



(12) 发明专利

(10) 授权公告号 CN 107870915 B

(45) 授权公告日 2021.08.17

(21) 申请号 201610847713.0

(22) 申请日 2016.09.23

(65) 同一申请的已公布的文献号
申请公布号 CN 107870915 A

(43) 申请公布日 2018.04.03

(73) 专利权人 伊姆西IP控股有限责任公司
地址 美国马萨诸塞州

(72) 发明人 黄坤武 陈超 张磊 刘晶晶
代洪涛

(74) 专利代理机构 北京市金杜律师事务所
11256
代理人 王茂华

(51) Int.Cl.
G06F 16/248 (2019.01)
G06F 16/838 (2019.01)

(56) 对比文件

US 2009052804 A1, 2009.02.26

US 2009052804 A1, 2009.02.26

US 2010115396 A1, 2010.05.06

CN 102087669 A, 2011.06.08

CN 103927358 A, 2014.07.16

审查员 罗必辉

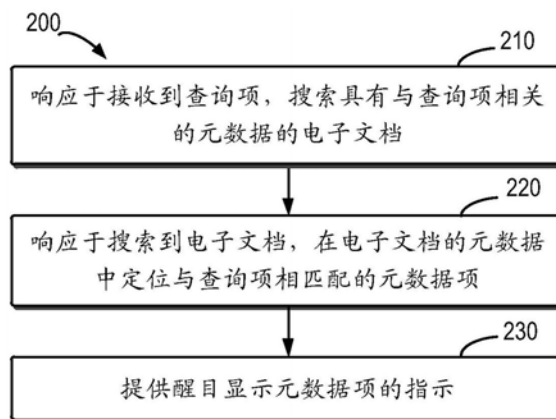
权利要求书3页 说明书9页 附图4页

(54) 发明名称

对搜索结果的指示

(57) 摘要

本公开的实施例提供了用于指示搜索结果的方案。公开了一种用于指示搜索结果的方法，其包括响应于接收到查询项，搜索具有与查询项相关的元数据的电子文档。该方法还包括响应于搜索到电子文档，在电子文档的元数据中定位与查询项相匹配的元数据项。该方法进一步包括提供醒目显示元数据项的指示。



1. 一种指示搜索结果的方法,包括:

从用户接收搜索请求,所述搜索请求包括:

元数据查询项,所述元数据查询项被请求为在多个电子文档的元数据内搜索,和
文本数据查询项,所述文本数据查询项被请求为在所述多个电子文档的非元数据文本
数据内搜索;

响应于接收到所述搜索请求,搜索所述多个电子文档中具有以下两者的电子文档:(a)
与所述元数据查询项相关的元数据和(b)与所述文本数据查询项相匹配的文本数据;

响应于通过所述搜索定位到特定电子文档,从所述电子文档的所述元数据中定位与所
述元数据查询项相匹配的元数据项,并且从所述电子文档的所述文本数据定位与所述文
本数据查询项相匹配的文本项;以及

提供醒目显示所述元数据项和所述文本项的指示,包括:

标识所述元数据中包括所定位的元数据项的第一部分,包括标识所述元数据内的第一
序列,所述第一序列包括所定位的元数据项以及所述元数据内在所定位的元数据项之前和
在所定位的元数据项之后的附加元数据,所述第一部分包括比所述电子文档的所有所述元
数据更少的元数据;

标识所述文本数据中包括所定位的所述文本项的第二部分,包括标识所述文本数据内
的第二序列,所述第二序列包括所定位的文本项以及所述文本数据内在所定位的文本项之
前和在所定位的文本项之后的附加文本数据,所述第二部分包括比所述电子文档的所有所
述文本数据更少的文本数据;

合并所述第一部分和所述第二部分;以及

提供合并后的所述第一部分和所述第二部分的表示的显示,在合并后的所述第一部分
和所述第二部分的所述表示中,所述元数据项和所述文本项被醒目显示,其中提供合并后
的所述第一部分和所述第二部分的所述表示的显示包括显示比所述电子文档的全部所述
元数据更少的元数据和比所述电子文档的全部所述文本数据更少的文本数据。

2. 根据权利要求1所述的方法,其中定位与所述元数据查询项相匹配的元数据项包括:

确定所述元数据查询项与所述元数据的多个数据字段中的至少一个数据字段相关联;
以及

在所确定的至少一个数据字段中定位与所述元数据查询项相匹配的元数据项。

3. 根据权利要求1所述的方法,其中定位与所述元数据查询项相匹配的元数据项包括:

处理所述元数据以获得索引,所述索引与所述元数据包含的至少一个元数据项相关
联;以及

基于获得的索引,定位与所述元数据查询项相匹配的元数据项。

4. 根据权利要求1至3中任一项所述的方法,其中所述元数据包括与所述用户的搜索意
图相关联的附加信息。

5. 根据权利要求1至3中任一项所述的方法,

其中所述搜索请求经由计算机化环境的输入设备接收;

其中所述电子文档和所述电子文档的所述元数据驻留在所述计算机化环境的存储器
内,所述元数据项响应于所述电子文档和所述元数据从所述存储器被访问而被定位;并且
其中提供醒目显示所述元数据项的所述指示包括使用所述计算机化环境的用户界面

来标识所述元数据项。

6. 根据权利要求1至3中任一项所述的方法,其中所述元数据选自以下:

所述电子文档的作者;

所述电子文档的格式;

所述电子文档的创建日期;

所述电子文档的更新日期;

描述所述电子文档的所述文本数据的一个或多个关键字;以及

关于所述电子文档的所述文本数据的一个或多个评论。

7. 一种用于指示搜索结果的装置,包括:

至少一个处理器;以及

包括计算机程序代码的至少一个存储器;其中

所述至少一个存储器和所述计算机程序代码被配置为与所述处理器一起促使所述装置:

从用户接收搜索请求,所述搜索请求包括:

元数据查询项,所述元数据查询项被请求为在多个电子文档的元数据内搜索,和

文本数据查询项,所述文本数据查询项被请求为在所述多个电子文档的非元数据文本数据内搜索;

响应于接收到所述搜索请求,搜索所述多个电子文档中具有以下两者的电子文档:(a) 与所述元数据查询项相关的元数据和(b) 与所述文本数据查询项相匹配的文本数据;

响应于通过所述搜索定位到特定电子文档,从所述电子文档的所述元数据中定位与所述元数据查询项相匹配的元数据项,并且从所述电子文档的所述文本数据定位与所述文本数据查询项相匹配的文本项;以及

提供醒目显示所述元数据项和所述文本项的指示,包括:

标识所述元数据中包括所定位的元数据项的第一部分,包括标识所述元数据内的第一序列,所述第一序列包括所定位的元数据项以及所述元数据内在所定位的元数据项之前和在所定位的元数据项之后的附加元数据,所述第一部分包括比所述电子文档的所有所述元数据更少的元数据;

标识所述文本数据中包括所定位的所述文本项的第二部分,包括标识所述文本数据内的第二序列,所述第二序列包括所定位的文本项以及所述文本数据内在所定位的文本项之前和在所定位的文本项之后的附加文本数据,所述第二部分包括比所述电子文档的所有所述文本数据更少的文本数据;

合并所述第一部分和所述第二部分;以及

提供合并后的所述第一部分和所述第二部分的表示的显示,在合并后的所述第一部分和所述第二部分的所述表示中,所述元数据项和所述文本项被醒目显示,其中提供合并后的所述第一部分和所述第二部分的所述表示的显示包括显示比所述电子文档的全部所述元数据更少的元数据和比所述电子文档的全部所述文本数据更少的文本数据。

8. 根据权利要求7所述的装置,其中所述至少一个存储器和所述计算机程序代码被配置为与所述处理器一起促使所述装置:

确定所述元数据查询项与所述元数据的多个数据字段中的至少一个数据字段相关联;

以及

在所确定的至少一个数据字段中定位与所述元数据查询项相匹配的元数据项。

9. 根据权利要求7所述的装置,其中所述至少一个存储器和所述计算机程序代码被配置为与所述处理器一起促使所述装置:

处理所述元数据以获得索引,所述索引与所述元数据包含的至少一个元数据项相关联;以及

基于获得的索引,定位与所述元数据查询项相匹配的元数据项。

10. 根据权利要求7至9中任一项所述的装置,其中所述元数据包括与所述用户的搜索意图相关联的附加信息。

11. 根据权利要求7至9中任一项所述的装置,

其中所述搜索请求经由计算机化环境的输入设备接收;

其中所述电子文档和所述电子文档的所述元数据驻留在所述计算机化环境的存储器内,所述元数据项响应于所述电子文档和所述元数据从所述存储器被访问而被定位;并且

其中提供醒目显示所述元数据项的所述指示包括使用所述计算机化环境的用户界面来标识所述元数据项。

12. 根据权利要求7至9中任一项所述的装置,其中所述元数据选自以下:

所述电子文档的作者;

所述电子文档的格式;

所述电子文档的创建日期;

所述电子文档的更新日期;

描述所述电子文档的所述文本数据的一个或多个关键字;以及

关于所述电子文档的所述文本数据的一个或多个评论。

13. 一种计算机可读介质,其上存储有机器可执行指令,所述机器可执行指令在被执行时使机器执行根据权利要求1至6中任一项所述的方法的步骤。

对搜索结果的指示

技术领域

[0001] 本公开的实施例涉及文本搜索,并且更具体地,涉及对搜索结果的指示。

背景技术

[0002] 随着互联网以及数据库技术的快速发展,对信息的搜索已经成为广泛存在的需求。搜索系统或搜索引擎可以用于帮助用户定位和检索感兴趣的信息。在基于文本的搜索过程中,搜索系统接收用户输入的查询项,利用查询项来搜索相关的电子文档。例如,如果用户期望查找与“self-driving vehicle”有关的内容,则搜索系统会根据接收到的查询项“self-driving vehicle”来搜索正文中包含“self-driving vehicle”的电子文档。

[0003] 搜索系统的搜索结果可以被提供给用户终端并且被呈现在终端的用户界面中。然而,由于搜索结果数量往往较大并且某些电子文档还可能包含大量的相关或不相关的内容,因此希望能够有一种对于搜索结果的呈现方式,使得用户能够从大量信息中快速且方便地定位出感兴趣的内容。

发明内容

[0004] 本公开的实施例提供一种用于指示搜索结果的方案。

[0005] 根据本公开的第一方面,提供了一种指示搜索结果的方法。该方法包括响应于接收到查询项,搜索具有与所述查询项相关的元数据的电子文档。该方法还包括响应于搜索到所述电子文档,在所述电子文档的所述元数据中定位与所述查询项相匹配的元数据项。该方法进一步包括提供醒目显示所述元数据项的指示。

[0006] 根据本公开的第二方面,提供了一种指示搜索结果的装置。该装置包括至少一个处理器和包括计算机程序代码的至少一个存储器。至少一个存储器和所述计算机程序代码被配置为与所述处理器一起促使该装置响应于接收到查询项,搜索具有与所述查询项相关的元数据的电子文档并且响应于搜索到所述电子文档,在所述电子文档的所述元数据中定位与所述查询项相匹配的元数据项。至少一个存储器和所述计算机程序代码被配置为与所述处理器一起进一步促使该装置提供醒目显示所述元数据项的指示。

[0007] 根据本公开的第三方面,提供了一种计算机程序产品。该计算机程序产品被有形地存储在非易失性计算机可读介质上并且包括机器可执行指令。机器可执行指令在被执行时使机器执行根据第一方面所描述的方法。

[0008] 提供发明内容部分是为了简化的形式来介绍对概念的选择,它们在下文的具体实施方式中将被进一步描述。发明内容部分无意标识本公开的关键特征或主要特征,也无意限制本公开的范围。

附图说明

[0009] 通过结合附图对本公开示例性实施例进行更详细的描述,本公开的上述以及其它目的、特征和优势将变得更加明显,其中,在本公开示例性实施例中,相同的参考标号通常

代表相同部件。

- [0010] 图1示出了可以在其中实施本公开的多个实施例的环境的示意图；
- [0011] 图2示出了根据一个实施例的指示搜索结果的方法的流程图；
- [0012] 图3示出了根据本公开的一个实施例的示例电子文档的示意图；
- [0013] 图4A至图4C示出了根据本公开的实施例的在终端处的用户界面的示意图；
- [0014] 图5示出了根据一个实施例的用于指示搜索结果的装置的框图；以及
- [0015] 图6示出了可以用来实施本公开的实施例的示例设备的框图。

具体实施方式

[0016] 下面将参考附图中示出的若干示例实施例来描述本公开的原理。虽然附图中显示了本公开的优选实施例，但应当理解，描述这些实施例仅是为了使本领域技术人员能够更好地理解进而实现本公开，而并非以任何方式限制本公开的范围。

[0017] 在本文中使用的术语“包括”及其变形表示开放性包括，即“包括但不限于”。除非特别申明，术语“或”表示“和/或”。术语“基于”表示“至少部分地基于”。术语“一个示例实施例”和“一个实施例”表示“至少一个示例实施例”。术语“另一实施例”表示“至少一个另外的实施例”。术语“第一”、“第二”等等可以指代不同的或相同的对象。术语“信息”、“数据”和“内容”可互换地使用。下文还可能包括其他明确的和隐含的定义。

[0018] 图1示出了可以在其中实施本公开的多个实施例的环境100的示意图。环境100包括搜索系统110，其用于提供搜索服务。搜索系统110可以连接到存储待搜索的电子文档的数据库120，以便从中检索出感兴趣的电子文档。如本文中所使用的，术语“电子文档”指的是任何机器可读格式的文件，包括但不限于pdf文件、txt文件、各种office文件、各种网页文件、电子邮件等等。数据库120可以是提供数据源的各类数据库、例如企业数据库、网站数据库、个人数据库等等。虽然仅示出一个数据库120，但是在其他示例中，搜索系统110可以访问多个存储电子文档的数据库。

[0019] 终端130可以利用搜索系统110查询期望访问的电子文档。例如，终端130可以将用户输入的查询项发送给搜索系统110。搜索系统110可以基于查询项在数据库120中查找相关的电子文档，并且然后将查找到的电子文档、文档的预览或者其可访问地址提供给终端130。终端130可以是任意类型的移动终端、固定终端或便携式终端。

[0020] 终端130和/或数据库120可以经由有线和/或无线连接而连接至搜索系统110。数据库120可以是搜索系统110的本地存储设备或者可以经由网络与搜索系统110通信。此外，虽然被单独示出，搜索系统110也可以被嵌入在终端130中。

[0021] 搜索系统110在将搜索结果、例如查找到的电子文档提供给终端130时，还可以指示搜索结果在终端130的用户界面中的呈现方式。为了使得用户可以从搜索结果中快速且方便地定位出感兴趣的内容，一种可能的方法是搜索系统110可以指示搜索结果中需要醒目显示的内容。

[0022] 在一些常规的用于醒目显示的方法中，搜索系统110会在搜索到的电子文档的正文(也被称为文本数据)中标识出与用户的查询项匹配的内容，并且指示终端130的用户界面对该内容进行醒目显示。对搜索结果这样的醒目显示使得用户可以更快注意到其感兴趣的内容或者判断搜索结果与查询信息的相关度。例如，如果搜索系统110搜索到与“self-

driving vehicle”有关的网页,则在将网页作为搜索结果呈现时,将该网页中出现的所有“self-driving vehicle”进行醒目显示。

[0023] 然而,这样的醒目显示方式还会缺失一些用户关心的内容。在很多情况中,用户不仅希望查找电子文档的文本数据,还希望获知未呈现在文本数据中的附加信息、诸如该文档的作者、格式、创建日期、更新日期、关键字、评论等等。这样的附加信息被统称为元数据,并且通常与电子文档相关联地被存储。例如,在文档协同处理的情况中,一篇电子文档可能会由多个用户共同编辑、修改和查看,并且不同用户可能会向与该电子文档相关联的元数据中添加很多有用的评论信息。搜索这篇电子文档的用户可能期望查看这部分元数据。

[0024] 发明人发现,虽然目前已经有搜索技术可以根据需要查找到包含用户感兴趣的元数据的电子文档,然而却缺乏提供对于元数据的醒目显示。因此,用户依然不能够从搜索结果中快速注意到与要查找的元数据内容或者判断出搜索到的电子文档的元数据部分与查询信息的相关度如何。

[0025] 在另外一些情况中,电子文档的文本数据可能以结构化数据格式、诸如可扩展标记语言(XML)格式被组织,并且与元数据类似地被存储、而不存在整体文本数据的概念。在这种情况下,对于这类电子文档的文本数据与元数据的醒目显示均会存在问题。

[0026] 根据本公开的示例实施例,提供了一种对搜索结果的呈现进行指示方案。该方案涉及对电子文档的元数据的醒目显示。图2示出了根据一个实施例的指示搜索结果的方法200的流程图。方法200可以由搜索系统、诸如搜索系统110来执行。

[0027] 在210,响应于接收到查询项,搜索具有与查询项相关的元数据的电子文档。通常,电子文档可以包含文本数据并且还具有相关联的元数据。文本数据是电子文档实际记载的内容。在一些情况中,电子文档可以是空白文档,并且因此文本数据部分是空。元数据可以与电子文档的文本数据相关联地存储(例如存储为单独的文件),并且可以指示与该电子文档有关的各类附加信息,诸如该文档的作者、格式、创建日期、更新日期、关键字、评论等等。

[0028] 应当理解的是,本公开的实施例对于电子文档的元数据具体包含的内容不作具体限制,因为这类数据可以根据需要被任意添加和删减。任何未被记载在电子文档的正文中、而是被另外以相关联的方式存储和提供的信息都可以被称为该电子文档的元数据。

[0029] 不同类型的元数据可以被存储在不同的数据字段中。例如,元数据可以包括与作者、格式、创建日期、更新日期、关键字和评论分别对应的数据字段。这些数据字段可以分别被修改、添加和更新。将会理解的是,不同电子文档的元数据可以不必包括全部的数据字段,而可以仅包括其中的一个或一些数据字段。图3图示了一个示例300,其包括记载文本数据的电子文档310以及它的相关联的元数据320。元数据320还包括创建时间字段322、评论字段324和关键字字段326。

[0030] 除了电子文档的文本数据之外,搜索系统110还可以访问电子文档的元数据并且确定元数据是否与查询项相关。查询项可以由用户输入,例如可以由终端130的用户输入并且由搜索系统110获得。查询项可以包括用户需要查询的关键字符、字、词、句子或字符串等,并且它的格式可以根据不同的搜索系统的配置而变化。搜索系统110接收到查询项之后,可以对查询项进行分析,以便能够从连接的数据库120中查找具有与查询项有关的内容的电子文档。在本文中,与查询项“相关”可以指该电子文档具有该查询项中的一个、一些或全部字符、字、词或它们的变形。

[0031] 在一些实施例中,接收到的查询项可以被标识为与元数据相关联并且因此搜索系统110可以直接从待搜索的电子文档相关联的元数据中查找与查询项有关的元数据。例如,终端130的用户在输入一个查询项时可以指定查询项是针对元数据或是针对文本数据的查询项。终端130可以提供特定的用户界面,其中给出不同的输入区域用于接收用户针对元数据或文本数据的查询项。如果从针对元数据的输入区域中接收到查询项,终端130可以标识该查询项与元数据相关联并且将其传送给搜索系统110。进一步地,查询项还可以被标识为与元数据的某个数据字段相关(例如为不同的数据字段提供输入区域),并且因此搜索系统110可以在每个待搜索的电子文档的元数据的特定数据字段中查找与查询项有关的数据。

[0032] 在一些实施例中,可以接收到针对元数据的多个查询项。根据用户或者搜索系统110对精确度的要求,搜索系统110可以搜索具有与全部查询项均相关的元数据的电子文档、或者可以搜索具有与其中一个或一些查询项相关的元数据的电子文档。

[0033] 在另外一些实施例中,搜索系统110还可以接收到针对元数据的(多个)第一查询项和针对文本数据的(多个)第二查询项,并且因此也可以查找满足一个或两个要求的电子文档、即具有与(多个)第一查询项有关的元数据和/或具有与(多个)第二查询项有关的文本数据的电子文档。备选地,查询项可以不被具体指定为与元数据相关联。因此,搜索系统110可以在待搜索的电子文档的正文内容和相关联的元数据中均进行查找。

[0034] 图4A图示了提供在终端130处的用于接收查询项的用户界面400。用户界面400包括用于接收针对正文的查询项的输入区域402和针对元数据的查询项的输入区域404。进一步地,在输入区域404中,还给出了针对不同类型的元数据(不同的数据字段)的特定输入框,包括针对作者字段的输入框406-1、针对创建日期字段的输入框406-2、针对评论字段的输入框406-3、针对关键字字段的输入框406-4等等。

[0035] 如果用户期望搜索正文包括“self-driving vehicle”、并且带有评论“prototype”和关键字“BI”的附加信息的电子文档,则可以在输入框402中输入“self-driving vehicle”、在针对评论的输入框406-3中输入“prototype”并且针对关键字的输入框406-4中输入“BI”。相应地,终端130可以获取针对正文的查询项“self-driving vehicle”和针对元数据中的评论字段和关键字字段的查询项“prototype”和“BI”并将它们提供给搜索系统110。

[0036] 应当理解的是,图4A仅示出了一种示例的用户界面和查询项的接收方式,并且在不同的实现中可以设计各种不同的用户界面用于接收查询项。在其他示例中,取代于提供对于不同查询项的多个输入区域,用户可以通过输入例如“查找正文‘self-driving vehicle’、带有评论‘prototype’和关键字‘BI’的电子文档”。终端130或搜索系统110可以通过分析该输入来确定用户的搜索意图,并且获得对应的查询项。

[0037] 搜索系统110可以采用多种搜索技术来针对电子文档的元数据(以及如果需要的话对文本数据)进行搜索,并且本公开的实施例在此方面不受限制。搜索系统110可以一旦在某个电子文档的元数据中查找到查询项(或者其变形),则确定该查询项与该元数有关。搜索系统110也可以在查找到一定数量的查询项的存在时,才确定该查询项与该元数据有关或者具有较高的相关性。本公开的实施例在确定元数据与查询项的相关性方面也不受限制。

[0038] 在本公开的实施例中,针对搜索系统110搜索到的具有与查询项有关的元数据的

电子文档,期望能够对元数据的某些内容进行醒目显示。因此,响应于在210搜索到具有与查询项有关的元数据的电子文档,在220,在电子文档的元数据中定位与查询项相匹配的元数据项。元数据项可以是元数据包括的一个或多个字符、字、词和/或句子等等。

[0039] 为了醒目显示的目的,需要首先定位元数据中需要被醒目显示的内容。通常,醒目显示是为了强调电子文档中用户关心的部分内容,从而使得用户能够快速且方便地注意到这部分内容。在搜索场景中,可以确定用户最关心的是与查询项匹配的内容。因此,当搜索系统110检索到与查询项有关的元数据之后,为了能够后续醒目显示这些元数据项,可以在元数据中确定与查询项相匹配的元数据项的位置。

[0040] 如本文中所使用的,“匹配”指的是查询项与元数据项完全相同或相似。例如,在相似的情况下,被定位的元数据项中的一个或一些字符、字或词可以是查询项的对应词的某种变型。举例而言,在语言为英语的电子文档中,元数据项可以包括查询项中的某个或某些单词的不同时态、单数或复数形式等等。通常,被定位的元数据项也是在搜索阶段搜索系统110查找到的用于确定元数据与查询项的相关性的项。在一些实施例,搜索系统110可以定位元数据中的与查询项匹配的一个或一些元数据项。备选地,搜索系统110可以定位出所有的匹配元数据项。

[0041] 在一些实施例,如上提及的,电子文档的元数据可以具有多个数据字段,并且查询项可能被标识为与其中的一个或一些数据字段相关联。因此,在定位元数据项之前,可以确定查询项与哪些数据字段相关联。如果确定查询项与其中的某个或某些数据字段相关联,则在所确定的数据字段中定位相匹配的元数据项。这样的方式可以提高定位效率和降低计算复杂度。例如,假设针对元数据的一个查询项“prototype”被标识为与评论字段相关联,则可以在该评论字段中定位匹配的元数据项。

[0042] 为了定位元数据,在一些实施例,搜索系统110可以处理元数据以获得索引。索引也可以被称为令牌(token),其可以与元数据包含的至少一个元数据项相关联。在一些示例中,为了获得索引或令牌,可能需要对元数据进行分词、语态分析等操作,以便恰当地理解元数据的实际含义并且丢弃对于含义没有贡献的助动词、虚词等等。例如,对于句子“Self-driving is quite hot in industry and a lot of technology companies have jumped into this field”,通过分词和语态分析,可以确定出索引“technology company”,其对应于该句子中的元数据项“technology companies”。搜索系统110可以基于索引来定位匹配的元数据项。例如,可以将查询项与索引进行比较。如果匹配,则查找索引对应的元数据项。

[0043] 为了获得索引而对元数据进行处理可以类似于对文本数据的常规处理。常规地,为了对文本数据进行醒目显示,也需要从文本数据中定位出与对应的查询项匹配的文本项。然而,由于文本数据通常较大,为了提高性能,通常会预先处理文本数据以获得索引,所获得的索引被存储(例如存储在数据库120中或搜索系统110可访问的其他存储空间中)以供在定位时使用。然而,这样需要牺牲存储空间。

[0044] 由于元数据的一些特点,可以不需要以牺牲存储空间为代价对元数据进行预先处理,而是可以根据需要实时处理。具体地,相比于文本数据而言,电子文档的元数据的大小通常较小,例如可能仅包括若干个词或句子。因此,即使实时处理元数据也不会降低处理速度。另外,对于元数据的多个不同数据字段而言,用户可能更关心且希望搜索其中的某些字

段而非全部字段。因此不必要对所有类型的元数据均继续预先处理和存储索引,以避免存储空间和处理资源的浪费。

[0045] 在一些实施例中,对于获得索引的处理也可以发生在搜索阶段。为了确定元数据是否与查询项有关,搜索系统110可以采用基于索引的搜索方式,即首先处理元数据以获得索引,并且检索与查询项对应的索引。在这些实施例中,搜索系统110在定位时也可以更方便地确定出对应的元数据项的位置。

[0046] 可以理解的是,如果搜索到的电子文档的文本数据中也包括与相应的查询项(与针对元数据的查询项相同或不同)相匹配的文本项,搜索系统110也可以定位出这些文本项。

[0047] 继续参照图2,在230,提供醒目显示被定位的元数据项的指示。搜索系统110可以指示请求检索的终端、例如终端130对被定位的元数据项进行醒目显示。如本文中所使用的,“醒目显示”可以指的是在视觉上强调包含在元数据中的所匹配的元数据项。可以使用多种方式对被定为的元数据项进行醒目显示。例如,可以对被定位的元数据施加加框属性、颜色属性、粗体属性、下划线属性或者在显示时能从其余元数据中区分出所定位的元数据项的其它属性。醒目显示的方式可以由搜索系统110指示给终端130、或者可以由终端130的用户预先选定。通过醒目显示的方式,用户能够方便地标识包含电子文档的元数据中的感兴趣内容。

[0048] 在一些实施例中,搜索系统110在向终端130提供搜索结果时,可以首先指示终端130呈现搜索结果的预览信息而不是全部、例如搜索到的电子文档的元数据的一部分。为了醒目显示与查询项匹配的元数据项,搜索系统110标识全部元数据中包括被定位的元数据项的一部分(也被称为第一部分),以便用于构成用于呈现的预览信息。除了元数据项之外,被标识的部分可以包括在元数据项的前面和/或后面的一些内容,以使得用户能够从预览信息中了解到元数据项的一些上下文。

[0049] 如果有多个被定位的元数据项,则可以相应地标识出多个部分以用于构成预览信息。这些部分可以是连续或非连续的。在一些示例中,还可以根据可呈现的预览信息的长度来确定标识出的那些部分的长度和/或数量。搜索系统110可以向终端130提供在所标识的(多个)部分的呈现中醒目显示元数据项的指示。根据这个指示,终端130可以在其用户界面中呈现被标识的部分,并且还将其中与查询项匹配的元数据项进行醒目显示。

[0050] 对于电子文档的文本数据,搜索系统110也可以类似地提供预览信息以及在预览信息中指示对匹配的文本项的醒目显示。具体地,在定位出电子文档的文本数据中与对应的查询项相匹配的文本项之后,搜索系统110可以标识出文本数据中包括被定位的文本项的部分(例如,第二部分)。搜索系统110可以将从元数据中标识出的第一部分和从文本数据在标识出的第二部分合并成预览信息,并且然后提供对合并后的第一和第二部分的呈现的指示。而且,在这样的呈现中要醒目显示匹配的元数据项和文本项。

[0051] 作为一个具体示例,搜索系统110接收到如图4A所示的查询项之后,查找到电子文档310。搜索系统110然后从电子文档310的文本数据和相关联的元数据320中标识出包括匹配的文本项和元数据项的部分以组成预览信息。搜索系统110向终端130提供指示,以指示在预览信息的呈现中醒目显示匹配的文本项和元数据项。图4B示出了终端130在接收到指示后对元数据项和文本项的醒目显示的用户界面410。

[0052] 如图所示,电子文档的一部分文本数据中出现的文本项“self-driving vehicles”和“self-driving vehicle”均被方框标记412-1和412-2分别醒目显示。此外,元数据的评论字段中元数据项“prototypes”和关键字字段中的元数据项“BI”也被方框标记412-3和412-4分别醒目显示。通过如用户界面410的呈现方式,用户不仅可以快速地注意到搜索到的电子文档的文本数据中感兴趣的内容,而且还可以注意到元数据中与查询项匹配的内容并且因此方便地判断出该电子文档的元数据部分也与查询项相关度较高。

[0053] 除了提供在预览信息中的醒目显示之外或者取而代之,搜索系统110可以向终端130提供在全部元数据的呈现中醒目显示匹配的元数据项的指示。在这样的情况中,如果存在的话,可以醒目显示更多的匹配元数据项。此时,虽然被呈现的内容更多,但是由于醒目显示的功能,用户仍然能够从大量的信息中方便地定位出那些匹配的内容。

[0054] 搜索系统110可以主动向终端130提供这样的指示。备选地或附加地,搜索系统110可以响应于请求或设置而提供这样的指示。在一些情况中,在呈现预览信息之后,用户可能期望查看电子文档的元数据的完整内容。用户可以通过点击电子文档的访问地址或者以其他方式发出请求。类似地,搜索系统110也可以提供在全部的文本数据的呈现中醒目显示匹配的文本项的指示。这样的指示也可以是主动或者响应于请求(设置)而被提供。

[0055] 举例而言,在图4B中,用户可以点击在用户界面410中显示的与预览信息对应的电子文件的访问地址414。响应于用户的操作,终端130可以向搜索系统110发出请求显示电子文档的完整文本数据和完整元数据内容的指示。应当理解的是,在其他示例中,可以针对电子文档的完整文本数据或完整元数据内容分别发起请求,并且相应地,搜索系统110可以以对应的指示进行响应。

[0056] 终端130接收到搜索系统110的指示之后,可以将完整内容(元数据和/或文本数据)呈现给用户,并且在这样的呈现中醒目显示匹配的项。图4C示出了终端130在电子文档的完整显示中对于匹配项的醒目显示的用户界面420。用户界面420可以包括对电子文档310的文本数据的呈现422和对元数据320的呈现424。在呈现424中,显示了与元数据的各个数据字段322-326对应的数据,包括创建日期432、评论434和关键字436。

[0057] 在呈现422中,除了在预览信息的呈现中被方框标记412-1和412-2醒目显示的文本项之外,还醒目显示了被方框标记412-5标识的另外的匹配文本项。在呈现424中,评论数据434和关键字数据436中匹配的元数据项也分别被方框412-3和412-4都强调出来。

[0058] 根据本公开的实施例,提供了对于搜索结果中的元数据进行醒目显示的方案。这样的方案弥补了已有的搜索结果呈现方案中的不足,使得用户能够对搜索结果的电子文档的各部分内容中的感兴趣项进行快速和方便的定位。此外,这样的方案还能够被应用到以各种结构化格式被存储的电子文档,使得对这类文档的文本数据和附加元数据的醒目显示都称为可能。

[0059] 图5示出了根据本公开的实施例的用于指示搜索结果的装置500。装置500可以用于实现图1的搜索系统110或者可以是搜索系统110的一部分。如图所示,装置500可以包括搜索部件510,用于响应于接收到查询项,搜索具有与查询项相关的元数据的电子文档。装置500还包括定位部件520,用于响应于搜索到电子文档,在电子文档的元数据中定位与查询项相匹配的元数据项。装置500进一步包括指示提供部件530,用于提供醒目显示元数据项的指示。

[0060] 在一些实施例中,定位部件520用于确定查询项与元数据的多个数据字段中的至少一个数据字段相关联,并且在所确定的至少一个数据字段中定位与查询项相匹配的元数据项。

[0061] 在一些实施例中,定位部件520用于处理元数据以获得索引,索引与元数据包含的至少一个元数据项相关联;并且基于获得的索引,定位与查询项相匹配的元数据项。

[0062] 在一些实施例中,装置500还包括标识部件,用于标识元数据中包括被定位的元数据项的第一部分。指示提供部件530用于提供在第一部分的呈现中醒目显示元数据项的指示。

[0063] 在一些实施例中,电子文档包括与接收到的另外的查询项相匹配的文本数据。在一些实施例中,定位部件520还用于在电子文档的文本数据中定位与另外的查询项相匹配的文本项,并且标识部件还用于标识文本数据中包括被定位的文本项的第二部分。装置500还包括合并部件,用于合并第一部分和第二部分。指示提供部件530还用于提供对合并后的第一部分和第二部分的呈现的指示,在对合并后的第一部分和第二部分的呈现中醒目显示元数据项和文本项。

[0064] 在一些实施例中,指示提供部件530用于提供在对元数据的呈现中醒目显示元数据项的指示。

[0065] 图6示出了可以用来实施本公开的实施例的示例设备600的示意性框图。设备600可以用于实现图1的搜索系统110或者可以是搜索系统110的一部分。如图所示,设备600包括中央处理单元(CPU)601,其可以根据存储在只读存储器(ROM)602中的计算机程序指令或者从存储单元608加载到随机访问存储器(RAM)603中的计算机程序指令,来执行各种适当的动作和处理。在RAM 603中,还可存储设备600操作所需的各种程序和数据。CPU 601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0066] 设备600中的多个部件连接至I/O接口605,包括:输入单元606,例如键盘、鼠标等;输出单元607,例如各种类型的显示器、扬声器等;存储单元608,例如磁盘、光盘等;以及通信单元609,例如网卡、调制解调器、无线通信收发机等。通信单元609允许设备600通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0067] 处理单元601执行上文所描述的各个方法和处理,例如方法200。例如,在一些实施例中,方法200可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元608。在一些实施例中,计算机程序的部分或者全部可以经由ROM 602和/或通信单元609而被载入和/或安装到设备600上。当计算机程序加载到RAM 603并由CPU 601执行时,可以执行上文描述的方法200的一个或多个步骤。备选地,在其他实施例中,CPU 601可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行方法200。

[0068] 本领域的技术人员应当理解,上述本公开的方法的各个步骤可以通过通用的计算装置来实现,它们可以集中在单个的计算装置上,或者分布在多个计算装置所组成的网络上,可选地,它们可以用计算装置可执行的程序代码来实现,从而可以将它们存储在存储装置中由计算装置来执行,或者将它们分别制作成各个集成电路模块,或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样,本公开不限制于任何特定的硬件和软件结合。

[0069] 应当理解,尽管在上文的详细描述中提及了设备的若干装置或子装置,但是这种

划分仅仅是示例性而非强制性的。实际上,根据本公开的实施例,上文描述的两个或更多装置的特征和功能可以在一个装置中具体化。反之,上文描述的一个装置的特征和功能可以进一步划分为由多个装置来具体化。

[0070] 以上所述仅为本公开的可选实施例,并不用于限制本公开,对于本领域的技术人员来说,本公开可以有各种更改和变化。凡在本公开的精神和原则之内,所作的任何修改、等效替换、改进等,均应包含在本公开的保护范围之内。

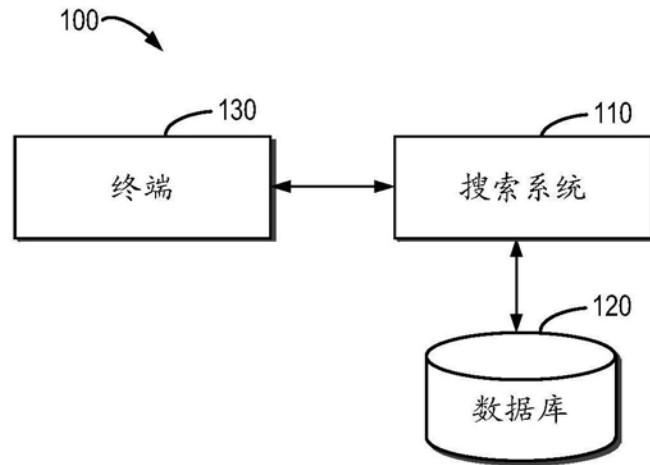


图1

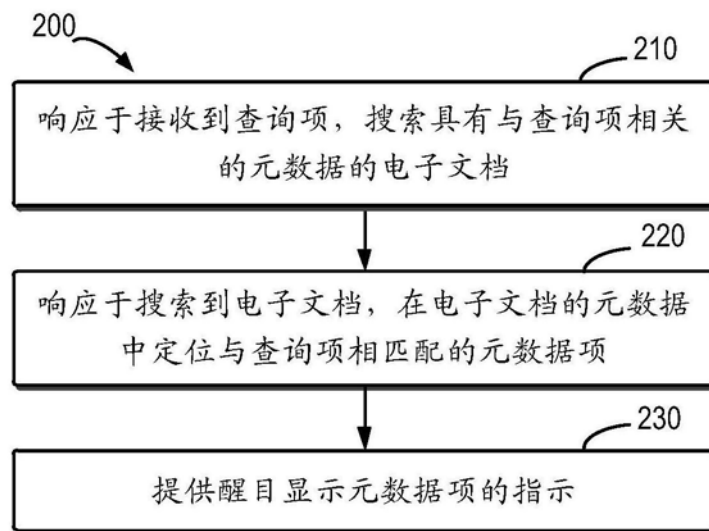


图2

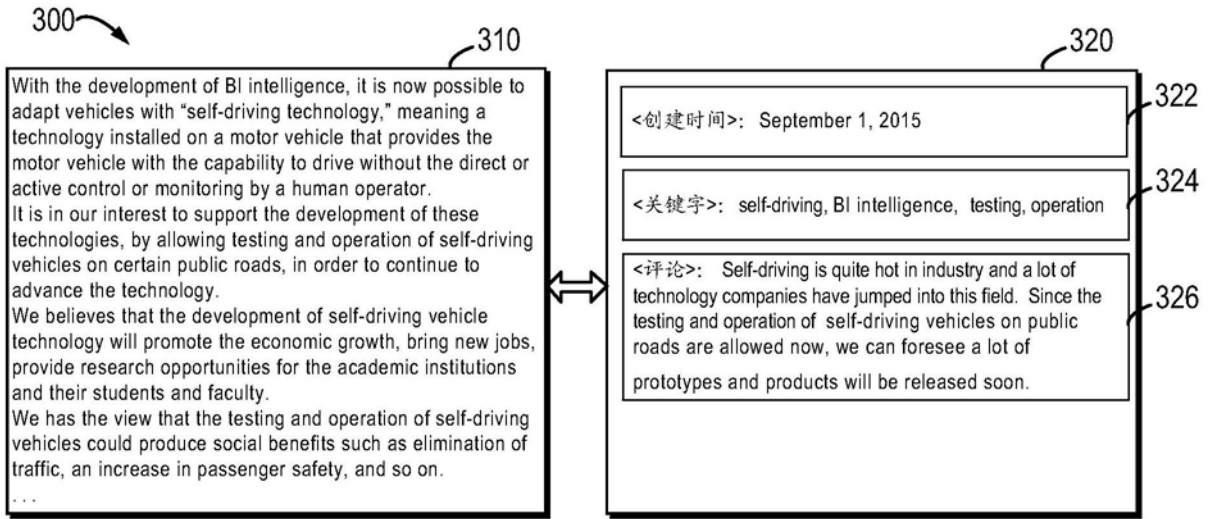


图3

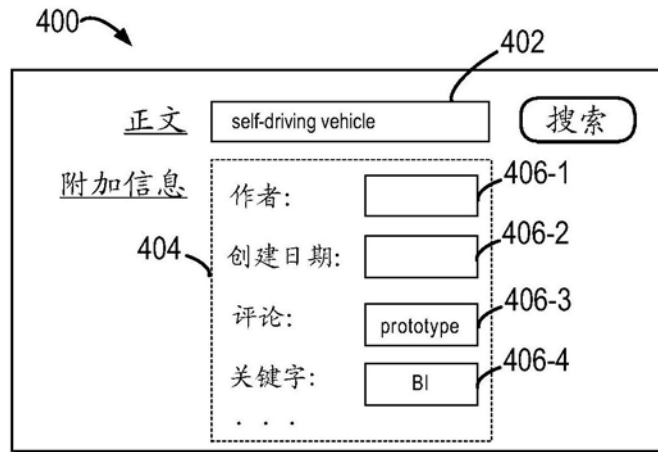


图4A

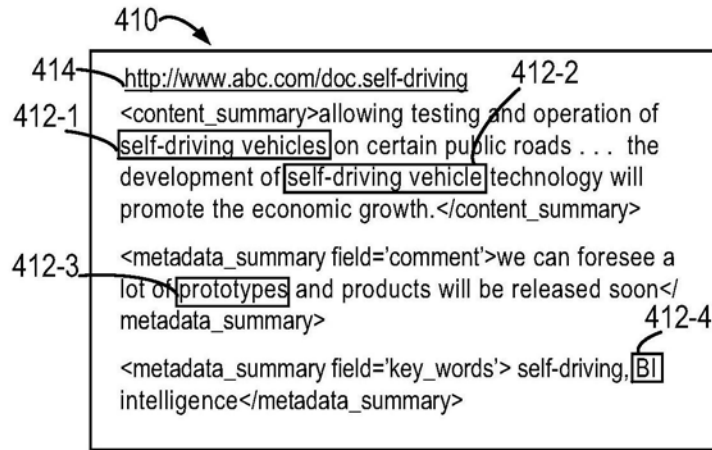


图4B

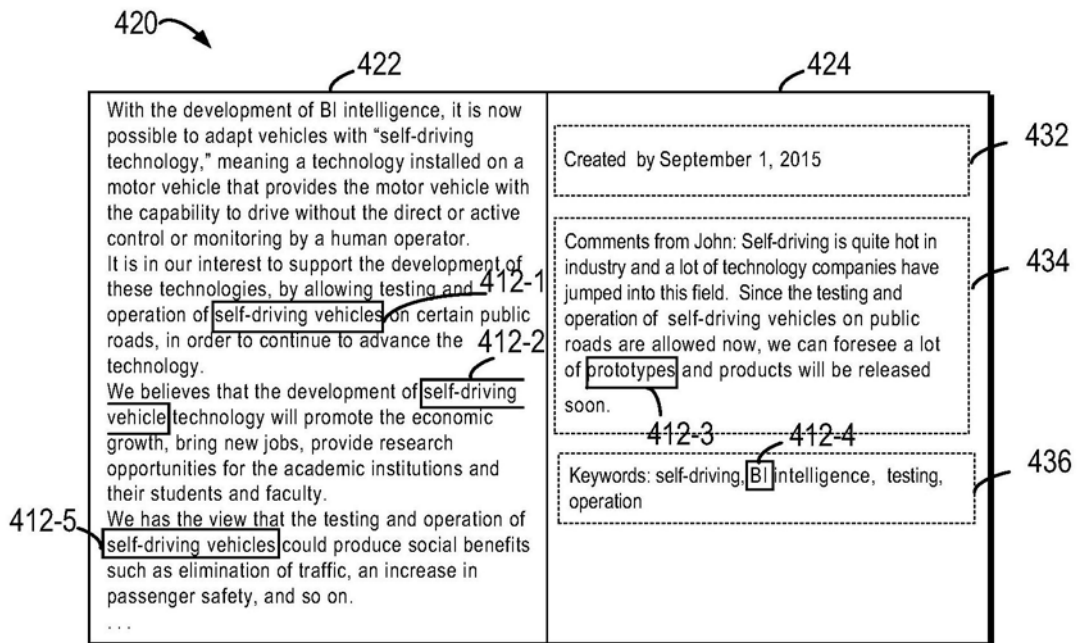


图4C

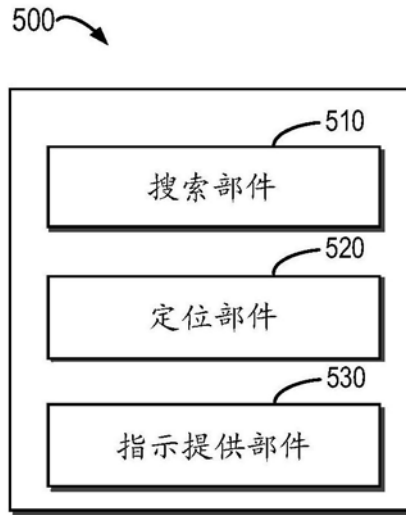


图5

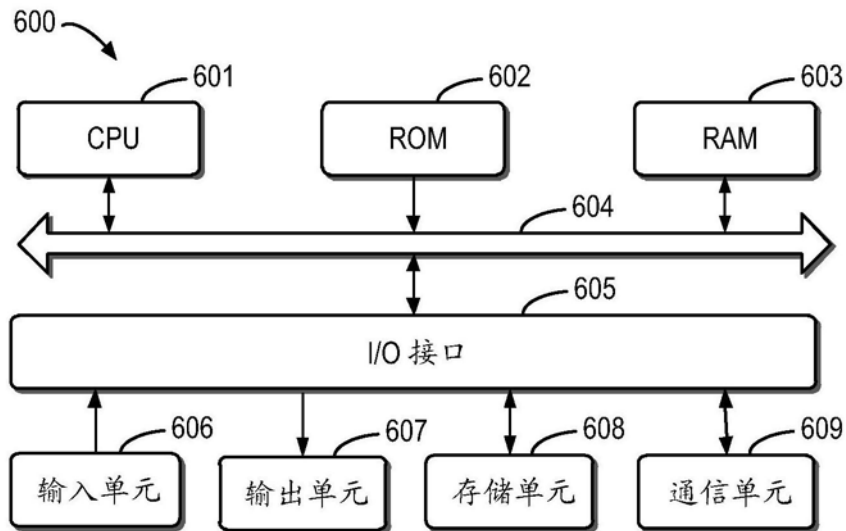


图6