

XiaoyangChen

Natural Language Processing & Text Mining

contact

5# Zhongguancun
south street
Beijing 100081
China

+86 156 5299 5998

ling032x@gmail.com
http://ling0322.info
Twitter: @ling0322
Github: ling0322

Languages

♥ C/C++
Python
Java, C#
Javascript
Go, Scala

Machine Learning

Maximum Entropy
Logistic regression
Perceptron, SVM
HMM, LDA

Libraries

Hadoop
MongoDB, LevelDB
node.js, tornado
cedar, darts, opal
MaltParser, ZPar
nltk, word2vec

Tools

♥ Sublime Text
Visual Studio
Eclipse, Vim
automake
gdb, valgrind
L^AT_EX

Education

- 2012-Now **Masters** of Science Beijing Institute of Technology
Major: Computer Science; Research Direction: Natural Language Processing
Main Courses: Computational linguistics, Machine learning, Data mining.
- 2008-2012 **Bachelor** of Science University of Science and Technology of Suzhou
Major: Computer Science;
Main courses: Algorithms, Operating system, Compilers, Network theory.

Internship

- 2014.5-Now **Baidu Inc.** Research and Development Engineer
HCI-BASE group, NLP department. Projects involved are:
- Entity tag extraction for Baidu knowledge graph.
- Improve Query-Title matching with dependency tree and machine translation.
- 2012.3 **Wandou Lab (wandoujia.com)** Front-end Engineer
Front-end group. Internal collaboration tools maintains.

Project Experience

- 2012-Now **MilkCat - Chinese Dependency Parsing Toolkit** Open source project
An Chinese open source dependency parsing toolkit. Implemented in C++. Provides interfaces for C, Python and Go.
- Segmenter: bigram & CRF, 2.5MB/s, F1 = 0.968
- POS Tagger: TnT & CRF (OOV only), 2MB/s
- Parser: arc-eager, zhang11, UAS=0.854 (beam-search)
- 2013-2014 **Research of critical technologies in sentiment analysis** Laboratory project
Supported by National Natural Science Foundation of China. My work includes:
- Web crawlers (for news and comments, Weibo, Tieba).
- Chinese new word extraction from large corpus.
- 2013 **Goldeye: Web data mining software.** Laboratory project
Data mining platform for web text, Deployed by Financial Magazine, My work includes:
- Data visualization (d3.js).
- Relationship analysis for entities and words.
- 2013-2014 **CoCoCat: Sina weibo crawler** Laboratory open source project
User profile, friends and connections, timeline and user search results. Written in python. My work includes:
- HTTP request module (with cookies) and web login module of Weibo user.
- User timeline and comments extraction module.