

# Projektseminar Informetrie (Vorlesung)

Hinweise nach den Begutachtungen  
Projekt 4/5/wasauchimmer...  
(auf jeden Fall das letzte für dieses Semester)

A word cloud in the shape of a heart, featuring the year 2014 and various positive wishes like 'HAPPY NEW YEAR!', 'GOOD LUCK', 'PEACE', 'LOVE', 'SUCCESS', 'HAPPINESS', 'GOOD HEALTH', 'MONEY', 'PROSPERITY', 'FRIENDSHIP', and 'GOOD FORTUNE'. The words are arranged in a way that they form the outline of a heart, with the year 2014 prominently displayed in the center. The words are in various sizes and orientations, creating a dynamic and celebratory feel. The colors are primarily shades of blue, green, and yellow, with some white text for contrast. The overall design is festive and optimistic, reflecting the spirit of the New Year.

# WE WANT YOU



Projekt X: Uncle Sam  
(optionales Projekt, Meldefrist: 15.01.14)

# Projekt X: Uncle Sam

- Wanted! Projektseminar IR braucht Unterstützung!
- Gesucht werden:
  - 2 SHK
  - Zeitraum: 01.03.2014 – 31.07.2014 (5 Monate)
  - Wochenarbeitsstunden: 8
  - fürstliche Entlohnung: 8,73 pro Stunde (damit > Mindestlohn!)
  - Tätigkeiten:
    - Korrektur von Hausaufgaben (hauptsächlich)
    - Bereitstellung von Materialien
    - E-Mail-Support
    - was uns sonst noch einfällt...
  - Voraussetzungen:
    - gute(!) Python-Kenntnisse
    - schon mal was von „Information Retrieval“ gehört...
    - Verlässlichkeit und Motivation
    - ~ 8 Stunden Zeit pro Woche







Rückmeldungen zu den Abgaben

# Rückmeldungen / Anmerkungen

- Umlaute können (und sollten) in Sphinx und im Quelltext genutzt werden
- Umgang mit eigenen API-Keys oder generell personalisierten Zugangsdaten
- Was nicht erforderlich ist:
  - Aufgabenstellung in eigene Abgabe kopieren
  - Wikipedia-Einträge kopieren



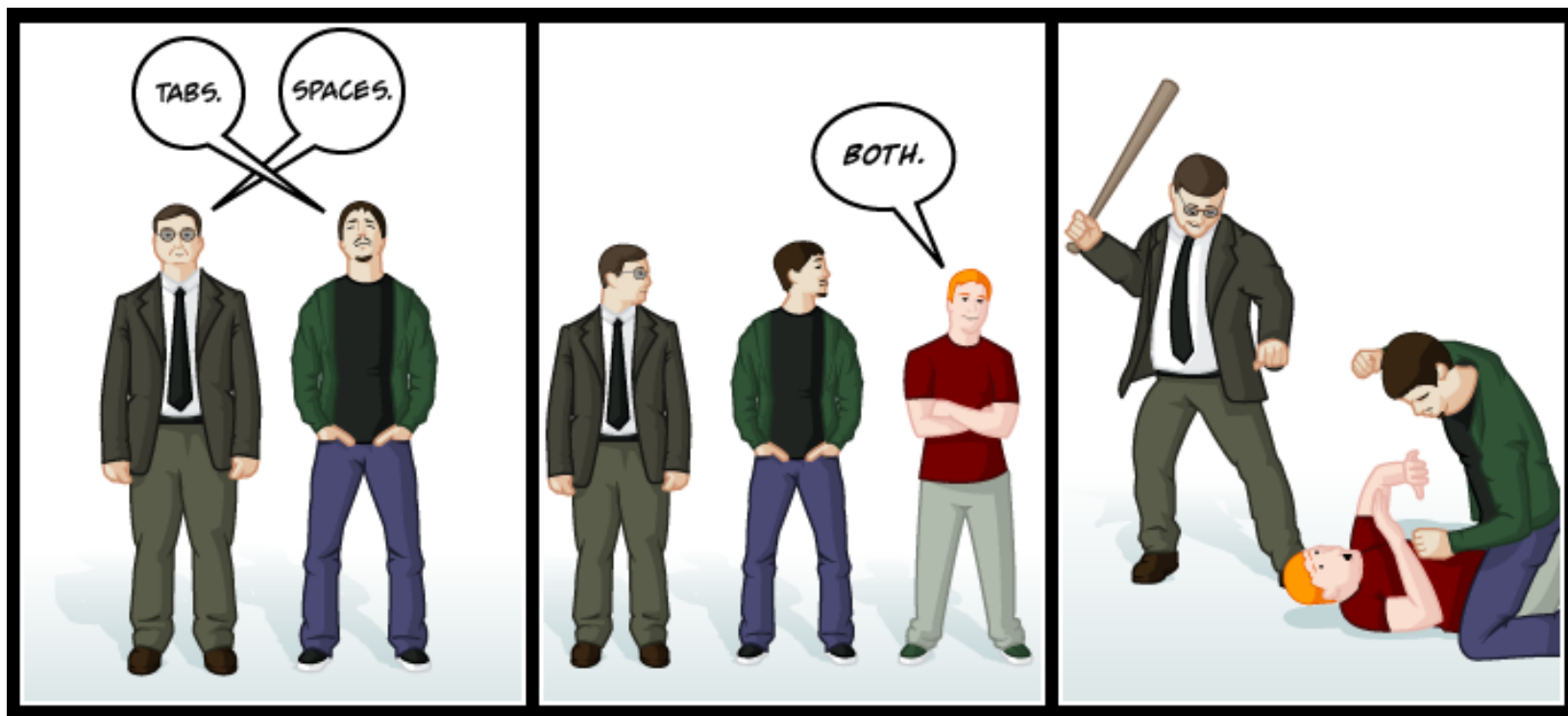
THE EXCLAMATION POINT!

www.exclamationpoint.com © Tony Prescott and Bruce Anthony

# Quiz: Was passiert hier (nicht)?



# Quiz: Was passiert hier (nicht)? - Auflösung







Projekt 5 (oder meinetwegen 4): Ausblick

## Projekt 5 (oder meinetwegen 4): Ausblick (alt)

- Aufgabenstellung: Wahlmöglichkeiten
- Teamzusammenstellungen: komplexe (Wahl-)Möglichkeiten bzw. Zusammenstellungsmöglichkeiten
  - Eigene Ideen für Projekte? Teile sie uns mit! (E-Mail)
- Bearbeitungszeit: 08.01.14 - 05.02.14
- mehr dazu in der ersten Vorlesung im Januar...



## Projekt 4/5/... :: Teamzuteilung

- eigene Teamzusammenstellung möglich, Möglichkeit (a)
- alternativ: weiterhin zufällige Zuteilung, Möglichkeit (b)
- in jedem Fall:
  - E-Mail an jeweiligen Übungsgruppencoordinierenden
    - Möglichkeit (a): E-Mail in CC an alle Teammitglieder; Teammitglieder in E-Mail-Text aufführen; passenden Betreff wählen (z.B. Projektseminar IM – Projekt 4/5 – Team)
    - Möglichkeit (b): E-Mail mit der Info, dass man zufällig zugeteilt werden soll
- Frist für diese Information: Montag, 13.01.2014 um 12:00 Uhr, CET
- Frist verpasst bzw. keine Rückmeldung?
  - zufällige Zuteilung zu einem Team





## Projekt 5: „Endkampf Kraken“



# Möglichkeit 1: Radiosender

- Herunterladen von Daten zu gespielten Songs
- Auswertung, z.B.:
  - Häufigkeit (Songs, Künstler)
  - Zeitpunkt (Uhrzeit, Tag)
  - Top X (Wochencharts, etc.)
  - Zeitverlauf (saisonale Auffälligkeiten)
  - Keywords in Titeln / Künstlernamen
- mögliche Quellen (teilweise mit API, teilweise ohne API)
  - <http://www.einslive.de/musik/playlists/>
  - <http://www.jumpradio.de/musik/suche/index.html>
  - etc.



## Möglichkeit 2: Börsenkurse

- Herunterladen von Börsendaten
- Auswertung:
  - Entwicklung bestimmter Börsenkurse im Vergleich
  - Bezug auf Ereignisse
  - historische Entwicklung
- mögliche Quellen (teilweise mit API, teilweise ohne):
  - <http://de.finance.yahoo.com/>
  - <http://www.google.com/finance>
  - <http://www.eoddata.com>
  - etc.



## Möglichkeit 3: Filmdaten

- Herunterladen von Metadaten zu Filmen (nicht die Filme selber!)
- Auswertung:
  - Durchschnittsbewertungen (Schauspieler, Regisseur, Studio)
  - Filmerscheinungen pro Jahr
- mögliche Quellen (teilweise mit API, teilweise ohne):
  - [www.rottentomatoes.com](http://www.rottentomatoes.com)
  - [www.themoviedb.org](http://www.themoviedb.org)
  - [www.imdb.com](http://www.imdb.com)
  - [www.amazon.com](http://www.amazon.com)
  - etc.



## Möglichkeit 4: Wetterdaten / Pegelstände

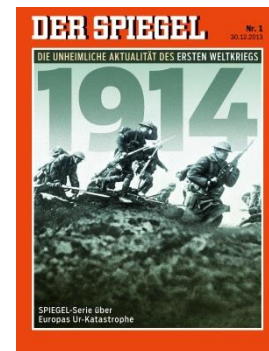
- Herunterladen von Wetterdaten / Pegelständen
- Auswertung:
  - zeitlicher Verlauf
  - Regenmenge, Sonnenstunden, ...
  - Auffälligkeiten
- mögliche Quellen (teilweise mit API, teilweise ohne):
  - [www.wetter.com](http://www.wetter.com)
  - <http://openweathermap.org/>
  - etc.





# Möglichkeit 5: Spiegel

- Herunterladen der Artikel aus dem Spiegel der letzten n Zeiteinheiten
- Auswertung:
  - relative Häufigkeit bestimmter Worte
  - Phänomene und Besonderheiten
  - Personennamen / Institutionen
  - Sprachwandel
- Quelle:
  - <http://www.spiegel.de/spiegel/print/>



## Möglichkeit 6: „Macht doch was ihr wollt!“

- eigene Idee umsetzen
- aber: vorher vorstellen und mit uns absprechen
- Frist: bis 13.01.2014 16:00 Uhr
- bei der Wahl einer anderen Möglichkeit: nicht melden!



# Aufgabenstellung für jede der Möglichkeiten

- Daten erheben (per API oder von den Seiten crawlen)
- Daten persistent abspeichern
- Daten auswerten (je nach Daten ergeben sich unterschiedliche Auswertungsmöglichkeiten; überlegt euch sinnvolle Ansätze)
- Auswertung visualisieren (matplotlib-Diagramme) und dokumentieren (Sphinx)
- Abgabe:
  - Dokumentation mit Beschreibung (mit Sphinx erstellt!)
    - welche Daten wie geholt wurden
    - welche Auswertungen gemacht wurden und warum diese gemacht wurden (Auswertungsfragen)
    - welche Ergebnisse dabei herausgekommen sind und wie diese **interpretiert** werden können (nicht nur ein Satz!)
  - heruntergeladene Daten (in welchem Format auch immer...)
  - lauffähiger(!) Quellcode in vernünftiger Aufteilung

*The  
End*