

KenLM with Python Bindings - Installation Manual

Platform Information and Prerequisites

1. **Platform** : Python2.7 on MacOSX Yosemite 10.10.4
2. **Github Repo** : [KenLM](#)
3. **Creator's notes**: [Link](#)

Setting up KenLM

Part 1: Download

```
wget http://kheafield.com/code/kenlm.tar.gz
tar -zxvf kenlm.tar.gz
sudo mv kenlm /opt/kenlm
```

Part 2: Building binaries

```
cd /opt/kenlm
sudo mkdir -p build
cd build
sudo cmake .. -DKENLM_MAX_ORDER=10
sudo make -j 4
```

The binaries will be in `/opt/kenlm/build/bin`. This path must be added to `$PATH`. This can be done by creating/editing `~/.bash_profile` and adding the following line:

```
export PATH=$PATH:/opt/kenlm/build/bin
```

Now the binaries should be accessible from anywhere, especially `lmplz` and `build_binary` which should be accessible from anywhere.

Part 3: Setting up the Python bindings

Open `/opt/kenlm/setup.py` file and change the line `-DKENLM_MAX_ORDER=6` to `-DKENLM_MAX_ORDER=10`.

```
sudo python setup.py install
```

You can test your installing by trying to `import kenlm` from anywhere on your machine.

Setting up model files used in our code

Create the following bash script in the same directory as the `LM-train-100MW.txt` file.

```
# Bash script called create_lm.sh
lmplz -o $1 --skip_symbols < LM-train-100MW.txt > LM_${1}gram.arpa
build_binary LM_${1}gram.arpa LM_${1}gram.klm
```

Build 2-6gram Language models(binary files)

```
bash create_lm.sh 2  
bash create_lm.sh 3  
bash create_lm.sh 4  
bash create_lm.sh 5  
bash create_lm.sh 6
```

Now just modify `settings.ini` to point to the directory holding these model files.

Done!