Spätherbstanlass des Datenschutzbeauftragten des Kantons Basel-Stadt
4. November 2014
Hotel Victoria, Basel

Herzlich willkommen!
Beginn 17:00 Uhr

«Die Hermeneutik von Ortsdaten» – oder: Was verraten meine Daten über mich?

Datenschutzbeauftragter des Kantons Basel-Stadt

Einstiegsfragen

Herzlich willkommen zum Spätherbstanlass des Datenschutzbeauftragten des Kantons Basel-Stadt.

Wir alle hinterlassen Datenspuren.

z.B. auf Ihrem iPhone – schauen Sie mal nach unter:
Einstellungen | Datenschutz | Ortungsdienste |
(dann ganz zuunterst:) Systemdienste | Häufige Orte

Aber wir haben ja nichts zu verbergen, oder?

Vielleicht lohnt es sich aber doch, genauer hinzuschauen: Was verraten diese Spuren über uns?

Was können wir dagegen unternehmen, damit wir nicht zum offenen Buch werden?

© 2014 DSB BS 2

«Die Hermeneutik von Ortsdaten» – oder: Was verraten meine Daten über mich?

Prof. Dr. Thomas Hofmann

Professor für Data Analytics,
Departement Informatik,
Eidgenössische Technische
Hochschule (ETH) Zürich
Zürich / Schweiz

[→]

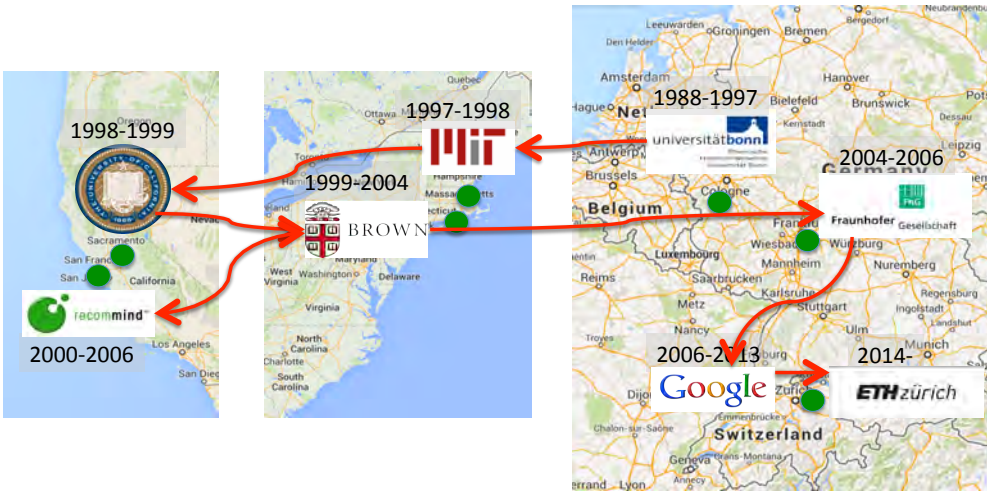
© 2014 DSB BS 3

«Die Hermeneutik von Ortsdaten» – oder: Was verraten meine Daten über mich?

Thomas Hofmann
Institute for Machine Learning
Department of Computer Science, ETH Zürich
Thomas.Hofmann@inf.ethz.ch

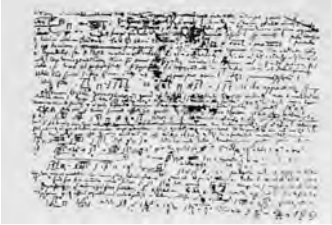
Herbstanlass des Datenschutzbeauftragten des Kantons Basel-Stadt
4. November 2014

Ortsdaten zur Person :-)



Hermeneutik Digitaler Daten

Hermeneutik (SCHLEIERMACHER, 1838):
die Kunst, die Rede eines anderen, vornehmlich die schriftliche, richtig zu verstehen



Hermeneutik digitaler Daten (2014):
die Kunst, die Daten eines anderen, vornehmlich die digitalen, richtig zu verstehen

EINLEITUNG

Fokus hier & heute: Ortsdaten

- Ortsdaten gesammelt über Smartphones, Autos, Navigationsgeräte, «Wearables» usw.

- GPS: Global Positioning System, satelliten-basiert, 3-5m Genauigkeit
- WPS: Ortung über WIFI-Netzwerke (SSID, MAC), interessant in Gebäuden und Innenstädten
- Triangulierung mittels Funktürmen: Genauigkeit abhängig von Layout, erfordert Pings/Anrufe

Ortsdaten via Smartphones

4.55 Milliarden Mobilgeräte
1.75 Milliarden Smartphones
(Quelle: eMarketer, Jan 2014)

Q4.2013:
219M Android/GOOG = 78%
50M iPhone/AAPL = 18%
8M Windows/MSFT = 3%

Wieviele lat/long Rohdaten werden gespeichert?
≈ (Tausend) pro Tag und Gerät
≈ (Milliarden) Geräte weltweit
≈ (Billionen) Datenpunkte pro Tag!

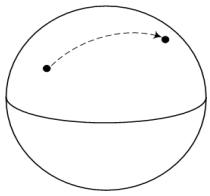
Mehr Ortsdaten ...

STUFEN-MODELL DER HERMENEUTIK

Rohdaten

Eine Folge von Wegpunkten, kodiert als sphärische Koordinaten (geographische Breite & Länge)

Latitude	Longitude
+47.661222	+9.176919
+47.661150	+9.176900
+47.661171	+9.177096
+47.661183	+9.177116
+47.661225	+9.177177
+47.661201	+9.177336
+47.661201	+9.177402
+47.661267	+9.177446
+47.661376	+9.177352
+47.661365	+9.177256
+47.661153	+9.176823
+47.661149	+9.176795
+47.661108	+9.176713
+47.661038	+9.176459
+47.661004	+9.175972
+47.660999	+9.175868
+47.661030	+9.175663



Rohdaten sind semantisch arm: diskretisierter Pfad auf einer Kugel

Was sagen meine Daten über mich? (Stufe 1)

«Ich habe mich entlang eines diskretisiert gemessenen Weges auf der Erdkugel von A nach B bewegt.»

A = +47.6612222, +9.176919
B = ...

Was für einen Aktionsradius habe ich?

Rundwege vs. einfacher Weg.

Ortsdaten + Zeitstempel

Hinzunahme von Zeitstempelinformationen:

wann war ich wo?

Timestamp	Latitude	Longitude
15 May 2013, 09:35:51 am +0200	+47.661222	+9.176919
15 May 2013, 09:37:01 am +0200	+47.661150	+9.176900
15 May 2013, 09:37:11 am +0200	+47.661171	+9.177096
15 May 2013, 09:37:12 am +0200	+47.661183	+9.177116
15 May 2013, 09:37:16 am +0200	+47.661225	+9.177177
15 May 2013, 09:37:22 am +0200	+47.661201	+9.177336
15 May 2013, 09:37:24 am +0200	+47.661201	+9.177402
15 May 2013, 09:37:29 am +0200	+47.661267	+9.177446
15 May 2013, 09:37:36 am +0200	+47.661376	+9.177352
15 May 2013, 09:37:40 am +0200	+47.661365	+9.177256
15 May 2013, 09:37:57 am +0200	+47.661153	+9.176823
15 May 2013, 09:37:59 am +0200	+47.661149	+9.176795
15 May 2013, 09:38:03 am +0200	+47.661108	+9.176713
15 May 2013, 09:38:14 am +0200	+47.661038	+9.176459
15 May 2013, 09:38:29 am +0200	+47.661004	+9.175972
15 May 2013, 09:38:32 am +0200	+47.660999	+9.175868
15 May 2013, 09:38:39 am +0200	+47.661030	+9.175663

Rückschluss über Fortbewegungsmittel

11.61 km/h
16.44 km/h
14.22 km/h
12.99 km/h

?

Was sagen meine Daten über mich? (Stufe 2)

«Ich habe mich mit einer mittleren Geschwindigkeit von 14 km/h von A nach B bewegt. Möglicherweise (gemütlich) mit einem Fahrrad.»

Wie sportlich bin ich?

Wie ausdauernd bin ich?

Habe ich angehalten?

Besitze ich ein Fahrrad?

Habe ich mich verfahren?

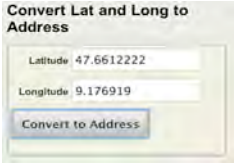
+ Strassenatlas

Lat/long Wegpunkte können auf Kartenpunkte abgebildet werden (z.B. Google Maps)

Heutige digitale Karten: Strassenadresse

= Inverse geo-coding

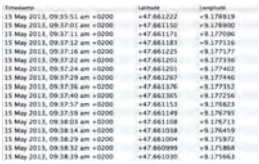
latlong.net



Semantische Potenzierung durch Verknüpfung von Datenquellen

Für sich genommen sind die Rohdaten nicht sehr aussagekräftig – was ist 47.366026, 8.539732 ?

Durch die Verknüpfung mit anderen Datenquellen, etwa einem Digitalatlas, erhalten die Daten eine neue Interpretationsebene. (heute: automatisch, *at scale*)



Personenbezogene Rohdaten

+

Allgemeine Hintergrunddaten

=

Angereicherte Daten, personenbezogen

Was sagen meine Daten über mich? (Stufe 3)

«Ich bin (wahrscheinlich) mit dem Fahrrad in Zürich von der Universitätsstrasse 6 via Central zur Seestrasse 19 gefahren. Losgefahren um 17:32 am 5. Mai 2014, 10‘ 25‘‘ Wegdauer.»

Ich war in Zürich.

Bin ich auf dem Rückweg von der Arbeit?

Respektiere ich Ampeln, Einbahnstrassen, usw.?

Bin ich ortskundig?

Habe ich mich vom Ziel aus weiterbewegt - wann und wohin?

Habe ich auf dem Weg irgendwo angehalten?

+ Identifikatoren

Die Aussagekraft von Ortsdaten hängt wesentlich von der Beobachtungsdauer ab: ein Tag vs. viele Tage, Jahre, ein Leben!

Wesentliche Frage: welche eindeutigen Identifikatoren (=IDs) werden mit den Daten abgespeichert?

Unique Device Identifier = UDID (iPhone)

2b6f0cc904d137be2e1730235f5664094b831186.

Identifizierung durch Login oder App

Identifizierung durch Browser Cookies

Einschub: Identifikatoren

IDs: Keys/Schlüssel nach denen Daten indiziert werden

Wichtige Fragen:

- In welchem Umfang repräsentiert eine ID die Totalität meiner Aktivitäten / Spuren?
- Wie anonym ist eine ID? (Gerät, Browser, Account, *Real Name*, Personendaten)

Das automatische mapping von IDs ist problematisch!

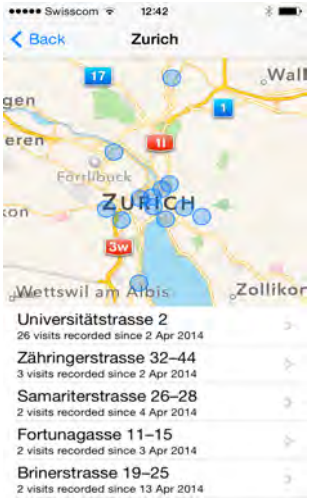
+ Identifikatoren

Identifikation von oft besuchten Punkten. Trivial (i.a.): Wohnadresse, Arbeitsort

Andere relevante Adressen in meinem Leben.

Reisetätigkeit, Lebensrhythmus

Bis hin zum Lesen einer Lebensgeschichte!



Was sagen meine Daten über mich? (Stufe 4)

«Ich bin wie jeden Dienstag mit dem Fahrrad in Zürich von meinem Arbeitsplatz in der Universitätsstrasse 6 via Central zu meinem Wohnort in der Seestrasse 19 gefahren. Am 5. Mai 2014 war ich 10.2 Minuten unterwegs und 1.5 Minuten schneller als im Durchschnitt.»

Dauer der Datensammlung / Beobachtung →

Wo wohne ich?

Wo arbeite ich?

Wann verlasse ich das Haus, wann komme ich zurück?

Wo fahre ich am Wochenende hin?

Bin ich umgezogen?

Welche Adressen suche ich häufig auf?

Gehe ich oft abends aus?

Bin ich ein Frühaufsteher?

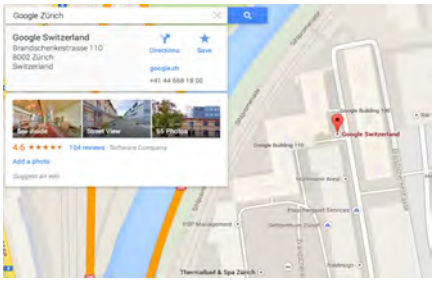
Habe ich die Arbeitsstelle gewechselt?

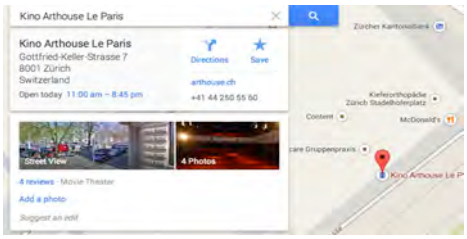
Wie oft bin ich am Arbeitsplatz?

+ «Semantische» Karten

Die Aussagekraft von personenbezogenen Ortsdaten hängt stark von dem Wissen über die Orte ab – was befindet sich an einer Adresse?

Grosses Problem bzgl. der Wahrung der Privatsphäre!





+ «Semantische» Karten

Die Aussagekraft von personenbezogenen Ortsdaten hängt stark von dem Wissen über die Orte ab – was befindet sich an einer Adresse?

Grosses Problem bzgl. der Wahrung der Privatsphäre!

Was sagen meine Daten über mich? (Stufe 5)

«Ich war diese Woche dreimal bei meinem Hausarzt XY.»
«Ich habe diese Woche insgesamt 3 Stunden im Café Henrici verbracht.»
«Ich habe einen 30 minütigen Besuch bei Ochsner Sport in der Bahnhofstrasse eingelegt.»
«Ich habe am Freitag um 19 Uhr den Film Still Life im Arthouse gesehen.»
«Ich war diese Woche zweimal im City Bad für ca. 50 Minuten schwimmen.»
«Ich war am Samstagabend bis 3 Uhr im Club Hive.»
«Ich lege Montag bis Donnerstag gegen 8 Uhr einen Stopp in der Kinderkrippe XXX ein.»
«Ich war zweimal innerhalb von 3 Tagen auf der Polizeistation in Pfäffikon.»

+ Sozialer Graph

Annahmen:

- Datensammlung erfolgt flächendeckend für einen Grossteil der Bevölkerung
- jedem Benutzer werden «wesentliche» Adressen zugeordnet
- Daten werden mit Social Graph verknüpft (z.B. Facebook)

Dann lassen sich Bewegungsdaten auf einer sozialen Ebene interpretieren

+ Soziale Informationen

Brainstorming ohne Anspruch auf «Realität»

- Bluetooth Sensing & Proximity – welches Smartphone wurde in der Nähe von welchem anderen Gerät gesehen?
- Anfragen über eine «Big Data»-Ortdatenbank: welche anderen Nutzer haben ein statistisch signifikante Überlappung in Bezug auf den Aufenthaltsort (Familie, Arbeitskollegen, Freunde, ...)
- Rekonstruktion von sozialen Graphen aus Kommunikationsdaten (Anrufe, E-Mail, ...)

Was sagen meine Daten über mich und Sie(!) ? (Stufe 6)

«Wie oft treffe ich mich mit meinem Sohn, wann und wo?»
«Mit wem gehe ich öfter gemeinsam Wandern oder Fahrradfahren?»
«Wie viel Zeit verbringe ich mit meiner Frau? Wie steht es um meine Ehe?» Gut! ☺
«Wer sind meine Arbeitskollegen?»
«Sehe ich noch Freunde aus meiner Schulzeit? »
«Gibt es Personen, mit denen ich mich in letzter Zeit signifikant häufig treffe?»
«Habe ich Urlaubsbekanntschaften gemacht?»
«Bin ich in Vereinen oder Organisationen aktiv?»

«Lässt sich aus gesammelten Datenspuren ableiten, dass Sie hier sind und meinen Vortrag hören?“

+ Prädiktive Modelle

Die letzte Stufe der Kunst der Dateninterpretation liegt darin den ~~Autor~~ Nutzer besser zu verstehen als er selbst.

Prädiktive Modelle, trainiert über grossen Datenmengen (kollektive Daten), erlauben Vorhersagen über zukünftiges Verhalten Einzelner.

Was sagen meine Daten über mein zukünftiges Ich? (Stufe 7)

«Wie wahrscheinlich ist es, dass ich auf eine Werbung reagiere? ..., dass ich einen Kauf tätige?»

«Was könnten meine nächsten Reiseziele sein?»

«Wann werde ich auf welcher Strecke mit meinem Auto unterwegs sein?»

«Bin ich empfänglich für politische oder religiöse Propaganda?»

«Wie interessant könnte eine andere Person für mich sein?» (Dating, Partnerschaften)

«Bin ich ein vertrauenswürdiger Mensch?»

Hermeneutik digitaler Daten

RÜCKGEWINNUNG DER DATENHOHEIT

Rückgewinnung der Datenhoheit:

1 - Spuren vermeiden

Mehr nutzbare Funktionalität zum ein- und abschalten der Datensammlung

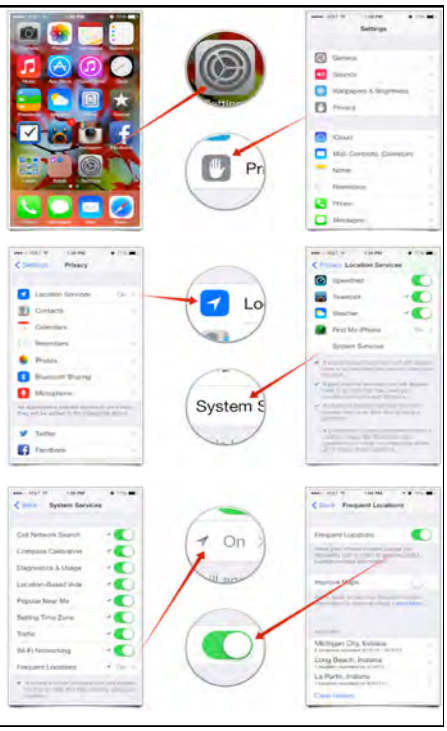
Beispiele:

- Anonymes Surfen in Chrome

Misslungene Beispiele

- Frequent Location Tracking ausschalten (iPhone)

Der lange Weg
zum «Opt-out».



Rückgewinnung der Datenhoheit:

2 – Daten ausdünnen

Reduktion der räumlichen Auflösung erschwert semantische Anreicherung und Verknüpfung mit sozialen Daten. Viele Dienste könnten trotzdem gut funktionieren.

Reduktion der zeitlichen Auflösung erschwert die Wiederidentifizierbarkeit.

Rückgewinnung der Datenhoheit:

3 - Spuren löschen (können und dürfen)

Manuelles Löschen der Rohdaten ist impraktikabel.

Programmatischer Zugang (APIs) ist erforderlich.

Das würde Möglichkeiten eröffnen, effektivere Privacy Werkzeuge zu entwickeln.

- z.B. lösche alle meine Ortsdaten nach 3 Tagen.

Der Gesetzgeber ist gefragt!

Rückgewinnung der Datenhoheit:

4 - Nutzung beschränken

Restriktivere Regelung bezüglich Weitergabe von Daten an Dritte

Grössere Transparenz, welche App/Organisation Daten sammelt, z.B.:

- Welche Daten will ich meinem Service-Provider anvertrauen?
- Welche Daten will ich an ein Verkehrsunternehmen (DB- oder SBB-App) weitergeben? usw.

Rückgewinnung der Datenhoheit:

5 - Verknüpfungen erschweren

Verwendung von applikations-orientierten IDs (z.B.: iPhone Vendor IDs)

Beschränkung der Persistenz von IDs

Anonymisierung

Problem: Oftmals kann Verknüpfung nachträglich wiederhergestellt werden (via Inferenz)

Mobilitätsspuren sind hochgradig individuell:
4 unabhängige Messpunkte reichen in 95% der Fälle aus, Nutzer aus einer Gruppe von 500k Menschen zu identifizieren!
(MONTJOYE ET AL., Nature 2013)

Rückgewinnung der Datenhoheit:

6 - Kryptographisches Rechnen

Homomorphische Verschlüsselung

- Erlaubt die Ausführung bestimmter Berechnungen auf den verschlüsselten Daten
- Erste Fortschritte im Bereich Cloud-Computing
- Aber: Was bewegt Unternehmen dazu, eine entsprechende Infrastruktur aufzubauen?

SCHLUSSBEMERKUNG

Schlussbemerkung

Personenbezogene Ortsdaten erlauben Einblicke in grundlegende Aspekte unserer Persönlichkeit.

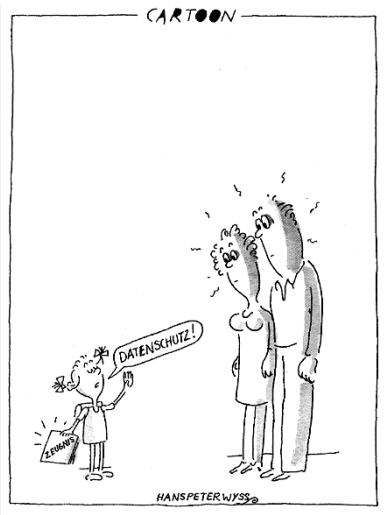
Durch Verknüpfung von Daten kommt es zu einer semantischen Anreicherung (Stufenmodell).

Wir leben im Wilden Westen der Daten. Soziale Praxis und Gesetzgebung können mit der technologischen Entwicklung nicht mithalten. Neue Mechanismen sind geboten.

Besten Dank für Ihr Interesse und Ihre aktive Teilnahme!

Wir sind gerne für Sie da – auch später bei konkreten Fragen.

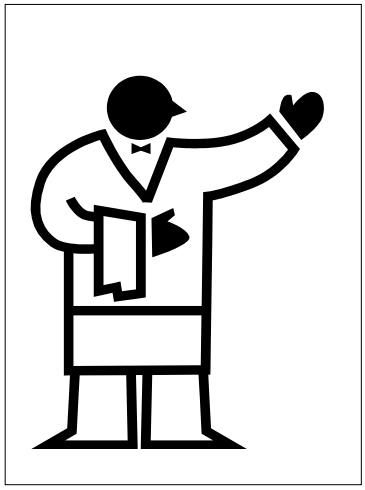
Datenschutzbeauftragter
des Kantons Basel-Stadt
Henric Petri-Strasse 15
4010 Basel
T 061 201 16 40
F 061 201 16 41
datenschutz@dsb.bs.ch
<http://www.dsb.bs.ch>



© 2014 DSB BS 43

Apéro

Die Folien werden morgen zum Download aufgeschaltet: <http://www.dsb.bs.ch/ueber-uns/Veranstaltungen.html>



Gerne laden wir Sie nun noch zum Apéro ein.

© 2014 DSB BS 44