

## 一种进行文本数据搜索的系统和方法



通信和信息

计算机/人工智能/数据处理和信息技术

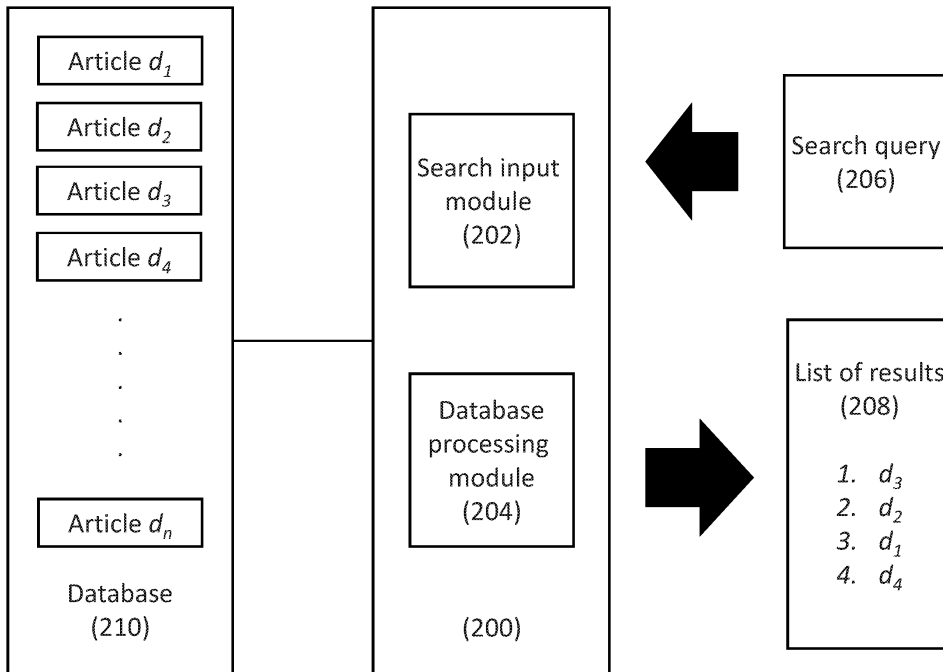


FIG. 2



发明人

徐睿云

陈海亮

赵建良教授

询问: [kto@cityu.edu.hk](mailto:kto@cityu.edu.hk)

### 机会

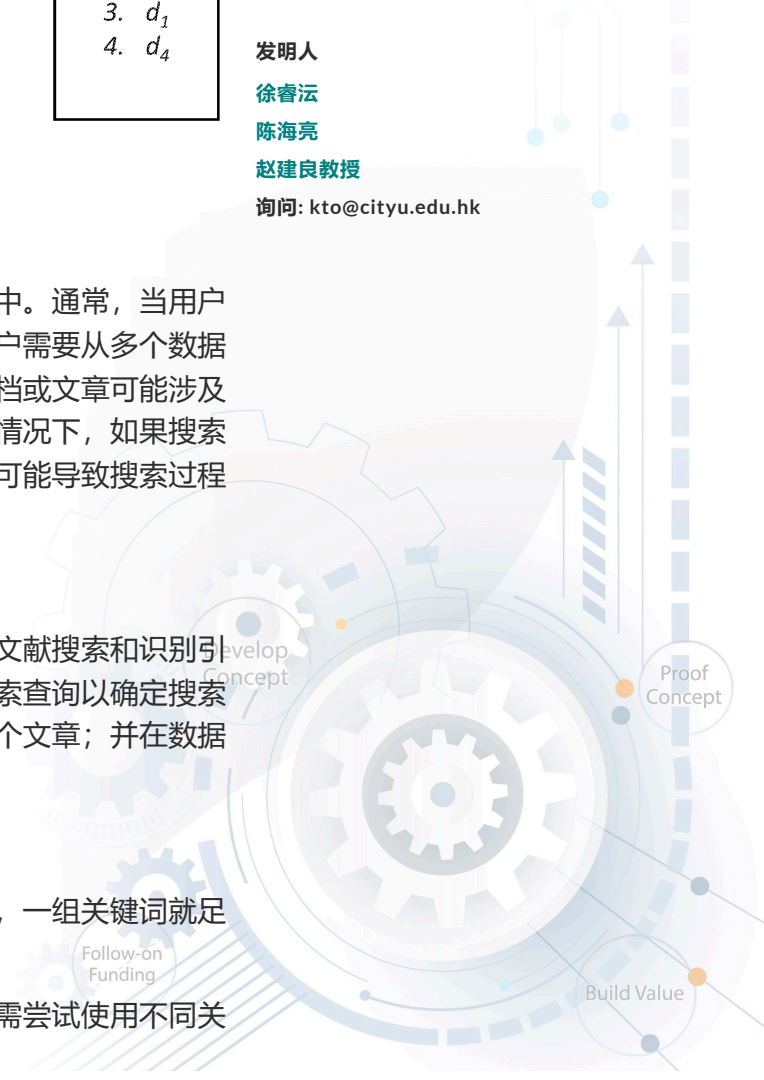
文本内容可能以数字形式包含在存储于电子数据库的文档中。通常，当用户需要检索存储在大型数据库中的文档中的文本数据时，用户需要从多个数据库中的多个文档中定位特定的文档。定位或搜索特定的文档或文章可能涉及将搜索查询与存储在文档中的信息相匹配。然而，在某些情况下，如果搜索查询没有准确地制定，可能难以定位一些存储的文档，这可能导致搜索过程变得耗时且效率低下。

### 技术

本发明涉及一种进行文本数据搜索的系统和方法，特别是文献搜索和识别引文数据。它包括接收与搜索主题相关的搜索查询；分析搜索查询以确定搜索主题的至少一个属性；处理至少一个属性和数据库中的多个文章；并在数据库中的多个文章中识别与搜索主题相关的一个或多个结果。

### 优势

- 无需精心撰写的摘要作为查询。由于句子中的词袋模型，一组关键词就足以作为查询输入。
- 不要求用户基于关键词来制定他们的搜索查询，用户无需尝试使用不同关键词的多次查询。



- 高效的方法，适合大规模实施。

## 应用

- 根据搜索查询定位或搜索特定的文档或文章
- 基于文本内容的引文
- 文献计量学与科学计量学

