

# WEI XU

✉ readonlyfile@hotmail.com · ☎ (+86) 1800-451-5446 · 🔗 <http://memeda.github.io>

## 🎓 EDUCATION

---

**HARBIN INSTITUTE OF TECHNOLOGY**, Harbin, China Sep. 2015 -- Present

*Master student* in Computer Science (CS), expected Jul. 2017

**HARBIN INSTITUTE OF TECHNOLOGY**, Harbin, China Aug. 2011 -- Jul. 2015

*B.S.* in Computer Science (CS), grads among top 10%

## 👤 EXPERIENCE

---

**Deep Learning for Chinese Word Segmentation, POS Tagging, NER** Mar. 2016 -- Present

C++ Main Project

GitHub repository : <https://github.com/memeda/sequence-labeling-by-nn>

The goal is to build industrial grade sequence labeling framework for NLP with high accuracy and speed.

- Developed unified models that integrated neural representation learning (MLP, LSTM) with global structural inference methods (CRF).
- Conducted extensive experiments on benchmark datasets and obtained better score than traditional structured perceptron-based implementations in LTP ([www.ltp-cloud.com](http://www.ltp-cloud.com)).

**Text Classification on Imbalanced Datasets** Mar. 2015 -- Jun. 2015

Python Graduation Project

GitHub repository : <https://github.com/memeda/GraduationDesignCodeForTextClassification>

Developed two systems respectively using boosting tree(robust) and SVM(high F1-score).

- Explored a variety of models with different features and classifiers on balanced and imbalanced datasets.
- Improved imbalanced data classification performance by sampling, SMOTE, etc.

**Baidu Inc.** Beijing, China Jun. 2014 -- Sep. 2014

C/C++ , Python , shell , MapReduce NLP Intern

Deployed knowledge base extraction system in Hadoop clusters independently.

- Improved shell program performance and verified kernel procedure's correctness.
- Deployed them in 5 Hadoop clusters simultaneously and processed data from 20K web databases.
- Obtained 200+ TB knowledge in one month.

## ⚙️ SKILLS

---

- Programming Languages: skilled in C++ , familiar with C/Python, have experiences on Shell, JS, Java.
- Development Experience: web(frontend), basic Linux scripts, GUI application(Qt/MFC/Java).
- Research Interest: text classification, sequence labeling(structure prediction), deep learning.

## ♥️ HONORS AND AWARDS

---

**National Scholarship** Oct. 2014

## 📄 MISCELLANEOUS

---

- GitHub: <https://github.com/memeda>
- Foreign Language: English - CET-6