

Pressemitteilung

Friedrich-Alexander-Universität Erlangen-Nürnberg

Gertraud Pickel

04.01.2001

<http://idw-online.de/de/news28694>

Forschungsprojekte
Informationstechnik, Wirtschaft
überregional

Anfrageoptimierung in Data-Warehouse-Systemen: Summendaten zur Selbstbedienung

Um Analysen vornehmen zu können, wie sie für einen Überblick der Entwicklung von Verkaufszahlen, für ein Urteil über den Erfolg von Marketingstrategien oder die Abschätzung von Trends erforderlich sind, ist es nicht nötig, alle Einzeldaten heranzuziehen. So ist beispielsweise nicht der Verkauf eines CD-Players zu einem konkreten Datum in einem bestimmten Laden von Interesse, sondern Informationen, in denen sich Kennzahlen zu solchen Verkäufen verdichtet wiederfinden. Am Lehrstuhl für Datenbanksysteme der Universität Erlangen-Nürnberg wird daran gearbeitet, die Vorberechnung derartiger Summendaten oder Aggregate so zu gestalten, daß sie den Ansprüchen für analytische Auswertungen optimal entsprechen. Die deutsche Forschungsgemeinschaft (DFG) hat ein Projekt von Prof. Dr. Hartmut Wedekind und Dipl.-Inf. Andreas Bauer mit dem Thema "Anfrageoptimierung in Data-Warehouse-Systemen durch inhaltsbasierte Aggregationen" zum Oktober 2000 genehmigt.

Als Data Warehouse wird eine Datenbank bezeichnet, die heterogene Daten aus verschiedenen Quellen in einer eigenen, homogenen, in sich stimmigen und in zeitlicher Abfolge geordneten Datenbasis zur Verfügung stellt. In diesem "Warenhaus" sind einlaufende Daten nicht einfach aufgestapelt; sie werden in eine Struktur gebracht, die es ermöglicht, auf gezielte Anfragen mit unterschiedlicher Thematik Antwort zu bekommen. Das Data-Warehouse-System enthält alle Komponenten, die für "Online Analytical Processing" (abgekürzt OLAP) erforderlich sind; es wählt Daten aus und bereitet sie einheitlich auf, fügt neue Daten hinzu, ohne die vorhergehenden Bestände zu verlieren, läßt Vergleiche, unterschiedliche Verknüpfungen und Gruppierungen zu und sorgt für die grafische Darstellung der Ergebnisse.

Inhaltsbasierte Aggregationen

Am Erlanger Lehrstuhl für Datenbanksysteme existiert seit 1994 eine Data Warehouse-Forschungsgruppe, die mehrere Industrie- und Forschungsprojekte durchgeführt hat. In einem vorhergehenden DFG-Projekt zur "Darstellung und kostenoptimierte Verarbeitung komplexer Tabellen in statistischen Datenbanksystemen" hat dieses Team unter anderem ein Verfahren zur Anfrageoptimierung erstellt, das nun aufgegriffen und weiterentwickelt wird. Mittels inhaltsbasierter Aggregationen soll eine neue Methode zur Anfrageoptimierung in Data-Warehouse-Systemen entstehen, die industriell eingesetzt werden kann.

Online Analytical Processing wird angewendet, um aussagekräftige Kennzahlen - im konkreten Fall zur Charakterisierung eines Marktes - aus einer Rohdatenbasis zu ermitteln, die in der Regel sehr groß ist. Täglich kann ein Datenvolumen anfallen, das ein bis zwei Milliarden Schreibmaschinenseiten entspricht. Damit bei Anfragen, die eine Aggregation verlangen, keine unzumutbaren Wartezeiten entstehen, müssen (Zwischen-)Ergebnisse vorberechnet werden.

Aus der Analyse des Benutzerverhaltens und der Datenstruktur soll in dem Forschungsprojekt eine möglichst kostenoptimale Menge von Aggregaten zur Vorberechnung ausgewählt werden, wobei durch eine Beschränkung auf die Speicherung der nutzbringenden Datenbereiche (partitionierte Materialisierung) Hot Spots in den von den Anfragen betroffenen Daten berücksichtigt werden. Anfragen unter Verwendung der Aggregate sollen zudem auf effiziente und transparente Weise ausführbar sein. Darüber hinaus muß es möglich sein, die Aggregate zu aktualisieren, ohne daß die Arbeit eines Anwenders beeinträchtigt wird, der zu gleicher Zeit Analysen vornimmt. Aus der Sicht dieses Benutzers soll der Datenbestand während eines Analysevorgangs stabil bleiben; neu hinzukommende Anwender dagegen haben schon Zugriff auf die "aufgefrischten" Daten aus einem veränderten Rohdatenbestand.

Grundlagen- und praxisorientiert



Seit Mai 1999 kooperiert der Lehrstuhl im Data Warehouse-Bereich mit der Gesellschaft für Konsumforschung (GfK) - Marketing Services in Nürnberg. Als Dienstleister der Konsumgüterindustrie erfaßt die GfK Handelsdaten und stellt ihren Kunden die aufbereiteten Informationen in verschiedenen Berichtsformen zu. Dieser Partner gibt für das neue Projekt die spezifischen Anforderungen vor, die bei der Anfrageoptimierung zu erfüllen sind, und testet die Forschungsergebnisse. Der vorgesehene Technologietransfer garantiert, daß das als Grundlagenforschung geplante Vorhaben gleichzeitig anwendungsorientiert bleibt.

Die Förderung der Deutschen Forschungsgemeinschaft umfaßt eine auf zwei Jahre befristete Stelle für einen wissenschaftlichen Mitarbeiter nach BAT IIa sowie zwei studentische Hilfskräfte, je 19h/Woche.

* Kontakt:

Prof. Dr. Hartmut Wedekind, Dipl.-Inf. Andreas Bauer
Lehrstuhl für Datenbanksysteme, Martensstraße 3, 91058 Erlangen
Tel.: 09131/85 -27892, -27901, Fax: 09131/85 -28854
E-Mail: hartmut.wedekind@informatik.uni-erlangen.de,
andreas.bauer@informatik.uni-erlangen.de