



INSTITUT FÜR INFORMATIK
SoSe 2009
Exposé zur Studienarbeit

Webinterface zur Datenintegration und Visualisierung im Projekt Dhaka INNOVATE

im Rahmen des DFG Schwerpunktprogramm 1233
„Megacities-Megachallenge – informal dynamics of global change“

Betreuer: Prof. Dr. Ulf Leser und Prof. Dr. Tobia Lakes
Autor: Eric Redlin
Datum: 29.09.2009

Hintergrund

Dhaka, die Hauptstadt von Bangladesch, gilt als eine der am schnellsten wachsenden Megastädte¹ der Welt und als eine der ärmsten zugleich. Durch das schnelle Wachstum von etwa 1,5 Mio. Einwohner im Jahr 1975 auf prognostizierte 16,8 Mio. Einwohner im Jahr 2015 und eine unzureichend organisierte bzw. fehlerhafte Städteplanung, sieht sich die Region Dhaka schwerwiegenden sozialen und ökologischen Problemen ausgesetzt.

So verursachen das Schmelzwasser aus dem Himalaya und die heftigen monsunalen Niederschläge jährliche Überschwemmungen der Stadt und fügen der aufgebauten Infrastruktur erheblichen Schaden zu. Als Folge bleibt das Wasser tage- bis wochenlang in den Straßen und führt zu erheblichen Belastungen der Gesundheit.

¹nach Definition der UN eine Stadt mit mehr als 10Mio. Einwohnern.

Auch die hohe Luftverschmutzung durch den Straßenverkehr, Industrieanlagen und die Verbrennung von Feuerholz bzw. alten Autoreifen haben Auswirkungen auf die Gesundheit der Menschen, aber auch auf das lokale und globale Klima.

Ein weiteres großes Problem stellt die verheerende Wohnsituation und die schlechten Lebensbedingungen für den Großteil der Bevölkerung dar. Die Zahl der in Slums lebenden Bevölkerung stieg bis 2005 auf 3,4 Mio. und wird weiter steigen. Diese verteilen sich auf über 4900 Slums und besetzte Gebiete und erreichen somit Ausmaße von Berlin. Die hohe Bevölkerungsdichte, die unzureichend errichteten Häuser und das Fehlen von sanitären Einrichtungen, Trinkwasseranschlüssen und Abwasseranlagen führen zu einer starken Belastung von Grundwasser, Boden und vor allem der Gesundheit der Bewohner in den informellen Siedlungen [1].

Unter anderem mit diesen Themen befasst sich das Schwerpunktprogramm 1233 *Megacities - Megachallenge: Informal dynamics of global change* der Deutschen Forschungsgemeinschaft (DFG). Das Teilprojekt *Dhaka - INNOVATE: Informal settlements, economic and environmental change, and public health - strategies to improve the quality of life in Dhaka* erarbeitet verschiedene Strategien um die Lebensqualität in Dhaka zu erhöhen [2]. Dabei werden die informellen Siedlungen mit den dort bestehenden Problemen und die Konsequenzen des unkontrollierten, schnellen Wachstums fokussiert. So werden in Bereichen des Klimas, der Wirtschaft und der Gesundheit eine Vielzahl von Daten erhoben. Daraus ergibt sich das Ziel die komplexen Zusammenhänge und Abhängigkeiten zwischen Wirtschaft, Gesundheit und Klima zu erfassen, diese auf Basis von räumlichen Daten zu analysieren und daraus Strategien für eine Verbesserung der Situationen in diesen Gebieten zu erarbeiten. Durch die Unterteilung in verschiedene Forschungsprojekte liegen die Daten in unterschiedlichen Formaten vor und können somit nur schwer zusammenhängend betrachtet werden. Darum ist eine zentrale Integration zur Erleichterung der Auswertung erforderlich.

Eine vergleichbare Aufbereitung dieser Problematik, allerdings in größerem Umfang, erfolgte im Rahmen einer Diplomarbeit. Als Ergebnis entstand die NaDiNe-Plattform, welche als Vernetzungsplattform der Helmholtz-Gemeinschaft Deutscher Forschungszentren im Bereich Naturkatastrophen fungiert. Der Zweck dieser Plattform besteht vor allem in der zentralen Bündelung und Bereitstellung von Daten und Informationen für einen großen Nutzerkreis.

Zielstellung

Ziel dieser Studienarbeit ist die Erstellung eines Webinterfaces, dass den Mitarbeitern des Dhaka-Projekts die zentrale Verwaltung von Daten ermöglicht,

die z.B. bei durchgeführten Umfragen im Gebiet Dhaka entstanden sind. Das auf objektorientiertem PHP basierende Portal bietet den am Projekt beteiligten Wissenschaftlern Funktionen wie

- dem Upload
- dem Download
- der Aktualisierung
- der Visualisierung

ihrer Daten. Auf Grundlage der vorhandenen Nutzer in der verwendeten PostgreSQL-Datenbank können diese Funktionen nach erfolgreichem Login genutzt werden. Für nicht registrierte Nutzer bietet das Portal die Möglichkeit der Visualisierung ausgewählter Daten, welche durch die Google Maps API umgesetzt wird.

Daten und Methodik

Zunächst muss das Web Interface den Wissenschaftlern einen Login ermöglichen. Wie bereits im Abschnitt Zielstellung beschrieben, erfolgt die Authentifizierung auf Basis der angelegten Nutzer aus der PostgreSQL-Datenbank. Nach erfolgreichem Login erschließen sich dem Benutzer verschiedene Möglichkeiten um seine Daten zu verwalten. Die Datengrundlage dieser Studienarbeit bilden dabei zwei CSV-Dateien und eine XML-Datei, welche beispielhaft für die Durchführung des Uploads ausgewählt wurden.

Die erste CSV-Datei² besteht aus 630 Spalten und 1008 Zeilen. Dabei handelt es sich um Daten aus durchgeführten Umfragen zur Gesundheitssituation in Dhaka ergänzt mit zusätzlichen Informationen, z.B. einer ID zur geographischen Positionierung. Diese ist in jedem Datensatz enthalten und kann durch die Auflösung dieser ID zu geographischen Koordinaten zur Visualisierung genutzt werden. Durch die Vielzahl an Spalten muss zusätzlich eine CSV-Datei mit Informationen über die Daten wie dem Kürzel und der Bezeichnung jeder Spalte hochgeladen werden. Anhand dieser Informationen wird in der Datenbank eine Tabelle angelegt, welche anschließend durch den Import der Daten aus der ersten CSV-Datei befüllt wird. Beim Upload wird definiert, durch welches Zeichen die Spalten voneinander getrennt sind und wie der Name der neuen Tabelle lauten soll. Außerdem wird festgelegt, wer auf die Daten lesend oder schreibend zugreifen darf und ob diese visualisiert werden können.

Der Aspekt der Integration von Informationen kommt dadurch zum Tragen, dass mehrere Nutzer ihre Daten hoch laden und diese anschließend anhand der Koordinaten miteinander in Verbindung gebracht werden können.

²Abkürzung für *Comma-Separated Values* bzw. *Character Separated Values*.

Des Weiteren wird auf Basis einer XML-Datei der Import von Metainformationen ermöglicht. Diese werden in ArcCatalog (bzw. ArcGIS, ArcView) verwendet um vorhandene Rasterdaten, die in BIL-Dateien³ vorliegen, zu beschreiben. Diese BIL-Dateien speichern Werte für einen Pixel. Sie sind also selbst keine Bilddateien, sondern organisieren Informationen über Bilddateien. Zu einem Pixel kann es ein oder mehrere Werte, so genannte Bänder, geben. In BIL-Dateien werden die Pixelinformationen für jedes Band und für jede Zeile eines Bildes gespeichert.⁴ In den XML-Dateien sind vor allem die Informationen der Tags relevant, deren Wert mit REQUIRED beginnt.

Durch die zentrale Datenhaltung wird für das Teilprojekt *Dhaka - INNOVATE* und den verschiedenen Themen der vier Unterprojekten gewährleistet, dass bei Bedarf jederzeit auf die aktuellen Daten zugegriffen werden kann. Dazu gehört neben dem bereits beschriebenen Upload von Daten auch das Löschen und Aktualisieren von Daten. Jeder kann also seine hoch geladenen Daten löschen bzw. aktualisieren. Letzteres erfolgt lediglich durch Löschen und erneutes Hochladen der geänderten Daten, da das automatische Vergleichen struktureller Veränderungen der Daten (wie z.B. das Hinzunehmen einer Spalte) einen unnötig höheren Implementationsaufwand bedeuten würden, der so auf recht einfache Weise vermieden werden kann.

Eine weitere Funktionalität stellt der Download dar. Je nachdem auf welche Daten der angemeldete Benutzer zugreifen darf, wird automatisch eine CSV-Datei mit den Daten und eine CSV-Datei mit der Struktur dieser erstellt, so dass die zwei neuen Dateien auch wieder zum Upload verwendet werden können.

Wie bereits erläutert enthalten die Daten der CSV-Datei eine ID zur geographischen Positionierung. Dadurch wird es zum Einen ermöglicht räumliche Verbindungen zwischen den Themenbereichen herzustellen, also für ein Haus oder eine Straße in Dhaka Informationen zur Gesundheit und Wirtschaft zu vereinen, zum Anderen können diese aber auch durch ArcGIS bzw. die Google Maps API visualisiert werden. Letzteres soll ebenfalls im Informationssystem ermöglicht werden und auch nicht angemeldeten Besuchern des Webinterface zur Verfügung stehen.

Dadurch entstehen für das Informationssystem zwei Anwenderkreise. Die erste Gruppe der *Wissenschaftler* kann alle Funktionen des Informationssystems nutzen. Dazu gehören das Hochladen, Aktualisieren, Herunterladen und Visualisieren von Daten. Die zweite Gruppe der *Laien* darf ausgewählte Daten visualisiert bekommen.

Nach der Fertigstellung des Portals erfolgt die Dokumentation und Diskussion der Ergebnisse.

³Abkürzung für *Band interleaved by line*.

⁴Für weitere Informationen zu BIL-Dateien siehe [3].

Literatur

- [1] Katrin Burkart, Oliver Gruebner, M.M.H. Khan, Ronny Staffeld:
Mega- city Dhaka. Geographische Rundschau International Edition
Vol. 4, 01/2008.

- [2] Flyer vom Projekt. *Dhaka - INNOVATE: Informal settlements, economic and environmental change, and public health - strategies to improve the quality of life in Dhaka*.

- [3] ESRI (Hg.): *ArcGIS Desktop 9.3 Help*. ESRI, 2009.
URL: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=BIL,_BIP,_and_BSQ_raster_files
(Abgerufen: 28.10.2009, 21:15Uhr)