科学论文（科学家活动）新颖性（novelty）的评估

主题---评估新颖性

检测特定语言模式--->获得句子级别的相似性--->定义新颖性

算法：评估某个作者再某个研究领域的个人贡献

现有的评估方式：同行评估—定性评估（审稿人）、定量指标——顶刊上发表的作品数量、总引用次数、期刊IF、某些论文最大引用量----🡪转到文章的内在结构（文章内容）

参考文献--->引文分析、词汇聚类(发现某一领域的研究人员)

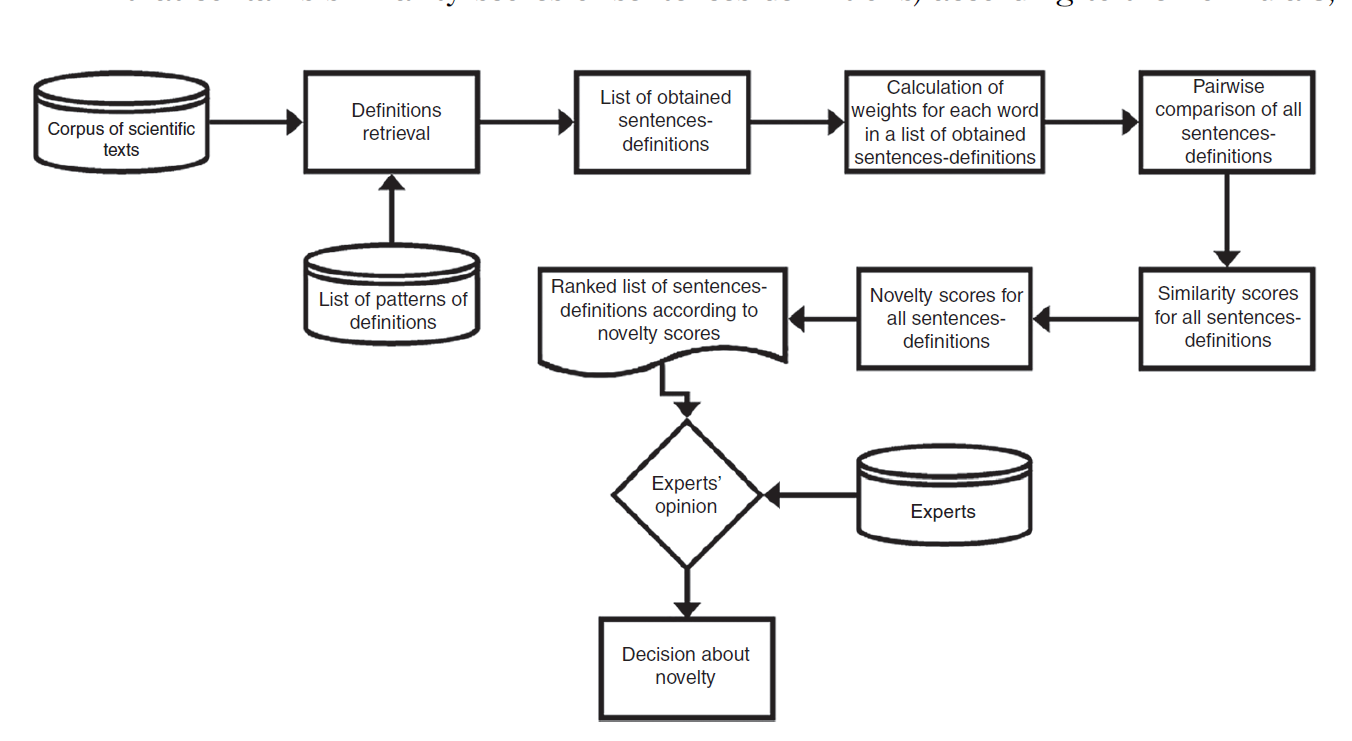
**信息检索**（研究领域内的资料收集，类似文献综述）---某个领域发生了什么新情况？解决了什么样的任务？应用了哪些方法？方法的优点是什么？谁提出了重要概念和定义？

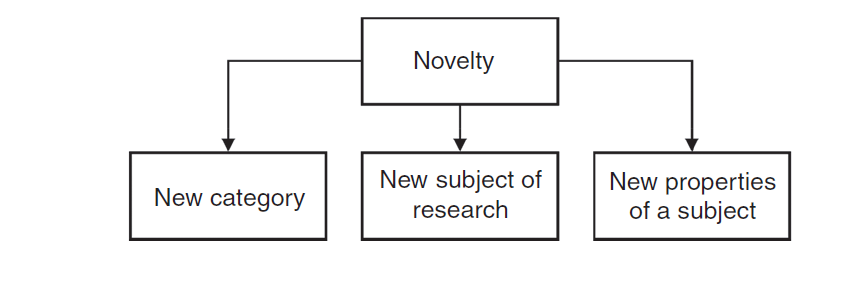
Novelty detection is the technique used to extract novel information from a set of relevant documents or from the same document in a given topic (query) (Sendhilkumar et al., 2013). 新颖性检测是用于从一组相关文档或给定主题（查询）中的同一文档中提取新颖信息的技术。

检测新文本步骤：预处理（删除停用词、词干、词性标注）、分类（将文档归入相关主题目录）、新颖性挖掘（三个层次：事件、句子、文档）

* 事件**event：**通常使用**向量空间模型、语言模型、词汇链**等几种模型来表示新的故事/文档。然后，通过计算基于统计或基于语义的相似性，将文档分组为簇。开始一个新簇的故事被标记为关于一个新主题的第一个故事，或者如果存在一个小于相似阈值的新颖阈值，并且故事与其最近簇之间的相似分数大于新颖阈值，则该故事将被标记为“旧”故事。
* **句子sentence：**句子级新颖性挖掘的研究工作旨在从一系列文档/句子中发现相关且新颖的句子。Harman和Soboroff为我们提供了TREC新颖性轨迹的概述，其基本任务是检测相关和新颖性句子（Harman, 2002; Soboroff and Harman, 2003; Soboroff, 2004）。句子的新颖性通常是根据出现在句子中的新词的数量来计算的。句子层面的新奇性检测主要分为两个步骤：相关句子检索和新奇句子提取。通常，句子与给定查询之间的高相似度分数会增加句子的相关性等级。**句子与所有之前看到的句子之间的高相似性分数会降低句子的新颖性排名。**

An algorithm of analysis of a subject of research：输入文档—定义检索





科学作品的新颖性可以从以下三点中找到（图1）：一个新的类别，描述了与其他类别的联系；描述与其他学科的关系并描述给定学科的特性的新研究学科；以及学科的新属性，并对其进行描述，并与已发现的属性建立连接