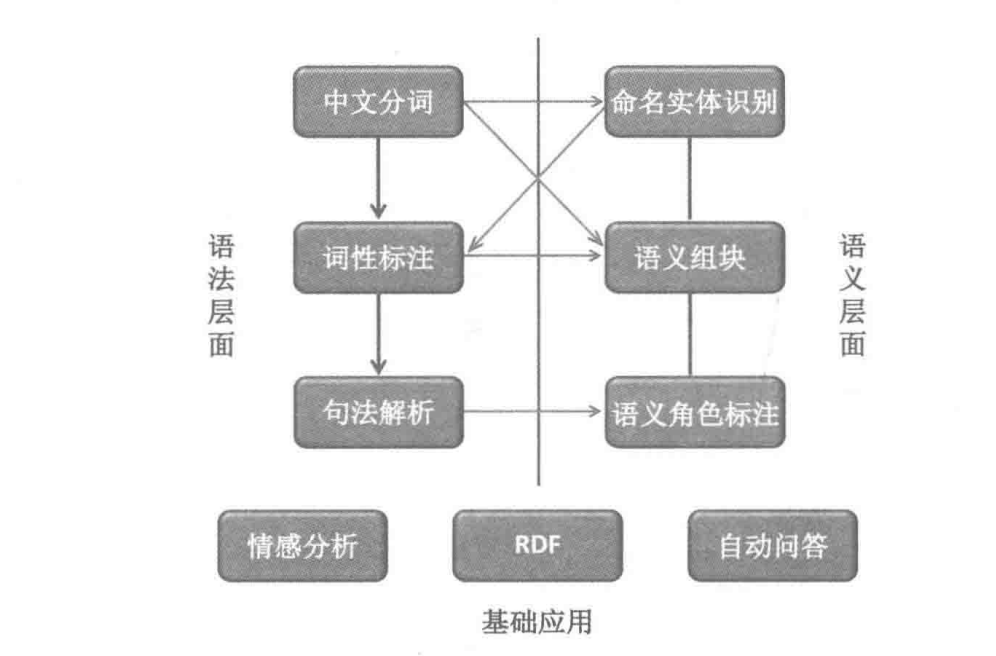
Nlp大致流程







<https://pan.baidu.com/share/link?shareid=1988562907&uk=2738088569#list/path=%2F&parentPath=%2F>

哈工大 <http://www.ltp-cloud.com/demo/>

斯坦福 <http://nlp.stanford.edu/>

哈工大ltp帮助文件:<http://ltp.readthedocs.org/zh_CN/latest/ltptest.html>

斯坦福NLP <https://stanfordnlp.github.io/CoreNLP/>

Python整合corenlp <https://github.com/Lynten/stanford-corenlp>

Python 整合corenlp 中文地址<https://blog.csdn.net/qq_35203425/article/details/80451243>

Word2Vec

1、Strip()删除空格及特殊符号

S=’hello world!’

s.strip()

s.strip(‘!’)

2、联接字符串+

3、查找字符串index，find，findall

4、大小写转换lower upper

5、比较，compile

6、分隔字符串split

正则表达式

. 匹配除换行符意外的任意字符

\d 匹配所有的数据

\D 除数字之外所有的字符

\s 所有空格 tab建 换行

\S 出空格 tab 换行\n之外的所有字符

\w 单个字符[A-Za-z0-9\_]任意一个

\W 非\w

{n} 重复次数\d{2,4} 匹配2~4个数字

？匹配前一个字符0次或者1次

\* 匹配前一个字符0次或者不限次

+ 匹配前一个字符至少出现1次 a(bc)+，bc至少出现1次

[] 指定的任意一个字符[ace]

^匹配字符串开头

$ 匹配结尾

| 或者

(……) 模式

<http://regexr.com>

import re库 re.compile(strPattern[,flag]) re.I 忽略大小写 re.M 匹配多行文本

<https://www.xuebuyuan.com/2042477.html>

<https://github.com/fxsjy/jieba>