Contents

[1 Kriterien 1](#_Toc381291063)

[2 Datenquellen 1](#_Toc381291064)

[2.1 Million Songs 1](#_Toc381291065)

[2.2 Website-Logs (GovClicks) 4](#_Toc381291066)

[2.3 GeoNames 4](#_Toc381291067)

[2.4 thingISBN 5](#_Toc381291068)

[2.5 1 Million Github Repos: Language 5](#_Toc381291069)

[2.6 Sourceforge 6](#_Toc381291070)

# Kriterien

* Verständlichkeit / Anschaulichkeit der Daten
* Interessante Algorithmen anwendbar (Simulationen?)
* GUI

# Datenquellen

<http://www.kdnuggets.com/datasets/index.html>

<http://datamob.org/datasets>

## Million Songs

<http://labrosa.ee.columbia.edu/millionsong/pages/example-track-description>

**artist\_mbid: db92a151-1ac2-438b-bc43-b82e149ddd50**

the musicbrainz.org ID for this artists is db9...

**artist\_mbtags: shape = (4,)**

this artist received 4 tags on musicbrainz.org

**artist\_mbtags\_count: shape = (4,)**

raw tag count of the 4 tags this artist received on musicbrainz.org

**artist\_name: Rick Astley**

artist name

**artist\_playmeid: 1338**

the ID of that artist on the service playme.com

**artist\_terms: shape = (12,)**

this artist has 12 terms (tags) from The Echo Nest

**artist\_terms\_freq: shape = (12,)**

frequency of the 12 terms from The Echo Nest (number between 0 and 1)

**artist\_terms\_weight: shape = (12,)**

weight of the 12 terms from The Echo Nest (number between 0 and 1)

**audio\_md5: bf53f8113508a466cd2d3fda18b06368**

hash code of the audio used for the analysis by The Echo Nest

**bars\_confidence: shape = (99,)**

confidence value (between 0 and 1) associated with each bar by The Echo Nest

**bars\_start: shape = (99,)**

start time of each bar according to The Echo Nest, this song has 99 bars

**beats\_confidence: shape = (397,)**

confidence value (between 0 and 1) associated with each beat by The Echo Nest

**beats\_start: shape = (397,)**

start time of each beat according to The Echo Nest, this song has 397 beats

**danceability: 0.0**

danceability measure of this song according to The Echo Nest (between 0 and 1, 0 => not analyzed)

**duration: 211.69587**

duration of the track in seconds

**end\_of\_fade\_in: 0.139**

time of the end of the fade in, at the beginning of the song, according to The Echo Nest

**energy: 0.0**

energy measure (not in the signal processing sense) according to The Echo Nest (between 0 and 1, 0 => not analyzed)

**key: 1**

estimation of the key the song is in by The Echo Nest

**key\_confidence: 0.324**

confidence of the key estimation

**loudness: -7.75**

general loudness of the track

**mode: 1**

estimation of the mode the song is in by The Echo Nest

**mode\_confidence: 0.434**

confidence of the mode estimation

**release: Big Tunes - Back 2 The 80s**

album name from which the track was taken, some songs / tracks can come from many albums, we give only one

**release\_7digitalid: 786795**

the ID of the release (album) on the service 7digital.com

**sections\_confidence: shape = (10,)**

confidence value (between 0 and 1) associated with each section by The Echo Nest

**sections\_start: shape = (10,)**

start time of each section according to The Echo Nest, this song has 10 sections

**segments\_confidence: shape = (935,)**

confidence value (between 0 and 1) associated with each segment by The Echo Nest

**segments\_loudness\_max: shape = (935,)**

max loudness during each segment

**segments\_loudness\_max\_time: shape = (935,)**

time of the max loudness during each segment

**segments\_loudness\_start: shape = (935,)**

loudness at the beginning of each segment

**segments\_pitches: shape = (935, 12)**

chroma features for each segment (normalized so max is 1.)

**segments\_start: shape = (935,)**

start time of each segment (~ musical event, or onset) according to The Echo Nest, this song has 935 segments

**segments\_timbre: shape = (935, 12)**

MFCC-like features for each segment

**similar\_artists: shape = (100,)**

a list of 100 artists (their Echo Nest ID) similar to Rick Astley according to The Echo Nest

**song\_hotttnesss: 0.864248830588**

according to The Echo Nest, when downloaded (in December 2010), this song had a 'hotttnesss' of 0.8 (on a scale of 0 and 1)

**song\_id: SOCWJDB12A58A776AF**

The Echo Nest song ID, note that a song can be associated with many tracks (with very slight audio differences)

**start\_of\_fade\_out: 198.536**

start time of the fade out, in seconds, at the end of the song, according to The Echo Nest

**tatums\_confidence: shape = (794,)**

confidence value (between 0 and 1) associated with each tatum by The Echo Nest

**tatums\_start: shape = (794,)**

start time of each tatum according to The Echo Nest, this song has 794 tatums

**tempo: 113.359**

tempo in BPM according to The Echo Nest

**time\_signature: 4**

time signature of the song according to The Echo Nest, i.e. usual number of beats per bar

**time\_signature\_confidence: 0.634**

confidence of the time signature estimation

**title: Never Gonna Give You Up**

song title

**track\_7digitalid: 8707738**

the ID of this song on the service 7digital.com

**track\_id: TRAXLZU12903D05F94**

The Echo Nest ID of this particular track on which the analysis was done

**year: 1987**

year when this song was released, according to musicbrainz.org

## Website-Logs (GovClicks)

<http://www.usa.gov/About/developer-resources/1usagov.shtml>

<http://1usagov.measuredvoice.com/>

{

"a": USER\_AGENT,

"c": COUNTRY\_CODE, # 2-character iso code

"nk": KNOWN\_USER, # 1 or 0. 0=this is the first time we've seen this browser

"g": GLOBAL\_BITLY\_HASH,

"h": ENCODING\_USER\_BITLY\_HASH,

"l": ENCODING\_USER\_LOGIN,

"hh": SHORT\_URL\_CNAME,

"r": REFERRING\_URL,

"u": LONG\_URL,

"t": TIMESTAMP,

"gr": GEO\_REGION,

"ll": [LATITUDE, LONGITUDE],

"cy": GEO\_CITY\_NAME,

"tz": TIMEZONE # in http://en.wikipedia.org/wiki/Zoneinfo format

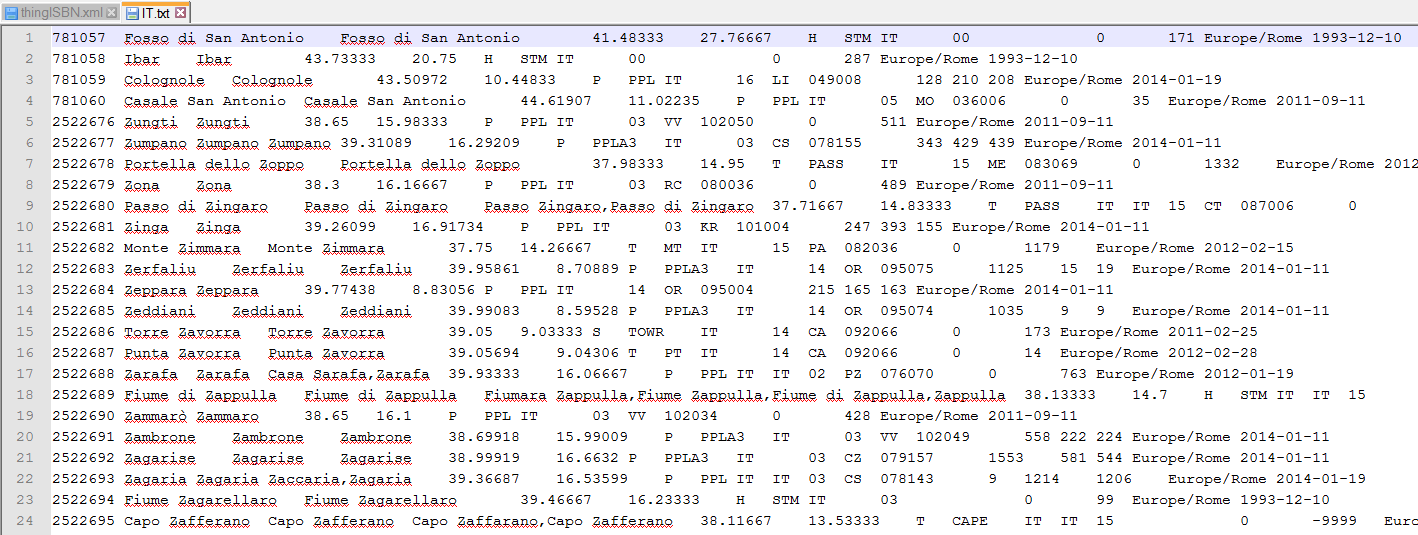
"hc": TIMESTAMP OF TIME HASH WAS CREATED,

"al": ACCEPT\_LANGUAGE http://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html#sec14.4

}

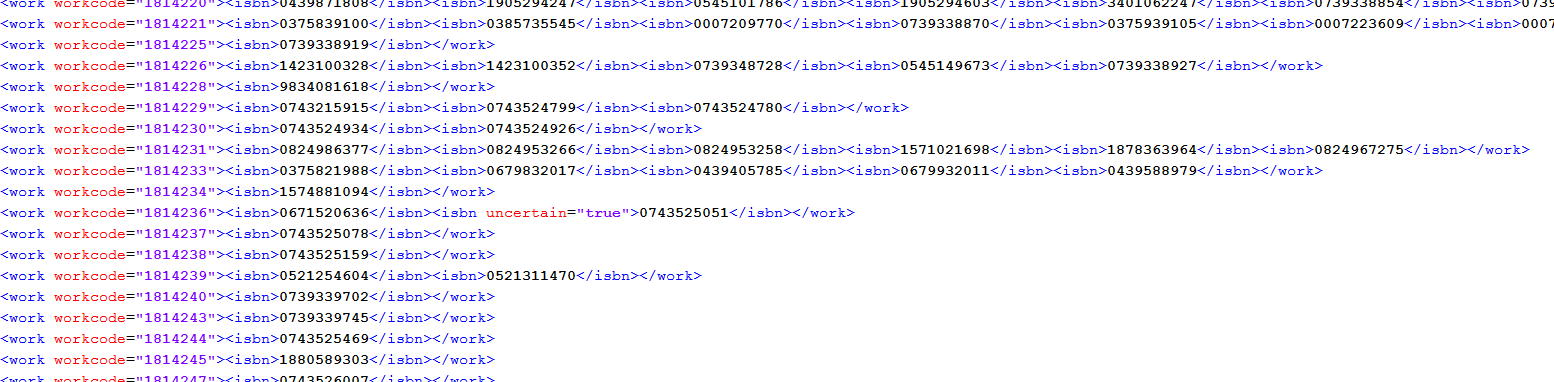
## GeoNames

<http://www.geonames.org/export/>



## thingISBN

<http://datamob.org/datasets/show/thingisbn>



## 1 Million Github Repos: Language

<https://bigml.com/gallery/models>

<https://bigml.com/user/czuriaga/gallery/model/52f3ac9c0c0b5e6e7c0020fc>



## Sourceforge

<http://www3.nd.edu/~oss/Data/data.html>

Kein direkter Download möglich, nur Zugriff auf SQL-Tabellen

