

12 wöchiges Praktikum beim Deutschen Forschungszentrum für Künstliche Intelligenz



Forschungsgruppe Sprachtechnologien (DFKI-LT)
Projekt "Smart Data for Mobility" (SD4M)

DFKI Projektbüro Berlin
Alt-Moabit 91c
10559 Berlin

Praxisbericht

Tom Oberhauser

Beuth Hochschule für Technik Berlin
Studiengang Bachelor Medieninformatik
Matrikelnummer 798158
Trierer Straße 11, 13088 Berlin
E-Mail: tom@devfoo.de

Betreuer

Prof. Christoph Knabe

Fachbereich VI - Informatik und Medien
Beuth Hochschule für Technik Berlin

Dr. Philippe Thomas

Forschungsgruppe Sprachtechnologie (DFKI-LT)
Deutsches Forschungszentrum für Künstliche Intelligenz

August 26, 2015

Inhaltsverzeichnis

1	Einleitung	1
1.1	Vorstellung des Praktikumsbetriebes	1
1.2	Weg zur Praktikumsstelle	1
2	Tätigkeitsbereiche und Aufgaben	3
2.1	Überblick	3
2.1.1	Das Projekt SD4M	3
2.2	Vorbereitung der Arbeitsumgebung	4
2.3	Einarbeitung	4
2.4	Aufgaben	4
2.4.1	Extraktion einer Straßenliste	4
2.4.2	Verknüpfung von Daten der Deutschen Bahn mit Daten aus OpenStreetMap	4
2.4.3	Datenaufwertung??	4
3	Fazit	5
3.1	Praktikum und Studium	5
3.2	Bewertung des Praktikums	5
4	Anlagen	6

Einleitung

1.1 Vorstellung des Praktikumsbetriebes

Das Deutsche Forschungszentrum für Künstliche Intelligenz GmbH, im folgenden DFKI genannt, wurde 1988 gegründet. Es unterhält Standorte in Kaiserslautern, Saarbrücken, Bremen und ein Projektbüro in Berlin. Mit seinen 478 Mitarbeitern sowie 337 studentischen Mitarbeitern erforscht und entwickelt das DFKI innovative Softwaretechnologien auf der Basis von Methoden der Künstlichen Intelligenz. Die notwendigen Gelder werden durch Ausschreibungen öffentlicher Fördermittelgeber wie der Europäischen Union, dem Bundesministerium für Bildung und Forschung (BMBF), dem Bundesministerium für Wirtschaft und Technologie (BMWi), den Bundesländern und der Deutschen Forschungsgemeinschaft (DFG) sowie durch Entwicklungsaufträge aus der Industrie akquiriert.¹

Ich absolvierte mein Praktikum innerhalb der *Forschungsgruppe Sprachtechnologie*, einer von 15 Forschungsgruppen² des DFKI, im Projektbüro Berlin. Die Gruppe wird geleitet durch Prof. Dr. Hans Uszkoreit.³

Meine Aufgabengebiete konzentrierten sich um das Projekt *”SD4M - Smart Data for Mobility”*. Das DFKI ist hier Teil eines Konsortiums aus 5 Partnern unter der Konsortialführung der *DB Systel GmbH*.⁴ Das Projekt *SD4M* wird in Abschnitt 2.1.1 auf Seite 3 näher erläutert.

1.2 Weg zur Praktikumsstelle

Herr Prof. Dr. habil. Alexander Löser aus dem Fachbereich VI der Beuth Hochschule für Technik Berlin machte mich auf den Praktikumsplatz aufmerksam. Durch seine Mitarbeit in Projekten im DFKI Projektbüro Berlin hatte er wargenommen, dass Bedarf und Interesse an Praktikanten und studentischen Mitarbeitern besteht und mich benachrichtigt. Ich habe mich daraufhin auf der Website des DFKI über die aktuellen Projekte informiert. Da ich mich sehr für Datenintegration interessiere und

¹<http://www.dfki.de/web/ueber>

²<http://www.dfki.de/web/ueber/orgaeinheiten>

³<http://www.dfki.de/lt/>

⁴<http://sd4m.net/konsortium>

ich vor meinem Studium bereits Berufserfahrung auf diesem Gebiet gesammelt habe, fand ich das Projekt *SD4m - Smart Data for Mobility* sehr interessant. Es umfasst zwei Themenkomplexe: zum einen die Verknüpfung unterschiedlicher Datenquellen und zum anderen Methoden des *Natural Language Processings* bzw. des *Text Minings*. Das Thema Text Mining wurde kurz im Modul Datenbanksysteme im zweiten Semester angeschnitten und es hatte mich auch sehr interessiert. Nach einem persönlichen mit Herr Uszkoreit, in dem ich mein Interesse für das Projekt darlegen konnte, kam es zur Vertragsunterzeichnung.

2.1 Überblick

Ich habe im Rahmen meines Praktikums am Projekt *SD4M - Smart Data for Mobility* mitgearbeitet. Meine konkrete Aufgabe war die Aufbereitung und Integration verschiedener Daten und Datenbanken, damit diese im Projekt Verwendung finden können. Meine Hauptdatenquelle waren die Geodaten des OpenStreetMap¹ Projekts.

2.1.1 Das Projekt SD4M

Das Projekt *Smart Data for Mobility*², im folgenden *SD4M* genannt, ist ein Verbundprojekt eines Konsortiums aus 5 Partnern und wird vom Bundesministerium für Wirtschaft und Energie gefördert. Das Konsortium besteht aus 4 Wirtschaftsunternehmen und dem DFKI als Forschungseinrichtung.

- DB Systel GmbH (Konsortialführung)
- Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
- idalab GmbH
-]init[AG für digitale Kommunikation
- PS-Team Deutschland GmbH Co. KG

Ziel des SD4M Projekts ist eine branchenübergreifende Serviceplattform, welche Daten der unterschiedlichen Mobilitätsanbieter (z.B. der Fahrplan der Deutschen Bahn) sowie öffentliche verfügbare strukturierte und unstrukturierte Daten (z.B. Twitter oder Facebook) miteinander verknüpft. Diese verknüpften Daten sind für Endnutzer, aber auch für Unternehmen oder die öffentliche Verwaltung von Interesse. In Abbildung 1 wird verdeutlicht, wie sich aus unstrukturierten Twitter-Daten Verspätungsinformationen für konkrete Verkehrsmittel extrahieren lassen. Diese können dann Endnutzern oder den Mobilitätsanbietern zur Verfügung gestellt werden.

¹<http://www.openstreetmap.org/>

²<http://sd4m.net/>

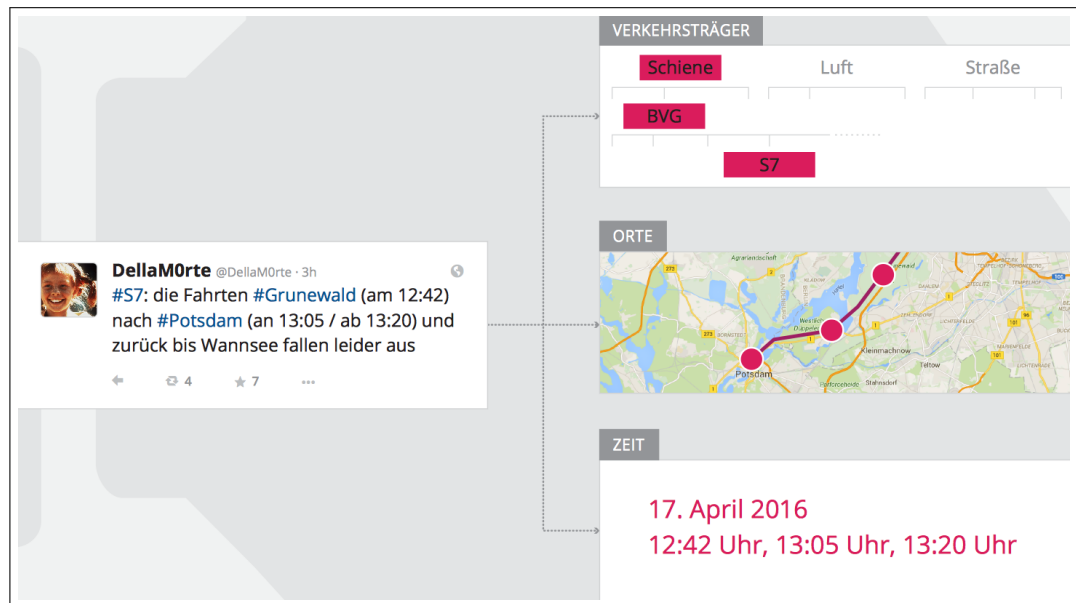


Abb. 1: Verknüpfung eines Tweets mit Fahrplandaten[@Ing16]

2.2 Vorbereitung der Arbeitsumgebung

Eigener Rechner, lokale Umgebung, DB ... warum?

2.3 Einarbeitung

2.4 Aufgaben

2.4.1 Extraktion einer Straßenliste

2.4.2 Verknüpfung von Daten der Deutschen Bahn mit Daten aus OpenStreetMap

2.4.3 Datenaufwertung??

Fazit

3.1 Praktikum und Studium

3.2 Bewertung des Praktikums

Anlagen

4

Online-Quellen

- [@Ing16] Ingo Schwarzer. *Smart Data For Mobility (SD4M) – Projekt-Praesentation*. 2016.
URL: <http://www.sd4m.net/sites/default/files/publications/%20SD4M-Pr%C3%A4sentation.pdf> (besucht am 10. Mai 2016) (zitiert auf Seite 4).

Declaration

You can put your declaration here, to declare that you have completed your work solely and only with the help of the references you mentioned.

Berlin, August 26, 2015

Tom Oberhauser