

Stichwortverzeichnis

A

ACADEMIA SINICA 60, 81
Accuracy 45, 156, 197
add one smoothing 55, 164
Ähnlichkeitsmaß 51
Äraname *siehe* Regierungsdevise
AIC *siehe* AKAIKES Informationskriterium
AKAIKES Informationskriterium 212, 220
Anachronismus 37, 196, 229, 250
ante-dating 114, 240
API *siehe* Application Programming Interface
Application Programming Interface 61, 85
AQUIN, Thomas von 4
Archaismus 49, 54
ASCII 69
Average Year of Lexicalization 8, 156, 210, 229, 241, 247
AYL *siehe* Average Year of Lexicalization

B

Bag of Words I, 47, 52, 54, 59, 94, 167
baihua wen 白話文 6, 20, 143, 148, 178
Baseline 45, 96, 158, 172, 175
Beamtenprüfung 19, 202
BERT 76, 83
Bibel 35, 112
bieming 別名 98, 100
Big5 69
bihui [zi] 避諱 [字] *siehe* Tabuzeichen
BoW *siehe* Bag of Words
Browser 8, 232
BUSA, Roberto 4

C

CHAO Yuan Ren 趙元任 110
CHEN Hanbo 陳翰伯 110
chengyu 成語 82, 146
Chinese Treebank 63, 79, 89
chronon 50, 107, 161
Classical Chinese 5
Codierung 69
collocation 45, 249

Computerlinguistik I, 60, 155, 244
COOK's distance 223
cosine similarity 51, 157, 169
CS *siehe* cosine similarity

D

danwei 單位 110
Datenbank 8
Datierung 35
deep learning 41
DENG Xiaoping 鄧小平 109
Dependency Parsing 75
DH *siehe* Digital Humanities
Diagnoseplot 221
Dialekte 6
difangzhi 地方誌 66, 71, 134, 155, 158, 162, 164
Digital Humanities 4, 4, 39, 61, 66, 219
DIRICHLET smoothing 55, 164, 166
Diversifikationsquotient 217
Django 233
duoyinzi 多音字 31, 113, 121, 139

E

encoding 69
Extensible Markup Language 62, 79

F

F-Score 75, 78
false positive 101
feature 47, 94, 156
Federalist Papers 4
Framework 233, 234
Frequency Weighted Average Year of Lexicalization .. 214
Fälschung 37, 38, 192, 196, 230, 246, 250

G

GB *siehe* guojia biaoqun 國家標準
GBK *siehe* guobiao kuozhan 國標擴展
Gewichtungskorrektur 183, 185, 214
ggplot2 5
Goldstandard 78

Stichwortverzeichnis

- guobiao kuozhan* 國標擴展 69
guojia biao zhun 國家標準 69
guwen 古文 6, 148
- H**
Hapax legomenon 158, 160
Heteronym *siehe duoyinzi* 多音字
Hidden-MARKOV-Modell 75, 83
HMM *siehe* Hidden-MARKOV-Modell
- I**
IK Analyzer 89
Imitation 38
Indexjahr 236
Information Retrieval 53
Internetsprache 19
Interpolation 54
Interpunktation 213, 219
Intertextualität 39, 175, 247
inverse document frequency 168
Invisible Hand 12
- J**
JACCARD similarity 52, 157, 161, 169, 171, 173, 178
JACCARD-Divergenz 178
Java 85, 89
JavaScript 234
JELINEK-MERCER smoothing 55, 166
jian tizi 簡[筒]體[体]字 57, 70, 115
Jieba 結巴 83
Julian Day 103
- K**
Kang Ri zhanzheng 抗日戰爭 110
Kangxi Zidian 康熙字典 114
Klassisches Chinesisch 5
KLD *siehe* KULLBACK-LEIBLER-Divergenz
Kommunistische Partei Chinas 109
Konkordanz 112, 150, 153, 236
Korpora 55
Korpus 62
Kosinus-Ähnlichkeit *siehe cosine similarity*
KPCh *siehe* Kommunistische Partei Chinas
KULLBACK-LEIBLER-Divergenz 52, 157, 177
Kulturrevolution 110
- L**
LAPLACE-smoothing 55, 164
Largest chronon minimum smoothing 159, 164
- Lexem-Modell* 160
Lexical dating 179
Lexikalisierung 14
Li Jinxi 黎錦熙 110
linear interpolation 55, 166
Linearregression 202, 210, 212, 230
Linguistik, forensische 37
Linguistische Datierung 35
Literary Chinese 5
LIU Xiang 劉向 142
Locus classicus ... 114, 123, 125, 133, 144, 150, 179, 184, 190, 194, 211, 217, 229, 248
LOEWE-Korpus 66, 134
Logarithmus 167
Lokalchronik *siehe difangzhi* 地方誌
Lü Shuxiang 呂淑湘 91, 110
- M**
machine learning 40
maximum matching 76, 87, 96
mean average error (MAE) 157
Meng xi bi tan 夢溪筆談 ... 123, 126, 138, 180, 185, 211
mySQL 8
- N**
Named Entity Recognition 56, 59, 63, 97, 244
Namenstabu *siehe* Tabuzeichen
Natural Language Processing 60, 63, 250
Nearest neighbour smoothing 166
Neologismus .. 13, 18, 29, 32, 48, 49, 54, 56, 107, 143, 144, 243, 245
Neologismusprofil 180, 186
NER *siehe* Named Entity Recognition
n-Gramm 188
n-Gramm 43, 47, 59, 61, 66, 91, 94, 102, 104, 156, 164, 173, 182, 183, 197, 200, 217, 230, 244, 249
nianhao 年號 *siehe* Regierungsdevis
NLLR *siehe* Normalized Log-Likelihood-Ratio
NLP *siehe* Natural Language Processing
NLTK 93
Normalized Log-Likelihood-Ratio . 51, 157, 164, 169, 175, 229, 237
- O**
Open Source 75, 81, 89, 250
Optical Character Recognition 48
Optical Character Recognition 69
OUYANG Xiu 歐陽修 20, 128, 223

P

Paoding's Knives 89
Part-of-Speech Tagging 25, 45, 63, 75
 PCA *siehe Principal Component Analysis*
 PIOTROWSKI-Gesetz 145, 245
 Plain Text 3, 64, 73, 118, 122, 230
 PoS-Tagging *siehe Part-of-Speech Tagging*
 Precision 78
 Principal Component Analysis 21, 41, 43
 Python 83, 92, 93, 96, 104, 121, 134, 233, 245

Q

Qianlong 乾隆 67, 68, 204

R

R [Korrelationskoeffizient] 199, 212
 R [Programmiersprache] 5
 Recall 78, 80, 88, 94
 RegEx 96, 98, 104, 121
 Regierungsdevisen 67, 103–105, 161, 189, 203, 246
 Regulärer Ausdruck *siehe RegEx*
 Renmin ribao 人民日報 109, 111
 RMRB *siehe Renmin ribao* 人民日報

S

SAYL .. *siehe Standardized Average Year of Lexicalization*
 Schriftsprachliches Chinesisch 5
 Scriptura continua 48, 74
 SHAKESPEARE, William 4, 41, 112
 Shiji 史記 190
 si zi 死字 113
 Slicing 184
 SLM *siehe Statistical Language Model*
 Smoothing 54, 164, 229
 Softwarebibliothek 5, 71, 76, 81, 83, 93
 SONG Qi 宋祁 223
 Sprachmodell *siehe Statistical Language Model*
 Sprachwandel ... I, II, 48, 49, 148, 161, 177–179, 243, 248
 SQL 8, 71, 99
 Standardized Average Year of Lexicalization ... 213, 214
 Stanford Segmenter 89
 Statistical Language Model I, 40, 43, 156, 229
 Stemming 69
 Stichprobe 116
 Stilometrie 2, 12, 21, 41, 61, 68
 stop words II, 69
 Stylochronometry 41, 210
 support vector machine 53, 249

T

Tabuzeichen 39, 70
 Tagging 98
 TEI 62
 temporal tagging 104
 temporal expression 46, 189
 Temporale Entropie 53, 157, 158
 Temporales Textprofil 180, 189
 term frequency 51
 Textdatierung 35
 tf-idf 53, 158
 Tokenizer 78
 tokens 59, 74, 78
 Topic modelling 39, 43, 61
 Trainingskorpus 40
 Transtemporalität 247
 TRENCH, Richard 112
 TTR *siehe type-token ratio*
 types 59, 161
 type-token ratio 217

U

unseen events 164, 229
 user interface 8, 232
 UTF-8 70

V

Visualisierung 234
 VITERBI-Algorithmus 75, 83

W

WAYL *siehe Frequency Weighted Average Year of Lexicalization*
 Weighted Average Year of Lexicalization 213
 wenhua da geming 文化大革命 110
 Wenlin 文林 87
 wenyan wen 文言文 5, 148
 word sense disambiguation 76, 139
 wusi yundong 五四運動 178

X

xinci 新詞 *siehe Neologismus*
 XML *siehe Extensible Markup Language*
 Xu xiu si ku quan shu 續修四庫全書 66
 XXSKQS *siehe Xu xiu si ku quan shu* 續修四庫全書

Y

yitizi 異體字 31, 67, 70, 70
 Yuan shan 原善 177, 208

Stichwortverzeichnis

Z

- Zhongjing* 忠經 39, 192, 224
Zhongyang yanjiu yuan 中央研究院 ..siehe ACADEMIA
SINICA
ZHOU Enlai 周恩來109
Zufallsgenerator 45, 158, 175, 232