HTWK Leipzig
Fachbereich IMN
Sommersemester 2013

Beleg im Fach Informationssysteme

Konzeption

Kurt Junghanns, B.Sc.
Philipp-Rosenthal-Straße 32
04103 Leipzig
kurt.junghanns@stud.htwk-leipzig.de

Marcel Kirbst, B.Sc.
Sieglitz 39
06618 Molau
marcel.kirbst@stud.htwk-leipzig.de

7. Mai 2013

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einleitung

Ziel dieses Belegs ist ein Data Warehouse zu erstellen und die Phasen des Data Warehousing zu durchlaufen. Als Datenquellen dienen dabei das soziale Netzwerk Facebook, Geodaten von OpenStreetMap sowie Wetterdaten.
Ziel ist die Gewinnung neuer Aussagen anhand der Korrelation dieser Daten.

2 Beschreibung der Datenquellen

2.1 Facebook-API

Facebook bietet für Entwickler APIs zur Abfrage öffentlicher Profildaten an. Mit einem eindeutigen Schlüssel ist es möglich diese Daten per HTTP abzufragen.
Die Daten werden im JSON-Format ausgeliefert.

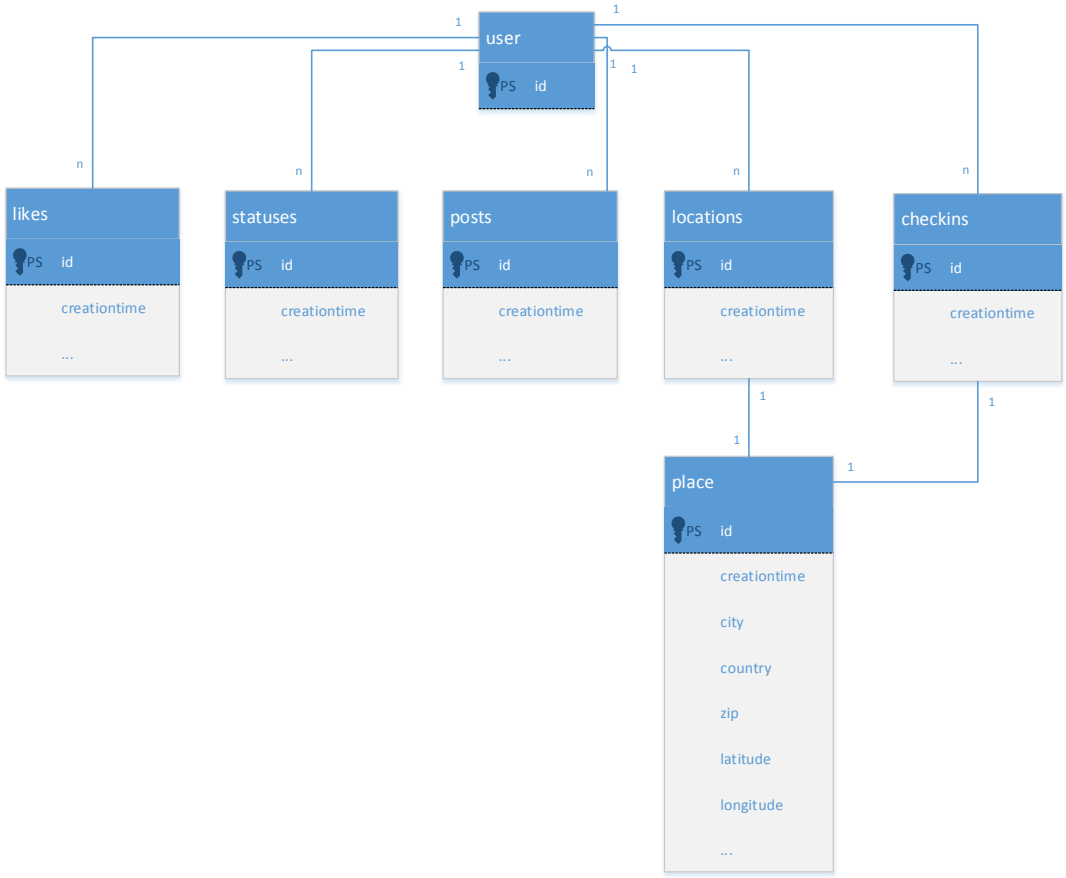

Abbildung 1: Facebook öffentliche Nutzerdaten

Abbildung 1 deutet den Charakter der zu erhaltenten Daten an.
Die aufgeführten Attribute und Entitäten sind für die Auswertung essentiell. Die Entitäten und deren Attribute sind jedoch optional.
Aus diesem Grund sollen im ETL-Prozess nur Datensätze mit vorhandenen Entitäten und Attributen ausgewertet werden.

Zur Abfrage der Nutzer muss eine Zeichenkette zur Filterung mitgegeben werden. Es findet eine Filterung des Nutzernamens statt.
Aus diesem Grund werden die weltweit häufigsten Namen als weitere Datenquelle

herangezogen und damit die Nutzerdaten erhoben.

Jede Anfrage liefert nur einen Ausschnitt der geforderten Daten und Verweise auf die restlichen Daten.

2.2 OpenStreetMap

OpenStreetMap bietet neben umfangreichen Kartenmaterial auch den Service, entsprechend einer übergebenen Auswahl die darin befindeten Orte mit diversen Daten mit HTTP abzurufen.
Mit diesem Service sollen fehlende Informationen von Nutzerdaten von Facebook in Bezug auf Positionen und Orte geliefert werden.

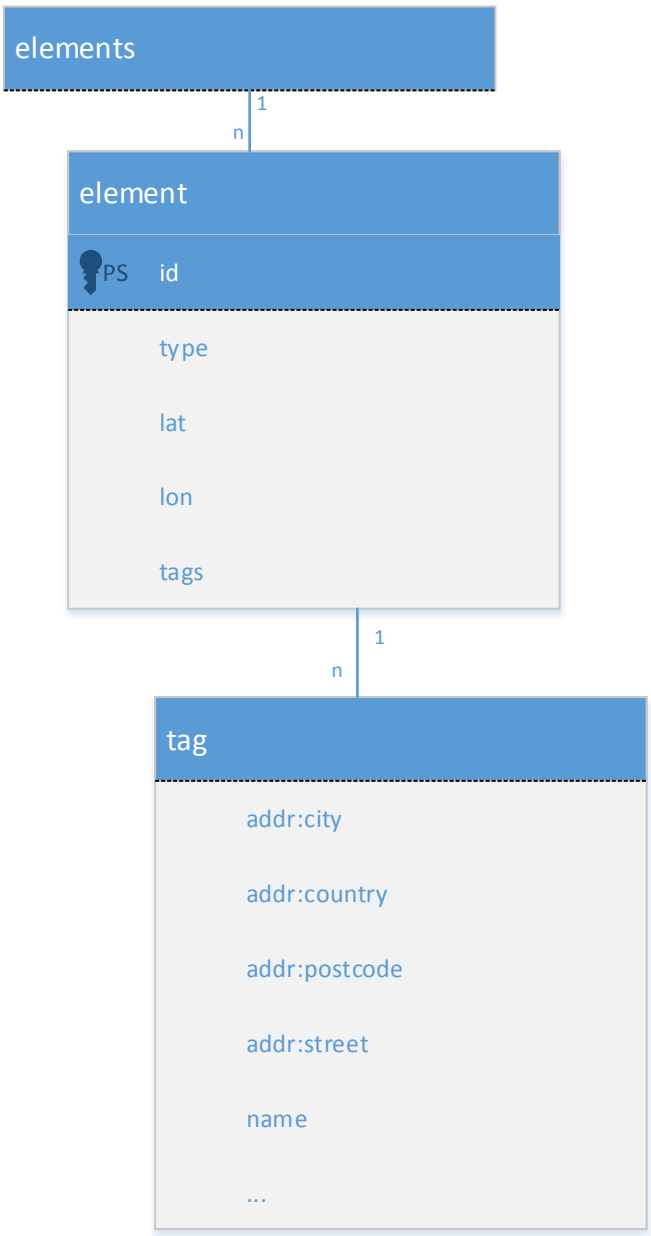

Abbildung 2: OpenStreetMap Orte

2.3 Wetter-API

Der Wetterdienst Weather Underground bietet angemeldeten Entwicklern per API Zugriff auf aktuelle und vergangene Wetter- und Klimadaten der ganzen Welt an. Auch hierbei werden die Daten mit HTTP abgefragt und im JSON-Format ausgeliefert.

Der Zugriff ist in der kostenlosen Version für Entwickler auf 10 Anfragen pro Minute

und 500 Anfragen pro Tag begrenzt.

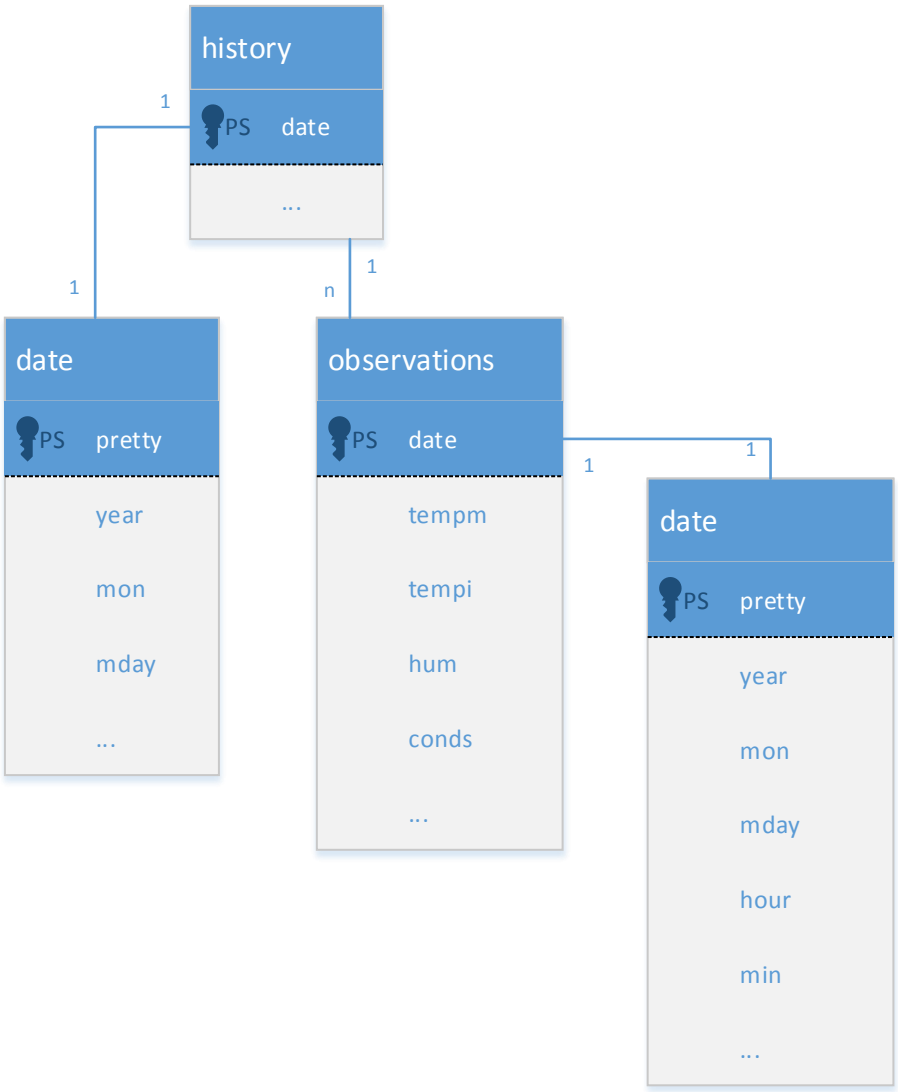

Abbildung 3: historische Wetterdaten von Weather Underground

Das obrige ERM stellt lediglich einen Teil der Granularität und Masse der Daten dar. Die bezeichneten Entitäten und Attribute sind für den ETL Prozess relevant.

2.4 Wetterstationen

Die Wetterstationen werden mit

Abbildung 4: Wetterstationen

3 Architektur des aufzubauenden Data Warehouse

Die Architektur wurde entsprechend der Vorlesung Informationssysteme entworfen.

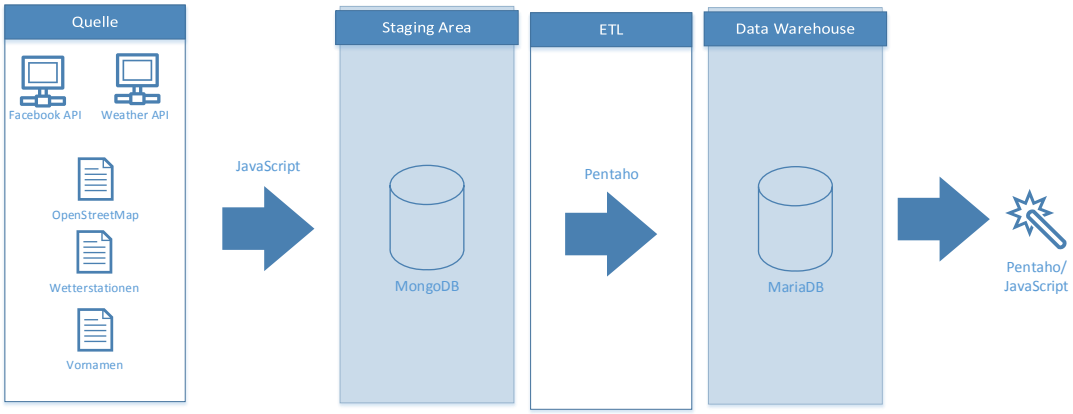

Abbildung 5: Architektur des Data Warehouse

Der Zugriff auf große Mengen von Daten mit Hilfe der APIs ist auf Grund des zerstückelten Erhaltes der Daten nur mit einer entsprechenden Logik möglich. Dazu ist es notwendig diese Fragmente zu sammeln.
Da die Datenquellen JSON als Format liefern und die Umwandlung in ein solches wenig aufwendig ist, wurde als Stating Area die schemafreie dokumentbasierte Datenbank MongoDB verwendet. Außerdem wird damit der heterogene Charakter der Quelldaten erhalten und betont.

Um den ETL-Prozess übersichtlich zu halten und dessen Definition zu erleichtern, wird das Framework Pentaho verwendet. Pentaho ist eine OpenSource Java Business-

Intelligence-Software, welche für die Bereiche ETL, Reporting, OLAP/Analysis und Data-mining geeignet ist.

Ein Metadata-Repository ist nicht vorgesehen. Die Definition der Tabellen der MariaDB dienen der Beschreibung der Daten.
Eine relationale Abbildung des Data Warehouse wurde gewählt, da die hinter dem Data Warehouse befindliche Geschäftslogik zumeist von relationalen Daten ausgeht.

Einfacher halber wird Pentaho zur Auswertung weiter verwendet, wodurch mit geringem Auswand entsprechende Visualisierung erzeugt werden können. Sollte sich Pentaho nicht eignen, wird zur Auswertung auf JavaScript zurückgegriffen.

4 Datenbankschemata

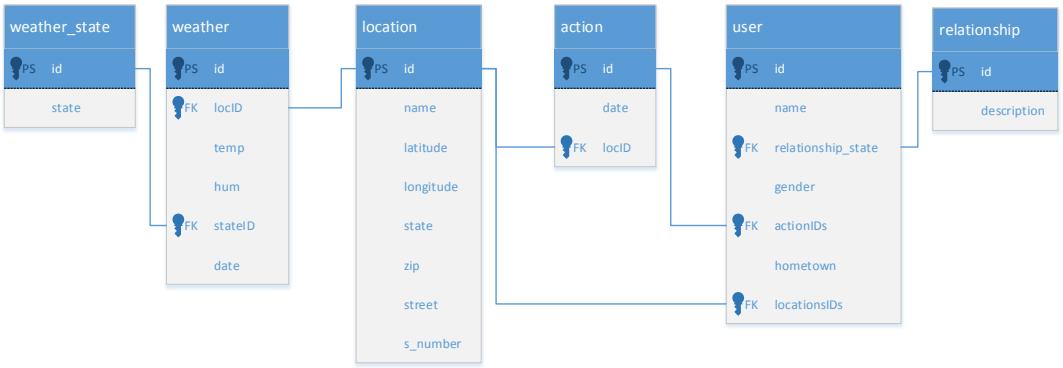

Abbildung 6: ERM des Data Warehouse

5 Beschreibung der anvisierten Analysen

6 Literatur- und Quellenverzeichnis

Literaturverzeichnis

[1] Kiumars Farkisch: *Data- Warehouse-Systeme kompakt*, Springer Verlag, 2011, ISBN: 978-3-642-21532-2

Quellenverzeichnis

[1] http://www.wunderground.com/weather/api/d/docs
Abrufbar am 07.05.2013

[2] http://developers.facebook.com/docs/reference/api/
Abrufbar am 07.05.2013

[3] http://www.pentaho.de/
Abrufbar am 25.04.2013

[4] http://wiki.openstreetmap.org/wiki/Overpass_API/Language_Guide
Abrufbar am 07.05.2013