|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 论文题目 | Opinion mining for software development；A systematic literature Review | | | |
| Paper URL | https://dl.acm.org/doi/10.1145/3490388 | | | |
| Project URL |  | | | |
| 综述/背景介绍 | 发展状况 | 原因 | 意义 | 关键词（速记词汇、信息索引词汇） |
| 发展状况：观念发掘(Opinion mining)是Dave于2003年提出,文本情感分析（也称为意见挖掘）是指用自然语言处理、文本挖掘以及计算机语言学等方法来识别和提取原素材中的主观讯息。通常来说在大部分的研究里Opinion mining和Sentiment analysis可以互换。  近年来，意见挖掘受到了软件工程研究者的广泛关注。研究已经看到意见挖掘在收集应用评论信息方面的使用，以了解开发者如何改进他们的产品和修改他们的发布计划。研究人员还应用意见挖掘技术来监测开发人员在开发活动中表达的情绪。意见挖掘也被用于评估软件产品的质量。  原因：鉴于所有这些研究，有必要对现有的意见挖掘技术及其在软件工程中的应用进行概述。通过这种方式，研究人员可以有一个基础来推进该领域，工具用户可以更好地了解他们如何应用现有技术以及它们的局限性是什么。 | | | 观念发掘、情感分析、软件工程 |
| 假设 |  | | |  |
| 方法描述(含图) | 整体搜索策略（如图1）：先采用关键词搜索（("opinion mining" OR "sentiment analysis" OR "emotion") AND ("software") AND ("developer" OR "development")）得出初始论文库，再对初始论文进行文献筛查并进行一轮滚雪球得到最终需要研究的论文库（整体搜素策略旨在保持高召回率，搜索到所有相关研究）  文献筛查：通过人工检查标题以及摘要，按照标准（如表2），用web程序将候选论文分成三类：  1纳入研究2弃之不用3次要研究（第三类论文指该文章是偏向综述类文章，将在滚雪球时来主要研究。）  每个文章都经过两位不同的研究人员分别判断类别，当有分歧时全体研究人员对分歧文章进行讨论得出结果。    完成上述搜索过程后得出185篇所选论文。  该文提出一个核心问题： How can opinion mining techniques support software development activities? 并基于该核心问题细分出6个小研究问题  重点关注以下6个研究问题.  RQ1哪些软件工程活动应用了SE？  RQ2采用了哪些工具？这些工具功能。  RQ3多少时间评估一次工具可靠性？面对不同的上下文，可能效果不同。  RQ4性能对比  RQ5评估数据集  RQ6碰到哪些问题  为回答RQ1到5，对被选择的文章提取信息如表3  后对提取的数据进行了手动编码：1方便检索2统一术语  首先通过检查表3中第12题的答案来确定是否包含该论文，因为如果论文没有通过全文过滤，我们要求检查员在这里做笔记。然后，我们确定:1)研究目的(例如，检测开发人员在软件工件中表达的情感/情绪/礼貌)，2)方法是否被定制，3)使用的工具，4)方法是否可用，5)意见挖掘技术的类型(例如，情绪极性分析)，6)工具是否应用于与其起源不同的上下文，7)& 8)方法的性能是否已被验证，10)数据集类型(例如，GitHub发布评论)，11)数据集是否可用。很少有研究被重复，因此没有从问题9中收集到有用的信息。  检查共有395篇论文（未）主要作者先对论文库的前23篇先进行实验性编码，并与剩下作者讨论，得出一致结论后，再将156论文均分给每个作者，一个人平均26个，从这一轮编码中得出统一术语，最后将剩下的216篇论文均分给每个作者。  为了回答RQ6，我们检查了提出或评估意见挖掘技术的论文，因为我们更感兴趣的是证据支持的问题/限制，而不是基于假设的问题/限制。每篇论文都由两位作者独立手工检查，他们在满足以下标准时提取见解:  •它们应该在结果，讨论或结论中明确指出。  •它们应该与在软件工程中定制/采用意见挖掘方法相关。  •它们应该有数据支持(即，那些没有证据的建议应该被丢弃)。  •它们不应该描述特定于工具的优化，如参数调优。  合并了作者提取的关注点/限制和丢弃重复的关注点/限制。也就是说，我们从同一篇论文中删除了相同的见解，但如果从不同的论文中提取，则保留了这些相似/相同的见解。然后根据主题相似度对提取的见解进行手动分类 | | |  |
| 实验设计 |  | | |  |
| 数据处理 | 输入 | 筛除特例 | 处理方式 | 关键词（速记词汇、信息索引词汇） |
| 关键词检索所得论文 | 人工根据标准筛除不符合的文献 | 分类处理 |  |
| 结论 | 该文进行了系统的文献综述，涉及185篇相关的论文并进行以下工作：  1提出了应用意见挖掘的软件开发活动的细粒度分类，并描述了这些活动是什么。  2总结了主题论文中公开可用的意见挖掘工具，并介绍这些工具功能  3调查了在其他研究中采用工具时是否评估了这些工具的性能，发现很少有研究人员在这些工具被用于不同于它们设计的领域时评估工具的性能。  4比较了这些工具在不同背景下的性能，以便读者找出哪种工具可能最适合他们的数据  5介绍了23个公开可用的软件相关数据集，这些数据集可用于评估和定制软件工程领域的新意见挖掘方法。  6强调了在采用和定制意见挖掘工具时可能会所面临的问题和局限性，并指出了可能的解决方案。  **今后工作方向：**  1应用在未使用SE的软工领域：如人力资源管理方面  2.监测开发人员情绪在人力管理方面的应用：现有多只是检测，而没有真正应用到管理上。  3.情感极性分析性能改进：1针对某领域，构建特殊词汇表。2对多个领域的数据集，设计自适应的分析工具  4验证用户反馈是否对当前软件有效：例如随着软件更新以前的反馈可能失效  5细粒度分类：需要更细的划分法，来降低开发人员的手动工作量 | | | |
| 局限性分析 | 一、在构建论文备选库时的威胁：  （1）论文质量不行：使用谷歌学术时，大量未经审查的论文，或是经过审查的文章仍然可能有潜在缺陷  （2）关键词搜索缺陷：会漏掉一些相关研究，采用一轮snowballing来尽量避免遗漏。  二、内部因素（搜素引擎）对可复制性的影响：使用的数据库不断索引更多的论文，并且它们像黑盒一样运作，意味着我们无法判断它们的搜索算法是否会在某个时候发生变化。然而，由于我们采取了所有返回的结果并进行了snowballing，问题是这些数据库动态地索引论文。即使使用相同的搜索策略，我们可能无法复制搜索结果。例，有些论文可能在数据库中的索引时间要远远晚于它们的实际出版日期。因此，即使出版日期范围保持不变，将来仍有可能找到更多的论文。  三、外部因素（其他语言）对普适性的影响：该文采用英语编纂，对其他语言的普适性不足  四、人力资源较少对结论正确性的影响：每一篇文章结论只有一名研究员完成 | | | |

1.论文总结：

（论文名中文）：软件开发中的意见挖掘;系统的文献综述

（论文名英文）：Opinion mining for software development；A systematic

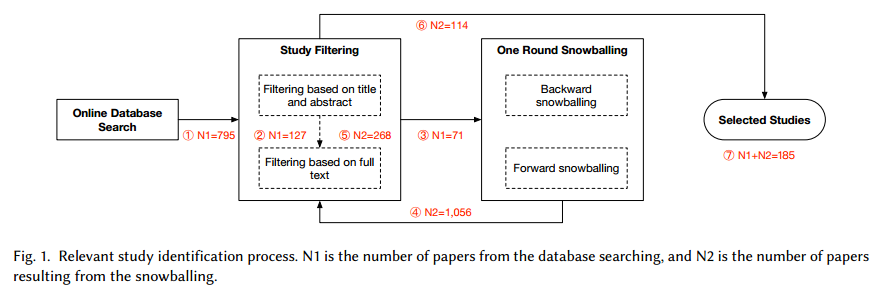
literature Review

（论文URL）：https://dl.acm.org/doi/10.1145/3490388

（总结）：该文是一篇总结回顾方面的综述论文，通过关键词检索与滚雪球的方式尽可能的找到情感分析领域的大部分文献，根据本文核心问题：情绪分析对软件工程领域发展起了什么样的作用？细分出六个研究问题，并根据六个研究问题提出所需要提取信息的表格，对被选择的文献进行信息提取，并编码提取后的信息，统一术语，建立索引。根据提出的信息完成对六个问题的回答，最后讨论了，研究可能的缺陷，给出了结论。并指明了未来研究方向。

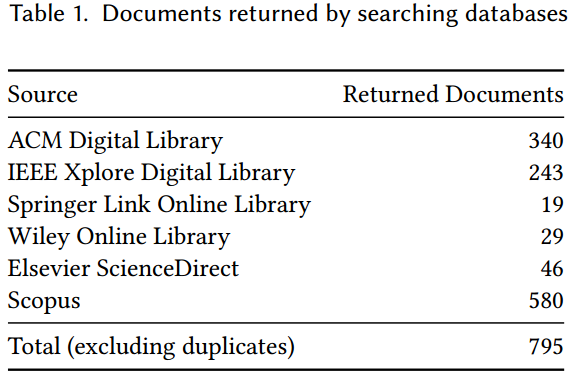
**重要图表：**

1. **研究方法：**

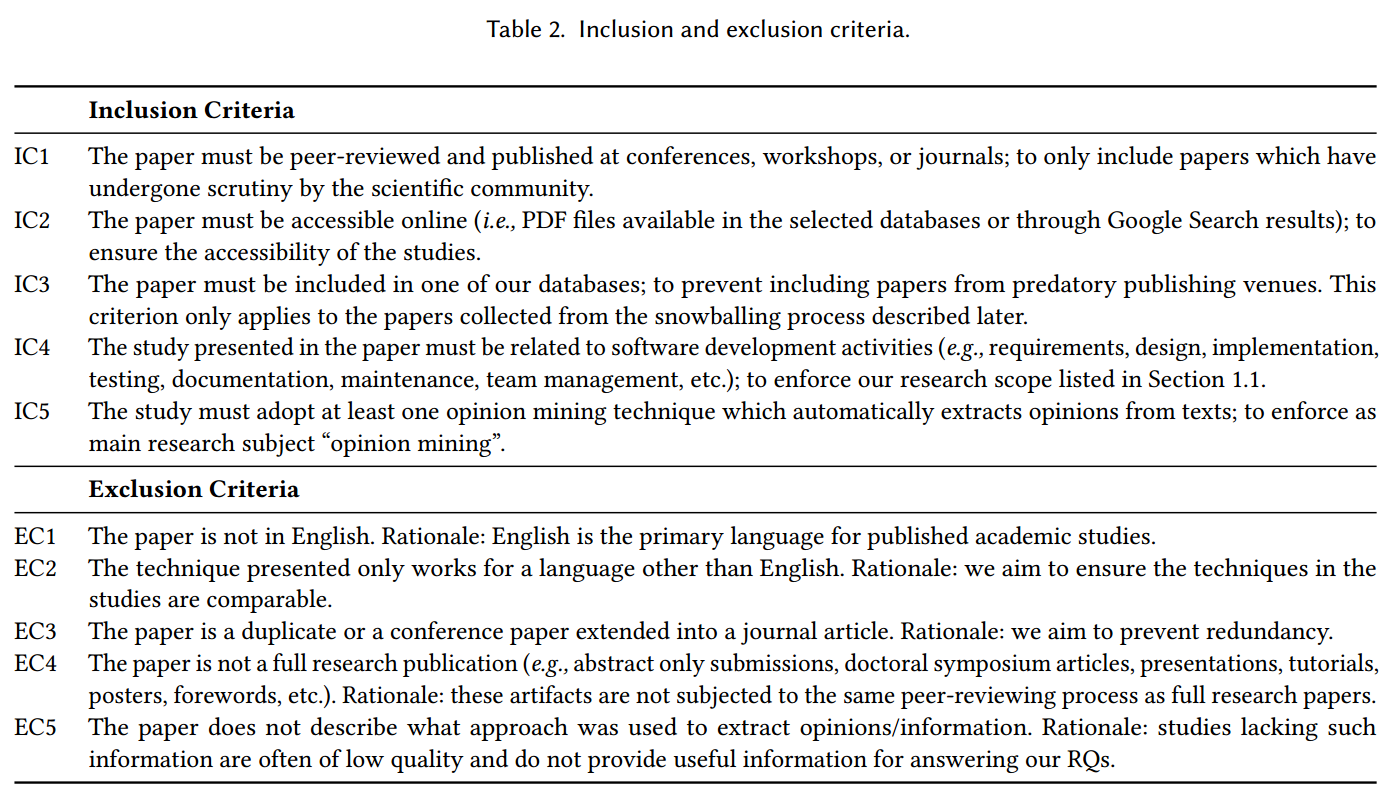


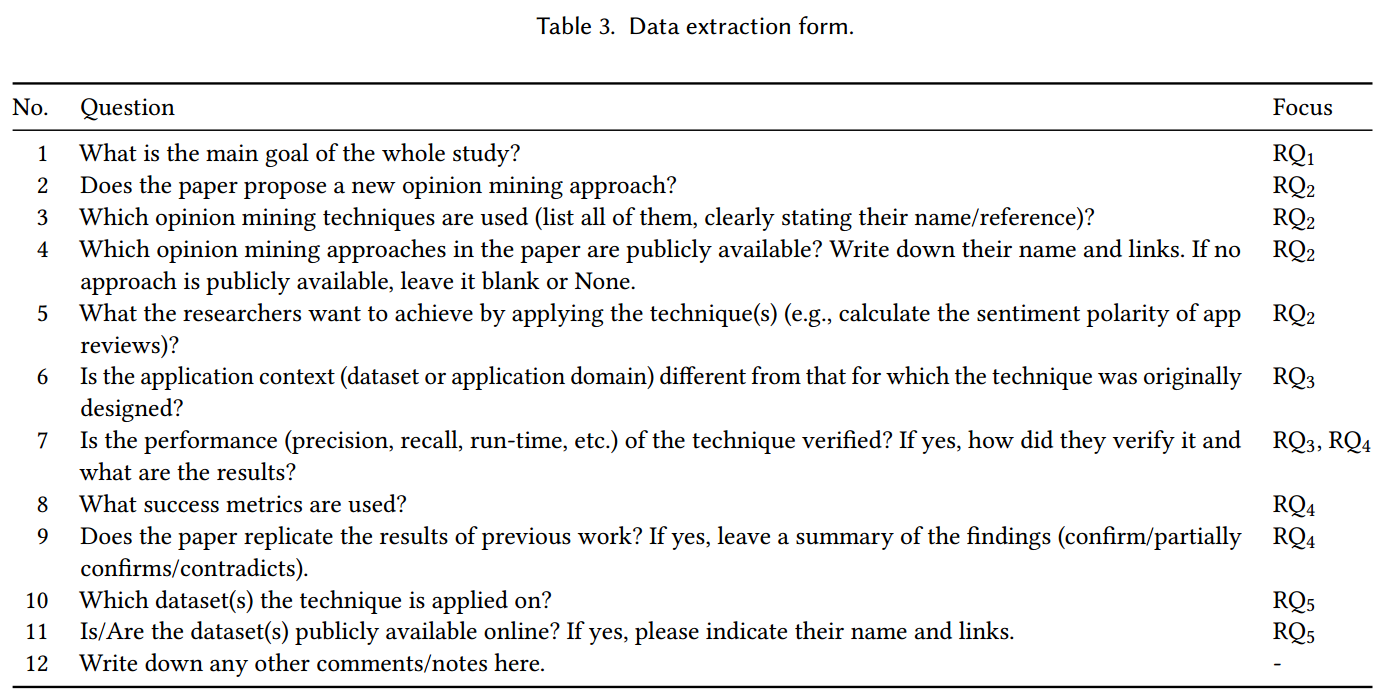
图一文献搜索策略

表一文献来源分布



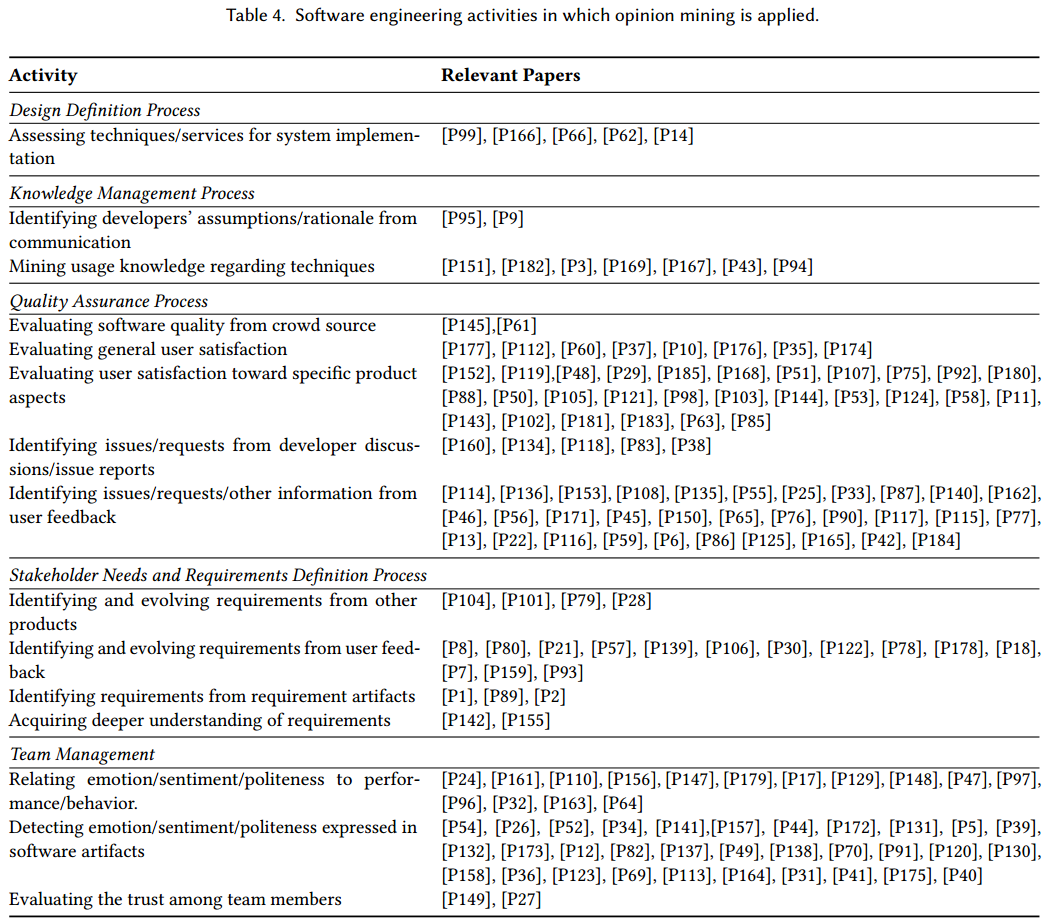
表二论文纳用标准



表三信息提取格式

1. **研究结果：**

表四观点发掘在软件工程领域方面的应用



表五观念发掘工具用于情感极性

