

请参阅此出版物的讨论、统计数据和作者简介：<https://www.researchgate.net/publication/370144276>

利用企业资源规划实施文本分析

国际模拟期刊文章:系统,科学与技术 · 2023 年 4 月  
DOI:10.5013/IJSSST.a.24.01.05

引文  
53

阅读  
834

1 作者:



苏吉特·辛格·巴瓦

5 篇出版物 190次引用

查看简介

# 通过企业资源规划实施文本分析

苏吉特·辛格·巴瓦

解决方案架构师 美国印第安纳  
纳州卡梅尔 Surjit.sb@gmail.com

摘要 -组织中非结构化数据量的不断增加导致人们采用文本分析来提取有价值的见解。在本文中,我们讨论通过企业资源规划 (ERP) 系统实施文本分析以改进决策流程。我们提供克服这些挑战的建议,并为组织在 ERP 系统中实施文本分析时提供遵循的路线图。

关键词 - 智能企业、ERP、文本分析、机器学习。

## 一、企业资源规划 (ERP)中的文本分析简介

企业资源规划 (ERP)[1] 被组织用来整合业务流程、管理资源和简化运营。它使组织能够管理和优化其业务功能,例如财务、库存、人力资源、客户关系管理等。

电子邮件以及企业每天生成的其他文本数据,越来越需要工具来帮助从这些数据中提取见解  
这个数据。这就是文本分析的用武之地。

文本分析[2]是一种使用自然语言处理 (NLP)和机器学习算法从非结构化文本数据中分析和提取有价值信息的技术。它可以帮助组织了解客户需求和偏好、识别新兴趋势、改进产品和服务等。

随着文档、客户反馈、社交媒体评论等非结构化数据量的不断增加,

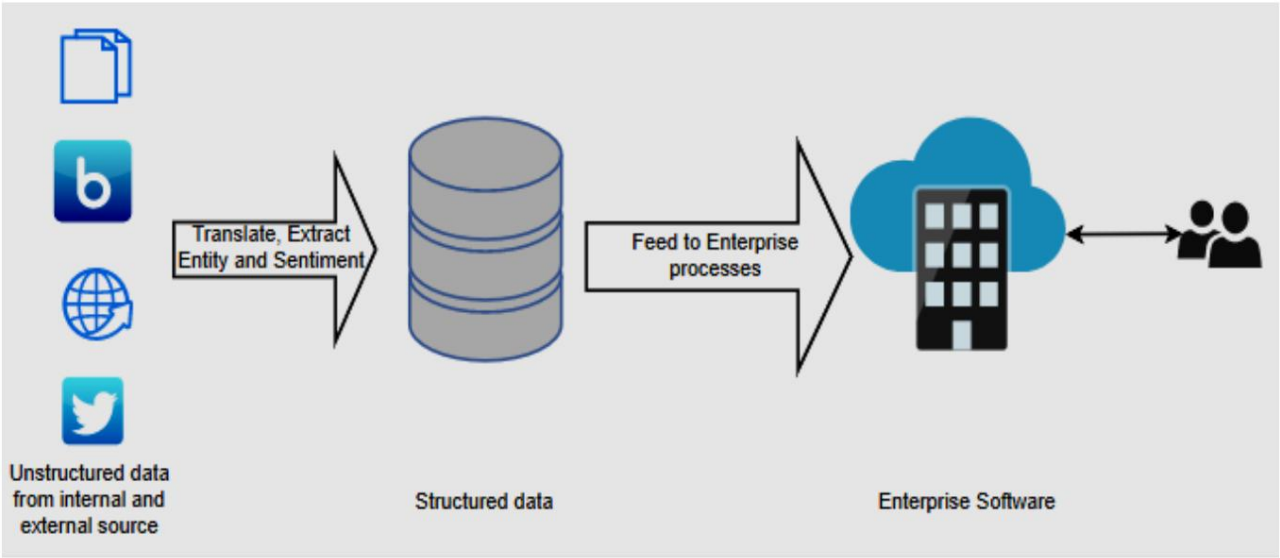


图 1. 检索非结构化数据

如图 1 所示,将文本分析与 ERP 系统集成可以为组织提供业务运营的全面视图,包括客户反馈、供应商绩效和员工情绪。它可以帮助组织做出明智的决策,提高客户满意度并推动业务增长。

在本文中,我们将探讨文本分析在 ERP 中的作用、其主要挑战、成功实施的策略、在 ERP 中使用文本分析的好处、实际用例以及 ERP 中文本分析的未来。

SURJIT SINGH BAWA:利用企业资源规划实施文本分析

二. 文本分析在 ERP 中的作用

如图 2 所示,借助企业资源规划 (ERP) 的文本分析,组织可以利用非结构化文本数据来深入了解其业务运营、客户和竞争对手。文本分析可以通过以下方式帮助组织:

客户洞察 [3]:文本分析可以通过分析客户反馈、社交媒体评论和其他非结构化文本数据来帮助组织了解客户需求和偏好。这种洞察力可用于改善客户服务、开展有针对性的营销活动以及优化产品和服务。



图 2. 文本分析和 ERP 流程

供应商绩效 [4]:文本分析可以通过从非结构化文本数据 (例如供应商发票、采购订单和合同)中提取信息来帮助组织分析供应商绩效。这可以帮助组织确定需要改进的领域,协商更好的价格并降低供应链风险。

员工情绪[5]:文本分析可以通过从非结构化文本数据 (例如员工调查、社交媒体评论和绩效评估)中提取信息来帮助组织分析员工情绪。

这种洞察力可以用来提高员工的参与度、保留率和生产力。

风险管理 [6]:文本分析可以通过分析新闻文章、社交媒体评论和监管文件等非结构化文本数据来帮助组织识别潜在风险。这种洞察力可用于降低风险、遵守法规和维护声誉。

竞争情报:文本分析可以通过分析非结构化文本数据 (例如行业报告、新闻文章、以及社交媒体评论。这一洞察可以用于

识别新兴趋势、监控竞争对手活动并制定战略计划。

III. 在 ERP 中实施文本分析的主要挑战

尽管文本分析可以提供有价值的见解,但通过企业资源规划 (ERP) 系统实施它可能会面临重大挑战 [7]。以下是一些关键挑战:

- 数据质量:文本分析依靠高质量数据来提供准确的见解。然而,非结构化文本数据的清理、处理和规范化可能具有挑战性。数据质量差会导致不准确的见解,从而导致错误的决策。

- 数据集成:将非结构化文本数据与 ERP 系统集成可能很复杂,特别是当数据存储在不同的系统中时。确保数据一致性、准确性和完整性可能具有挑战性,从而导致集成问题。

- 数据量:文本数据可能非常庞大,因此存储、处理和分析起来非常困难。组织需要投资于强大的基础设施和可扩展的解决方案来处理大量数据。

- 数据安全:文本数据可能包含敏感信息,例如客户详细信息、财务信息和专有业务信息。组织必须确保采取数据安全措施来保护数据

防止未经授权的访问、入侵和网络威胁。

- 变更管理:在 ERP 系统中实施文本分析需要改变业务流程、工作流和用户角色。组织必须投资变更管理计划,以确保利益相关者做好准备并愿意接受这些变更。

应对这些挑战需要采取全面的方法,包括强大的数据管理流程、技术投资、组织变革管理和人才获取策略。通过应对这些挑战,组织可以利用文本分析的强大功能来获得有价值的见解并推动业务增长。

IV. 非结构化数据源

如图 3 所示,非结构化数据有多种来源,这些数据分为两组:内部和外部。内部数据由组织生成,外部数据由客户、员工和供应商在组织外部的平台上生成。



图 3. 数据来源

**文件:**每个企业都会生成与合同、销售、发票、产品设计、产品质量、制造和采购相关的文件。这些文件包含与公司销售、制造和采购的产品相关的重要信息。

**社交媒体:**社交媒体在营销产品时发挥着重要作用,客户将分享他们的经验,也可能在社交媒体上向其他人推荐产品,这些数据可以用于营销、与客户联系并识别产品的潜在问题。

**电子邮件:**当今数字世界中的大部分沟通都是通过客户和销售团队、供应商和采购团队、人力资源和员工之间的电子邮件完成的。

五、文本分析方法

这些电子邮件可能包含有关产品反馈、产品质量问题、供应商问题、员工反馈的信息。

确定非结构化数据的来源后,了解分析和使用这些数据的步骤非常重要。

**客户支持:**支持团队是公司与客户之间的重要接口,他们收集有关产品问题、退货、购买问题和反馈的信息。

下面解释了图 4 中所示的一些最重要的步骤。

他们通常由代理手动登录支持系统。

**词袋:**在文本分析中,词袋模型是一种将文本数据表示为单个单词或“标记”集合的方法,而不考虑它们的顺序或上下文。它被称为单词“包”,因为该模型仅跟踪文档中每个单词的频率,而不跟踪它们的位置或与其他单词的关系。为了创建词袋表示,首先对文本进行预处理,以删除标点符号、数字和停用词(诸如“the”、“and”、“a”等没有太多含义的常用词)。

**调查:**许多公司进行调查,以确定员工和客户面临的问题,这些调查包含用户手动添加的信息,可以通过文本分析提取和使用这些信息。

**客户反馈:**客户通常会在购买产品的网站上留下反馈,这可能是电子商务、第三方销售或直接面向客户的销售,所有这些信息都很有价值,

然后,对文本中的每个单词进行计数并将其表示为数据矩阵中的单独特征,每个单词的频率就是其在相应列中的值。然后,该矩阵可以用作机器学习算法的输入,用于执行情感分析、文本分类或主题建模等任务。虽然词袋模型简单且计算效率高,但它们有一些局限性。他们无法捕捉短语或句子的含义,他们将所有单词视为同等重要,并且不擅长处理拼写错误或错误。

与产品有关。

**评论:**员工通常会在社交媒体或其他有关公司的平台上撰写反馈或分享他们的经验,这些数据通常由现任或前员工手动输入。

SURJIT SINGH BAWA :通过企业资源规划实施文本分析

生僻词。通常使用更高级的 NLP 技术,例如词嵌入或语境化表示来解决这些限制并提高文本分析任务的准确性。

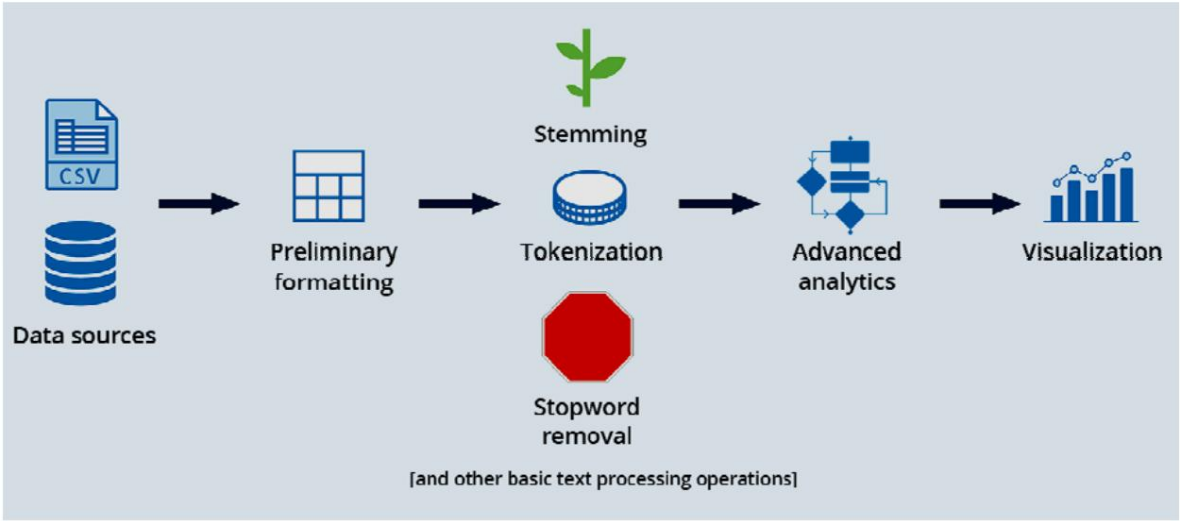


图 4. 文本分析方法[10]

词干提取:词干提取是一种文本分析技术,涉及将单词简化为词根或基本形式,以便将具有相似含义的单词组合在一起。它通常用于自然语言处理 (NLP),以预处理文本数据以供分析。词干提取过程涉及从单词中删除后缀和前缀,以便将它们与具有相同词根的其他单词匹配。例如,单词“running”、“runners”和“run”可以提取为“run”。词干提取可用于各种 NLP 任务,包括信息检索、分类和情感分析。通过减少文本语料库中唯一单词的数量,词干提取还可以帮助提高用于分析大量文本数据的机器学习算法的效率。但是,需要注意的是,词干提取也会给文本分析带来错误,因为某些单词可能被错误地提取词干并失去其原意。

- 删除任何不需要的字符,例如标点符号  
标记或数字。
  - 将文本分割成单独的单词或短语。
  - 将所有标记转换为小写或大写
- 使文本标准化。
- 删除任何停用词,即没有太多含义的常用词,例如“the”、“and”或“a”。

文本 不同编程语言中存在多种可用于标记化的工具和库,例如 Python 中的 NLTK (自然语言工具包)或 Java 中的 CoreNLP。

标记化是文本分析中重要的预处理步骤,因为它可以对文本进行进一步分析,例如  
例如计算每个标记的频率、识别最常见的单词或为机器学习模型生成特征。

因此,谨慎使用词干提取并结合其他文本分析技术以确保结果的准确性非常重要。

标记化:在文本分析中,标记化是将一段文本分解为称为标记的单独单元的过程。这些标记可以是单词、短语甚至字符,具体取决于分析所需的粒度级别。分词是许多自然语言处理任务中的重要步骤,例如文本分类、情感分析和机器翻译。它

停用词删除:在文本分析中,停用词删除是预处理文本数据时使用的常用技术。停用词是一种语言中经常出现但没有太多含义的单词,例如“the”、“and”、“a”等。它们通常从文本数据中删除,因为它们会给分析和分析增加噪音。没有提供太多有用的信息。停用词删除的过程涉及识别给定语言中的停用词列表并将其从文本数据中删除。这可以通过创建特定于域或应用程序的自定义停用词列表来手动完成,也可以通过使用各种 NLP 库中可用的预定义停用词列表来自动完成。去除停用词可以提高准确性

有助于标准化输入数据,使其更易于处理和分析。标记化过程通常涉及以下步骤:

SURJIT SINGH BAWA:通过企业资源规划实施文本分析

以及各种文本分析任务（如文本分类、情感分析和主题建模）的效率。通过删除常用词，停用词删除可以帮助将分析重点放在文本中更有意义和更相关的词上。

然而，值得注意的是，停用词删除并不总是必要或适合每个文本分析任务。在某些情况下，停用词可能携带重要信息或上下文，删除它们可能会导致有价值信息的丢失。因此，评估停用词删除对当前特定任务的影响并就是否在分析中包含或排除停用词做出明智的决定非常重要。

聚类分析:聚类分析是一种文本分析技术，涉及根据内容或相似性将相似的文档或单词分组到聚类中。聚类分析的目标是识别大量文本数据中的模式或结构，这对于各种文本分析任务（例如文本分类、主题建模或信息检索）非常有用。

文本分析中有多种聚类分析方法，例如层次聚类、K 均值聚类或基于密度的聚类。每种方法都有自己的优点和缺点，具体取决于特定的任务和数据集。

在层次聚类中，文本数据根据其相似性被递归地划分为更小的簇，直到每个簇仅包含几个文档或单词。这种方法对于探索数据结构和识别有意义的子组非常有用，但它的计算成本也很高，并且可能无法很好地扩展到非常大的数据集。

在K 均值聚类中，文本数据根据其相似性被划分为预定义数量的聚类，每个聚类由该聚类的质心（或中心）表示。这种方法比层次聚类更有效，并且适用于较大的数据集，但它需要提前指定聚类数量，而这可能并不总是事先知道的。

在基于密度的聚类中，文本数据根据其密度和连通性进行聚类，聚类定义为高密度区域被低密度区域分隔。这种方法可用于识别不同大小和形状的聚类，但它也可能对参数的选择很敏感，并且可能不适用于非常稀疏或高维的数据。

总体而言，聚类分析是文本分析中发现大集中的模式和结构的强大工具

文本数据，并且可以与其他技术结合使用，例如降维，特征

提取或情感分析，以获得对数据更深入的了解。

例如，文本分析可用于生成词的强度和关系如图5和图6所示。



图 5. 词语强度

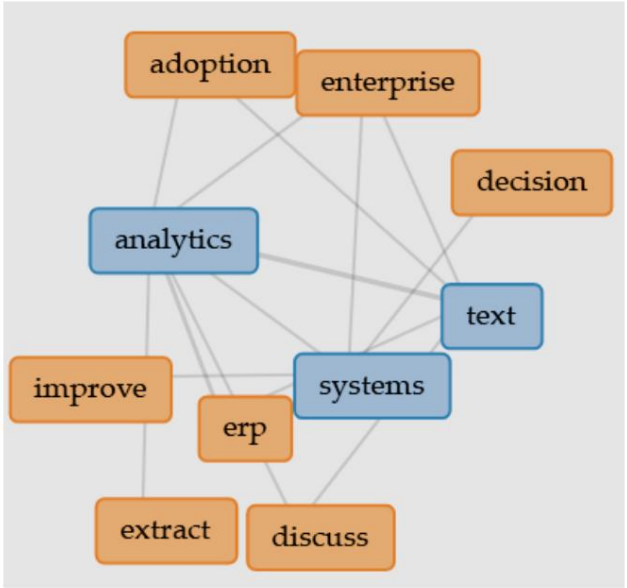


图 6. 词语关系

六.在 ERP 中成功实施文本分析的策略

使用企业资源规划 (ERP) 系统实施文本分析 [8] 可能具有挑战性，但通过正确的策略，组织可以成功实施该技术。以下是在 ERP 中成功实施文本分析的一些策略：

- 定义明确的目标:组织应明确定义其实施文本分析的目标

SURJIT SINGH BAWA:通过企业资源规划实施文本分析

ERP 系统。这包括识别数据源、用例和预期结果。

- 制定数据策略 :组织应制定全面的数据策略,包括数据质量、集成、存储和安全性。该策略应与组织的整体数据管理策略保持一致。

投资正确的技术 :组织应投资能够处理大量文本数据、提供准确见解并与 ERP 系统集成的技术。这包括自然语言处理 (NLP) 工具、机器学习算法和可视化工具。

-从小处着手,逐步扩大规模 :组织应从小型试点项目开始,以测试技术并完善实施策略。试点成功后,应逐步扩大规模,同时考虑组织影响、变更管理和资源需求。

-跨职能协作 :在 ERP 系统中实施文本分析需要跨职能协作,包括 IT、数据管理、业务分析和业务运营。组织应促进跨职能协作,以确保实施符合业务目标。

-衡量和监控进度 :组织应建立绩效指标来衡量文本分析实施的成功与否。他们还应定期监测进展情况,找出差距和需要改进的领域,并采取必要的纠正措施。

七.文本分析的现实用例

ERP

文本分析可以在企业资源规划 (ERP) 系统中以多种方式应用,以提供有价值的见解并改进业务流程 [9]。以下是 ERP 中文本分析的一些实际用例:

- 客户反馈情绪分析 :文本分析可用于分析客户反馈和情绪,以深入了解客户需求 and 偏好。例如,制造商可以使用文本分析来分析客户对其产品的评论,以确定需要改进的领域。

- 欺诈检测和预防 :文本分析可用于分析非结构化文本数据,例如电子邮件和聊天日志,以识别潜在的欺诈和其他金融犯罪。例如,保险公司可以使用文本分析通过分析这些索赔相关的文本数据来识别欺诈性索赔。

-预测性维护 :文本分析可用于分析来自维护日志和其他来源的文本数据,以预测设备故障和维护需求。例如,发电厂可以使用文本分析来预测一台设备何时可能发生故障,从而使他们能够主动执行维护。

-质量控制和检查 :文本分析可用于分析检查报告中的文本数据,以识别与质量问题相关的趋势和模式。例如,制造商可以使用文本分析通过分析生产过程相关的文本数据来确定质量问题的根本原因。

-供应链优化 :文本分析可用于分析来自供应商、物流提供商和其他来源的非结构化文本数据,以优化供应链运营。例如,零售商可以使用文本分析来分析供应商发票,以识别差异并协商更好的条款。

- 员工敬业度和满意度 :文本分析可用于分析员工情绪和敬业度水平,以确定需要改进的领域。例如,人力资源部门可以使用文本分析来分析来自调查和其他来源的员工反馈,以识别与员工满意度相关的趋势和模式。

八. ERP 中文本分析的未来

文本分析在企业资源规划 (ERP) 中的前景光明。以下是未来可以预期的一些潜在进步:

-增强与人工智能和机器学习的整合 :  
文本分析可以与人工智能 (AI) 和机器学习 (ML) 技术集成以增强其功能。这可以使 ERP 系统自动识别非结构化文本数据中的模式和趋势,并根据该数据进行预测。

-增强的自然语言处理 (NLP) :随着 NLP 技术的不断进步,文本分析在分析和解释自然语言方面将变得更加复杂。这可以使 ERP 系统能够理解更复杂的语言结构,更准确地识别情感,并识别文本数据中更微妙的细微差别。

-更加重视上下文分析 :随着文本分析技术的不断发展,人们将更加重视上下文分析。这可以使 ERP 系统更好地理解文本数据的使用背景,并根据该数据做出更准确、更相关的预测。

SURJIT SINGH BAWA:通过企业资源规划实施文本分析

- 改进数据可视化:随着组织继续收集越来越多的数据,以有意义的方式可  
视化这些数据变得越来越重要。文本分析可以与数据可视化工具集成,以提供文  
本数据的交互式 and 视觉吸引力表示,使组织能够一目了然地获得见解。

- 更加关注隐私和安全:随着 ERP 系统收集的数据量不断增长,人们将更加  
关注隐私和安全。开发文本分析技术时需要考虑隐私和安全因素,以确保敏感数  
据得到保护和合乎道德的使用。

总体而言,ERP 中文本分析的未来可能会更加复杂、与其他技术集成以及更  
加重视隐私和安全。通过利用文本分析,组织可以从非结构化文本数据中获得有  
价值的见解,从而使他们能够做出更好的决策、降低成本并推动业务增长。

IX. 在 ERP 中采用文本分析的结论和建议。

总之,文本分析可以通过提供见解和改进企业资源规划 (ERP) 系统内的业  
务流程,为组织提供显著的竞争优势。然而,为了在 ERP 中成功实施文本分析,还  
必须克服一些关键挑战。

考虑在 ERP 中采用文本分析的组织应考虑以下建议:

- 确定明确的业务目标:在 ERP 中实施文本分析之前,确定与组织总体目标一  
致的明确业务目标非常重要。这有助于确保文本分析的使用方式能够增加价值并  
推动业务增长。

- 投资正确的工具和技术:  
在 ERP 中实施文本分析需要正确的工具和技术,包括软件和硬件。投资符合组织  
特定需求和目标的正确工具非常重要。

- 确保数据质量:文本分析依赖于高质量数据。在 ERP 中实施文本分析之前,  
确保数据干净、准确和完整非常重要。

- 建立强大的数据治理框架:在 ERP 中实施文本分析需要强大的数据治理框  
架,其中包括数据访问、安全和隐私的政策和程序。这有助于确保以合乎道德和负  
责任的方式使用数据。

- 投资培训和教育:文本分析是一项复杂的技术,需要专业技能和知识。投资培  
训和教育很重要,以确保员工拥有有效使用文本分析所需的技能和知识。

通过遵循这些建议,组织可以在 ERP 中成功实施文本分析,并通过利用来自  
非结构化文本数据的宝贵见解来获得竞争优势。

参考

[1] 克劳斯·赫尔穆特·迈克尔·罗斯曼和盖伊·盖博。  
“什么是ERP?”信息系统前沿 2 (2000):141-162。

[2] 杰弗里·罗克韦尔。“文本分析到底是什么?”  
文学与语言计算 18,第 2 期 (2003):209-  
219。

[3] 加拉格尔·康纳·伊奥根·弗雷和凯文·柯兰。  
“将情感分析和文本分析应用于客户体验评论,以了解客户的真实想法。”  
国际数据仓储和挖掘杂志 (IJDWM) 15,第 1 期。 4 (2019) :21-47。

[4] Sabbagh, Ramin。“用于制造供应商分类的语义文本分析技术。” (2018  
年)。

[5] 大卫·戈德伯格和诺埃尔·扎曼。“人力资源管理中员工不满的文本分析。”  
(2018)。

[6] Haile, Issayas M。“金融机构中的数据分析师:文本分析如何帮助风险管理。”  
博士论文,科罗拉多技术大学,2020 年。

[7] Ittoo,Ashwin 和 Antal van den Bosch。“行业文本分析:挑战、需求和趋  
势。”工业计算机 78 (2016): 96-107。

[8] Anandarajan,Murugan,Chelsey Hill 和托马斯·诺兰。“实用文本分  
析。”最大化文本数据的价值。(分析和数据科学的进展。  
卷。 2。) ,施普林格 (2019) :45-59。

[9] Miner, Gary, John Elder IV, Andrew Fast, Thomas Hill,  
Robert Nisbet 和 Dursun Delen。非结构化文本数据应用的实  
用文本挖掘和统计分析。Academic Press,2012 年。[10] [https://  
www.softwareadvice.com/resources/  
what-is-  
text-analysis/](https://www.softwareadvice.com/resources/what-is-text-analysis/)。