Meteor 1.4 开源工具使用

1、不加任何参数运行：java -Xmx2G -jar meteor-*.jar

可以输出帮助信息，包含如下一些参数。

-l: language

-t: task (rank util adq hter li tune，选择其中之一，默认为 rank)

rank: 根据 WMT09 和 WMT10 的人工排序调参

adq: 用 NIST MT2009 的分数调参

hter： 用 GALE P2 和 P3 的 HTER 数据调参

li: 不依赖与语言的参数

util: -ch 时使用

-norm: tokenize/ normalize punctuation and lowercase

-lower 如果-norm，则不需要

-p: ‘alpha beta gamma delta’ 用户自己设定的参数

-m: ‘module1 module2 …’ 匹配时使用的模块（exact stem synonym paraphrase）

-w: ‘weight1 weight2 …’ 对匹配时使用的各个模块指定权重

-r: refCount 参考译文的个数（仅在 plain 格式时需要）

-x: beamSize (缺省 40)

-s wordListDirectory (if not default for language)

-d synonymDirectory (if not default for language)

-a paraphraseFile (if not default for language)

-f filePrefix 输出文件的前缀 (default 'meteor')

-q 句子级分数在 stderr 输出，最终分数在 sedout 输出，没有其他信息（仅在 plaintext 使用）

-ch Character-based precision and recall

-noPunct 计算分数时不考虑标点

-sgml 输入文件为 SGML 格式

-mira 输入文件为 MIRA 格式 (Use '-' for test and reference files)

-vOut Output verbose scores (P / R / frag / score)

-ssOut Output sufficient statistics instead of scores

-writeAlignments 输出打分过程中的对齐信息

其中-p,-m,-w 在赋值时需要加上单引号。 但是-p 我使用时不管用，可以在代码中赋值.

meteor 在打分时，经常用的是-l 和-norm 两个参数，不指定输入文件格式时，默认为 plaintext，比如：

```
java –Xmx2G –jar meteor-*.jar test reference –l en –norm
```

2、支持的语言：

meteor 支持 UTF-8 编码的各种语言，只是某些语言没有 stem, synonym 和 paraphrase 信息。

Fully supported languages:

Language	Exact Match	Stem Match	Synonym Match	Paraphrase Match	Tuned Parameters
English	Yes	Yes	Yes	Yes	Yes
Arabic	Yes	No	No	Yes	Yes
Czech	Yes	No	No	Yes	Yes
French	Yes	Yes	No	Yes	Yes
German	Yes	Yes	No	Yes	Yes
Spanish	Yes	Yes	No	Yes	Yes

Partially supported languages:

Language	Exact Match	Stem Match	Synonym Match	Paraphrase Match	Tuned Parameters
Danish	Yes	Yes	No	No	LI
Dutch	Yes	Yes	No	No	LI
Finnish	Yes	Yes	No	No	LI
Hungarian	Yes	Yes	No	No	LI
Italian	Yes	Yes	No	No	LI
Norwegian	Yes	Yes	No	No	LI
Portuguese	Yes	Yes	No	No	LI
Romanian	Yes	Yes	No	No	LI
Russian	Yes	Yes	No	No	LI
Swedish	Yes	Yes	No	No	LI
Turkish	Yes	Yes	No	No	LI

对于上面没有提到的语言(比如汉语)，在指定语言时用 other 表示，系统会自动选择 exact 匹配方式（需要自己根据需要提前做分词，token，转小写等工作）。

3、单独用 Aligner（单语的词对齐）

```
Usage: java -Xmx2G -cp meteor-*.jar Matcher <test> <reference> [options]
    -l  language    可以选择的语言: en da de es fi fr hu it nl no pt ro ru
se tr
    -m 'module1  module2 …'
    -t type    对齐类型（coverage VS accuracy），可以选择 maxcov 或 maxacc
    -x beamSize
    -d synonymDirectory
    -a paraphraseFile
```

4、 单独用 Stemmer

```
Usage: java -cp meteor-*.jar Stemmer  language < in  > out
Languages: en da de es fi fr hu it nl no pt ro ru se tr
```

单独用 Aligner 和 Stemmer 时，输入文件为 plain 格式，不包含分词，token，转小写功能，如果需要，用户自己提前做好。