden können. Bei der Integration von externen Daten über Wrapper können z.B. Abbildungen externer Tabellen auf vom Datenbanksystem ansprechbare Foreign Tables definiert werden.

Digitale Bibliotheken und Content Management

Das Tutorium »Digitale Bibliotheken und Content Management« wurde von Prof. Dr. Andreas Heuer von der Uni Rostock gehalten. Das Hauptziel des Tutoriums war die ausführliche Erläuterung der Themen Digitale Bibliotheken und Content Management, insbesondere die Bezüge

zur Thematik Datenbanksysteme. Der Inhalt gliedert sich in die folgenden sechs Teile: Überblick über Digitale Bibliotheken und Content Management, Phasen des Content Management, Rahmenarchitektur für Digitale Bibliotheken, Anwendungsszenarien, Softwarewerkzeuge und Anwendungsprojekte.

Im ersten Teil wurden die folgenden Konzepte definiert: Content, Content einer digitalen Bibliothek, Content-Management-Systeme und Web-Content-Management-Systeme. Hierbei wurden die Unterschiede zwischen klassischer Bibliothek und digitaler Bibliothek hervorgehoben. XML-Datenbanksysteme wurden als eine geeignete Plattform zur Verwaltung der Daten und Dokumente, zur Suche nach Dokumenten usw. bewertet. In »Phasen des Content Management« wurden die Systeme dargestellt, die am Lebenszyklus des Content Managements beteiligt sind, d.h. Systeme zum Sammeln von Content, Redaktionssysteme, Repositories/Content-Verwaltung, Workflow-Systeme und Administrationssysteme und die Architektur von Web-Content-Management-Systemen. In »Rahmenarchitektur für Digitale Bibliotheken« wurden die Begriffe Services beim Autor, Verreiber, Vermittler und Kunden genauer erklärt und die Softwarekomponenten im Überblick präsentiert.

Zwei interessante Anwendungsszenarien, die digitale Universitätsbibliothek und der Kauf von Musik und Videos über das Internet, haben die bereits dargestellten Konzepte veranschaulicht. Für Implementierungszwecke wurden die folgenden Technologien empfohlen: XML, ORDBMS und SQL-99/MM. Am Ende wurden einige wichtige Anwendungsprojekte kurz vorgestellt.

Web Services Infrastructure

Im dritten Tutorium gab Prof. Leymann von IBM Deutschland (Böblingen) eine Einführung in den Bereich Web-Services und einen Einblick in aktuelle Entwicklungen.

Nach einer kurzen Einführung der grundlegenden Begriffe wurde zunächst der Zugriff auf Web-Services über SOAP behandelt. Dabei wurde ausführlich auf die Möglichkeit eingegangen, eine Nachricht über mehrere Intermediaries zu verschicken, die jeweils unterschiedliche

Teile der Nachricht auswerten können. In den folgenden beiden Abschnitten wurden die Definition von Web-Services mit WSDL und das Auffinden mit UDDI und WSIL erklärt. Im Abschnitt Security wurde WS-Security zur Verschlüsselung von (Teilen von) Nachrichten sowie WS-Trust zum Aufbau und zur Verwaltung von Vertrauensbeziehungen behandelt. Anschließend wurde gezeigt, wie mit WS-Policy Eigenschaften wie Sicherheit und Antwortzeit definiert werden können, die dann bei der dynamischen Auswahl eines Dienstes berücksichtigt werden. Einen großen Teil des Vortrages hat dann die Aggregation von Diensten zu Business-Prozessen mittels der von Microsoft und IBM entworfenen Sprache BPEL eingenommen. Zum Abschluss wurden die Besonderheiten der Grid Services von der Open Grid Services Architecture beschrieben. Insgesamt wurde ein sehr interessantes Tutorium zum Thema und zu den aktuellen und anstehenden Entwicklungen angeboten.

Zugriff auf XML in Datenbanken: XQuery und SQL/XML

Im vierten Tutorium zeigte Dr. Harald Schöning von der Software AG, Darmstadt, verschiedene Möglichkeiten zum Zugriff auf XML-Daten in Datenbanken auf. Zunächst wurde XQuery ausführlich und umfassend vorgestellt. Anschließend ging der Dozent noch auf die zwei Komponenten von SQL/XML ein: einen XML-Datentyp mit Funktionen und eine Abbildung von SQL-Tabellen auf XML. Alle Konzepte und Sprachelemente wurden durch passende Beispiele vertieft. Die Eigenschaften der vorgestellten Standards wurden abschließend zusammengefasst und ihre Eignung bewertet.

Für dieses Tutorium gibt es ein passendes Buch des Referenten: XML und Datenbanken – Konzepte und Systeme, erschienen im Carl Hanser Verlag, ISBN 3-446-22008-9.

3 Eingeladene Vorträge

Web-Services: Distributed Applications without Limits

Der wissenschaftliche Teil der BTW und der KiVS wurde durch den eingeladenen Beitrag von Prof. Leymann zum Thema »Web-Services: Distributed Applications Without Limits« eröffnet. Der Vortrag

gab einen Überblick über Web-Services und stellte aktuelle Entwicklungen vor. Web-Services sind an sich nichts Neues, sondern nur eine weitere Möglichkeit zum verteilten Rechnen. Das Neue ist, dass es von allen großen Herstellern unterstützt wird. Damit besteht Hoffnung, dass die Heterogenitätsprobleme in aktuellen Systemen reduziert werden können. Dazu werden Web-Services als virtuelle Komponenten betrachtet, die über einen Service-Bus aufgerufen werden. In der Zukunft soll ein Business Grid die Aufrufe an einen beliebigen geeigneten Service weiterleiten. Mehrere Web-Services können mit BPEL zu komplexen Prozessen zusammengefasst werden, wobei auch abstrakte Prozesse modelliert werden können, die z.B. nur den außerhalb eines Unternehmens sichtbaren Teil eines Workflows zeigen.

Data Integration: A Status Report

Die Datenintegration ist ein Problemfeld, der die Datenbankforschungsgemeinschaft schon seit vielen Jahren beschäftigt. Alon Halevy von der University of Washington, Seattle, begann seinen Vortrag mit einem kurzen Überblick über die Entwicklungen der letzten Jahre auf diesem Gebiet. Der Schwerpunkt des weiteren Vortrags lag auf Peer-data Management und Schema Matching / Mapping. Peer-data Management ist eine flexible Architektur für die Verwaltung und Verarbeitung von verteilten Datenbeständen. Jeder Peer ist dabei sowohl Server als auch Client. Mit Hilfe von lokalem Schema Matching / Mapping können Daten zwischen den Peers ausgetauscht und damit globale Anfragen beantwortet werden, d.h., jeder Peer stellt für die anderen einen logischen Mediator dar. Der letzte Teil des Vortrags war dem Schema Matching gewidmet. Es wurde ein neuer Ansatz vorgestellt, bei dem der Matcher aus einem Korpus von Schematas und Matches lernt und sich so mit der Zeit selbst verbessert.

Ein sehr guter Vortrag, der einen kompakten Überblick über die Entwicklungen der letzten Jahre auf dem Gebiet der Datenintegration gab, der aber auch zeigte, welche Probleme momentan angegangen werden und wo noch Forschungsbedarf besteht.

Information System Architectures: From Art to Science

Nach einer ausführlichen Vorstellung des Dozenten Peter Lockemann von der Uni Karlsruhe durch Gerhard Weikum folgte ein interessanter Vortrag mit dem vielversprechenden Titel »Information System Architecture: From Art to Science«. Die vorgestellte Vorgehensweise zum Architektorentwurf besteht darin, nach der Zerlegung in funktionale Einheiten die wichtigste Dienstqualität des zu entwerfenden Systems in der Hierarchie nach unten zu propagieren und an der Basis zur Verfügung zu stellen. Diese Vorgehensweise wurde mit Hilfe eines Beispiels, eines relationalen Datenbankmanagementsystems mit Performanz als wichtigster Dienstqualität, beschrieben und verifiziert. Im Anschluss wurde der vorgetragene Ansatz interessiert und konträr diskutiert.

4 Wissenschaftliches Programm

Anfrageverarbeitung

Während sich der erste Langbeitrag dieser Sitzung mit der Anfrageoptimierung in relationalen DBMS befasste, lag der Schwerpunkt der nachfolgenden Beiträge, ein Lang- und ein Kurzbeitrag, auf der Anfrageverarbeitung unter Integration unterschiedlicher XML-Dokumente bzw. Datenquellen.

Götz Graefe (Microsoft) machte als namhafter Vertreter dieses Forschungsgebiets den Auftakt zu dieser Sitzung. Damit war ein interessanter Vortrag garantiert, der einen Einblick in aktuelle Forschungsinteressen der Datenbankhersteller gab. Es wurde deutlich gemacht, dass die Verarbeitung von geschachtelten Anfragen viele interessante, aber bisher ungeklärte Fragen aufwirft. In seinem Vortrag hat er unterschiedliche Alternativen in den Bereichen I/O, Datenfluss und Kontrollfluss aufgezeigt und ist dabei insbesondere auf die Themen Caching, Batching, Iteratoren und Parallelität eingegangen.

Thema des zweiten Langbeitrags (Uni Magdeburg) war die Integration heterogener Datenquellen mit schwach gekoppelten Domänen. Da in solchen Szenarien nicht die strukturelle Korrespondenz, sondern die semantische Korrespondenz im Mittelpunkt steht, wurde ein konzeptbasiertes Integrationsmodell vor-

gestellt, dessen globales Schema als Ontologie auf Basis von RDF Schema realisiert ist. Die Objekte aus den lokalen Modellen der Quellsysteme werden während der Verarbeitung einer Anfrage auf die entsprechenden globalen Konzepte abgebildet (local as view). Als Anfragesprache dient ein um Funktionen für den Umgang mit Konzepten erweitertes XQuery. Konzeptschema und Quelldaten stehen hierfür als virtuelle XML-Dokumente zur Verfügung.

Unter dem vielversprechenden Titel »Catalog Integration Made Easy« (Uni Freiburg) wurde ein Verfahren zur Integration von elektronischen Katalogen durch adaptive Auswertung von XPath-Anfragen vorgestellt. Dazu wurden drei Transformationen vorgeschlagen: keine Transformation, Generalisierung von Subanfragen und Eliminierung von Subanfragen. Diese können mehrfach und in unterschiedlicher Kombination auf den XPath-Ausdruck angewandt werden. Daher wurde anschließend eine Bewertung dieser Transformationsfolgen vorgenommen, um eine Ordnung auf ihnen definieren zu können.

XML

Die Attraktivität der XML-Technologie und ihre Auswirkung auf den Datenbankbereich wird durch die Gruppe von vier Forschungsergebnissen (zwei Lang- und zwei Kurzbeiträge) nachgewiesen, die in dieser Sitzung dargestellt wurden.

Ein interessanter Ansatz auf Basis des »Chunking von XML-Dokumenten«, um den Filterprozess für eine Menge von eingetragenen Subskriptionen zu beschleunigen, die auf XPath-Ausdrücken basieren, wurde im ersten Langbeitrag (Uni Erlangen) vorgeschlagen. Der »eXtrakt«-Prototyp und die Leistungsmessungen heben die Leistungsfähigkeit dieses Frameworks hervor.

Der zweite Langbeitrag (Uni Passau, SAP) befasste sich mit einer Erweiterung der relationalen Datenbanksysteme um einen XML-Archivierungsoperator, der die gesamten Archivierungsprozesse auf der Datenebene ausführt und die Daten als XML-Dokumente abzulegen erlaubt.

Im ersten Kurzbeitrag (Uni Freiburg) wurde eine logische, transparente Datenmodellierung für verbundene Dokumente vorgeschlagen. Die Anfragen werden in Standard-XPath gegen das logische Mo-

dell formuliert. Zusätzliche Attribute mit dem dbxlink (database-xlink) namespace spezifizieren das Mapping von XLinks zum logischen Modell.

Der letzte Beitrag (Uni Ulm, EML Heidelberg) stellte eine flexible und effiziente Modellierung, Verwaltung und Abfragemöglichkeit von zeitlichen Informationen durch eine Erweiterung der XPath-Anfragesprache, T-XPath genannt, vor. Diese ist voll abwärtskompatibel zu XPath und stellt für zeitliche Anfragen eine Reihe neuer Operationen und Funktionen zur Verfügung.

Information Retrieval

Der Bereich des Information Retrieval wurde in dieser Sitzung durch einen Lang- und vier Kurzbeiträge dargestellt. Das Problem der »Ähnlichkeitsanfragen«, das im Multimedia- oder Bioinformatikbereich einen größeren Wert gewann, wurde in dem ersten und einzigen Langbeitrag (Uni Bayreuth) adressiert. Es wurde ein Ansatz zur Übertragung einer Rangordnung, die für Objekte eines bestimmten Typs definiert ist, auf verbundene Objekte vorgestellt, mit dem Vorteil, dass die Semantik der Übertragung und der entsprechende Algorithmus als allgemein gültig betrachtet werden können. Die prototypische Implementierung und ihre Benutzungsoberfläche macht deutlich, dass er auch zur grafischen Formulierung von Anfragen geeignet ist, und zeigt, dass die erreichbare Performance konkurrenzfähig ist.

Im ersten Kurzbeitrag (Uni Magdeburg) wurde eine Methode zur Spezifikation von Relevanzgewichten für Anfrage-terme in komplexen Anfragen vorgestellt. Dieser Multi-Level-Gewichts-Ansatz wurde durch zwei Alternativen der Gewichtung repräsentiert. Bei der impliziten Gewichtung wird jedem atomaren Anfrageterm ein Gewicht zugeordnet. Die explizite Gewichtung gibt dem Nutzer die Freiheit, sowohl atomare Terme als auch Teilanfragen explizit mit einem Gewicht zu versehen.

Der zweite Kurzbeitrag (Uni Saarbrücken) befasste sich mit der automatischen Klassifikation von Web-Dokumenten unter einer vorgegebenen Taxonomie. Es wurden Möglichkeiten zur Generierung von Featurevektoren unter Berücksichtigung der Besonderheiten von Web-Dokumenten für vektorbasierte Verfahren des

maschinellen Lernens am Beispiel von Support Vector Machines beschrieben.

Der nächste Kurzbeitrag (Uni Halle) stellte einen Ansatz vor, der versucht, durch eine Kombination aus konventioneller Ähnlichkeitssuche und interaktiver Visualisierung die semantische Lücke zwischen Benutzer und System zu überbrücken. Für die gleichzeitige Visualisierung des Musikstücks wurde eine neue Feature-Extraktionsmethode für Musikdaten entwickelt.

Der letzte Beitrag (Uni Saarbrücken) motivierte, dass Ontologien mit Kategorien und unterschiedlichen semantischen Verhältnissen nützlich sind, die Ähnlichkeitssuche auf XML-Daten zu verbessern. Der vorgestellte Ansatz wurde in der XXL Search-Engine implementiert.

Konsistenz und Standards

Diese Sitzung wurde von je zwei Teams von den Universitäten Lübeck und Jena bestritten.

Das XML-Objects-Projekt XOBÉ an der Uni Lübeck beschäftigt sich mit der Integration von XML und Java. Ein XOBÉ-Programm arbeitet mit XML-Objekten, nur für die Kommunikation mit der »Außenwelt« werden die Daten in XML-Dokumente umgewandelt. Dieser Ansatz erlaubt, die Validität von XML-Objekten innerhalb eines Programms zur Übersetzungszeit zu prüfen. Es wurde ein Algorithmus zur Prüfung der Korrektheit von Anweisungen, die auf XML-Objekte zugreifen, vorgestellt. Dieser Ansatz wurde mit anderen XML-Java-Verknüpfungen wie JDOM, JAXB usw. verglichen.

Aus der Zusammenarbeit der Uni Jena mit IBM Böblingen ergab sich der nächste Beitrag zum SQL/MM-Spatial-Standard. Dieser Standard wurde entwickelt, um räumliche Daten in relationalen Datenbanksystemen speichern und abfragen zu können. Im Vortrag wurde der Standard kurz vorgestellt und kritisch hinterfragt. Zum Beispiel stehen Multi-Datentypen in der Typhierarchie des Standards nicht in Beziehung zu den entsprechenden Einfach-Datentypen, was bei der Weiterverwendung von Zwischenergebnissen zu komplexen SQL-Statements führen kann. Anschließend wurden einfache räumliche Anfragen an eine den Standard implementierende Datenbank vorgestellt.

Der dritte Beitrag stammte wiederum aus Lübeck. In diesem Kurzbeitrag wurde ein System vorgestellt, das dem Anwender durch natürlichsprachliche Erklärungen hilft, Transaktionen und damit Änderungen des Datenbestands auf einem hohen Abstraktionsniveau nachzuvollziehen. Gezeigt wurde, wie Transaktionen in einem geeigneten Format aufgezeichnet und mit Hilfe von Schablonen in natürlicher Sprache übersetzt werden können.

Ein weiterer Kurzbeitrag, diesmal aus Jena, beschäftigte sich mit dem Thema »Integritätsbedingungen für komplexe Objekte in objektrelationalen Datenbanksystemen«. Objektrelationale Datenbanksysteme haben erweiterbare Typsysteme mit z.T. komplex geschachtelten Datenstrukturen, für die keine Integritätsbedingungen formuliert werden können, da diese momentan auf der Attributebene von Tabellen ansetzen. Es wurde ein Vorschlag für eine Ergänzung der SQL-Norm vorgestellt, der sich an vorhandenen Sprachmitteln orientiert.

Indexstrukturen

Diese Sitzung setzte sich aus 2 Lang- und 2 Kurzbeiträgen zusammen, die thematisch recht breit gefächert waren.

Der erste Langbeitrag (LMU München) gab einen Einblick in das Konzept relationaler Indexe. Der Vortrag begann mit einem kurzen Vergleich des relationalen mit dem nativen und generischen Ansatz. Dabei hat sich herausgestellt, dass der relationale Ansatz als einziger eine vollkommen plattformunabhängige Variante darstellt, die einen geringen Implementierungs- und Wartungsaufwand aufweist und sich trotzdem als sehr effizient erweist. Anschließend wurde anhand von Beispielen ein anschaulicher und verständlicher Überblick über die unterschiedlichen Speicherschemata relationaler Indexe, deren Eigenschaften, Vor- und Nachteile gegeben.

Der zweite Langbeitrag (TU München) adressierte die Problematik der Speicherung von XML-Daten in RD-BMS. Hierfür wurde ein multidimensionaler Ansatz gewählt, der ein XML-Dokument in Pfade, Werte und Dokumenten-IDs trennt. Zur Repräsentation der Pfade innerhalb eines XML-Dokuments wurde Multidimensionales Hierarchisches Clustering (MHC) als Nummerierungsschema verwendet. Anschließend

wurden die Ergebnisse detaillierter Messungen unter Verwendung verschiedener Indextypen präsentiert. Diese beinhalten die Zahl logischer und physischer Seitenzugriffe aufgeschlüsselt nach Selektion und Projektion. Dabei hat sich gezeigt, dass sich einige Indextypen bei der Selektion als besonders vorteilhaft erwiesen haben, andere bei der Projektion.

Im ersten Kurzbeitrag (Uni Hannover) wurde die Notwendigkeit eines generischen Indexing Framework für ORD-BMS, das die Implementierung neuer Indextypen erleichtern soll, deutlich gemacht. Es basiert auf generalisierten Suchbäumen (GiST). Durch objektorientierte Spezialisierung können daher schnell neue Indextypen für bestimmte UDTs und Operatoren entwickelt werden. Im Vortrag wurde der für die Oracle-Datenbank entwickelte Prototyp OraGist vorgestellt und die Leistungsfähigkeit von benutzerdefinierten Indizes anhand einiger Messungen bestätigt.

Ein neues Verfahren, um das Routing zwischen Knoten in Peer-to-Peer-Systemen zu verbessern, kommt von der Uni Magdeburg. Bisherige Verfahren bilden jeden Knoten auf einen n-dimensionalen Schlüsselbereich ab, wobei ein Knoten nur die direkten Nachbarn seines Schlüsselbereichs kennt und Nachrichten an Nachbarn in Richtung des Ziels weiterleitet. Daher wurde vorgeschlagen, zusätzlich zu den Nachbarn weitere Knoten in einem Kontaktdaten-Cache zu halten und damit die Zahl der Hops pro Operation zu reduzieren; des Weiteren wurde eine Lokaltäterhaltende Schlüsselabbildung vorgeschlagen. Eine Beschleunigung des Routing konnte durch entsprechende Messungen nachgewiesen werden.

Verteilte Systeme und Web-Services

Das Themengebiet dieser Sitzung war breit gefächert: Der erste Beitrag behandelte die Personalisierung von Web-Services, die folgenden beiden Beiträge betrachteten die Abfrage von Daten in verteilten Systemen und der letzte Beitrag stellte die Besonderheiten beim Handel mit digitalen Gütern vor.

Im ersten Langbeitrag dieser Sitzung (Uni Passau) wurden Technologien vorgestellt, mit denen sich Web-Services auf Vorgaben der Aufrufer einstellen können, was insbesondere im Business-to-Consumer-Bereich interessant ist. Dazu werden

von der Plattform Parameter im Header eines Requests ausgewertet und darauf basierend an geeignete Services weitergegeben.

Im zweiten Langbeitrag (Uni Ulm, Daimler-Chrysler) wurde beschrieben, wie Produktdaten aus einer verteilten Datenbank effizient abgefragt werden können. Je nach Verteilung der Daten müssen bei einer Anfrage unter Umständen einzelne Datenquellen mehrfach angefragt werden. Um den damit verbundenen Aufwand zu reduzieren, werden zusätzliche Referenzen gespeichert. Diese geben zu den Daten einer Quelle an, auf welche weiteren Datenquellen diese verweisen.

Im ersten Kurzbeitrag dieser Sitzung (Uni Saarbrücken) wurde vorgestellt, wie mit Ontologien und Web-Services im »Deep Web« nach Informationen gesucht werden kann.

Der letzte Kurzbeitrag (Uni Oldenburg) behandelte dShops, Onlineshops für den Verkauf von digitalen Waren. Zunächst wurden die Eigenschaften von digitalen Bibliotheken und Onlineshops beschrieben, um dann auf die besonderen Anforderungen der dShops einzugehen.

Konsistenz, Replikation, Workflow

Im einzigen Langbeitrag dieser Sitzung (Uni Erlangen) wurde untersucht, welche Datensätze bei Anfragen auf Datenströmen von mehreren Sensoren verwendet werden sollen. Neben möglichst aktuellen Daten kann es hier auch notwendig sein, alte Daten zu verwenden. In diesem Beitrag wurde eine Möglichkeit zur Quantifizierung der Konsistenz und ein entsprechender Algorithmus vorgestellt.

Im ersten Kurzbeitrag (Uni Leipzig) wurden die Forschungsergebnisse und weitere Vorhaben zum WebFlow-Projekt vorgestellt. Mit WebFlow sollen kooperative Workflows zwischen unterschiedlichen Unternehmen unterstützt werden. Dabei wird die Robustheit der Workflows durch Ausführungsbedingungen und eine flexible, regelbasierte Behandlung von Ausnahmen deutlich verbessert.

Die Modellierung und Abwicklung von Datenflüssen in unternehmensübergreifenden Prozessen wurde im zweiten Kurzbeitrag (Uni Kaiserslautern, Uni Hamburg, Daimler-Chrysler) behandelt. Dabei wurde zunächst untersucht, welche Möglichkeiten es bei der Weitergabe der

Daten gibt. Anschließend wurde die Realisierung mit aktueller EAI-Technologie beschrieben.

Die letzten beiden Kurzbeiträge dieser Sitzung behandelten die Replikation von Daten. Im Beitrag der Uni Oldenburg wurde ein mittels J2EE-Technologie implementierter Replikationsmanager, der das verwendende Replikationsverfahren an die aktuelle Situation anpassen kann, vorgestellt. Im Beitrag von der Uni Jena wurde eine nutzerdefinierte Replikation für neue mobile Anwendungen vorgeschlagen. In der Beispielanwendung geben Benutzer aktuelle Berichte aus ihrer Umgebung in ihren PDA ein, die dann durch das Reiseinformationssystem Hermes verteilt werden.

Datenbanktechnologie in den Lebenswissenschaften

In molekularbiologischen Experimenten entstehen umfangreiche Datensammlungen. Im Rahmen der Genomforschung ist eine wesentliche Anforderung an die Informationstechnologie, Werkzeuge bereitzustellen, die es ermöglichen, implizite Zusammenhänge und Korrelationen in Datensätzen unterschiedlicher methodischer Herkunft aufzudecken.

Zuerst stellte eine Arbeitsgruppe der Humboldt-Universität zu Berlin ein System vor, das von der Aufbereitung der Daten bis zur Visualisierung der Ergebnisse einen integrierten Lösungsansatz bereitstellt. U.a. lassen sich die Daten und Ergebnisse grafisch visualisieren und es kann in ihnen interaktiv navigiert werden. Zum Abschluss wurde die Implementierung des beschriebenen Lösungsansatzes kurz demonstriert.

Der zweite Vortrag dieser Sitzung war ein »Heimspiel«: Von der Uni Leipzig wurde ein System zur Speicherung und Verwaltung von Daten, die bei Microarray-Experimenten anfallen, vorgestellt. Das Microarray-Verfahren wird angewendet, um über die Menge an mRNA in einer Zelle zu bestimmen, welche Gene aktuell aktiv sind. Dabei lassen sich Tausende von Genen simultan messen, weswegen bei jedem Experiment große Datenmengen anfallen. Verschiedene Forschergruppen an der Uni Leipzig führen solche Experimente durch. Anhand einer Sammlung von Anforderungen wurden bestehende Systeme untersucht und als nicht ausreichend bewertet, weshalb ein

neues System mit einem Data-Warehouse-Ansatz implementiert wurde.

5 Industrieprogramm

Integration und Konfiguration

Im ersten Beitrag hat die CronideSoft AG ihre Internet Operating Platform (IOP) vorgestellt. IOP ist ein Enterprise Framework das im Wesentlichen auf den Standards UML, XML und Java aufbaut.

Im zweiten Vortrag wurde das Problem der Disaggregation in einem verteilten System bei der GfK beschrieben. Bei Fehlern oder auch nur zur Überprüfung der Ergebnisse möchte man herausfinden können, welche Daten bei einer Aggregation verwendet wurden. Dazu wurden im Vortrag mehrere Ansätze vorgestellt und verglichen.

In einem Kurzbeitrag von der Firma Swiss Re wurde Information Integration aus Sicht der Praxis beschrieben. Das Problem in der Praxis ist, dass es existierende, funktionierende Systeme gibt, die nicht einfach ersetzt werden können. Deshalb werden zunächst nur die wichtigsten Entities modelliert und in einem Repository gespeichert. Bei der Entwicklung neuer Anwendungen wird dieses Modell dann schrittweise verfeinert.

Der letzte Kurzbeitrag stammte vom Datenbankhersteller Oracle. Anhand eines Beispiels wurde gezeigt, wie mit Tools von Oracle ein Data Warehouse erstellt und verwendet werden kann.

XML und Tuning

Das Thema XML war nicht nur im wissenschaftlichen Programm vertreten, sondern auch im Industrieprogramm in Form von zwei Langbeiträgen und einem Kurzbeitrag. Den Schluss bildete ein Kurzbeitrag zum Thema Datenbank-Tuning.

Im ersten Vortrag (Infonyte) wurde der XML-Prozessor Infonyte-DB vorgestellt, der eine skalierbare Verarbeitung von XML-Dokumenten garantiert. Dies wird ermöglicht durch eine persistente DOM-Implementierung, mit der auch große Datenvolumen ressourcenschonend und effizient verarbeitet werden können, während bei anderen DOM-Implementierungen der Dokumentenbaum in der Regel komplett im Hauptspeicher aufgebaut wird. Neben DOM, XQL, XPath und XSLT soll in Zukunft auch XQuery als Anfragesprache angeboten werden.

Mit der Problematik, Änderungsoperationen auf XML-Dokumenten in eine Anfragesprache einzubinden, befasste sich der Langbeitrag der Software AG. Zuerst wurden die bestehenden Ansätze betrachtet. Danach wurde gezeigt, wie in Tamino die Anfragesprache XQuery um zusätzliche Elementaroperationen zum Einfügen, Ändern, Ersetzen und Löschen von Knoten im Dokumentenbaum erweitert wurde. Anhand einiger Beispiele wurde anschaulich dargestellt, welche Konflikte bei der Durchführung von Änderungen in XML-Dokumenten auftreten können. Es wurde gezeigt, wie diese Konflikte momentan in Tamino gelöst bzw. vermieden werden.

Vom Datenbankhersteller Oracle kam ein Kurzbeitrag, der den Zuhörern näher bringen sollte, wie die Speicherung und Verarbeitung von XML-Daten in Oracle9i realisiert ist. Behandelt wurden die Abbildung von XML auf objektrelationale Datenbankobjekte, die XML-DB-Architektur, der Zugriff über Webschnittstellen, die Anfrageverarbeitung und die Generierung von XML aus objektrelationalen Daten. Aufgrund der begrenzten Zeit konnten leider nicht alle Folien detailliert besprochen werden; die Vortragende stand aber am Informationsstand im Foyer für weitere Fragen und einer Demo zur Verfügung.

Der letzte Beitrag befasste sich mit der automatischen Konfiguration des IBM-Datenbanksystems DB2. Nach einigen allgemeinen Worten zum »Autonomic Computing« wurde der DB2 Configuration Advisor demonstriert, der automatisch die wichtigsten Performance-kritischen Parameter für eine Datenbank setzt. Dazu bekommt der Benutzer 9 Fragen gestellt, die die Anforderungen des Anwendungsszenarios an das DBMS spezifizieren. Aus den Antworten, den gemessenen Systemparametern und den Heuristiken, die auf den Erfahrungen einer großen Zahl von Datenbankadministratoren und -experten basieren, werden automatisch die Parameterwerte berechnet. Die vorgestellten Testergebnisse haben gezeigt, dass ein von einem Experten getuntetes Datenbanksystem zwar ein (minimal) besseres Ergebnis liefert als der DB2 Configuration Advisor, bedenkt man jedoch wie viel Jahre Erfahrung hinter einem solchen Experten stecken und wie viel Zeit ein solches Tuning »von Hand« in Anspruch nimmt im Vergleich

zur Beantwortung der 9 Fragen des Configuration Advisor, dann ist der Configuration Advisor ein durchaus hilfreiches Werkzeug, das sich nicht nur an unerfahrene Datenbankadministratoren richtet, sondern auch einen guten Ausgangspunkt für ein anschließendes Feintuning bietet.

Data Warehousing und Indexierung

Die letzte Sitzung im Industrieprogramm setzt sich aus drei unterschiedlichen Langbeiträgen zusammen.

Im ersten Vortrag (cundus AG) wurde ein Überblick über die logische Datenmodellierung im SAP Business Information Warehouse gegeben. Es wurden das erweiterte Star-Schema und die unterschiedlichen Modellierungsvarianten für hierarchische Dimensionsstrukturen behandelt. Die einzelnen Modellierungsalternativen wurden unter verschiedenen Kriterien bewertet, um eine Entscheidungshilfe für die Wahl der unter den gegebenen Anforderungen passendsten Variante geben zu können.

Im zweiten Beitrag hat die Transaction Software GmbH auf die Notwendigkeit mehrdimensionaler Indexe, besonders im Data-Warehouse-Bereich, hingewiesen. Anhand einiger Beispiele wurde kurz die Funktionsweise des Multidimensionalen Hierarchischen Clustering (MHC) erklärt. Anschließend wurde gezeigt, wie dieses Clustering-Verfahren in der Datendefinitionssprache des eigenen Datenbanksystems Transbase verwendet wird und wie es in die Anfrageverarbeitung integriert wurde. Als ein in diesem Zusammenhang sinnvolles Optimierungskonzept wurde das »hierarchical pre-grouping« vorgestellt, das unmittelbar nach dem Zugriff auf die Faktentabelle eine Gruppierung auf der entsprechenden hierarchischen Ebene vornimmt, um damit die Zahl der Tupel zu verringern.

Der letzte Beitrag stammt von Goetz Graefe, der damit nicht nur im wissenschaftlichen Programm, sondern auch im Industrieprogramm vertreten war. Er betonte nicht nur, dass er in seinem Vortrag etwas zeigen wolle, was jeder selbst nachmachen könne, er forderte die Zuhörerschaft geradezu auf, die Idee der partitionierten B-Bäume zu Hause einmal selbst auszuprobieren. Für die Partitionierung wird jeder Tabelle, auf der ein partitionierter B-Baum-Index erzeugt werden soll, eine zusätzliche führende

Schlüsselspalte hinzugefügt. Bei jedem Einfügevorgang werden nun die hinzukommenden Daten in einer neuen künstlichen Partition gespeichert; dies wird dadurch erreicht, dass die Daten einen neuen Wert in der führenden Schlüsselspalte erhalten. Erst nach und nach werden die Partitionen miteinander verschmolzen. Das heißt, der Gesamtaufwand ist zwar der gleiche wie bei einem nicht partitionierten B-Baum-Index, die alten Partitionen werden jedoch für den Einfügevorgang nicht gesperrt und die Daten in der neuen Partition sind nach kürzester Zeit indiziert und verfügbar. Dies ist in der verzögerten Reorganisation des Index begründet. Im nicht partitionierten Fall könnte eine Tabelle mit Cluster-Index aufgrund der Reorganisation des gesamten Index für längere Zeit gesperrt sein.

6 Dissertationspreise

Zum zweiten Mal in der Geschichte der BTW wurden Dissertationspreise von je 2.500 Euro verliehen.

Ein Preis ging an Reinhard Braumandl (Uni Passau) für seine Dissertation »Quality of Service and Optimization in Data Integration Systems«. In seiner Dissertation hat er ein global agierendes Datenintegrationssystem entwickelt, das eine internetweite Anfrageauswertungsschicht über den Datenquellen definiert, Techniken zur Realisierung von Dienstgütereinverabreden einbezieht und damit eine interaktive Anfrageverarbeitung ermöglicht. Das heißt, es werden mehrere Dienstgütereigenschaften überwacht, um auf Abweichungen von der Dienstgütereinverabreden entsprechend reagieren zu können.

Der zweite Dissertationspreis wurde an Jens Lechtenböcker (Uni Münster) für seine Arbeit »Data Warehouse Schema Design« verliehen. In dieser Dissertation wird Data-Warehouse-Design entsprechend einem Prozessmodell durchgeführt, wie es vom traditionellen Datenbankentwurf her bekannt ist. Es werden daher ebenfalls die folgenden Entwurfsphasen unterschieden: Anforderungsanalyse und Spezifikation, konzeptueller Entwurf und physikalischer Entwurf. Zum Zwecke des logischen Data-Warehouse-Entwurfs wird ein Schematransformationsprozess vorgeschlagen, um eine relationale Implementierung der konzeptuellen Schemata mittels einer

Menge von update-unabhängigen materialisierten Sichten zu realisieren. Um Update-Unabhängigkeit zu erzwingen, wird eine Lösung, die auf dem Begriff der Sichertergänzung basiert, vorgeschlagen, und die Ausdrücke zum Berechnen von Ergänzungen für monotone Sichten werden gegeben.

7 Web-Adressen

Tagungsseite BTW2003:
<http://www.btw2003.de>

Elektronischer Tagungsband:
<http://www.btw2003.de/proceedings/proceedings.shtml>

Bildergalerie:
<http://www.btw2003.de/bildervonderBTW.shtml>

Dipl.-Inf. Tobias Kraft
 Dr.-Ing. Carmen Constantinescu
 Dipl.-Inf. Nicola Hönle
 Dipl.-Inf. Frank Wagner
 Universität Stuttgart – IPVS
 Abteilung Anwendersoftware
 Universitätsstr. 38
 70569 Stuttgart
 {vorname.nachname}@informatik.uni-stuttgart.de
<http://www.informatik.uni-stuttgart.de/ipvs/as/>



Verleihung des Dissertationspreises:
 v.l.n.r.: Prof. Leymann (IBM, Sponsor), Prof. Heuer (Sprecher GI-Fachbereich Datenbanken und Informationssysteme), Dr. Braumandl (Preisträger), Dr. Lechtenböcker (Preisträger), Prof. Weikum (Programmkomitee-Vorsitzender), Prof. Rahm (Tagungsleiter)