

在上述fp1、fp2、adg和timex2norm文件夹中，对于一个给定的文档，我们能够看到这个文档的.sgm源文件以及.ag.xml和.apf.xml的标注。换句话说，对于每一个新闻专题来说，上述每一个文件夹中都包含一个源文本(.sgm文件)的相同副本以及相关标注的不同版本(.ag.xml、.apf.xml)。注意的是，在许多情况下，对于一个给定的源文本，如果在两个标注阶段的后一个阶段中没有做任何更改，那么两个标注阶段会产生相同的转

Filelist文件包含了对于每一个文件的单词统计信息和标注状态。

如下为完整标注文件和它们对应的源文件的路径：

```
*/timex2norm/*sgm
*/timex2norm/*apf.xml
```

接下来是每一种文件类型的内容格式。对于大多数用户来说，最重要的文件是.sgm文件和.apg.xml文件。

Source Text (.sgm) Files

- These files contain the source text data in an SGML format; they use UTF-8 encoding and UNIX-style line termination.

AG (.ag.xml) Files

- These are annotation files created with the LDC's annotation toolkit. These files have been converted to the corresponding .apf.xml files.

ACE Program Format (APF) (.apf.xml) Files

- These files are in the official ACE annotation file format. ACE format is derived by means of a routine format conversion process, so that the underlying annotation content of the two files is equivalent. See section 8 for more details.

ID table (.tab) Files

- These files store mapping tables between the IDs used in the ag.xml files and their corresponding apf.xml files.

关于APF的一些说明（懒得翻，以后有需要的时候再翻一下）

- Offsets

APF uses the offset counting method traditionally used in previous ACE evaluation programs:

- 1) Each (UTF-8) character, not byte, is counted as one.
- 2) Each newline character is counted as one. (The .sgm files use the UNIX-style end of line characters.)

3) SGML tags are *not* counted towards offsets. (Please note that the AG files included in this release do count SGML tags in offsets.)

4) SGML entities are counted in terms of each character in the entities. For example, "&" is counted as five characters, not as one character.

- TIMEX2

The timex2 element represents TIMEX2 time expression annotations. Its optional attributes, such as "VAL" and "MOD", represent the TIMEX2 normalization values.

- TYPE, LDCTYPE and LDCATR in entity_mention

The TYPE attribute in entity_mention stores the official ACE entity mention type, and the LDCTYPE and LDCATR attributes store the attributes used in the LDC's annotation process.

- Name in entity_attributes

The "name" element in entity_attributes stores the heads of "NAM"-type mentions as in the previous years. In response to George Doddington's request, we have added the NAME attribute to the "name" element. The NAME attribute stores slightly normalized versions of the names where:

- \n is replaced with a space
- multiple spaces are reduced to one space
- " (double quote) is removed

- Example:

```
<entity_attributes>
  <name NAME="United States">
    <charseq START="4242" END="4254">United
States</charseq>
  </name>
</entity_attributes>
```

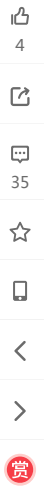
- Nickname metonymy

Nickname metonyms are indicated with METONYMY_MENTION="TRUE" in entity_mentions. "NAN"-type entity mentions marked as nickname metonymy do not give rise to name elements.

- Cross-type metonymy

"Cross-type" metonyms are represented with relations of the type METONYMY. The METONYMY type relations do not have relation_mentions. The METONYMY type relations are automatically generated after the annotation process, and are the only kind of relation annotations that appear in this corpus.

- For more details, please refer to the APF V5.1.2 DTD.



👍 点赞 4 ☆ 收藏 ➦ 分享 ...



taolusi

发布了27 篇原创文章 · 获赞 50 · 访问量 21万+



私信



亿速云高防服务器防御1000G攻击

亿速云防攻击服务器，20+行业领袖视频推荐 增强防CC? 防1000G DDOS BGP\

广告 亿速云

想对作者说点什么

巴斯光甜

3个月前

几千美元的东西，(╯▽╯)"

tailor_long

3个月前

您好，最近做科研需要用到ACE2005语料库，但是我们没有LDS的账号，请问您能不能给我发一份ACE2005的数据集呀！万分感谢：2649139694@qq.cc

qq_38866447

3个月前

作者，你好，请问这个数据集怎么预处理

登录

查看 35 条热评

4

35

赏

ACE 2005数据集（介绍2）

阅读数 3123

以下内容来自<https://catalog.ldc.upenn.edu/LDC2006T06>ACE 2005 Multilingual ... [博文](#) | 来自：[技术践行者](#)

ACE 2005数据集（介绍1）

阅读数 1283

以下内容来自CMU的William Cohen小组（<http://curtis.ml.cmu.edu/w/courses/in...> [博文](#) | 来自：[技术践行者](#)

ACE2005系列1 - 基本数据理解

阅读数 206

参考博文：ACE2005系列1 - 基本数据理解 [博文](#) | 来自：[u014765...](#)

求一个ACE RDC 2005中文标注语料库

求一个ACE RDC 2005中文标注语料库，用作语义关系抽取，论文需要用。感激不尽！204... [论坛](#)

服务器被攻击咋办？
10大游戏公司都在用这防御
立即咨询

亿速云高防服务器送防御增强防CC
亿速云高防服务器，20+行业领袖视频推荐 前10大游戏公司CEO鼎力支

广告 亿速云

关于生成csv文件并下载的问题

有一个需求，将查询到的list先导出成csv文件，然后再将这个csv文件从服务器下载到本地... [论坛](#)

ACE2005系列2 -- 数据预处理

阅读数 587

ACE2005系列2 – 数据预处理ACE2005数据预处理，指使用Standford NLP Tools对... [博文](#) | 来自：[carrie_03...](#)

ace语料，关系抽取可以用

ace_multilang_tr中的其中一个文件，从LDC上面下载下来 [下载](#)

ACE2005数据集文件夹说明

阅读数 847

参考：ACE 2005 文件格式常用数据集 [博文](#) | 来自：[qq_2503...](#)

ACE数据类型

阅读数 536

ACE基本数据类型 ACE_INT16,ACE_UINT16 16位整数（有... [博文](#) | 来自：[game_ru...](#)

ACE2005数据介绍

阅读数 376

ACE2005数据介绍overallEnglish partrelationchinese partarabic partoverallACE2... [博文](#) | 来自：[bobobe...](#)

关系抽取常用的数据集和工具

阅读数 7202

参考文献：张春云.实体关系抽取算法研究[D].北京邮电大学,2015.数据集：1.MUC关... [博文](#) | 来自：[长弓smile...](#)



Tina Tang

72篇文章

排名:千里之外

关注



Wbq_Zero

343篇文章

排名:千里之外

关注



御风而行Carrie

6篇文章

排名:千里之外

关注



Winky_D

46篇文章

排名:千里之外

关注

举报

https://blog.csdn.net/taolusi/article/details/80812597

4/8

<div>数据集划分</div> <div>使用数据集时，一般将其分为三段：训练集、验证集、测试集。测试集决不能使用测...博文 来自：明月清风</div>	阅读数 239	
<div>提取数据集的问题</div> <div>现在数据库Info表有以下字段：tel_number tel_type other_number call_time calls_tim...论坛</div>		<div><div>👍4</div><div>🔗</div><div>💬35</div><div>☆</div><div>📱</div><div><</div><div>></div><div>👑</div></div>
<div>ACE 2005 Data Prep 数据预处理</div> <div>以下内容来自https://github.com/mgormley/ace-data-prepACE 2005 Data Prep...博文 来自：技术践行者</div>	阅读数 782	
<div>上载和下载CSV文件</div> <div>sap中把txt、excel、文件上载到内表中，txt和csv速度最快。excel文件导出的csv是...博文 来自：weixin_3...</div>	阅读数 13	
<div>ACE全套资料</div> <div>包含ACE开发技术蓝皮书，ACE技术论文集上篇,ACE程序员教程中篇,ACE应用实例下篇,四本书。</div> <div>11-05</div> <div>下载</div>		
<div>请问后缀名为ace的是什么文件。用什么软件打开。谢谢。</div> <div>-论坛</div>		
<div>ACE事件抽取任务之事件类型定义</div> <div>ACE定义的8大类事件类型和33小类事件类型原博文介绍很详细，这里不再亲手编写。...博文 来自：qq_4419...</div>	阅读数 802	
<div>求助对话语料哪里找？</div> <div>最近在寻找对话语料，尤其通用领域的尤其难找找到几个链接http://blog.csdn.net/z...博文 来自：lb521200...</div>	阅读数 3975	
<div>ACE下载地址</div> <div>ACE 下载地址http://download.dre.vanderbilt.edu/博文 来自：tissar的博...</div>	阅读数 1898	
<div>ACE库入门(中文)下载</div> <div>ACE库入门(中文) 相关下载链接：//download.csdn.net/download/wshjjobless/210988...论坛</div>		
<div>SemEval2010_任务8_实体关系抽取数据集下载</div> <div>SemEval2010 任务8 实体关系抽取数据集 已经标注的语料 相关下载链接：//download.c...论坛</div>		
<div>ACE事件类型及模板</div> <div>事件类型及子类型每个事件类型和子类型都有自己的一组潜在参与者角色，用于在其...博文 来自：CLOUD</div>	阅读数 2918	
<div>ace各个版本下载地址</div> <div>ace各个版本下载地址</div> <div>04-21</div> <div>下载</div>		
<div>COAE2014，NLPCC2014数据集汇总下载</div> <div>包含中文倾向性分析评测(COAE2014)语料资料，COAE2014评测数据集和NLPCC2014评...论坛</div>		
<div>数据集标注</div> <div>window版本的图像标注工具，下载解压后直接可用。使用简单，直接上手，标签可自定义，输出文件为xml格式。</div> <div>08-05</div> <div>下载</div>		
<div>关系抽取（分类）总结【转载】</div> <div>366次阅读关系抽取(分类)总结文章目录基本介绍FullySupervisedLearning相关文献...博文 来自：微澜同学</div>	阅读数 1万+	
<div>sighan 2006 MSRA命名实体语料(BIO格式)下载</div> <div>2006年sighan命名实体识别任务语料，MSRA提供。已经转成BIO格式，可直接用于NER...论坛</div>		
<div>事件本体以及突发事件语料库--CEC (Chinese Emergency Corpus)</div> <div>1、什么是本体？本体最初是一个哲学上的概念，十多年前被引入计算机领域中作为知...博文 来自：以家为家...</div>	阅读数 6047	
<div>中文语料库中文语料库中文语料库</div> <div>中文语料库中文语料库中文语料库中文语料库中文语料库中文语料库</div> <div>04-20</div> <div>下载</div>		<div><div>🔊</div><div>举报</div></div>

基于 ACE语料库的 信息抽取与文本分类		阅读数 232
一、背景介绍1.事件抽取：根据ACE中的定义，事件由事件触发词（Trigger）和描述... 博文 来自：一位浪漫...		
香港城市大学语料库；微软亚洲研究院语料库；北京大学语料库；		06-14 4
香港城市大学语料库（1 772 202 字，训练集）、微软亚洲研究院语料库（1 089 050 字，训练集）、北京大学语料库（1 ... 下载		
利用动态多池化CNN实现事件提取		阅读数 4155 35
利用动态多池化CNN实现事件提取本篇博客主要介绍论文[Event Extraction via Dyna... 博文 来自：wang_CJ		
对corpus数据处理的认识		阅读数 283
下面是代码：from sklearn import model_selection, preprocessing, linear_model,... 博文 来自：CSDN_LY...		
R各个包里面的数据集列表		阅读数 9193
Package Item Title csv docdatasetsAirPassengersMonthly Airline Passenger Nu... 博文 来自：四纪启悟 ...		
【分享】Delicious 数据集：2005年9月(Delicious Dataset:September ...		阅读数 4119 赏
这个是delicious的data数据集，包含用户、网址和标签数据。This dataset contains ... 博文 来自：数据堂-科...		
Entities and Entities's Relation Extraction of ACE 2005		阅读数 887
Entity Detection and Recognition The ACE Entity Detection and Recognition ta... 博文 来自：HH的专栏		
怎样解压后缀为：*.ace 的文件？		
我再网上下了一个软件他的后缀为 *.ace 请问该如何解压？？？？		论坛
130 个相见恨晚的超实用网站，一次性分享出来		阅读数 41万+
相见恨晚的超实用网站持续更新中。。。		博文 来自：爪白白的...
Java学习的正确打开方式		阅读数 38万+
在博主认为，对于入门级学习java的最佳学习方法莫过于视频+博客+书籍+总结，前... 博文 来自：程序员宜...		
python json java mysql pycharm android linux json格式		



taolusi

TA的个人主页 >

原创

27

粉丝

42

获赞

50

评论

51

访问

21万+

等级: 博客 5

周排名: 6万+

积分: 2299

总排名: 3万+

勋章: 

关注

私信



LOUIS VUITTON

LaTeX自定义新命令 (newcommand、def)


Latex强制图片位置


SSH设置别名访问远程服务器


tf.device()指定tensorflow运行的GPU或CPU设备


Linux grep、tail命令的混合使用


分类专栏

 专利检索与分析

 Ubuntu

 Natural Language P...

 C/C++32篇

 Python32篇

展开

归档

2019年4月6篇

2019年3月3篇

2019年2月9篇

2019年1月1篇

2018年10月3篇

2018年8月12篇

2018年7月61篇

2018年6月19篇

展开

热门文章

Ubuntu下创建新用户
阅读数 24199

gdb断点的删除
阅读数 12469

Python通过pandas将多个csv文件进行合并
阅读数 12302

TensorFlow使用gpu
阅读数 10718

手把手教你使用EndNote X8 详细笔记
阅读数 9668

最新评论

Linux下删除指定文件夹下指定后...
weixin_43789932：您好，"."表示从当前目录开始递归查找，那如果希望仅在当前目录进行该如...

ACE 2005 文件格式
weixin_38950807：几千美元的东西，(￣▽￣)

ACE 2005 文件格式
wql2014302721：您好，最近做科研需要用到ACE2005语料库，但是我们没有LDS的账号，请问...

ACE 2005 文件格式
qq_38866447：作者，你好，请问这个数据集怎么处理

4

35

赏

举报

https://blog.csdn.net/taolusi/article/details/80812597

7/8

ACE 2005 文件格式
weixin_43106157 : [reply]qq_37830928[/reply]
你好，我也想拿数据实现一下。请问有没有ace...



LOUIS VUITTON

QQ客服

kefu@csdn.net

客服论坛

400-660-0108

工作时间 8:30-22:00

[关于我们](#) | [招聘](#) | [广告服务](#) | [网站地图](#)

京ICP备19004658号 经营性网站备案信息

公安备案号 11010502030143

©1999-2020 北京创新乐知网络技术有限
公司 网络110报警服务

北京互联网违法和不良信息举报中心

中国互联网举报中心 家长监护 版权申诉

4

35

举报