M2ASR NP Package Documentation

Dong Wang   xx
CSLT, Tsinghua University   xxxxxx University
Beijing 100084, <u>China</u>   xxxx, <u>xxx</u>

January 18, 2017

Abstract

This is a document for M2LP package.

1 M2ASR introduction

M2LP is part of the M2ASR projected, and it is an open-source package for multi-minority langauge processing tasks.

Multi-Minority Automatic Speech Recognition (M2ASR) is an NSFC-supported project, with the aim of providing speech recognition service for minority nations of China. We are planning to investigate, develope and architect the ASR service within 5 years. We will pulish all the tools, data, model, and service for free.

M2ASR is a multi-party project involing CSLt@Tsinghua Unviersity, Xinjiang University, Northwest University for Nations (NUW). M2LP provides the basic functions for langauge processing, including text normalization, morphology segmentation, langauge modeling, and all the related resources, functions, and tools.

2 M2LP pacakge

The basic design principle is reusable and modulation. The functions will be implemented as functions written by c/c++, python or perl, and the recipe will be a shell script that integrating modules by piple line.

DIRECORY

- rc : multilingual langauge resource

- src : source code for LP

- md : model released