

本科论文版检测报告（详细版）

报告编号: BD-20190518-C7892ED3-XX

检测时间: 2019-05-18 18:50:27

题名: 地方

作者: 发阿发

检测范围:
☒ 中国学术期刊数据库
☒ 中国学术网页数据库

☒ 中国学位论文全文数据库
☒ 中国专利文献数据库

☒ 中国学术会议论文数据库
☒ 中国优秀报纸数据库

检测结果

% 总相似比: 18.12%

检测字数: 12273

参考文献相似比: 0.00%

排除参考文献相似比: 18.12%

可能引用本人已发表论文相似比: 0.00%

辅助排除本人已发表论文相似比: 18.12%

单篇论文最大相似比: 2.93%

相似片段分布图



注: 绿色区域为参考文献相似部分, 蓝色区域为本人已发表论文相似部分, 黄色区域为本人学位论文相似部分, 红色区域为其他文献相似部分

相似作者分布

序号	作者	相似比
1	赵京胜	2.93%
2	北京邮电大学	2.59%
3	张荣杰	1.48%
4	蒋莎	1.20%
5	段康康	1.17%
6	杨廷	0.93%
7	王沛	0.86%
8	杨云峰	0.86%
9	肖其伟	0.84%
10	北京百度网讯科技有限公司	0.50%
11	王超	0.46%
12	张胜男	0.42%
13	江苏大学	0.37%
14	刘爽	0.35%
15	北京工商大学	0.33%
16	诸葛庆子	0.29%
17	吕昭	0.26%
18	刘海燕	0.26%
19	陈求兴	0.24%
20	李亚松	0.22%
21	夏玉芹	0.22%
22	苏向东	0.20%
23	肖斌	0.16%
24	黄妩	0.16%
25	许晔	0.15%
26	濮文强	0.15%
27	曹顺良	0.13%

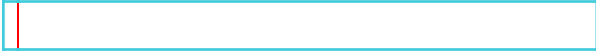

28	戴光荣	0.11%
29		0.11%
30	卜倩	0.11%
31	孙晔琦	0.09%

相似文献列表

序号	相似比	题名	作者	文献类型	来源	发表时间	是否引用
1	2.93%	基于自然语言处理的能源领域知识图谱	赵京胜 等	期刊论文	《信息技术与信息化》	2018-05-02	否
2	2.07%	201510919129.7	北京邮电大学	专利论文	中文专利全文数据库	2015-12-09	否
3	1.48%	感知与现实人际关系网络：经典文学作品分析	张荣杰	学位论文	闽南师范大学	2018-06-01	否
4	1.20%	适用于高校学院级办公系统的推荐技术的研究	蒋莎	学位论文	电子科技大学	2013-06-30	否
5	1.17%	基于隐马尔科夫模型的文本分类器的设计与实现	段康康	学位论文	北京交通大学	2016-06-01	否
6	0.93%	电信运营商智能客服应用模式分析与实践	杨廷 等	期刊论文	《山东通信技术》	2018-06-30	否
7	0.86%	文本倾向性分析中的评价对象抽取与消歧研究	王沛	学位论文	哈尔滨工业大学	2013-06-01	否
8	0.86%	基于汉字输入的词频统计方法研究	杨云峰	学位论文	汕头大学	2014-05-30	否
9	0.84%	Claroline学习管理系统的二次开发与应用研究	肖其伟	学位论文	华南师范大学	2011-03-01	否
10	0.52%	201610645598.9	北京邮电大学	专利论文	中文专利全文数据库	2016-08-06	否
11	0.50%	201410101751.2	北京百度网讯科技有限公司	专利论文	中文专利全文数据库	2014-03-17	否
12	0.46%	基于社交关系的职位推荐系统的架构与实现	王超	学位论文	华中科技大学	2015-05-27	否
13	0.42%	基于深度学习的虚假评论检测的研究与设计	张胜男	学位论文	重庆大学	2018-04-01	否
14	0.37%	201710044816.8	江苏大学	专利论文	中文专利全文数据库	2017-01-19	否
15	0.35%	基于大数据的情绪分析方法研究	刘爽	学位论文	曲阜师范大学	2018-06-02	否
16	0.33%	201510511588.1	北京工商大学	专利论文	中文专利全文数据库	2015-08-17	否
17	0.29%	基于主题模型的水利信息分类方案设计	诸葛庆子 等	期刊论文	《水利信息化》	2018-09-29	否
18	0.26%	基于OpenFlow的报文分类算法研究与应用	吕昭 等	期刊论文	《计算机工程与科学》	2014-05-30	否
19	0.26%	国内共享单车同质化问题研究	刘海燕	学位论文	湖南大学	2018-03-19	否
20	0.24%	基于爬虫的网络文本挖掘研究与应用	陈求兴	学位论文	上海交通大学	2016-06-01	否
21	0.22%	基于文本挖掘的用户评论分类解析系统的设计与实现	李亚松	学位论文	北京邮电大学	2014-12-30	否
22	0.22%	基于Python的简单文本情感分析	夏玉芹 等	期刊论文	《阴山学刊（自然科学版）》	2018-12-20	否
23	0.20%	基于深度学习和知识策略的蒙古文古籍识别研究	苏向东	学位论文	内蒙古大学	2016-03-01	否
24	0.16%	网络热点话题实时发现技术研究与应用	肖斌	学位论文	北京邮电大学	2013-11-29	否
25	0.16%	经济转型背景下中小企业市场营销模式创新探析	黄妩	期刊论文	《中国市场》	2018-10-11	否
26	0.15%	天津联通公司IT项目需求管理研究	许晔	学位论文	南开大学	2010-05-01	否
27	0.15%	基于Django框架的关键词排名监控系统设计	濮文强 等	期刊论文	《微型机与应用》	2017-03-24	否
28	0.13%	RE-OEM:一种半结构化生物数据的信息抽取模型	曹顺良 等	期刊论文	《计算机应用研究》	2007-12-14	否
29	0.11%	[转载]王克非教授:中国英汉平行语料库的设计与研制	戴光荣	学术网文	科学网	2013-11-12	否

30	0.11%	正则表达式	学术网文	百度百科	1900-01-01	否
31	0.11%	交通场景中的运动目标跟踪方法研究	卜倩	学位论文	东北师范大学	2010-05-01
32	0.09%	旅游景点现状分析——基于旅游网站评论数据	孙晔琦	学位论文	云南财经大学	2018-06-05

相似片段详情




1	送检文献片段	相似比：0.13%	相似文献片段
			[期刊论文]电信运营商智能客服应用模式分析与实践 杨廷 等-2018-06-30-《山东通信技术》-是否引用：否
	计算机与人类通过自然语言的方式进行通信， 意味着计算机需要理解自然语言背后的蕴含的意义 ，同时将处理后的任务以自然语言的形式表达出		和方法。自然语言处理是人工智能中最困难的问题之一。 实现人机间的自然语言通信， 意味着要使计算机既能理解自然语言文本的意义 ，也能以自然语言文本来表达给定的意图、思想等。前者称为自然语言理解，后者称为自然语
2	送检文献片段	相似比：0.09%	相似文献片段
			[期刊论文]电信运营商智能客服应用模式分析与实践 杨廷 等-2018-06-30-《山东通信技术》-是否引用：否
	此给计算机理解自然语言带来了严峻的挑战。由于 计算机难以理解自然语言的真正意图 ，导致自然语言生成更加困难。不论计算机		和方法。自然语言处理是人工智能中最困难的问题之一。 实现人机间的自然语言通信，意味着要使 计算机既能理解自然语言文本的意义 ，也能以自然语言文本来表达给定的意图、思想等。前者称为自然语言理解，后者称为自然语
3	送检文献片段	相似比：0.24%	相似文献片段
			[期刊论文]电信运营商智能客服应用模式分析与实践 杨廷 等-2018-06-30-《山东通信技术》-是否引用：否
	然语言或者计算机生成含义恰当的自然语言， 都是十分困难的任务 。基于当前学术领域与工业界的情况，恰当的、精准的自然语言处理系统，仍处于初级阶段。在一些特定的领域中， 自然语言处理系统的表现尚可 ：多种语言的翻译系统、检索系统、实施语音翻		括自然语言理解和自然语言生成两个部分。 无论实现自然语言理解还是自然语言生成， 都是十分困难的 。从现有的理论和技术看，通用的、高质量的 自然语言处理系统 仍然是较长期的努力目标，但是针对一定应用，具有相当自然语言处理能力的实用系统已经出现，有些已商品化，甚至
4	送检文献片段	相似比：0.16%	相似文献片段
			[期刊论文]经济转型背景下中小企业市场营销模式创新探析 黄毓-2018-10-11-《中国市场》-是否引用：否

义 自改革开放以来，我国经济迅猛发展， 企业已经成为推动我国经济发展的重要力量 。然而随着时代的进步，自然语言处理得到了研		，对我国的国民经济的推动发挥着积极的作用。中小型企业数量明显增加，规模不断扩大，目前中小型企业 已经成为我国经济发展的重要推动力量 。但是大多数的中小企业规模都不大，可用的资源也是非常有限，抗风险能力也比较弱，在市	
5	送检文献片段	相似比：0.11%	相似文献片段
		[学位论文]文本倾向性分析中的评价对象抽取与消歧研究 王沛-2013-06-01-哈尔滨工业大学-是否引用：否	
经济发展的重要力量。然而随着时代的进步， 自然语言处理得到了研究领域的广泛重视 。随着一系列技术和方法的提出，自然语言处理		述 条件随机场模型（CRF models）最早由J. Lafferty等人提出，之后在 自然语言处理领域得到广泛的应用 。 条件随机场模型是基于最大熵模型发展而来的。和最大熵一样，它也是一个判别模型。	
6	送检文献片段	相似比：2.07%	相似文献片段
		[专利论文]201510919129.7 北京邮电大学-2015-12-09-中文专利全文数据库-是否引用：否	
，但是达到理想的结果不是一件容易的事情，主要 存在两个方面的问题 ：一方面，迄今为止的语法都限于分析一个孤立的句子，上下文关系和谈话环境对本句的约束和影响还缺乏系统的研究，因此分析歧义、词语省略、代词所指、同一句话在不同场合或由不同的人说出来所具有的不同含义等问题，尚无明确规律可循，需要加强语用学的研究才能逐步解决。另一方面，人理解一个句子不是单凭语法，还运用了大量的有关知识，包括生活知识和专门知识，这些知识无法全部贮存在计算机里。因此一个书面理解系统只能建立在有限的词汇、句型和特定的主题范围内；计算机的贮存量和运转速度大大提高之后，才有可能适当扩大范围。 1.2 研究内容与成果 21世纪信			
7	送检文献片段	相似比：0.13%	相似文献片段
		[学位论文]国内共享单车同质化问题研究 刘海燕-2018-03-19-湖南大学-是否引用：否	

成为当下商业活动的重要组成。大数据技术、云计算及物联网等信息通信技术的出现，难以满足海量数据的价值挖掘和内在分析。面		与此同时，在移动互联网的促成之下，“互联网+交通”利用移动互联网、大数据开发、云计算及物联网等新兴的信息通信技术，将传统的交通运输业与互联网新兴产业完美的进行融合，形成了“线上资源分配合理，线下	
8	送检文献片段	相似比：0.74%	相似文献片段
			[学位论文]感知与现实人际网络：经典文学作品分析 张荣杰-2018-06-01-闽南师范大学-是否引用：否
	的数据，必须深度分析才能够获取所需信息，以构架数据分析平台及系统快速完成数据文学家主观虚构的文学世界，来源于现实，反映了现实世界，因此每个时期的作品或多或少都有着那个时期特有的特点。在文学作品中，人物的优缺点会被放大，这是作者为了将对现实世界的想法赋予到作品中的人物身上，通过对文学作品进行分析不仅能够了解文学家		文学家主观虚构的文学世界，来源于现实，反映了现实世界，因此每个时期的作品或多或少都有着那个时期特有的特点。在文学作品中，人物的优缺点会被放大，这是作者为了将对现实世界的想法赋予到作品中的人物身上。而反映了文学家主观建构的感知人际网络与客观现实人际网络，是否存在一定的联系？能否
9	送检文献片段	相似比：0.15%	相似文献片段
			[学位论文]天津联通公司IT项目需求管理研究 许晔-2010-05-01-南开大学-是否引用：否
	论文组织结构 本文的组织结构如下： 第一章，绪论，主要包括课题的研究背景与意义、研究内容与成果以及论文的组织结构。 第		分析、需求实施与验证、需求后评估等几个方面。具体来说，本论文的主要内容如下： 第一章，绪论：包括本课题的研究背景及意义，本课题的研究内容及框架，相关理论回顾等内容，其中相关理论回顾部分包括项目管理理论
10	送检文献片段	相似比：0.11%	相似文献片段
			[学位论文]交通场景中的运动目标跟踪方法研究 卜倩-2010-05-01-东北师范大学-是否引用：否
	，绪论，主要包括课题的研究背景与意义、研究内容与成果以及论文的组织结构。 第二章，背景介绍，主要包括神经网络		各章内容安排如下： 第一章引言。介绍了本论文的选题背景和意义、国内外研究现状、本文研究内容及成果和论文组织结构。 第二章常用目标检测，跟踪特征及跟踪方法，相关理论基础。 第三章各种矩特征的
11	送检文献片段	相似比：0.73%	相似文献片段





<div> <div></div> <div></div> </div>		[期刊论文]基于自然语言处理的能源领域知识图谱 赵京胜 等-2018-05-02-《信息技术与信息化》-是否引用：否	
4.5分类算法提取文本中的关键字[5]。2004 年，Mihalcea 和 Tarau 利用特征词在共现窗格中出现的构造词与词的图模型，并结合 Google 公司提出的 PageRank 算法，提出 TextRank 关键词提取算法[6]。 2012 年，Google 率先提出		ion ” 系统 [8]。而图模型是一种完全无监督技术，也是近期关键词提取的热点。2004 年，Mihalcea 和 Tarau 利用特征词在共现窗格中出现的构造词与词的图模型，并结合 Google 公司提出的 PageRank 算法，提出 TextRank 关键词提取算法 [9]。夏天将词向量聚类加权与 TextRank 算法相结合，提高了单篇文档关键词的准确	
12	送检文献片段	相似比：2.21%	相似文献片段
<div> <div></div> <div></div> </div>		[期刊论文]基于自然语言处理的能源领域知识图谱 赵京胜 等-2018-05-02-《信息技术与信息化》-是否引用：否	
xtRank 关键词提取算法[6]。 2012 年，Google 率先提出知识图谱（Knowledge Graph）的概念，将知识图谱应用到搜索引擎中。运用自然语言处理提取实体和实体间的关系，并以此分别作为图谱的节点和边，构建知识图谱。知识图谱的关键技术主要有知识抽取、知识表示、知识融合以及知识推理技术等。文献 [7] 运用 KNN 算法与条件随机场模型，对 Twitter 文本数据中实体进行抽取。陈立玮等人 [8] 提出了一种协同训练方法，通过向传统模型中引入 N-Gram 特征进行协同训练，改善了弱监督关系抽取模型的效果。Li 等人根据给定的少量种子链接，利用概念标注方法，发现新的链接，最终实现了知识的扩充 [9]。 2.2 Jieba Jieba[1		进行了总结分析，并预测了未来自动关键词抽取的发展和将要面临的挑战 [11]。 2012 年，Google 率先提出知识图谱（Knowledge Graph）的概念，将知识图谱应用到搜索引擎中。运用自然语言处理提取实体和实体间的关系，并以此分别作为图谱的节点和边，构建知识图谱。知识图谱的关键技术主要有知识抽取、知识表示、知识融合以及知识推理技术等。文献 [12] 运用 KNN 算法与条件随机场模型，对 Twitter 文本数据中实体进行抽取。 陈立玮等人 [13] 提出了一种协同训练方法，通过向传统模型中引入 N-Gram 特征进行协同训练，改善了弱监督关系抽取模型的效果。Li J 等人根据给定的少量种子链接，利用概念标注方法，发现新的链接，最终实现了知识的扩充 [14]。但是，知识图谱大部分工作是由人工处理的，与自然语言结合的少，而能源领域的知识图谱	
13	送检文献片段	相似比：0.35%	相似文献片段
<div> <div></div> <div></div> </div>		[学位论文]基于大数据的情绪分析方法研究 刘爽-2018-06-02-曲阜师范大学-是否引用：否	

<p>]是一个广泛使用的开源自然语言处理框架，支持三种分词模式：</p> <p>(1) 精确模式，试图将句子最精确地切开，适合文本分析；</p> <p>(2) 全模式，把句子中所有的可以成词的词语都扫描出来，</p>	<p>的分词器，但是它使用方便，接口清晰，容易上手使用，准确性能够达到常规使用的水平。并且支持三种分词模式：</p> <p>(1) 精确模式，试图将句子最精确地切开，适合文本分析；</p> <p>(2) 全模式，把句子中所有的可以成词的词语都扫描出来，速度非常快，但是不能</p> <p>第3章 抑郁情</p>
--	---

14	送检文献片段	相似比：0.52%	相似文献片段
			<p>[专利论文]201610645598.9 北京邮电大学-2016-08-06-中文专利全文数据库-是否引用：否</p>
	<p>精确地切开，适合文本分析；(2) 全模式，把句子中所有的可以成词的词语都扫描出来，速度非常快，但是不能解决歧义；(3) 搜索引擎模式，在精确模式的基础上，对长词再次切分，提高召回率，适合用于搜索引擎分词。Jieba分词包在这里运用到了数据结构里</p>		
15	送检文献片段	相似比：0.50%	相似文献片段
			<p>[专利论文]201410101751.2 北京百度网讯科技有限公司-2014-03-17-中文专利全文数据库-是否引用：否</p>
	<p>组合的问题。</p> <p>图2.1 Trie树</p> <p>Jieba利用TF-IDF (Term Frequency-Inverse Document Frequency, 词频-逆文件频率) 算法进行关键词的提取，该算法是一种统计方法，用以评估一个词语对</p>		
16	送检文献片段	相似比：0.29%	相似文献片段
			<p>[学位论文]文本倾向性分析中的评价对象抽取与消歧研究 王沛-2013-06-01-哈尔滨工业大学-是否引用：否</p>
	<p>词频-逆文件频率) 算法进行关键词的提取，该算法是一种统计方法，用以评估一个词语对于一个文件集或一个语料库中的一份文件的重要程度，其原理可概括为：一个词语在一篇文章中出现</p>		<p>ocument frequency) 是一种信息检索和文本挖掘领域的常用加权技术。它是一种统计方法，用以评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度。</p> <p>这个方法认为字词的重要性随着它在一个文件中出现的次数成正比增加，但同时会随着</p>
17	送检文献片段	相似比：0.07%	相似文献片段





<div> <div></div> <div></div> </div>		[学位论文]文本倾向性分析中的评价对象抽取与消歧研究 王沛-2013-06-01-哈尔滨工业大学-是否引用：否	
中的一份文件的重要程度，其原理可概括为：一个词语在 一篇文章中出现 次数越多，同时 在所有文档中出现 次数越少，越能够代表		个文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率增加而下降。 tf-idf主要思想是如果某个词在 一篇文章中出现 的频率高，即tf较高，同时它在其他文章中 出现频率 较低，即df较低，反过来idf较高，则认为	
18	送检文献片段	相似比：0.10%	相似文献片段
<div> <div></div> <div></div> </div>		[学位论文]基于深度学习和知识策略的蒙古文古籍识别研究 苏向东-2016-03-01-内蒙古大学-是否引用：否	
reNLP包含许多的斯坦福的nlp工具， 包括：词性标注器、命名实体的识别器 、解析器（句子与语法结构）、指代消解器（就		ollobertandWeston将卷积深度信念网应用于自然语言处理的多种任务， 包括词性标注 ，切分， 命名实体识别 ，语义对象识别【1011。ThomasDeselaers等将类似的模型应用于机器翻	
19	送检文献片段	相似比：0.19%	相似文献片段
<div> <div></div> <div></div> </div>		[学位论文]文本倾向性分析中的评价对象抽取与消歧研究 王沛-2013-06-01-哈尔滨工业大学-是否引用：否	
实体的识别器、解析器（句子与语法结构）、 指代消解器 （就是在 篇章中确定代词指向哪个名词短语的问题 ）、情感分析器、引导模式学习器、开放信息提取		量的实体。 4.3 评价对象消歧 评价对象消歧问题有点类似指代消解问题。 指代消解是在篇章中确定代词指向哪个名词短语的问题。 在自然语言处理领域，指代消解问题目前仍然是一个非常巨大的挑战。在一些语法规范的文	
20	送检文献片段	相似比：0.13%	相似文献片段
<div> <div></div> <div></div> </div>		[期刊论文]电信运营商智能客服应用模式分析与实践 杨廷 等-2018-06-30-《山东通信技术》-是否引用：否	
计算机与人类通过自然语言的方式进行通信， 意味着计算机需要理解自然语言 背后的 蕴含的意义 ，同时将处理后的任务以自然语言的形式表达出		和方法。自然语言处理是人工智能中最困难的问题之一。 实现人机间的自然语言通信， 意味着要使计算机既能理解自然语言文本的意义 ，也能以自然语言文本来表达给定的意图、思想等。前者称为自然语言理解，后者称为自然语	
21	送检文献片段	相似比：0.09%	相似文献片段

<div> <div></div> <div></div> </div>		<p>[期刊论文]电信运营商智能客服应用模式分析与实践 杨廷 等-2018-06-30-《山东通信技术》-是否引用：否</p>	
<p>此给计算机理解自然语言带来了严峻的挑战。由于计算机难以理解自然语言的真正意图，导致自然语言生成更加困难。不论计算机</p>		<p>和方法。自然语言处理是人工智能中最困难的问题之一。</p> <p>实现人机间的自然语言通信，意味着要使计算机既能理解自然语言文本的意义，也能以自然语言文本来表达给定的意图、思想等。前者称为自然语言理解，后者称为自然语</p>	
22	送检文献片段	相似比：0.24%	相似文献片段
<div> <div></div> <div></div> </div>		<p>[期刊论文]电信运营商智能客服应用模式分析与实践 杨廷 等-2018-06-30-《山东通信技术》-是否引用：否</p>	
<p>然语言或者计算机生成含义恰当的自然语言，都是十分困难的任务。基于当前学术领域与工业界的情况，恰当的、精准的自然语言处理系统，仍处于初级阶段。在一些特定的领域中，自然语言处理系统的表现尚可：多种语言的翻译系统、检索系统、实施语音翻</p>		<p>括自然语言理解和自然语言生成两个部分。</p> <p>无论实现自然语言理解还是自然语言生成，都是十分困难的。从现有的理论和技术看，通用的、高质量的自然语言处理系统仍然是较长期的努力目标，但是针对一定应用，具有相当自然语言处理能力的实用系统已经出现，有些已商品化，甚至</p>	
23	送检文献片段	相似比：0.86%	相似文献片段
<div> <div></div> <div></div> </div>		<p>[学位论文]基于汉字输入的词频统计方法研究 杨云峰-2014-05-30-汕头大学-是否引用：否</p>	
<p>研究的内容。语料是构成语料库的基本单元。所以，人们简单地用文本作为替代，并把文本中的上下文关系作为现实世界中语言的上下文关系的替代品。我们把一个文本集合称为语料库（Corpus），当有几个这样的文本集合的时候，我们称之为语料库集合（Corpora）。然后对语料进行预处理，在一个完整的中文自</p>		<p>基本概念</p> <p>语料：通常，在统计自然语言处理中实际上不可能观测到大规模的语言实例。所以，人们简单地用文本作为替代，并把文本中的上下文关系作为现实世界中语言的上下文关系的替代品。我们把一个文本集合称为语料库（Corpus），当有几个这样的文本集合的时候，我们称之为语料库集合（Corpora） [7]。在实际应用中，词频统计都是在某一范围的语料库里进行的。倘若对所有的语料进行词频统</p>	
24	送检文献片段	相似比：0.84%	相似文献片段
<div> <div></div> <div></div> </div>		<p>[学位论文]Claroline学习管理系统的二次开发与应用研究 肖其伟-2011-03-01-华南师范大学-是否引用：否</p>	

图3.1 UltraEdit工具界面			
UltraEdit 是一套功能强大的文本编辑器，可以编辑文本、十六进制、ASCII 码，完全可以取代记事本（如果电脑配置足够强大），内建英文单字检查、C++ 及 VB 指令突显，可同时编辑多个文件，而且即使开启很大的文件速度也不会慢。			
正则表达式，又称规则表达式，是计算机			
25	送检文献片段	相似比：0.37%	相似文献片段
			[专利论文]201710044816.8 江苏大学-2017-01-19-中文专利全文数据库-是否引用：否
而且即使开启很大的文件速度也不会慢。 正则表达式，又称规则表达式，是计算机科学的一个概念。 正则表达式是一种字符串的匹配模式，描述的是某一类字符串的共同特征，通常被用来检索、替换那些符合某个模式（规则）的文本。所谓			
26	送检文献片段	相似比：0.15%	相似文献片段
			[期刊论文]基于Django框架的关键词排名监控系统设计 濮文强 等-2017-03-24-《微型机与应用》-是否引用：否
某一类字符串的共同特征，通常被用来检索、替换那些符合某个模式（规则）的文本。所谓模式，就是模板样式或模具样式。正如符		数据抓取的正则表达式正则表达式是计算机科学的一个概念。正则表通常被用来检索、替换那些符合某个模式（规则）的文本。对于信息的抓取，原理是对相应的 HTML 源代码采用正则表达式去检索所需信息。	
27	送检文献片段	相似比：0.10%	相似文献片段
			[期刊论文]基于OpenFlow的报文分类算法研究与实现 吕昭 等-2014-05-30-《计算机工程与科学》-是否引用：否
高度抽象性词语来描述事物的共同特征一样，一个正则表达式正是某一类字符串的高度抽象，用来描述这类字符串的共同特征。也就是说，		ash函数的运算是并行的，时间开销为1。 5 基于有限自动机的通配报文分类 算法正则表达式是对字符串操作的一种逻辑公式，是用事先定义好的一些特定字符及这些特定字符的组合，组成一个规则字符串，用这个规则	
28	送检文献片段	相似比：0.11%	相似文献片段
			[学术网文]正则表达式 -1900-01-01-百度百科 ：http://baike.baidu.com/view/94238.html-是否引用：否

用来描述这类字符串的共同特征。也就是说，一个正则表达式代表了某类字符串的一个集合，而正则表达式相当于对该字符串集合的特征性		，组成一个“规则字符串”，这个“规则字符串”用来表达对字符串的一种过滤逻辑。	
		给定一个正则表达式和另一个字符串，我们可以达到如下的目的：	
		1. 给定的字符串是否符合正则表达式的过滤逻辑（称作	
29	送检文献片段	相似比：0.16%	相似文献片段
	<div></div>		[期刊论文]基于OpenFlow的报文分类算法研究与实现 吕昭 等-2014-05-30-《计算机工程与科学》-是否引用：否
	达式相当于对该字符串集合的特征性质描述。正则表达式还可看作是对字符串操作的一种逻辑公式，其构造方法和创建数学表达式的方法差不多，		ash函数的运算是并行的，时间开销为1。 5 基于有限自动机的通配报文分类 算法正则表达式是对字符串操作的一种逻辑公式，是用事先定义好的一些特定字符及这些特定字符的组合，组成一个规则字符串，用这个规则
30	送检文献片段	相似比：0.13%	相似文献片段
	<div></div>		[期刊论文]RE-OEM:一种半结构化生物数据的信息抽取模型 曹顺良 等-2007-12-14-《计算机应用研究》-是否引用：否
	式还可看作是对字符串操作的一种逻辑公式，其构造方法和创建数学表达式的方法差不多，也就是用普通字符（如字母a到z、数字0到		具体的字符，如表达式/[A—z]/将会与A—z范围内任何一个大写字母相匹配。 构造正则表达式的方法和创建数学表达式的方法一样，也就是用多种元字符与操作符将小的表达式结合在一起来创建更大的表达式。正则表达式的
31	送检文献片段	相似比：0.24%	相似文献片段
	<div></div>		[学位论文]基于爬虫的网络文本挖掘研究与应用 陈求兴-2016-06-01-上海交通大学-是否引用：否
	差不多，也就是用普通字符（如字母a到z、数字0到9等）和事先定义好的一些特定字符（专业术语称之为元字符），以及这些字符的组合，组成一个特定的规则字符串。而所谓特定的规则，即是正则；因此特定的规		则表达式是对字符串操作和模式匹配的一种逻辑表达公式，实现对字符串的一种过滤逻辑。就是用事先约定好的一些特定字符、及这些特定字符的组合，组成一个“规则字符串”，例如，“\d”匹配0到9数字，“\w”匹配单词字符。 (2) XPath XP

32	送检文献片段	相似比：0.11%	相似文献片段
	<div></div>		[学术网网][转载]王克非教授:中国英汉平行语料库的设计与研制 戴光荣-2013-11-12-科学网: http://blog.sciencenet.cn/blog-331736-741028.html -是否引用: 否
	著作进行初步整理后,得到相对规范的本文,然后对规范的文本进行分词和词性标注。 3.5.1 Jieba框架对文本进行		汉双语时文构成,采样年代为2000-2001年,共计54万词;该库实现句级对齐,并对英汉语文本都进行了分词和词性标注.台湾辅仁大学初步建立了范本财经英日汉平行语料库,收集语料约10万句对.最近,香港
33	送检文献片段	相似比：0.42%	相似文献片段
	<div></div>		[学位论文]基于隐马尔科夫模型的文本分类器的设计与实现 段康康-2016-06-01-北京交通大学-是否引用: 否
	a自带了一个叫做dict.txt的词典,里面有2万多条词,包含了词条出现的次数(这个次数是于作者自己基于人民日报语料等资源训练得出来的)和词性。该框架具有如下特征。 1. 基于前缀词		要生成DAG就必须有语料库的辅助了,结巴分词自带了一个叫做dict. 仅t的词典,里面有2万多条词,包含了词条出现的次数(这个次数是于作者自己基于人民日报语料等资源训练得出来的)和词性。语料库的有3列,第一列是词,第二列是词频第三列是词性。在程序中初始化语料库的动作
34	送检文献片段	相似比：0.33%	相似文献片段
	<div></div>		[专利论文]201510511588.1 北京工商大学-2015-08-17-中文专利全文数据库-是否引用: 否
	出来的)和词性。该框架具有如下特征。 1. 基于前缀词典实现高效的词图扫描,生成句子中汉字所有可能成词情况所构成的有向无环图(DAG),其主要过程分为两步:(1)根据dict.		
35	送检文献片段	相似比：0.35%	相似文献片段
	<div></div>		[学位论文]基于隐马尔科夫模型的文本分类器的设计与实现 段康康-2016-06-01-北京交通大学-是否引用: 否
	ieba自带一个dict.txt的词典,里面有2万多条词,包含了词条出现的次数和词性(作者基于人民日报语料等资源训练得出来)。trie树结构的词图扫描,说的就是把这		要生成DAG就必须有语料库的辅助了,结巴分词自带了一个叫做dict. 仅t的词典,里面有2万多条词,包含了词条出现的次数(这个次数是于作者自己基于人民日报语料等资源训练得出来的)和词性。语料库的有3列,第一列是词,第二列是词频第三列是词性。在程序中初始化语料库的动作
36	送检文献片段	相似比：0.39%	相似文献片段
	<div></div>		[学位论文]基于隐马尔科夫模型的文本分类器的设计与实现 段康康-2016-06-01-北京交通大学-是否引用: 否

查找速度快的优势)；(2)对待分词句子，根据dict.txt生成的trie树，生成DAG，通俗的讲，就是将句子根据给定的词典进行查词典操作，生成所有可能的句子切分。jieba在DAG中记录的是句子中某个词		的语料库数据结构用marshal序列化，然后存放在系统的临时目录。对待分词句子，根据dict.txt生成的trie树，生成DAG，通俗的说，就是对待分词句子根据给定的词典进行查词典操作，生成几种可能的句子。切分采用动态规划查找最大概率路径，找出基于词频的最大切分组合，最后使用yield	
37	送检文献片段	相似比：0.40%	相似文献片段
		[学位论文]适用于高校学院级办公系统的推荐技术的研究 蒋莎-2013-06-30-电子科技大学-是否引用：否	
行查词典操作，生成所有可能的句子切分。jieba在DAG中记录的是句子中某个词的开始位置，从0到n-1(n为句子的长度)，每个开始位置作为字典的键，value是个list，其中保存了可能的词语的结束位置(通过查		此有向无环图未必能转化成树，但任何有向树均为有向无环图)，由此形成切分词图。 形成有向无环图的具体做法是记录句子中某个词的开始位置，从0到N-1(N为句子的长度)，每个开始位置作为字典的键，值是一个list，其中保存了可能的词语的结束位置。通过查字典得到词，开始位置加上词语的长度就得到结	
38	送检文献片段	相似比：0.16%	相似文献片段
		[学位论文]网络热点话题实时发现技术研究 肖斌-2013-11-29-北京邮电大学-是否引用：否	
作为字典的键，value是个list，其中保存了可能的词语的结束位置(通过查字典得到词，开始位置+词语的长度得到结束位置)。			
39	送检文献片段	相似比：0.13%	相似文献片段
		[学位论文]适用于高校学院级办公系统的推荐技术的研究 蒋莎-2013-06-30-电子科技大学-是否引用：否	
可能的词语的结束位置(通过查字典得到词，开始位置+词语的长度得到结束位置)。 2. 动态规划查找最大概率路径，找		为字典的键，值是一个list，其中保存了可能的词语的结束位置。通过查字典得到词，开始位置加上词语的长度就得到结束位置。例如：{0: [1, 2, 3]} 这样一个简单的有向无环图，就是表示0位置开始，在1	
40	送检文献片段	相似比：0.22%	相似文献片段
		[期刊论文]基于Python的简单文本情感分析 夏玉芹 等-2018-12-20-《阴山学刊(自然科学版)》-是否引用：否	
开始位置+词语的长度得到结束位置)。 2. 动态规划查找最大概率路径，找出基于词频的最大切分组合。首先查找待分词句子中已经切分好的词语(全		现高效的词图扫描，生成句子中汉字所有可能成词情况所构成的有向无环图(DAG)，动态规划查找最大概率路径，找出基于词频的最大切分组合。jie-ba 可以添加自定义词库或者删除“无效词”(stop-words)。“	

41	送检文献片段	相似比：0.14%	相似文献片段
	<div></div>		[学位论文]适用于高校学院级办公系统的推荐技术的研究 蒋莎-2013-06-30-电子科技大学-是否引用：否
	概率路径，找出基于词频的最大切分组合。首先 查找待分词句子中已经切分好的词语 （全模式下的分词list），得出查找该词语出现的频率（次数/总数），		满足用动态规划求解所要求的最优子结构性质和无后效性。 具体的实现步骤如下： 1、先 查找待分词句子中已经切分好的词语 ，并查找该词语出现的频率；如果没有这个词，就把词典中出现频率最小的那个词语的频率作
42	送检文献片段	相似比：0.22%	相似文献片段
	<div></div>		[学位论文]基于文本挖掘的用户评论分类解析系统的设计与实现 李亚松-2014-12-30-北京邮电大学-是否引用：否
	切分好的词语（全模式下的分词list），得出 查找该词语出现的频率（次数/总数），如果没有该词（基于词典一般都是有的），就把词典中出现频率最小的那个词语的频率作		
43	送检文献片段	相似比：0.22%	相似文献片段
	<div></div>		[学位论文]适用于高校学院级办公系统的推荐技术的研究 蒋莎-2013-06-30-电子科技大学-是否引用：否
	，如果没有该词（基于词典一般都是有的）， 就把词典中出现频率最小的那个词语的频率作为该词的频率 。然后根据动态规划查找最大概率路径的方法，		1、先查找待分词句子中已经切分好的词语，并查找该词语出现的频率；如果没有这个词， 就把词典中出现频率最小的那个词语的频率作为该词的频率 。 2、根据动态规划查找最大概率路径的方法，对句子从左往右计算最大概率。在实际查
44	送检文献片段	相似比：0.29%	相似文献片段
	<div></div>		[期刊论文]基于主题模型的水利信息分类方案设计 诸葛庆子 等-2018-09-29-《水利信息化》-是否引用：否
	频率最小的那个词语的频率作为该词的频率。然后 根据动态规划查找最大概率路径的方法，对句子从右往左反向计算最大概率 （这里反向是因为汉语句子的重心经常落在后面（右边），因为通常情况下形容词太多，后面的才是		值为 value，里面存放了可能的词语结束位置，并将这些成词情况构成有向无环图， 根据动态规划查找最大概率路径的方法 ，水利文本处理 对句子采用从右往左的方式计算反向计算最大概率 ，由于汉语句子常常将重心落在后面，因此反向计算比正向计算的正确率更高。最后由最大概
45	送检文献片段	相似比：0.09%	相似文献片段
	<div></div>		[学位论文]旅游景点现状分析——基于旅游网站评论数据 孙晔琦-2018-06-05-云南财经大学-是否引用：否

因为汉语句子的重心经常落在后面（右边），因为 通常情况下形容词太多 ，后面的才是主干。因此，从右往左计算，正确率要高于从左		（某些资料中可能是从左往右计算，反向是由于，由汉语构成的句子的重心常常落在后面，即落在右边，而且 通常情况下，形容词太多 ，句子后面的才为主干，所以，反向计算，正确率要高于正向计算，类似于逆向的最大匹配）	
46	送检文献片段	相似比：0.32%	相似文献片段
	<div></div>		[学位论文]基于社交关系的职位推荐系统的架构与实现 王超-2015-05-27-华中科技大学-是否引用：否
	<p>）*Max（P（倒数第一个词）依次类推，最后得到最大概率路径，得到最大概率的切分组合。</p> <p>3. 采用了基于汉字成词能力的HMM模型对未登录的词进行分词。利用HMM模型将中文词汇按照BEMS四个</p>		<p>小的词语的词频作为该词的词频，然后根据最大概率路径方法，对句子反向计算最大概率，最后得到最大概率路径，得到最大概率的切分组合</p> <p>c) 对于未登录词，采用了基于汉字成词能力的HMM模型，使用了Viterbi算法[39]。</p> <p>基于作者采用的HMM模型了，中文词汇</p>
47	送检文献片段	相似比：0.35%	相似文献片段
	<div></div>		[学位论文]基于深度学习的虚假评论检测的研究与设计 张胜男-2018-04-01-重庆大学-是否引用：否
	力的HMM模型对未登录的词进行分词。利用 HMM模型将中文词汇按照BEMS四个状态来标记 ，B是开始begin位置，E是end结束位置，M是middle中间位置，S是single单独成词的位置。jieba采用（B,E,M,S）这四种状		组合。3）对于DAG中不记录在词典dict.txt中词，使用HMM模型进行分词。 HMM模型是对中文词汇按照BEMS四个状态进行标记，B为开始位置，E为结束位置，M为中间位置，S为单独成词的位置。 即（B,E,M,S）四种状态可对任何中文词语进行标注。如“虚假评论”，则可标注
48	送检文献片段	相似比：0.14%	相似文献片段
	<div></div>		[学位论文]基于社交关系的职位推荐系统的架构与实现 王超-2015-05-27-华中科技大学-是否引用：否
	置，S是single单独成词的位置。jieba采用（B,E,M,S） 这四种状态来标记中文词语 ，比如北京可以标注为BE，即北/B京/E		le，单独成词的位置，没有前，也没有后。也就是说，他采用了状态为（B,E,M,S） 这四种状态来标记中文词语 ，比如武汉可以标注为BE，即武/B汉/E，表示武是开始位置，汉是结束
49	送检文献片段	相似比：0.31%	相似文献片段
	<div></div>		[学位论文]适用于高校学院级办公系统的推荐技术的研究 蒋莎-2013-06-30-电子科技大学-是否引用：否

, E, M, S) 这四种状态来标记中文词语，比如北京可以标注为 BE，即北/B京/E，表示北是开始位置，京是结束位置，中华民族可以标注为 BMME，就是开始，中间，中间，结束。 3.5.2		le，单独成词的位置。我们采用状态为 (B, E, M, S) 这四种状态来标记中文词语。例如：“研究”可以标注为 BE，即研/B究/E，表示“研”是开始位置，“究”是结束位置，“路由算法”可以标注为 BMME，就是开始、中间、中间、结束。 基于 HMM 模型，本文对大量字典词语进行了分	
50	送检文献片段	相似比：0.07%	相似文献片段
	<div></div>		[学位论文]基于深度学习的虚假评论检测的研究与设计 张胜男-2018-04-01-重庆大学-是否引用：否
	中华民族可以标注为 BMME，就是开始，中间，中间，结束。 3.5.2 Jieba 框架对文本进行		M, S) 四种状态可对任何中文词语进行标注。如“虚假评论”，则可标注为 BMME，即开始，中间，中间，结束。语料训练的同时要统计位置到单字的发射概率、位置转换概率以及词语到某种状态开头的概
51	送检文献片段	相似比：0.10%	相似文献片段
	<div></div>		[学位论文]基于深度学习和知识策略的蒙古文古籍识别研究 苏向东-2016-03-01-内蒙古大学-是否引用：否
	reNLP 包含许多的斯坦福的 nlp 工具，包括：词性标注器、命名实体的识别器、解析器（句子与语法结构）、指代消解器（就		ollobertandWeston 将卷积深度信念网应用于自然语言处理的多种任务，包括词性标注，切分，命名实体识别，语义对象识别【1011。ThomasDeselaers 等将类似的模型应用于机器翻
52	送检文献片段	相似比：0.19%	相似文献片段
	<div></div>		[学位论文]文本倾向性分析中的评价对象抽取与消歧研究 王沛-2013-06-01-哈尔滨工业大学-是否引用：否
	实体的识别器、解析器（句子与语法结构）、指代消解器（就是在篇章中确定代词指向哪个名词短语的问题）、情感分析器、引导模式学习器、开放信		量的实体。 4.3 评价对象消歧 评价对象消歧问题有点类似指代消解问题。 指代消解是在篇章中确定代词指向哪个名词短语的问题。在自然语言处理领域，指代消解问题目前仍然是一个非常巨大的挑战。在一些语法规范的文
53	送检文献片段	相似比：0.13%	相似文献片段
	<div></div>		[学位论文]国内共享单车同质化问题研究 刘海燕-2018-03-19-湖南大学-是否引用：否

成为当下商业活动的重要组成。大数据技术、云计算及物联网等信息通信技术的出现，难以满足海量数据的价值挖掘和内在分析。面		。与此同时，在移动互联网的促成之下，“互联网+交通”利用移动互联网、大数据开发、云计算及物联网等新兴的信息通信技术，将传统的交通运输业与互联网新兴产业完美的进行融合，形成了“线上资源分配合理，线下	
54	送检文献片段	相似比：0.74%	相似文献片段
	<div></div>		[学位论文]感知与现实人际关系网络：经典文学作品分析 张荣杰-2018-06-01-闽南师范大学-是否引用：否
	的数据，必须深度分析才能够获取所需信息，以构架数据分析平台及系统快速完成数据文学家主观虚构的文学世界，来源于现实，反映了现实世界，因此每个时期的作品或多或少都有着那个时期特有的特点。在文学作品中，人物的优缺点会被放大，这是作者为了将对现实世界的想法赋予到作品中的人物身上，通过对文学作品进行分析不仅能够了解文学家		文学家主观虚构的文学世界，来源于现实，反映了现实世界，因此每个时期的作品或多或少都有着那个时期特有的特点。在文学作品中，人物的优缺点会被放大，这是作者为了将对现实世界的想法赋予到作品中的人物身上。而反映了文学家主观建构的感知人际网络与客观现实人际网络，是否存在一定的联系？能否

- 说明：
1. 送检文献总字数=送检文献的总字符数，包含汉字、非中文字符、标点符号、阿拉伯数字（不计入空格）
 2. 总相似比=送检论文与检测范围全部数据相似部分的字数/检测总字符数
 3. 参考文献相似比=送检论文与其参考文献相似部分的字数/检测总字符数
 4. 辅助排除参考文献相似比=总相似比-参考文献相似比
 5. 可能引用本人已发表论文相似比=可能抄袭本人已发表文献的字数/检测总字符数
 6. 辅助排除本人已发表论文相似比=总相似比-可能引用本人已发表论文相似比
 7. “单篇文献最大相似比”：送检文献与某一文献的相似比高于全部其他文献
 8. “是否引用”：某一相似文献是否被送检文献列为其参考文献