

Philipp Nowak – Projektseminar I4: „Angewandte Informationswissenschaft“

1. Welche Themen im Bereich der Informationswissenschaft und Sprachtechnologie interessieren mich? (z.B. Informetrie, Social Media, Gamification, Information Retrieval, Computerlinguistik, ...)
Welche(r) Aspekt(e) genau?

→ Die Bearbeitung der Projekte des Projektseminars I3 „Informetrie“ haben mir letztes Semester viel Spaß gemacht, weshalb ich mit dem Bereich „Informetrie“ antworten würde.

2. Welche Programmiersprachen und Tools liegen mir bzw. welche würde ich gerne ausprobieren? (z.B. Python, Java, Prolog, Flask, D3, Django, ...)

→ Auf jeden Fall Python.

3. Wie könnte ich mich in Form eines Projektes mit dem gewählten Thema auseinandersetzen? (z.B. Datenanalyse, Visualisierung, App, Text Mining, Suchmaschine, Wissensordnung, ...)

→ Ich habe dabei an Visualisierung und Analyse eines Datensatzes gedacht. Konkret würde ich gerne mit einem öffentlichen Datensatz der US-Regierung von data.gov arbeiten. Eine ausführliche Dokumentation würde per Sphinx erfolgen.

Bsp. 1: <http://catalog.data.gov/dataset/crime-safety-2010-2013>

Kompakter Datensatz über Verbrechen und Sicherheit in Baltimore von 2010 bis 2014. Haben Verbrechen in der Zeit ab- oder zugenommen? In welchen Bezirken geschehen die meisten Verbrechen, in welchen die wenigsten? Gibt es Unterschiede in der Verbrechensart bei der Jahresentwicklung oder Bezirken? Eigene Recherche: Wie kann dies begründet werden – gibt es Gemeinsamkeiten in Wohlstand/Armut oder Lage der Bezirke in der Stadt?

Bsp. 2: <http://catalog.data.gov/dataset/census-data-languages-spoken-in-chicago-2008-2012-9ad5d/resource/1e83077e-8573-48ff-9b40-b5c5c716e4c4>

Kompakter Datensatz über gesprochene nicht-englische Sprachen in Chicago. Welchen Sprache werden am meisten gesprochen? Welche Sprachen dominieren die meisten Bezirke? Eigene Recherche: Warum ist das so? Hat es mit der Lage der Bezirke zu tun; mit der Ethnik der Einwohner, mit Wohlstand/Armut?

Bsp. 3: <http://catalog.data.gov/dataset/open-data-crm-2016>

Sehr großer Datensatz über Verbrechen, die in diesem Jahr in Los Angeles verzeichnet wurden. Welche Verbrechensarten fanden am meisten statt; welche in welchem Bezirk? Eigene Recherche: Wie kann dies begründet werden? ... Einschränkung auf gewissen Zeitraum (da Datei sehr groß)? Falls die Größe einer Datei keine Probleme bereitet, wäre wohl auch eine Analyse die betreffenden Monate über möglich.

4. Ist dies in der Bearbeitungszeit (ca. 4-8 Wochen) realisierbar?

→ Ja. Am meisten würde mich Stand jetzt Bsp. 1 reizen.