

USER MANUAL

INSTALLATION AND TOOL DESCRIPTIONS

1. Installation

Prerequisites

The following software are required to run the system:

- bash (if run on Linux systems)
- csh (if run on Linux systems)
- Perl version 5.6 or later.
- Java VM 1.5 or later (required by Stanford Parser)
- WordNet 3.0 (required by WordNet::Similarity)
- Text::Similarity (required by WordNet::Similarity)
- WordNet::QueryData: (required by WordNet::Similarity)
- WordNet::Similarity

1.1. Linux

Install bash, csh, perl and Java

In most Linux systems, these packages will be installed by default. If not, one can use the following commands to search and download them accordingly:

```
$apt-cache search package_name
```

```
$apt-get install package_name
```

Install WordNet 3.0

```
$cd /tmp
```

```
$wget http://wordnet.princeton.edu/3.0/WordNet-3.0.tar.gz
```

```
$tar -zxvf WordNet-3.0.tar.gz
$cd /tmp/WordNet-3.0
$./configure
($./configure --prefix=/tmp/wn3.0 # non-root user)
$make
$make install
```

Install WordNet::QueryData

This assumes that you have already installed Perl.

For root user:

```
$sudo perl -MCPAN -e shell
>cpan install WordNet::QueryData
...
cpan> quit
```

For non-root user:

Download and unpack the package:

```
$cd /tmp
$wget http://search.cpan.org/CPAN/authors/id/J/JR/JRENNIE/ \
WordNet-QueryData-1.46.tar.gz
$tar -zxvf WordNet-QueryData-1.46.tar.gz
```

Set the WNHOME environment variable to the WordNet3.0 location:

```
$export WNHOME=/tmp/wn3.0
```

Then, install as usual:

```
$perl Makefile.PL PREFIX=/tmp/wnqd1.46
$make
$make test
$make install
```

Install Text::Similarity

For root user:

```
$sudo perl -MCPAN -e shell
>cpan install Text::Similarity
...
cpan> quit
```

For non-root user:

Download and unpack the package:

```
$cd /tmp
$wget http://search.cpan.org/CPAN/authors/id/J/JA/JASONM/ \
Text-Similarity-0.02.tar.gz
$tar -zxvf Text-Similarity-0.02.tar.gz
```

Install:

```
$perl Makefile.PL PREFIX=/tmp/ts0.02
$make
$make test
$make install
```

Install WordNet::Similarity

For root user:

```
$sudo perl -MCPAN -e shell
>cpan install WordNet::Similarity
...
cpan> quit
```

For non-root user:

Download and unpack the package:

```
$cd /tmp
```

```
$wget http://search.cpan.org/CPAN/authors/id/T/TP/TPEDERSE/ \
WordNet-Similarity-2.01.tar.gz

$tar -zxvf WordNet-Similarity-2.01.tar.gz
```

Set the PERL2LIB environment variables (could be set in .bashrc file):

```
export \
PERL5LIB="/tmp/ws2.01/lib/perl5/site_perl/5.8.8:/tmp/wnqd1.46/ \
lib/perl5/site_perl/5.8.8:/tmp/ts0.02/lib/perl5/site_perl/5.8.8"
```

Install:

```
$perl Makefile.PL PREFIX=/tmp/ws2.01

$make

$make test

$make install
```

I.2. Windows

Install WordNet 2.0

Download and install the package from the website:

<http://wordnet.princeton.edu/oldversions>

Install WordNet::QueryData 1.38

Download and unzip the package from the link:

<http://people.csail.mit.edu/jrennie/WordNet/WordNet-QueryData-1.38.tar.gz>

Copy the file "QueryData.pm" into the directory "C:\Perl\site\lib\WordNet" (create the WordNet directory if it is not existed).

Install Text::Similarity 0.07

Download and unzip the package from the link:

<http://archive.cpan.cz/authors/id/T/TP/TPEDERSE/Text-Similarity-0.07.tar.gz>

Open the unzipped folder, copy the directory "Text" which includes the files "OverlapFinder.pm", "Similarity.pm" and "Similarity" into "C:\Perl\site\lib\Text" (create the Text directory if it is not existed).

Install nmake

Download the package from the link:

<http://support.microsoft.com/default.aspx?scid=kb;en-us;Q132084>

Run the downloaded executable to extract the files. Copy both the NMAKE.EXE and the NMAKE.ERR file to the "C:\Perl\bin" directory.

Install WordNet::Similarity 0.16

Download and unzip the package from the link:

[http://backpan.perl.org/authors/id/T/TP/TPEDERSE/WordNet-Similarity- \ 0.16.tar.gz](http://backpan.perl.org/authors/id/T/TP/TPEDERSE/WordNet-Similarity-0.16.tar.gz)

Type the following commands in cmd prompt (note to cd into the current directory):

```
>perl Makefile.pl  
>nmake  
>nmake test  
>nmake install
```

2. Tool descriptions

Followings are the descriptions of the scripts used for the experiments of this work.

Script's name	Description
extract_questions.pl	A Perl script to extract user questions in desired format from log files since the log files contain other information, such as: markers, answers, etc.
extract_TREC_questions.pl	A Perl script to extract questions in desired format from TREC context track data since they are in XML format.
input_types.pl	A Perl script to help improve the process of annotating questions extracted from log files.
annotate_with_treetagger.pl	A Perl script to parse questions to get the lemmas and POS information using TreeTagger utility.
convert_to_question_features.pl	A Perl script to convert questions into question classification features as described in chapter V.1.3
convert_to_FU_question_features.pl	A Perl script to convert questions into topic shift and topic continuation features as described in chapter V.2.2
naïve_bayes.pl	A Perl script of the implementation of Naïve Bayes algorithm.
learn_with_naive_bayes.pl	A Perl script to learn a Naïve Bayes model from a provided training corpus.
classify_testset_with_naive_bayes.pl	A Perl script to evaluate a testing corpus on a trained Naïve Bayes model.
classify_cross_validation_with_naive_bayes.pl	A Perl script to evaluate a corpus using cross-validation technique using Naïve Bayes.