

Felix Naumann, Jens Bleiholder, Melanie Weis

Eine Übung zur Vorlesung Informationsintegration

Zusammen mit einer neu gestalteten, 4-stündigen Vorlesung zum Thema „Informationsintegration“ hat unsere Arbeitsgruppe im Sommersemester 2004 eine ebenfalls neuartige Übung zum gleichen Thema angeboten. Die Vorlesung setzt Datenbankbasiswissen voraus und behandelt weiterführende Themen wie verteilte Datenbankarchitekturen, föderierte Datenbanken, Anfrageplanung, Anfragesprachen, Data Cleansing, Schema Management uvm.

Inspiriert durch den Artikel „*Learning About Data Integration Challenges from Day One*“ von Alon Halevy [Halevy 2003] ließen wir Studenten in der Übung in Einzelarbeit Datenbanken entwerfen, die daraufhin in Gruppenarbeit zu „Firmen“ integriert wurden, um eine WWW Einkaufsplattform zu realisieren. Schließlich trieben die Firmen mittels Web Services untereinander Handel, um Bestellungen befriedigen zu können. So lernten Studenten hautnah die Probleme bei der Integration autonom entworfener Datenbanken und bei dem online Datenaustausch mit anderen Firmen kennen. Trotz anfänglicher Gewöhnungszeit an dieses kommunikationsintensive Lehrkonzept waren die Motivation der Studenten und insgesamt der Lernerfolg groß.

In diesem Artikel stellen wir das Konzept der Übung und unsere Erfahrungen bei der Durchführung dar, und wollen so andere Arbeitsgruppen motivieren und dabei unterstützen, diese spannende Übung durchzuführen.

Das Konzept

Die hier vorgestellte Übung ist eng angelehnt an die Vorschläge aus [Halevy 2003] und die von Alon Halevy und seiner Gruppe freundlicherweise zur Verfügung gestellten Materialien. Die ursprünglichen Aufgaben richten sich an Studenten einer Datenbankgrundvorlesung, also an Studenten, die das relationale Modell, SQL etc. noch nicht kennen. In der Zielgruppe unseres Kurses hingegen hatten Studenten bereits Grundkenntnisse im Datenbankbereich. Entsprechend wurden gewisse Übungselemente gekürzt oder herausgenommen und andere stärker betont. Die Übung ist in drei Phasen unterteilt:

- 1. Schemaentwurf und Datenbankgenerierung:** Basierend auf der Beschreibung dreier verschiedener Anwendungsdomänen sollen die Studenten in Einzelarbeit während der Übungsstunde ein relationales Schema entwerfen. Die Domänen *Rechnung*, *Versand* und *Produkte* sind so gewählt, dass später Gruppen gebildet werden können, in denen jede der Domänen jeweils einmal vertreten ist. Die in der Übungsstunde erstellten Schemata sollen daraufhin direkt und unverändert in jeweils eine Datenbank eingetragen und mit Daten befüllt werden.
Die Erstellung der Schemata in Einzelarbeit und unter zeitlichem Druck innerhalb der Übungsstunde stellt sicher, dass (wie im wirklichen Leben) die Schemata nicht perfekt sind und nicht aufeinander abgestimmt sind. Änderungen im Schema, auch in den folgenden Phasen, sind nur nach einem schriftlichen Antrag mit Begründung erlaubt. Als zweite Aufgabe dieser Phase soll jeder Student ein einfaches Web-basiertes Frontend zu ihrer/seiner Datenbank erstellen.
- 2. Firmengründung, Webshop und Managementanfragen:** Zu Beginn dieser Phase gründen jeweils drei Studenten (jeweils ein Vertreter jeder Domäne) eine virtuelle Firma, entwerfen ein Logo und erstellen eine Firmen-Homepage. Die weitere Arbeit gliedert sich in zwei Teile. Der erste und umfangreichere Teil ist die Erstellung eines Webshops, der es neuen Kunden erlaubt, sich zu registrieren und einzuloggen, die angebotenen Waren zu durchstöbern, gezielt nach Waren zu suchen, einen Warenkorb zu füllen und schließlich den Inhalt des Warenkorbes unter Angabe von Zahlungs- und Lieferbedingungen zu bestellen. Um dies erfolgreich zu realisieren müssen zunächst

die Schemadifferenzen überwunden werden. Beispielsweise enthalten alle Domänenbeschreibungen Angaben über Lagerhäuser. In einem integrierten System muss nun entschieden werden, welche der jeweiligen Relationen mit welchen Attributen verwendet werden soll. Weitere Elemente wie z.B. der Warenkorb, der in keiner Domänenbeschreibung auftaucht, müssen geeignet hinzugefügt werden. Das Anlegen neuer Kundendaten und Bestellungen erfordert gleichzeitigen schreibenden Zugriff auf mehrere Datenbanken.

Die zweite Aufgabe dieser Phase ist die Erstellung eines Web-basierten Frontends, das auf Knopfdruck einige *decision support* Anfragen beantwortet. Diese Anfragen benötigen Daten aus mindestens zwei verschiedenen Datenbanken, so dass Join-Techniken über mehrere Datenbanken entwickelt werden müssen.

3. **Web Services und Handel:** Das Ziel der letzten Phase ist eine Web Service-basierte Interoperation der einzelnen Firmen auf einem virtuellen Marktplatz. Zunächst soll jede Firma einige Web Services mit vorgegebenen Signaturen zur Annahme von Bestellungen, zum Kaufen bei anderen Web Services und zur Protokollierung dieser Transaktionen bereitstellen. Zum Testen dieser Phase wird ein zentrales Produktlager zur Verfügung gestellt, in dem Firmen innerhalb eines bestimmten Budgets per Web Service einkaufen können. Zentral durch die Übungsleiter gesteuert wird ein bestimmter Prozentsatz der Produkte bei den Firmen zurückgekauft. Ziel der Firmen ist es, einen möglichst großen Gewinn zu erzielen. Wenn eine Firma ein Produkt während des Rückkaufs nicht anbieten kann, kann sie das Produkt (ebenfalls per Web Service) bei einer der anderen Firmen einkaufen und weiterverkaufen. Die initiale Einkaufsstrategie und die Preisgestaltung obliegen dabei den einzelnen Firmen.

Den Abschluss der Übung bilden die Erstellung eines kurzen Erfahrungsberichts jeder Firma und die Prämierung der am Marktplatz erfolgreichsten Firma.

Zur Durchführung

Die Übung umfasste zwölf zweistündige Termine. Davon wurden vier Termine für Phase 1, drei Termine für Phase 2, und fünf Termine für Phase 3 genutzt. Während der Übungsstunden vermittelten wir den Studenten das für die jeweiligen Aufgaben nötige Wissen, insbesondere stellten wir neue Technologien vor. Die letzte Stunde vor dem Abgabetermin einer Phase war zur Diskussion vorgesehen, um auf konkrete Probleme bei der Bearbeitung der Aufgaben eingehen zu können und den Austausch zwischen den einzelnen Gruppen zu fördern. Auch außerhalb der vorgesehenen Übungstermine waren die Studenten nicht auf sich allein gestellt, da sie uns jederzeit ansprechen konnten. Dieses Angebot wurde von den Studenten genutzt, und sehr geschätzt.

Technischer Hintergrund der Implementation waren eine DB2 Datenbank, ein Apache Webserver und Tomcat als Servletcontainer. Aus Sicherheitsgründen war der Zugriff nur innerhalb der Universität möglich. Als Programmiersprache wurde Java vereinbart, die Interaktion zwischen Webseiten und Datenbank erfolgte mittels Java Server Pages (JSP). Es wurde nur ein einziger DB2 Benutzer für alle Studenten eingerichtet, was die Administration erleichterte, doch auch das Risiko barg, dass Studenten schreibenden Zugriff auf die Relationen anderer Gruppen hatten.

Um die Webseiten zu erstellen, stellten wir jedem Studenten ein eigenes Homeverzeichnis auf dem Apache Webserver zur Verfügung. Die Rechte wurden so vergeben, dass Studenten zunächst nur auf ihr Verzeichnis Zugriff haben, doch für Phase zwei und drei ihren Gruppenmitgliedern Rechte selbst erteilen können. Um Schemata in DB2 zu implementieren und die Datenbanken mit Daten zu befüllen wurde Aqua Data Studio (www.aquafold.com) empfohlen, und zur Realisierung der Web Services wurde den Studenten das Java Web Services Developer Pack (JWSDP) von Sun zur Verfügung gestellt. Wir schränkten die zu benutzenden

Technologien auf diese Weise ein, um technische Probleme möglichst gering zu halten, und um den Studenten technischen Support garantieren zu können.

Phase 1: Schemaentwurf und Datenbankgenerierung

Nach einer Einführung in die allgemeine Organisation der Übung wurde in der ersten Stunde wiederholt, wie man von einer natürlichsprachlichen Beschreibung einer zu modellierenden Domäne zu einer benutzbaren Datenbank gelangt. Dies umfasste relationalen Datenbankentwurf und SQL. Dieses Wissen wurde während der zweiten Übungsstunde angewandt, in der jeder Student ein Datenbankschema zu einer der Domänen Inventar, Rechnung und Versand entwarf. Informationen zu den Domänen und dem weiteren Verlauf der Übung wurden den Studenten in der ersten Stunde vorenthalten, um Absprachen beim Schemadesign zu vermeiden. Denn wie auch im echten Leben sollten Probleme beim Integrieren mehrerer Schemata auftreten. Die Geheimhaltung des Übungsverlaufs führte bei den Studenten zu anfänglicher Ratlosigkeit, wurde aber akzeptiert. Dennoch entstand eine gewisse und ungewollte Prüfungsatmosphäre, die durch unsere Hilfestellung bei Fragen, und das Angebot zur Fragestellung an die Gruppe nicht aufgelöst werden konnte. Der Rest der ersten Phase beinhaltete zum einen, das modellierte Schema zu implementieren und die Datenbank mit sinnvollen Daten zu befüllen. Zum anderen wurden vorgegebene parametrisierte Anfragen über eine Webseite mittels JSP an die Datenbank gestellt, und deren Ergebnisse in HTML Form dargestellt. Zu diesem Zweck wurden JSP und JDBC Grundlagen in der dritten Übungsstunde vorgestellt. Um die Rechte der Studenten auf der Datenbank auf Tabellen erstellen und ändern beschränken zu können, wurden ihnen von uns angelegte Datenbanken zugewiesen. Die Diskussionsstunde wurde dazu genutzt, eine Datenbank pro Domäne vorzustellen, und die Lösung zu diskutieren / mit anderen zu vergleichen. Die Abgabe erfolgte generell zum vereinbarten Termin.

Phase 2: Firmengründung, Webshop und Managementanfragen

Phase 2 begann mit der Firmengründung und der Erläuterung der Funktionalität des Webshops und der Management Anfragen. Die Schwierigkeit beider Aufgaben war, dass alle drei Schemata (Versand, Inventar und Rechnung) verwendet werden mussten, was die Studenten vor typische Integrationsprobleme stellte. Insbesondere haben wir bei der Formulierung der Managementanfragen darauf geachtet, dass mindestens zwei Datenbanken benötigt werden, so dass es nicht genügt, eine einfache SQL Anfrage zu stellen. Jede Firma bestand aus drei Studenten, von denen jeder eine unterschiedliche Domäne in Phase 1 bearbeitet hat. Da 18 Studenten an der Übung teilnahmen, gab es sechs Webshops, unter denen wir den Besten prämierten. Der technische Teil von Phase 2 beschäftigte sich mit Java Beans und JSP. Während der Diskussionsstunde vor Abgabe wurden Probleme diskutiert, auf die die Studenten bei der Integration stießen, und ihre Lösungsansätze. Ungewohnt für die Studenten war es, innerhalb der Gruppen eng zusammenzuarbeiten. Die sonst übliche Aufteilung von Teilaufgaben unter den Gruppenmitgliedern war kaum möglich. Entsprechende Kritik („zu kommunikationsintensiv!“, „zu schwierig einen gemeinsamen Termin zu finden!“) nahmen wir natürlich gelassen hin, bestätigte sie doch den tieferen Sinn der Aufgaben.

Phase 3: Web Services und Handel

Die meisten technischen Probleme gab es in der dritten Phase, da sich das Zusammenbringen von Java, Tomcat und WebService-Kit als schwierig herausstellte. Die Aufgabe wurde in zwei Teilen bearbeitet. Im ersten Teil sollten die Studenten die Rumpfwebservice bereitstellen, deren Signatur vorgegeben war, um die Kommunikation zwischen den Webservices für Testzwecke zu gewährleisten. Darauf folgte die Implementierung der vollen Funktionalität im zweiten Teil. Eine Übungsstunde wurde dafür verwendet, die Aufgabenstellung und Konzepte von Web Services vorzustellen. In einer weiteren Übungsstunde befassten wir uns mit der Implementierung und Nutzung von Web Services. Die Verwendung des JWSDK erleichterte

die Umsetzung erheblich, da insbesondere die Kommunikation, also die Funktionsaufrufe nicht selbst programmiert werden mussten. Die Studenten konnten sich so auf die Funktionalität, die Einkaufsstrategie und das Verhalten beim Einkauf bei anderen Services konzentrieren. Es stellte sich beim Testen heraus, dass die Einkaufsstrategien der Webshops nur wenig Interaktivität und wenig Zwischenhandel auslösten, da jeder mehr verdienen wollte, als der andere. Leider konnten nur wenige Tests durchgeführt werden, da das Aufrufen der Web Services wegen einiger ineffizienter Implementierungen viel Zeit in Anspruch nahm.

Evaluation

Die Arbeit der Studenten wurde mit bestanden oder nicht bestanden bewertet. Da die jeweiligen Phasen der Übung aufeinander aufbauen, war Voraussetzung für das Bestehen der Übung die „hinreichend“ erfolgreiche Bearbeitung aller Phasen. Diese Art der Evaluation stellte sich als zu subjektiv und für Studenten schwer nachvollziehbar heraus, so dass wir in kommenden Veranstaltungen ein Punktesystem einführen werden.

Zur Evaluation der Übung haben wir einen Feedbackbogen entworfen, bei dem wir den Aufbau der Übung, Zeitaufwand für die einzelnen Phasen, die Teamarbeit und den Lerneffekt in den benutzten Technologien bewerten ließen. Positive und negative Kritik, sowie Verbesserungsvorschläge konnten ebenfalls eingebracht werden.

Insgesamt war die Evaluation der Übung positiv. Generell war ein Lerneffekt vorhanden, und die Projektarbeit hat Spaß gemacht. Kritikpunkte waren, dass es zuviel Programmierung ohne viel Nachdenken gab (z. B. Daten in Datenbank einfügen und HTML schreiben) und die Übung zu wenig mit der in der Vorlesung behandelten Themen zu tun hatte. Viele Studenten wünschten sich zudem Zugriff auf DB2 und den Webserver außerhalb der Universität um auch von zu Hause aus die Übungen zu bearbeiten.

Mit Hilfe der Evaluation haben wir die Übung für zukünftige Semester modifiziert. Die Programmierarbeit wird den Studenten zum Teil abgenommen, indem wir die Datenbanken komplett vorgeben und implementieren, und ein Webshop Template mit HTML Seiten zur Verfügung stellen. Um dem Vorlesungsstoff noch näher zu kommen, werden einige Verfahren und Algorithmen am Webshopszenario implementiert.

Material

Wir stellen das Übungsmaterial (Aufgabenblätter und Folien) unter <http://www.informatik.hu-berlin.de/mac/> zur Verfügung, weisen jedoch darauf hin, dass wir für das kommende Wintersemester 2004/2005 die Aufgaben überarbeiten um so die gesamte Übung „runder“ zu gestalten und die Integrationsaspekte stärker zu betonen.

Literatur

[Halevy 2003] Alon Y. Halevy: Learning about data integration challenges from day one. SIGMOD Record 32(3): 16-17 (2003)