



(12) 发明专利

(10) 授权公告号 CN 109690528 B

(45) 授权公告日 2023. 07. 14

(21) 申请号 201780035153.1

(22) 申请日 2017.03.28

(65) 同一申请的已公布的文献号
申请公布号 CN 109690528 A

(43) 申请公布日 2019.04.26

(30) 优先权数据
15/092,999 2016.04.07 US
15/266,573 2016.09.15 US

(85) PCT国际申请进入国家阶段日
2018.12.06

(86) PCT国际申请的申请数据
PCT/US2017/024527 2017.03.28

(87) PCT国际申请的公布数据
W02017/176505 EN 2017.10.12

(73) 专利权人 里德爱思唯尔股份有限公司雷克
萨斯尼克萨斯分公司
地址 美国俄亥俄州

(72) 发明人 R·D·米勒 J·A·迈尔斯
T·J·弗拉斯康 G·梅拉
R·C·海内斯 P·维特默 Y·李

(74) 专利代理机构 上海专利商标事务所有限公
司 31100

专利代理师 侯颖嫒 钱慰民

(51) Int.Cl.
G06F 16/34 (2019.01)

(56) 对比文件
CN 1934569 A, 2007.03.21
CN 102625937 A, 2012.08.01
CN 1533163 A, 2004.09.29
US 2014149403 A1, 2014.05.29
WO 2012129062 A2, 2012.09.27
US 2013304720 A1, 2013.11.14
JP 2004199696 A, 2004.07.15
US 2012204104 A1, 2012.08.09
US 2014365886 A1, 2014.12.11
US 2008071738 A1, 2008.03.20
JP 2014532239 A, 2014.12.04
赵华军等. 网页搜索结果聚类与可视化. 《南
京大学学报(自然科学版)》. 2010, (第05期),
吕月娥等. 基于信息类别的元搜索引擎设
计. 《微计算机信息》. 2007, (第21期),

审查员 张媛媛

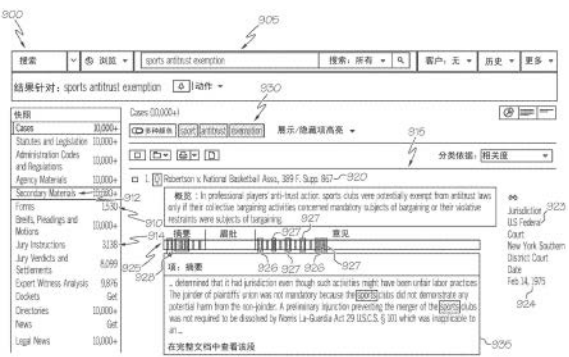
权利要求书3页 说明书17页 附图17页

(54) 发明名称

用于提供可视化结果列表的系统以及方法

(57) 摘要

本发明公开了用于响应于电子搜索请求来显示可视化结果列表的系统和方法。方法包括：访问从搜索查询得到的多个搜索结果中的每一个搜索结果的元数据；基于该元数据，利用针对第一单元和第二单元中的经分组的搜索项中的一个或多个搜索项的每一个搜索项的第一指示符和第二指示符来注释每个搜索结果中的一个或多个位置；以及显示包括了多个搜索结果和每个搜索结果的对应的命中图案的可视化结果列表。命中图案包括第一指示符和第二指示符。



1. 一种显示可视化结果列表的方法,所述方法包括:

由处理设备接收搜索查询,其中,所述搜索查询包括多个搜索项;

由所述处理设备将所述多个搜索项分组为多个单元,其中,所述多个单元中的每一个单元包括所述多个搜索项中的相关的一个或多个搜索项;

由所述处理设备来确定所述多个单元的数目是否大于单元的阈值数目;

当所述多个单元的所述数目大于所述阈值时:

基于每个单元的特性来向每个单元分配一个或多个点,来利用单元减少分值对所述多个单元中的每一个单元进行评分,其中每个单元的所述特性包括:由提交所述搜索查询的用户使用的搜索项、与具有语义相似性的一个或多个文档相关联的搜索项、与特定法律话题相关的搜索项、与特定文档语料库相关的搜索项、以及一个或多个其他用户所进行的搜索相关的搜索项,

基于所述多个单元中的每一个单元的所述单元减少分值,来对所述多个单元进行排名,以获取经排名的多个单元,以及

选择具有最高的单元减少分值的一定数目的所述经排名的多个单元,其中所述一定数目对应于所述阈值;

由所述处理设备访问对应于所述搜索查询的多个搜索结果中的每个文档的元数据;

由所述处理设备基于所述元数据,利用针对所述多个单元中的第一单元中的所述一个或多个搜索项的每一个搜索项的第一指示符和针对所述多个单元中的第二单元中的所述一个或多个搜索项的每一个搜索项的第二指示符来注释每个文档中的一个或多个位置;以及

由所述处理设备来显示包括所述多个搜索结果以及所述第一指示符和所述第二指示符的可视化结果列表;以及

在所述可视化结果列表中显示针对每个文档的命中图案,其中所述命中图案包括:

一个或多个部分,所述一个或多个部分对应于所述文档内的部分;

第一个或多个散列,所述第一个或多个散列对应于每个第一指示符;以及

第二个或多个散列,所述第二个或多个散列对应于每个第二指示符,其中所述第一个或多个散列和所述第二个或多个散列定位在所述命中图案中对应于每个文档中所述搜索项的位置的位置,并且

所述第一个或多个散列是与所述第二个或多个散列视觉上可区分的。

2. 一种显示可视化结果列表的方法,所述方法包括:

由处理设备接收搜索查询,其中,所述搜索查询包括多个搜索项;

由所述处理设备将所述多个搜索项分组为多个单元,其中,所述多个单元中的每一个单元包括所述多个搜索项中的相关的一个或多个搜索项;

由所述处理设备访问对应于所述搜索查询的多个搜索结果中的每个文档的元数据;

由所述处理设备基于所述元数据,利用针对所述多个单元中的第一单元中的所述一个或多个搜索项的每一个搜索项的第一指示符和针对所述多个单元中的第二单元中的所述一个或多个搜索项的每一个搜索项的第二指示符来注释每个文档中的一个或多个位置,其中所述第一单元中最高排名的搜索项是由所述处理设备已经确定为相对于其他项更可能与所述搜索查询相关的项;以及

由所述处理设备来显示可视化结果,所述可视化结果列表包括:

所述多个搜索结果,以及

针对所述多个搜索结果中的每一个文档的对应的命中图案,其中,针对每个搜索结果的所述命中图案包括:

所述文档的一个或多个部分,

第一个或多个散列,所述第一个或多个散列对应于每个第一指示符,以及

第二个或多个散列,所述第二个或多个散列对应于每个第二指示符,所述第一个或多个散列和所述第二个或多个散列定位在所述命中图案内对应于所述文档中的所述搜索项的所述一个或多个位置的一个或多个位置,所述第二个或多个散列是与所述第一个或多个散列视觉上可区分的,

框,所述框围绕所述第一个或多个散列和所述第二个或多个散列中的至少一者,并且所述框相对于所述命中图案的长度是可移动的,

以致于所述框能够移动到所述命中图案上的特定位置,以及

覆盖区域,所述覆盖区域位于所述命中图案的下方,所述覆盖区域显示来自所述文档的对应于所述框的定位的文本,并且高亮被显示文本内的搜索项,使得所述文本中对应于由所述框在所述命中图案上指示的所述特征位置的部分在所述覆盖区域内是可见的,并且所述搜索项被高亮。

3. 如权利要求2所述的方法,其特征在于,利用所述第一指示符和所述第二指示符来注释所述一个或多个位置包括:利用针对所述第二单元中的一个或多个排名最高的搜索项中的每一个搜索项的所述第二指示符来注释一个或多个排名最高的位置。

4. 如权利要求2所述的方法,其特征在于,所述命中图案经由所述第一个或多个散列和所述第二个或多个散列显示所述第一单元和所述第二单元中排名最高的搜索项。

5. 如权利要求3所述的方法,其特征在于,所述命中图案经由所述第一个或多个散列和所述第二个或多个散列显示所述第一单元和所述第二单元中排名最高的搜索项。

6. 如权利要求5所述的方法,其特征在于,所述文本包含与由所述框围绕的所述第一个或多个散列和所述第二个或多个散列中的所述至少一个相对应的注释。

7. 如权利要求5所述的方法,其特征在于,所述框进一步是相对于所述命中图案可移动的,以围绕所述第一个或多个散列和所述第二个或多个散列中的至少一个的第二区域,并且所述覆盖区域显示来自所述文档的对应于由所述框围绕的所述第二区域的经更新的文本。

8. 如权利要求6所述的方法,其特征在于,所述框进一步是相对于所述命中图案可移动的,以围绕所述第一个或多个散列和所述第二个或多个散列中的至少一个的第二区域,并且所述覆盖区域显示来自所述文档的对应于由所述框围绕的所述第二区域的经更新的文本。

9. 如权利要求2到8中的任一项所述的方法,其特征在于,所述命中图案进一步包括一个或多个命中集群指示符,所述一个或多个命中集群指示符对应于所述第一个或多个散列和第二个或多个散列的第一个或多个集群,所述第一个或多个集群相对于所述第一个或多个散列和所述第二个或多个散列的第二个或多个集群更紧密地与所述多个搜索项相关联。

10.一种用于显示可视化结果列表的系统,所述系统包括:

处理设备;以及

非瞬态处理器可读存储介质,所述非瞬态处理器可读存储介质与所述处理设备进行沟通,其中,所述非瞬态处理器可读存储介质包括一个或多个编程指令,所述一个或多个编程指令在被执行时使所述处理设备完成如前述权利要求中的任一项所述的方法。

11.一种显示可视化结果列表的计算系统,所述计算系统包括:

输入/输出硬件,所述输入/输出硬件用于从用户接收搜索查询,其中,所述搜索查询包括多个搜索项;

处理逻辑,所述处理逻辑耦合到所述输入/输出硬件,用于:

将所述多个搜索项分组为多个单元,其中,所述多个单元中的每一个单元包括所述多个搜索项中的一个或多个搜索项,

确定所述多个单元的数目是否大于单元的阈值数目,

当所述多个单元的所述数目大于所述阈值时:

基于每个单元的特性来向每个单元分配一个或多个点,来利用单元减少分值对所述多个单元中的每一个单元进行评分,其中每个单元的所述特性包括:由提交所述搜索查询的用户使用的搜索项、与具有语义相似性的一个或多个文档相关联的搜索项、与特定法律话题相关的搜索项、与特定文档语料库相关的搜索项、以及一个或多个其他用户所进行的搜索相关的搜索项,

基于所述多个单元中的每一个单元的所述单元减少分值,来对所述多个单元进行排名,以获取经排名的多个单元,以及

选择具有最高的单元减少分值的一定数目的所述经排名的多个单元,其中所述一定数目对应于所述阈值,

访问对应于所述搜索查询的多个搜索结果中的每个文档的元数据,以及

基于所述元数据,利用针对所述经排名的多个单元中的第一单元中的所述一个或多个搜索项的每一个搜索项的第一指示符和针对所述经排名的多个单元中的第二单元中的所述一个或多个搜索项的每一个搜索项的第二指示符来注释每个文档中的一个或多个位置;以及

显示器,所述显示器耦合到所述处理逻辑,用于显示包括所述多个搜索结果以及所述第一指示符和所述第二指示符的可视化结果列表。

用于提供可视化结果列表的系统以及方法

[0001] 相关申请的交叉引用

[0002] 本申请要求提交于2016年9月15日的题为“Systems and Methods for Providing a Visualizable Results Lists(用于提供可视化结果列表的系统以及方法)”的美国专利申请S/N.15/266,573的优先权,该申请要求提交于2016年4月7日的题为“Systems and Methods for Providing a Visualizable Results Lists(用于提供可视化结果列表的系统以及方法)”的美国临时专利申请S/N.15/092,999的优先权,以上申请的相应的全部内容通过引用被包含在本文中。

[0003] 背景

技术领域

[0004] 本说明书总的来说涉及提供搜索结果,并且更具体地涉及利用可视物来提供搜索结果以辅助用户确定哪些结果紧密地匹配用户的所期望的需求的系统以及方法。

技术背景

[0005] 当前,一旦在特定的数据库中进行搜索,就会向研究者呈现根据某一准则被排名的搜索结果(诸如日期、命中(hit)次数、经确定的相关度、结果类型等)。然而,该结果的经确定的排名可能不对应于研究者的所期望的排名方法。另外,该结果可能不向研究者提供根据研究者的所期望的排名方法进行重新排序的可视化方式。甚至,该结果不向用户提供来自该结果中的每次命中的足够的有关信息以使得用户能够确定命中是否是相关的。

[0006] 相应地,存在对于如下的系统以及方法的需求:该系统以及方法提供具有高延展性的搜索结果,且以对应于特定研究者的所期望的排名方法的对结果进行重新排序和重新排名的可视化方式提供搜索结果,并且向用户提供来自结果列表中的每次命中的足够的有关信息,以使得用户可以决定命中是否与他或她的搜索相关。

发明内容

[0007] 在一个实施例中,响应于电子搜索请求来显示可视化结果列表的方法包括:由处理设备来访问从搜索查询得到的多个搜索结果中的每一个搜索结果的元数据;由所述处理设备基于所述元数据,利用针对所述搜索查询中的一个或多个经分组的搜索项中的每一个搜索项的第一指示符和针对所述搜索查询中的第二单元中的一个或多个经分组的搜索项中的每一个搜索项的第二指示符来注释每个搜索结果中的一个或多个位置;以及由所述处理设备来显示包括了所述多个搜索结果和每个搜索结果的对应的命中图案的可视化结果列表。所述命中图案包括所述第一指示符和所述第二指示符。

[0008] 在另一个实施例中,响应于搜索请求来显示可视化结果列表的方法包括:由处理设备接收包括了多个搜索项的搜索查询;由所述处理设备将所述多个搜索项分组为多个单元,其中,所述多个单元中的每一个单元包括所述多个搜索项中的相关的一个或多个搜索项;由所述处理设备访问对应于所述搜索查询的多个搜索结果中的每个文档的元数据;由

所述处理设备基于所述元数据,利用针对所述多个单元中的第一单元中的所述一个或多个搜索项的每一个搜索项的第一指示符和针对所述多个单元中的第二单元中的所述一个或多个搜索项的每一个搜索项的第二指示符来注释每个文档中的一个或多个位置;以及由所述处理设备来显示可视化结果列表。可视化结果列表包括所述多个搜索结果和对于所述多个搜索结果中的每个文档对应的命中图案。所述命中图案包括所述文档中的一个或多个部分,第一个或多个散列对应于每个第一指示符,且第二个或多个散列对应于每个第二指示符。所述第一个或多个散列和所述第二个或多个散列定位在所述命中图案内对应于所述文档中的所述搜索项的所述一个或多个位置的一个或多个位置。

[0009] 在另一个实施例中,用于显示可视化结果列表的系统包括处理设备和与所述处理设备通信的非瞬态处理器可读存储介质。所述非瞬态处理器可读存储介质包括一个或多个编程指令,所述一个或多个编程指令当被执行时使得所述处理设备接收包括多个搜索项的搜索查询并且将所述多个搜索项分组为多个单元。所述多个单元中的每一个单元包括所述多个搜索项中的相关的一个或多个搜索项。非瞬态处理器可读存储介质进一步包括一个或多个编程指令,所述一个或多个编程指令当被执行时使得所述处理设备访问对应于所述搜索查询的多个搜索结果中的每一个文档的元数据,基于所述元数据,利用针对所述多个单元中的第一单元中的所述一个或多个搜索项的每一个搜索项的第一指示符以及针对所述多个单元中的第一单元中的所述一个或多个搜索项的每一个搜索项的第二指示符来注释每一个文档中的一个或多个位置,并且显示可视化结果列表。可视化结果列表包括所述多个搜索结果和对于所述多个搜索结果中的每个文档对应的命中图案。所述命中图案包括所述文档中的一个或多个部分,第一个或多个散列对应于每个第一指示符,且第二个或多个散列对应于每个第二指示符。所述第一个或多个散列和所述第二个或多个散列定位在所述命中图案内对应于所述文档中的所述搜索项的所述一个或多个位置的一个或多个位置。

[0010] 鉴于以下具体描述并结合附图将更完整地理解由本文中描述的实施例所提供的这些特征和附加特征。

附图说明

[0011] 在附图中阐述的实施例在性质上为说明性和示例性的,并不意图限制权利要求所定义的主题。可以在结合以下附图阅读时理解说明性实施例的以下详细描述,其中相同的结构用相同的参考标号表示,并且在附图中:

[0012] 图1是根据本文中所示出和所描述的一个或多个实施例的针对用于提供可视化结果列表的系统的说明性计算网络的示意性描绘;

[0013] 图2是来自图1的服务器计算设备的示意性描绘,进一步展示了根据本文中所示出和所描述的一个或多个实施例可以用于提供可视化结果列表的硬件和软件;

[0014] 图3描绘根据本文中所示出和所描述的一个或多个实施例的生成和呈现可视化结果显示的说明性概览的框图;

[0015] 图4描绘了根据本文中所示出和所描述的一个或多个实施例的用于搜索的说明性用户界面的屏幕截图;

[0016] 图5描绘了根据本文中所示出和所描述的一个或多个实施例的将项分组为单元的

说明性方法的流程图；

[0017] 图6描绘了根据本文中所示出和所描述的一个或多个实施例的选择最高得分单元的说明性方法的流程图；

[0018] 图7描绘了根据本文中所示出和所描述的一个或多个实施例的单元分值的说明性表格；

[0019] 图8描绘了根据本文中所示出和所描述的一个或多个实施例的显示利用单元来注释的结果的说明性方法的流程图；

[0020] 图9描绘了根据本文中所示出和所描述的一个或多个实施例的显示用单元所注释的搜索结果的说明性用户界面的屏幕截图；

[0021] 图10描绘了根据本文中所示出和所描述的一个或多个实施例的说明性搜索结果的详细的屏幕截图；

[0022] 图11描绘了根据本文中所示出和所描述的一个或多个实施例的说明性完整文档视图的屏幕截图；

[0023] 图12描绘了根据本文中所示出和所描述的一个或多个实施例的显示用单元所注释的搜索结果的替代性的说明性用户界面的屏幕截图；

[0024] 图13描绘了根据本文中所示出和所描述的一个或多个实施例的具有经重新排序的搜索结果的图12的替代性用户界面的屏幕截图；

[0025] 图14描绘了根据本文中所示出和所描述的一个或多个实施例的具有弹出显示的图12的替代性用户界面的屏幕截图；

[0026] 图15描绘了根据本文中所示出和所描述的一个或多个实施例的图12的替代性用户界面中的说明性搜索结果的详细的屏幕截图；

[0027] 图16描绘了根据本文中所示出和所描述的一个或多个实施例的具有多个扩展的条的图12的替代性用户界面的屏幕截图；

[0028] 图17描绘了根据本文中所示出和所描述的一个或多个实施例的具有原型变量控制器弹出框的搜索结果的图12的替代性用户界面的屏幕截图。

具体实施方式

[0029] 总的来说参考附图，本文中所描述的实施例针对向用户提供基于用户的搜索查询的搜索结果的可视化列表的系统以及方法。特别地，本文中所描述的系统和方法提供了从查询返回的文档的集合以及在汇总列表级别和/或文档级别描绘了“命中图案”的可视表示。本文中所描述的系统和方法进一步提供使用户能够可替代地基于一个或多个文档属性来对结果列表进行排名的一个或多个交互特征。另外，本文中所描述的系统和方法提供了来自用户所检索的每个文档的有关信息作为搜索的结果以使得用户能够更容易地确定文档是否与用户寻求获取的信息密切相关。

[0030] 如本文中将更详细地描述的，使用本文中所描述的系统和方法的研究者可以输入搜索短语（布尔(Boolean)或非布尔(non-Boolean)的)并且接收对应于搜索短语的结果列表，其可以是来自搜索短语的一个或多个令牌的使用的结果。对于结果列表中的每个结果，本文中所描述的系统和方法可以提供至文档的链接以及提供了关于文档内的搜索项的位置、文档内的搜索项的特定组的位置、文档部分等的附加信息的命中图案。此外，响应于一

个或多个输入,本文中所描述的系统和方法可以向研究者提供文档摘录,其中,该文档摘录与命中图案相关。可以在不需要用户单独地访问每个文档的情况下提供这样的摘录。如此,研究者能够在不必花费过多的时间阅读文档的剩余部分的情况下快速地确定文档是否与他/她的搜索密切相关。

[0031] 如从本公开中将显而易见的,本文中所描述的系统和方法可以特别地配置为,在通过提供特别地配置为准确提供有关信息(研究者需要该有关信息来作出关于搜索结果的知情的决定并且确定结果中所发现的特定的文档是否是相关的)的可延展的用户界面来实施搜索时,增加研究者的生产率。即,以对于用户是直观的方式提供结果,并向用户提供在必须获取所需的信息时更改显示的能力,其不同于之前的显示结果的方式(之前的显示结果的方式仅允许基于特定预设条件对结果列表进行重新分类和/或过滤)。另外,本文中所描述的系统和方法通过在无需附加的计算功率的情况下定制计算设备向特定的用户提供搜索结果的方法从而改进了计算设备的机能。如此,计算设备同时变成针对多个用户中的每一个用户特别定制的,而无需额外计算功率来这样做。

[0032] 虽然本文中所描述的系统和方法可以指与法律研究(包括判例法、法规、和/或法律论文研究)相关的特定的示例,但是应当理解到,这样的示例仅是说明性的并且本公开不限于这样的实例。相反,本文中所描述的系统和方法可以实施于任何研究领域,无论是否在本文中对这样的领域进行特别地描述。例如,可以在学术研究、科学研究、一般的基于网络的研究等的领域中实施该系统和方法。

[0033] 如本文中所使用的,术语“研究者”一般指使用本文中所描述的系统和方法来进行搜索的任何用户。如此,术语“研究者”和“用户”可以在本文中互换地使用。研究者的示例可以包括但不限于法律研究者、学术研究者、科学研究者、一般研究者等。

[0034] 如本文中所使用的,“文档”一般指搜索结果内的文本的集合。文档可以对应于电子文档或物理文档,但是其可以在用户界面(诸如网络浏览器屏幕等)内作为文本向用户显示。如此,文档一般是文档结果的集合内的单个搜索结果,如本文中所使用的这样的术语。说明性文档可以包括例如特定报道者所报道的那样的法庭意见的文本。

[0035] 现在参考附图,图1描绘了说明性计算网络,该说明性计算网络描绘根据本文所示出和所描述的一个或多个实施例的用于提供搜索界面以及可视化结果列表的系统的部件。如图1所展示的,计算机网络10可以包括广域网(WAN)(如,互联网)、局域网(LAN)、移动通信网络、公共服务电话网络(PSTN)、个人局域网(PAN)、城域网(MAN)、虚拟专用网络(VPN)和/或其他网络。计算机网络10一般可以配置成用于电子地连接一个或多个计算设备和/或其组件。说明性计算设备可以包括但不限于用户计算设备12a、服务器计算设备12b以及管理员计算设备12c。

[0036] 用户计算设备12a一般可以用作用户(例如,研究者)与连接到计算机网络10的其他组件之间的接口。因此,用户计算设备12a可以用于执行一个或多个面向用户的功能,如,接收来自用户的一个或多个输入或者将信息提供给用户,如本文中更详细描述的那样。相应地,用户计算设备12a可以至少包括显示器和/或输入硬件,如本文中更详细描述的那样。另外,图1中所包括的是管理员计算设备12c。如果服务器计算设备12b需要监督、更新或校正,管理员计算设备12c可以被配置成用于提供所需要的监督、更新和/或校正。管理员计算设备12c还可以用于将附加数据输入到存储在服务器计算设备12b上的语料库中。

[0037] 服务器计算设备12b可以从一个或多个源接收数据,存储数据,索引数据,搜索数据,和/或将数据以搜索结果、可视表示(诸如命中图案、文档、来自一个或多个文档的摘录等)的形式提供给用户计算设备12a。

[0038] 应当理解的是,虽然用户计算设备12a和管理员计算设备12c被描绘为个人计算机并且服务器计算设备12b被描绘为服务器,但是这些是非限制性示例。更具体地,在一些实施例中,任何类型的计算设备(例如,移动计算设备、个人计算机、服务器等等)可以用于这些组件中的任何组件。另外,虽然这些计算设备中的每个计算设备在图1中被展示为单件硬件,但是这也仅是示例。更确切地,用户计算设备12a、服务器计算设备12b和管理员计算设备12c中的每一个可以表示多个计算机、服务器、数据库、部件和/或类似物。

[0039] 图2描绘了来自图1的服务器计算设备12b,进一步展示了用于从一个或多个源接收数据、存储数据、索引数据、搜索数据、和/或提供数据的系统。另外,根据本文中所示出和所描述的实施例,服务器计算设备12b可以包括用于搜索数据的、被具体化为硬件、软件、和/或固件的非瞬态计算机可读介质。虽然在一些实施例中,服务器计算设备12b可以配置为具有必要硬件、软件、和/或固件的通用计算机,但是在一些实施例中,服务器计算设备12b还可以配置为用于执行本文中所描述的功能而专门设计的专用计算机。

[0040] 同样如图2中所示,服务器计算设备12b可以包括处理器30、输入/输出硬件32、网络接口硬件34、数据存储部件36(其可以存储分析数据38a、文档语料库38b、和其他数据38c)以及非瞬态存储器部件40。存储器部件40可以配置为易失性和/或非易失性计算机可读介质,并且如此,可以包括随机存取存储器(包括SRAM、DRAM、和/或其他类型的随机存取存储器)、寄存器、致密盘(CD)、数字通用盘(DVD)、和/或其他类型的存储部件。另外,存储器部件40可以配置为存储各种处理逻辑(诸如操作逻辑42、搜索逻辑44、列表显示逻辑45、和/或详细的显示逻辑48(作为示例,其中的每一个逻辑可以被具体化为计算机程序、固件、或硬件))。本地接口50也包括在图2中并且可以被实现为总线或其他接口以促进在服务器计算设备12b的组件之中的通信。

[0041] 处理器30可以包括被配置成用于接收和执行指令(诸如来自数据存储组件36和/或存储器组件40)的任何处理组件。输入/输出硬件32可以包括监视器、键盘、鼠标、打印机、照相机、麦克风、扬声器、触摸屏和/或用于接收、发送和/或呈现数据的其他设备。网络接口硬件34可以包括任何有线或无线联网硬件,如调制解调器、LAN端口、无线保真(Wi-Fi)卡、WiMax卡、移动通信硬件和/或用于与其他网络和/或设备通信的其他硬件。

[0042] 应当理解的是,数据存储部件36可以在服务器计算设备12b本地和/或远离服务器计算设备而驻留并且可以配置成用于存储一条或多条数据并且选择性地提供对所述一条或多条数据的访问。如图2中所示,数据存储部件36可以存储分析数据38a、文档语料库数据38b、和/或其他数据38c,如本文中更详细描述的那样。

[0043] 存储器组件40中所包括的逻辑包括运算逻辑42、搜索逻辑44、列表显示逻辑46、和/或详细的显示逻辑48。操作逻辑42可以包括用于管理服务器计算设备12b的组件的操作系统和/或其他软件。搜索逻辑44可以配置为从图形用户界面内的用户输入生成搜索查询。列表显示逻辑46可以生成和显示从搜索逻辑44揭露的搜索命中或文档的列表以及用于用户操控的可视表示,如以下所详细描述的那样。详细的显示逻辑48可以配置为生成和显示关于列表显示逻辑46所生成和显示的列表的特定元素的附加的细节,包括文档、命中图案、

和/或文档内的命中列表,如以下所详细描述的那样。

[0044] 应当理解的是,图2中所展示的部件仅是说明性的并且不旨在限制本公开的范围。更具体地,虽然图2中的部件被展示为驻留在服务器计算设备12b内,但是这是非限制性示例。在一些实施例中,这些组件中的一个或多个组件可以驻留在服务器计算设备12b外部。类似地,虽然图2涉及服务器计算设备12b,但是其他部件(诸如用户计算设备12a和管理员计算设备12c)可以包括相似的硬件、软件、和/或固件。

[0045] 图3描绘了提供可视化结果列表的说明性方法的框图。在步骤300处,接收查询。查询一般是由用户输入的一个或多个搜索项(例如,单词)。更具体地,用户可能期望发现与特定主题、特定的一组情形等有关的一个或多个搜索结果(例如,文档等)。如此,用户输入查询作为一组一个或多个查询项,用户相信该一组一个或多个查询项将导致搜索和用户所期望的搜索结果的检索。查询可以以现在已知或以后开发的任何格式(诸如,例如自然语言查询、布尔查询等)。自然语言查询是针对以自然语言(诸如,英语、西班牙语、法语、德语等)阐述的数据的请求。如此,用户可以通过单词或短语的任何组合来输入自然语言查询。在一些实施例中,用户可以以问题的形式来输入自然语言查询,即,用于可能期望检索将会回答被输入的问题的搜索结果。布尔查询是针对被阐述作为指示了项之间的关系的一个或多个特定的语法所分开的一个或多个项的数据的请求。说明性语法可以包括单词“和”(“AND”)以及“或”(“OR”)。

[0046] 为了接受查询,系统(或者其组件,诸如计算设备)可以提供用户界面中的搜索领域,其中,可以经由用户界面设备由用户将查询输入搜索领域。例如,如图4中所示,用户计算设备可以提供搜索领域作为在现实设备等上被显示给用户的用户界面400内的用户可修改的文本框405。用户可以经由一个或多个用户输入设备(诸如本文中描述的输入/输出硬件32(图2))来将文本输入到文本框405中。被输入到文本框405中的查询的非限制性示例可以是项“sports(体育)”、“antitrust(反垄断)”、和“exemption(豁免)”。这样的查询可以指示例如用户期望检索关于与体育相关的反垄断豁免的搜索结果。

[0047] 在一些实施例中,除用户可修改文本框405以外,计算设备还可以经由用户界面400提供附加的用户可选择的菜单项410以进一步调整用户的搜索查询。这样的用户可选择的菜单项410可以例如允许用户指定特定的位置以搜索查询项(诸如在所有可用的源中、仅在特定的数据库和/或文档类别内、在文档的特定的部分内等)。如图4中具体地示出的,用户可选择的菜单项410可以呈现为列出各种选项以供用户选择的下拉框。一旦用户已经将文本输入文本框405中并且选择了用户可选择菜单项410,则用户可以触发要被提交的搜索查询。这样的触发可以是例如当用户敲击键盘上的“输入”按钮、选择经由用户界面400呈现的选择按钮415等。

[0048] 应当理解到,图4中呈现的用户界面400仅是接收查询的一种可能的方式的说明性示例。如此,在不背离本公开内容的范围的情况下,可以构想到并且可能存在接收查询的其他方式。例如,查询可以是经由语音输入等的输入。

[0049] 再次参考图3,系统可以基于查询执行搜索,如在步骤305处所示的那样。可以在接收到搜索查询之后的任何时间完成执行搜索,并且执行搜索可以与关于图3在本文中所描述的一个或多个过程同时地发生。可以经由现在已知或以后开发的任何搜索方法来完成搜索的执行,并且不被本公开限制。

[0050] 在步骤310处,系统可以确定经接收的查询是否是布尔查询。这样的确定可以通过分析查询和确定查询是否包含一般被识别为布尔语法的特定的语法来完成。如本文中之前所描述的,这些语法可以包括但不限于诸如“和”以及“或”的单词。在一些实施例中,仅当特定的语法是以特定的格式(例如,全部大写、某些连接符、布尔语法等)时可以将查询识别为是布尔的。

[0051] 如果该确定指示了查询是布尔的,则可以基于查询的结构来对查询的各种单词进行分组。更具体地,可以将由“或”项分开的单词或短语识别为能够为了显示搜索结果的目的的被分组在一起作为单个元素的替代性搜索项。即,如果用户输入以下的布尔短语“basketball OR football(篮球“或”足球)”,则系统可以搜索含有两个单词之一的文档(对于布尔查询而言是通常的);然而,为了显示结果的目的,系统可以将单词“篮球”和“足球”分组在一起,就好像它们是用于提供可视化搜索结果的目的的同一项、与特定意思相关联的同一单元、和/或可互换的项,如本文中更详细地描述的那样。

[0052] 如果该确定是查询不是布尔的,则系统可以跳过步骤315并直接前进至步骤320。另外,一旦针对布尔查询完成步骤315,则系统可以前进至步骤320。在步骤320处,基于认识到为了显示结果的目的某些项是同义的或可互换的,对搜索项(例如,单词和短语)分组。即,类似于如关于步骤315描述的那样,仅是为了显示结果的目的而完成分组,且不影响搜索。如此,以通常的方式来进行搜索(即,通过现在已知或以后开发的搜索的任何方式来实现搜索)。

[0053] 可以在步骤320处通过访问一个或多个数据库对查询的项进行分组以确定分组。这样的数据库包括但不限于:通配符同义词词典数据库325、同义词数据库330、和短语数据库335。通配符同义词词典数据库325是包含了为了如本文所述的那样对搜索结果进行分组的目的而被识别为同义词的通配符项的数据库。例如,如果查询包括项“sport(运动)”,则通配符同义词词典数据库325可以包括与项“sport*”、“sport!”、“!sport”、“*sport”等相关的通配符的项,包括但不限于“sports(体育)”、“sporting(体育运动)”、“sportsmanship(体育道德)”、“sporty(体育的)”、“sportscast(体育转播)”、“sportswear(运动装)”、“multisport(多项目运动)”、“transport(运输)”、“passport(护照)”等。同义词数据库330是包含了单词的经识别的同义词,包括可能不是精确的同义词但是传达了与该项相似的意思的单词。例如,对于项“sport(运动)”,同义词数据库330可以包括精确的同义词(诸如“athletics(体育竞技)”、“game(比赛)”、“recreation(娱乐)”、“pastime(消遣)”等),以及传达了相似意思的其他单词(诸如“football(足球)”、“basketball(篮球)”、“tennis(网球)”、“hockey(曲棍球)”、“baseball(棒球)”等)。短语数据库335是包含了经识别的短语的数据库,即使这样的短语不受常用短语指示符(诸如,例如问号)约束。例如,短语数据库中的某些短语可以允许认识到当单词“basketball(篮球)”和“game(比赛)”在查询中连续出现时,尽管这些单词没有与短语指示符一起被输入,也将这些单词视作短语“basketball game(篮球比赛)”。在一些实施例中,短语数据库335可以是特定于特定领域(正在该特定领域中搜索查询)的。例如,如果搜索与法律数据库相关(例如,搜索判例法、法规、法条、法律论文等的搜索引擎),则短语数据库335可以是特定于特定法律术语(诸如“fee simple(无条件继承的不动产权)”、“beyond a reasonable doubt(无可置疑原则)”等)。这样的项可能被识别为或可能不被识别为法律情境之外的短语。在图5中示出为了显示结果的目的的

关于查询项的分组附加的细节

[0054] 在步骤500处,可以从查询中确定项。项可以是单词、短语等。如果查询是布尔查询,则可以由语法项/单词之一来将每个项分开。因此,例如,如果布尔查询是“antitrust exemptions AND sports OR basketball OR football(反垄断豁免“和”体育“或”篮球“或”足球)”,则可以确定项包括短语“antitrust exemptions(反垄断豁免)”与单词“sports(体育)”、“basketball(篮球)”以及“football(足球)”中的一个单词。如果查询是自然语言查询,除非某一标点符号出现在查询中,则可以初始认为查询中的每一个单词是一项。因此,如果自然语言查询是“what“antitrust exemptions”exist in sports, particularly football and basketball?(存在于体育中、特别是足球和篮球中的“反垄断豁免”是什么?)”,则将会认为查询中的每个单词是一项,除了“antitrust exemption”以外,因为其被约束在引号之间。

[0055] 在步骤505处,为了分组的目的,选择经确定的项。例如,如果确定来自查询的项是“antitrust exemptions”和“sports”,则可以选择“antitrust exemptions”或“sports”。如将在本文中更详细描述的那样,由于为了分组的目的,最终将选择查询中的所有的项,所以在步骤505处选择一项不限于特定的选择方法。例如,可能由于项出现在查询的第一位或由于项是查询中的下一项(当之前的项已经被选择时)所以选择该项。

[0056] 在步骤510处,可以作出关于该项是否是噪声单词或噪声短语的确定。噪声单词/短语一般是非关键的单词/短语,该非关键的单词/短语若被搜索时由于其共通性所以将可能不会缩窄搜索结果。如此,噪声单词和噪声短语是对于识别文档的主题而言一般无帮助的项和短语。噪声单词的非限制性示例包括“a(一)”、“the(该)”、“an(一个)”、“of(的)”、“to(至)”、“on(上)”等。如果项是噪声单词或噪声短语,则可以在步骤515处将其舍弃,并且过程可以返回至步骤505。一般将噪声单词和噪声短语舍弃,因为用户当看见搜索结果时可能将不对噪声单词/短语感兴趣,如本文中更详细地描述的那样。

[0057] 如果项不是噪声单词或噪声短语,则过程可以前进至步骤520。在步骤520处,可以在一个或多个数据库中搜索项。如以上之前所描述的那样,数据库一般可以是为了显示结果的目的特定于确定类似项的分组的数据库。即,数据库是用于确定查询分组的特定的数据库(包括通配符同义词词典数据库、同义词数据库、以及短语数据库)。

[0058] 在步骤525处,作出关于是否在一个或多个数据库中找到了针对该项的匹配的确定。如果没有找到匹配,则在步骤530处返回该项作为未分组的项。即,没有将该项与其他项分组到一起以用于如本文中所描述的那样显示搜索结果的目的。相反,将该项单独地指示为与其他单元不同,如本文中更详细地描述的那样。

[0059] 如果找到匹配,则在步骤535处对该项的同义词的列表进行解析,并且在步骤540处将其与剩余的项进行对比以确定匹配是否存在。例如,如果查询包括项“sports”、“basketball”、和“football”,并且在数据库中找到了针对项“sports”的匹配,则可以对数据库中的项“sports”的同义词进行解析,以确定这些同义词中的任何一个同义词是否匹配项“basketball”和“football”。如果基于步骤545处确定没有找到匹配,则在步骤530处返回该项作为未分组的项。

[0060] 如果基于步骤545处确定在查询和同义词列表中的剩余的项之间找到匹配,则在步骤550处对项进行分组。即,当显示搜索结果时,经分组的项将一起表现为相似颜色的指

示符等,如本文中所描述的那样。然而,为了实施搜索的目的,仍将分开地应用该项。在非限制性示例中,如果项“football”和“basketball”同时出现在“sports”的同义词的列表中时,将会把全部三个项分组到一起以使得搜索结果仅利用单种颜色的指示符显示这三个项的分组,而不是将三个项中的每一个项单独显示为不同颜色的指示符。

[0061] 在步骤555处,可以作出关于在查询中是否留有未被分组到一起或者返回为未分组的附加的项的确定。如果留有附加的项,则过程可以返回至步骤505。如果不留有附加的项,则可以在步骤560处提供经分组和未分组的(一个或多个)列表。提供(一个或多个)列表一般可包括将该列表存储在存储器中以用于进一步的访问,如将在本文中更详细地描述的那样。

[0062] 再次参考图3,一旦已经将项进行分组或确定不进行分组,就可以在步骤340处作出关于查询是否包含高于或低于特定阈值的数个单位的确定。单元指的是一组项或单个项(如果该项未被分组)。这样的确定可以是可期望的,因为过量的单元(例如,高于阈值数目的数量)可能导致结果列表表现为是凌乱的(即,太多不同颜色的指示符),并且可能使用户难以找到与用户的搜索目标密切相关的某些结果。虽然步骤340指5个单元的阈值,但是这样的数目仅是说明性的且本公开不限于此。即,在不背离本公开的范围的情况下,阈值可以是任何数目的单元。说明性的阈值可以是例如2个单元、3个单元、4个单元、5个单元、6个单元、或7个单元。

[0063] 如果查询具有大于该阈值的单位的数目,则可以在步骤360处对该查询进行分析并且可以确定单位减少分值。如本文中所使用的,单位减少分值是基于一个或多个准则分配至每个单元的分值。在一些实施例中,具有较高单元减少分值的单元相比于具有较低单元减少分值的单元更有可能是为了可视化搜索结果的目的而被显示的单元。

[0064] 可以通过访问一个或多个数据库和基于包含在数据库中的信息分配分值来确定每个单元的单元减少分值。说明性的数据库包括但不限于用户分析数据库365、结果元数据库370、个人用户历史数据库375、话题分类数据库380、文档语料库数据库385、以及相似用户历史数据库390。

[0065] 还参考图6,描绘关于减少分值的确定的附加的细节。在步骤605处,选择单元用于评分,并且在步骤610处作出关于该单元是罕见的单元或是用户常常使用的单元的确定。即,可查阅各种搜索历史(诸如用户的搜索历史)以确定该单元是否是过去已经被使用的项或一组项。可以通过访问一个或多个数据库(诸如,例如用户分析数据库365和/或个人用户历史数据库375)以及将数据库条目与来自该单元的一个或多个项进行对比来查阅搜索历史。如果该单元不是用户常常使用的(即,其是罕见项),则在步骤615处将一个或多个点分配给该单元。被分配的点的数目不受本公开限制,并且可以基于任何评分方法。在一个非限制性示例中,如果该单元(或者来自于其的搜索项)是罕见的话,则可以分配1个点。在另一个非限制性示例中,可以根据该单元过去被使用的频率的程度来分配点,以使得较频繁地被使用的单元相比于较不频繁地被使用的单元接收到更少的点。图7描绘了基于特定项组合被使用的频率来分配的点的说明性表格。在一些实施例中,如果该单元常常被用户使用,则可能不分配点。

[0066] 再次参考图3和图6,在步骤620处,作出关于该单元是否与具有语义上的相似性的项紧密相关联的确定。例如,如果单元包括项“adverse possession(逆权侵占)”,并且从搜

索结果中检索到的元数据指示大部分搜索结果涉及不动产(其在语义上与“adverse possession”相似,并且项“adverse possession”与其他不动产术语紧密接近地使用(例如,使用在同一个句子、同一个短语、同一个段落等中)),则可以确定该项具有高值,因为其与语义上相似的项在局部上(例如,在空间上)紧密接近。可以通过访问一个或多个数据库(诸如,例如结果元数据数据库370、话题分类数据库380、和/或文档语料库数据库385)作出这样的确定。可以分配较高值的项(例如,与语义上相似的项接近的项)分配,在步骤625处,可以将一个或多个点分配给单元。与以上描述的类似,被分配的点的数目不受本公开限制,并且可以基于任何评分方法。如果该单元不与语义上相似的项紧密接近,则可以不分配点。

[0067] 在一些实施例中,特别是在其中搜索是法律相关的搜索(例如,法规搜索、判例法搜索、法律论文搜索等)的实施例中,可以作出关于单元是否与特定的法学话题相关的确定。例如,可以将包含项“adverse possession”的单元识别为与不动产法相关而不是与另一个法律话题相关。可以例如通过搜索一个或多个数据库(诸如,话题分类数据库380和/或文档语料库数据库385)以及将数据库中的条目与来自该单元的一个或多个项进行对比来作出这样的确定。如果该单元与特定的法律话题相关,则可以在步骤635处将一个或多个点分配给该单元。与以上描述的类似,被分配的点的数目不受本公开限制,并且可以基于任何评分方法。如果该单元不与特定的法律话题相关,则可以不分配点。

[0068] 在步骤640处,作出关于该单元是否与文档语料库有关的确定。例如,如果该单元包含项“basketball”和“football”,则可以将其识别为与体育文档语料库有关。可以例如通过搜索一个或多个数据库(诸如,话题分类数据库380和/或文档语料库数据库385)以及将数据库中的条目与来自该单元的一个或多个项进行对比来作出确定。在一些实施例中,可以根据它们以什么程度与在其他文档内紧密接近的其他查询项共同出现来对来自话题分类数据库380和/或文档语料库数据库385的项进行加权。如果该单元与文档语料库相关,则可以在步骤645处将一个或多个点分配给该单元。与以上描述的类似,被分配的点的数目不受本公开限制,并且可以基于任何评分方法。如果该单元不与文档语料库相关,则可以不分配点。

[0069] 在步骤650处,作出关于该单元是否与已经被其他用户搜索的单元有关的确定。例如,如果多个用户已经搜索了项“basketball”、“football”、和“sports”,则可以确定包括了这些项中的一个或多个项的这样的单元与其他用户的搜索的单元相关。可以例如通过搜索一个或多个数据库(诸如,相似用户历史数据库390)以及将数据库中的条目与来自该单元的一个或多个项进行对比来作出确定。如果该单元与相似用户的搜索相关,则可以在步骤655处将一个或多个点分配给该单元。然而,这样的确定不受本公开限制,并且可以在不背离本公开的范围的情况下使用利用了其他资产的分析。与以上描述的类似,被分配的点的数目不受本公开限制,并且可以基于任何评分方法。如果该单元不与相似用户的搜索相关,则可以不分配点。

[0070] 在步骤660处,可以把被分配给该单元的点的数目记录并登录到文件中。因此,在步骤655中,可以作出关于是否要将点分配给附加的单元的确定。如果要将点分配给附加的单元,则可以在步骤605处重复该过程直到已经给所有的单元分配了点为止。

[0071] 如果附加的单元存在,则在步骤670处,可以根据包含了对分配给每个单元的点的数目的记录的文档的分值,将该文档进行重新排序和重新排名。例如,可以将具有较大数量

的点的单元排在具有较少数量的点的单元前面。然而,在不脱离本公开的范围的情况下,可以完成其他的记录方法。例如,记录可以根据特定公式而不是点之和。在步骤675处,可以将多达阈值量的最高得分单元选择为要显示的单元。例如,如果阈值是5个单元,则可以选择前5个单元。例如,如果阈值是3个单元,则可以选择前3个单元。

[0072] 回到图3,在步骤395处,保持最高得分单元。一般应当理解到,步骤395可以等同于图6的步骤675。虽然步骤395指示了得分最高的5个单元,但是应当理解到,阈值可以大于或小于5个单元,如之前描述的那样。因此,如果阈值是3,则仅保持最高得分单元中的3个单元。

[0073] 一旦单元的数目等于或小于阈值,则可以在步骤345处通过访问结果元数据数据库355来检索搜索结果中的每个文档的元数据。结果元数据数据库355一般是包含了出现在搜索结果中的文档中的每一个文档的元数据的数据库。即,与本文中所描述的步骤(例如,在步骤305处)同时被执行的搜索返回一个或多个结果。在一些实施例中,结果元数据数据库355可以是之前本文中所描述的相同的结果元数据数据库370。将结果中的每一个的元数据存储存储在结果元数据数据库355中。图8提供了关于检索元数据的附加的细节。

[0074] 在步骤805处,选择一个单元,并且在步骤810处,访问结果元数据数据库。可以经由接口来访问结果元数据数据库,如本文中所描述的那样。在步骤815处,对搜索结果的元数据进行搜索,以在步骤820处确定来自搜索结果中的每个文档内的单元的每个项的一个或多个位置。一旦已经确定了一个或多个位置,则可以在步骤825处对搜索结果中的每个文档进行注释以指示(一个或多个)位置。可以在步骤805处针对单元中的每一个重复该过程,如在步骤830处对附加的单元是否存在的确定所指示的那样。如此,将在每个文档中通过唯一的注释(例如,第一指示符、第二指示符等)对每个单元进行注释,以使得单元彼此基于注释是可区分的。一旦全部的单元已经用于搜索结果元数据并且注释了结果列表,则在步骤835中可以利用注释来显示结果。再次参考图3,可以将经注释的结果显示作为步骤350处的呈现显示的一部分。

[0075] 在图9中描绘了可能呈现的说明性显示。显示一般包括具有搜索栏905、结果快照部分910、和/或结果显示915的用户界面900。搜索栏905可以包含关于图4在本文中描绘的各种部件。

[0076] 结果快照部分910一般可以提供对划分成多种类型的文档的搜索结果的概览。例如,图9中示出的结果快照部分910将搜索结果划分成类别912(诸如:cases(案例); statutes and legislation(法规和法律); administrative codes and regulations(行政准则和规章); administrative materials(行政资料); secondary materials(次级法源性文献); forms(形式); briefs, pleadings, and motions(辩护状、诉状和动议); jury instructions(对陪审团的指示); jury verdicts and settlements(陪审团裁决及和解); expert witness analysis(专家证人分析); dockets(诉讼事件表); directories(目录); news(新闻); legal news(法律新闻); scientific(科学); company and financial(企业与金融); 以及 web categories(网络类别))。类别912是非限制性示例。如此,在不脱离本公开的范围的情况下,可以呈现其他的类别。另外,结果快照部分910可以展示呈现于类别912中的每一个类别中的搜索命中的数目914。例如,结果快照部分910指示了已经将1530个搜索结果分类成形式。在一些实施例中,结果快照部分910可以是交互式的以使得用户能够选择

要展示的特定的结果。例如,如果用户仅期望看到与法规和法律相关的搜索结果,则用户可以选择(例如,点击)类别912中的“statutes and legislation(法规和法律)”,这可以使得结果显示915仅显示来自该特定类别的搜索结果,如本文中更详细地描述的那样。

[0077] 结果显示915一般包括依据用户所提交的查询而完成的搜索的结果的列表。可以将每个搜索结果920显示在结果显示915内的一条或多条线上。被显示的每个搜索结果920的内容不受本公开限制,并且能够是任何类型和任何数量的内容,包括标题、引用、关于详细描述的文本文本、关于概览的文本文本等。在其中对法律数据库进行搜索的实施例中,每个搜索结果920可以包括例如案例名称921和/或报道者引用922。每个搜索结果920所包括的其他信息可以包括例如管辖区域923、日期924等。在一些实施例中,每个搜索结果920可以是用户可扩展的以用于揭露附加的信息。例如,用户可以选择(例如,点击)搜索结果920中的链接以显示关于该结果的附加的文本文本(例如,完整文档等)和/或将鼠标指针悬停在搜索结果920的一部分上以查看附加的信息。如此,术语“搜索结果”和术语“文档”可以在本文中互换地使用。即,搜索结果920可以是如本文中所述的那样的被发现作为搜索的结果的文档。

[0078] 作为本文中关于图3、图6和图8所描述的步骤的结果,每个搜索结果920还可以包括命中图案925。总的来说,命中图案925向用户提供可视化,该可视化指示了搜索结果920内的文档的经编程确定的最相关的部分的各种搜索项的位置。如图9中所示,命中图案925可以包括搜索结果的各种部分的明细。例如,如果搜索结果920是法庭意见,则命中图案925可以显示法庭意见的各种部分(诸如摘要部分、眉批(例如,关键字)部分、意见部分、和/或各种其他部分)。这样的部分仅是说明性的,并且应当理解的是,在不背离本公开范围的情况下可以使用替代性的、附加的、或更少的部分。在该部分中的一个或多个部分内,命中图案925还可以显示一个或多个散列926。散列926一般可以对应于已经在文档中被注释的指示符,并且指示了每个部分内的搜索项的位置,这允许用户快速地和容易地看出搜索项出现在文档中的哪个位置以及搜索项的集群。即,搜索项的集群可以指示对于用户而言比较感兴趣的文档的一个区域而不是与其他项隔离的单个搜索项。

[0079] 如本文中所述的,可以显示多达特定阈值数目的单元的数目。为了在命中图案925内将每个单元彼此区分,散列926可以是颜色编码的或以其他方式包含根据每个散列926表示的单元的区别特征。如此,用户界面900可以进一步包括指示了哪种颜色对应于每个单元的可选择的图例930。例如,如图9中所示,图例930可以指示针对项“sport(运动)”的第一颜色、针对项“antitrust(反垄断)”的第二颜色、和针对项“exemption(豁免)”的第三颜色,可以对其进行切换以便显示搜索结果920内的对应的高亮处,如本文中更详细地描述的那样。如此,在仅显示了关于每个搜索结果920的某种信息的第一视图中(例如,被发现作为搜索的结果的多个案例中的每一个案例的高级快照),利用对应的颜色将每个搜索结果920内的某些项突出显示,该颜色还可以指示:项出现在关于文档的哪个位置;项是聚集的或是散布的;以及项相对于彼此在哪个位置出现。另外,当单元包含多个项时,将利用相同的区别特征对该项中的每一个项进行编码(例如,项“football”、“basketball”和“sports”全部可以具有相同的颜色散列926,并且可以以相同的方式突出显示在每个搜索结果920内)。

[0080] 在各种实施例中,在如图9中描绘的第一视图内,散列和经突出显示的项可以仅指示文档中的排名最高的项、文档的特定部分等。即,“排名最高的”项是已经被确定为更可

能与特定用户的搜索相关的项,并可以在搜索结果920内注释该“排名最高的”项,如本文中更详细地描述的那样(例如,图8中的步骤825)。因此,排名最高的项可以保证在不包含搜索结果920的全部信息的特定搜索结果920的高级视图中向用户显示,如下文中所描述的那样。在一些实施例中,可以更进一步地利用区别特征928区分第一排名最高的部分。第一排名最高的项一般可以是来自全部排名最高的项的一项,该项最可能与用户的搜索相关(即,比其他排名最高的项具有更高的排名),其可以如本文中更详细地描述的那样被确定。即,将第一排名最高的项设置为与剩余的排名最高的项分离,以使得用户的注意力被吸引到该第一排名最高的项上。用于指示第一排名最高的项的目的的区别特征928不受本公开限制并且可以是任何区别特征。例如,如图9中所示,区别特征928可以是星形等。

[0081] 在一些实施例中,被显示在图9中的命中图案925可能不指示搜索结果920内的每一个命中,特别是在搜索结果920包含大量的命中的实例中。相反,命中图案925可以仅显示预定数目的“最相关”命中(诸如,被确定为相对于其他命中而言对于用户是最可能感兴趣的命中、排名最高的项等)。在这样的实施例中,当用户选择搜索结果920以查看整个搜索结果920时,可以在命中图案925中提供全部的命中,如本文中更详细地描述的那样。

[0082] 虽然图9总的来说涉及使用了有颜色的散列926以区分文档中的单元的颜色编码的命中图案925,但本公开不限于此。例如,在一些实施例中,命中图案925可以对于散列926利用各种不同的底纹以在单元之间进行区分,或者可以在不背离本公开的范围的情况下利用在单元之间进行区分的其他方式。

[0083] 在一些实施例中,如果用户期望获取关于搜索结果920中的文档的附加的信息,但是不想访问整个文档并滚动遍历整个文档以找到命中,则用户可以选择(例如,点击或将鼠标悬停在其上)散列926和/或散列926的集群(例如由框927约束的集群)中的一个或多个散列以使覆盖区域935出现在结果显示915上。覆盖区域935可以展示例如一种文档的文本的摘录版本,在该摘录版本中,搜索项被突出显示,以使用户能够快速确定摘录区域是否与搜索有关并且在不必回到之前的屏幕来浏览、重新运行搜索等的情况下转向另一些部分或文档。

[0084] 在一些实施例中,框927可以相对于命中图案是可移动的,以使得用户能够选择性地框927移动至命中图案925上的所期望的位置。例如,框927是一种滑块型框,用户能够致动该滑块型框并且沿命中图案925移动该滑块型框以使得该滑块型框包围命中图案925内的感兴趣的特定区域。即,用户能够将框927滑动至感兴趣的特定区域(诸如包含了散列926的特定集合的区域、具有散列926的特定的颜色/底纹的集合的区域、具有散列926的不同的颜色/底纹的特定分组的区域、利用区别特征928指示的区域等)。可以实时地更新覆盖区域435的内容以使得覆盖区域435显示对应于框927所约束的各种散列926的搜索结果920的经更新的文本。即,系统配置为:允许用户基于散列926的集群(或其他因素)来经由框927立刻选择搜索结果920内的一个或多个感兴趣的特定区域,并立刻看见搜索结果920内的对应的文本,而不必从结果显示915中的全部的结果中选择搜索结果920、等待搜索结果920的全文被显示、并随后滚动至感兴趣的特定区域。在一些实施例中,框927可以仅当特定文档被完全打开(即,不是对搜索结果的列表内的文档的预览)时是可操作的。

[0085] 框927的初始位置(即,在用户移动框927之前框927所在的位置)可以指示:例如已经基于单位的集群被确定为“最佳摘录”的搜索结果920的摘录部分、搜索结果920的特定部

分、经典和/或标志性案例中公认的搜索结果920的一部分、包含公共政策的讨论的搜索结果920的一部分、具有超出内容的阈值的相关性的搜索结果920的一部分等。在一些实施例中,被显示在覆盖区域935内的文本的一部分可以与剩余的文本在视觉上相区别(例如,突出显示、加粗文本等)或者使其突出(例如,使字符实体突出,使短语突出,使文档参考突出等)。在由用户选择包含多个不同单元的集群的实施例中,覆盖区域935中突出显示的区别特征(例如,颜色)可以对应于用于特定单元的散列926的颜色,如图10中所示。如此,用户能够快速地区分覆盖区域935中的项以获取所需的信息。

[0086] 参考图9,在一些实施例中,图例930可以包含将允许用户选择或取消选择某些单元的复选框或类似物。例如,用户可以从图例930取消选择特定单元(例如,通过点击该单元),这可以使对应于该单元的命中图案925中的散列926消失。类似地,将不再在覆盖区域935中突出显示对来自该单元的项的突出显示。然而,应当理解到,从图例930选择或取消选择特定单元仅影响结果的显示,并且不更改实际搜索(即,来自该单元的项仍呈现为搜索项,并且搜索结果仍然将项包含在其中)。

[0087] 在一些实施例中,用户可以期望看见整个文档而不是来自文档的特定摘录。在这样的实施例中,用户可以选择(例如,点击)搜索结果920(例如,文档标题等),这可能导致完整文档视图1100,如图11中所示。在一些实施例中,用户可以点击或以其它方式选择覆盖区域935,这可以将如图11所示的完整文档视图1100打开至在覆盖区域935中包含了经摘录的文本的文档内的部分。与图9和图10中所示的结果视图类似,图11的完整文档视图1100也显示了具有散列926的命中图案925,并且以对应于散列926的颜色来突出显示文档显示区域940内的搜索项。然而,如上文中所描述的,在图9中所显示的命中图案925不指示搜索结果920内的全部命中的实施例中(例如,因为搜索结果920具有大量的命中),图11中所提供的视图可以更新命中图案925以显示搜索结果920内的全部命中。如此,向用户提供关于搜索结果920内的命中的更多的细节。

[0088] 在一些实施例中,完整文档视图1100可以进一步在命中图案925上或毗邻命中图案925处显示一个或多个命中集群指示符929,该一个或多个命中集群指示符929对应于指示了相比搜索结果920内的其他命中而言与特定搜索结果更加紧密地相关联的命中的散列926的集群。如此,命中集群指示符929可以允许用户发现可能更紧密地对应于正被搜索的信息的搜索结果920的区域。如此,用户可以移动框927以使得其约束由命中集群指示符929指示的散列926的区域,并且查看搜索结果920内的对应的文本。结果,用户可以能够更快地定位搜索结果920内的感兴趣的区域,而不必花费太多的时间和/或努力滚动遍历搜索结果920以找到这样的区域。

[0089] 作为提供如本文中关于图9-11所描述的用户界面900的结果,用户可以能够更好地使选择和/或排名每个搜索结果920用于显示的原因可视化,使基于用户输入的搜索项与用户最相关(或更相关)的搜索结果920的部分可视化,使可以让一个特定搜索结果在结果的列表上排名比另一个特定搜索结果要高的原因可视化,使超过阈值的摘录的数量可视化,和/或使摘录中的项的数量可视化(即,越高的项的数目导致越高的排名)。在排名的原因原本不会对用户引起注意的实例中,这是尤其有可能的。另外,用户界面900指示了特定搜索结果920的内容相对于特定用户可以事实上在他/她的研究中如何使用该内容的差异。

[0090] 应该理解到,关于图9-11所描述的用户界面900仅是说明性的,并且利用了从本文

中描述的步骤中获取到的信息以向用户提供经排序的搜索结果的其他的用户界面被构想到并且是可能的。例如,在一些实施例中,如图12中所示,替代性用户界面1200可以包括针对一个或多个搜索结果1220中的每一个搜索结果的替代性命中图案部分1225。更具体地,替代性命中图案部分1225可以是指示了对于每个单元的一个或多个搜索结果1220中的每一个搜索结果中的命中的总数目的条形图,该柱状图也被称为“迷你图”可视化。即,第一条1250可以指示针对第一单元的命中的总数目,第二条1255可以指示针对第二单元的命中的总数目,以及第三条1260可以指示针对第三单元的命中的总数目。条越大(即,条的长度越长),特定搜索结果中呈现的命中的数目越多。例如,如图12中所示,搜索结果“J.S.v.Blue Mt.Sch.Dist.”指示替代性命中图案部分1225,其中第二条1255具有相比于第一条1250和第三条1260更多数量的命中。如图例1230所指示的,第二条1255对应于搜索项“school(学校)”。如此,用户一旦查看替代性命中图案部分1225就可以认识到搜索结果“J.S.v.Blue Mt.Sch.Dist.”经常地使用项“school”。在一些实施例中,可以根据条1250、1255、1260的相应的比例来确定它们的尺寸。即,如果搜索结果使用项“school”的频率是项“internet(因特网)”的两倍,则第二条1255(对应于项“school”)可以是第三条1260(对应于项“internet”)的两倍高。在一些实施例中,条1250、1255、1260可能涉及针对整个文档的命中、特定部分或多个部分中的命中等。

[0091] 再参考图13,在一些实施例中,替代性用户界面1200可以通过在替代性命中图案部分1225中指示对哪些项更感兴趣来向用户提供对搜索结果1220进行重新排序的能力。即,如果用户期望看见在该结果的顶部包含了最常提到的项“school”的搜索结果1220,则用户可以选择重新排序按钮1265中的对应的一个重新排序按钮(例如,重新排序按钮1265的中心的一个重新排序按钮),这将相应地对搜索结果1220进行重新排序。类似地,如果用户期望对搜索结果1220进行重新排序以指示哪些搜索结果在该结果的顶部包括了最常提到的项“school”和项“internet”,则用户可以选择重新排序按钮1265中的对应的两个重新排序按钮,这将相应地对搜索结果1220进行重新排序。应该理解到,重新排序按钮1265还可以用于对搜索结果1220进行重新排序以在该结果的顶部示出最小数量的命中,并还可以用于对搜索结果1220进行重新排序以在结果的顶部示出具有分布最均匀的数量项的结果(例如,各种项呈现为基本上彼此相同的数量)的结果。在一些实施例中,重新排序按钮1265还可以用于恢复默认相关度排名或原始相关度排名(即,在初始被提供给用户的初始的相关度排名)。

[0092] 现在参考图14,在一些实施例中,替代性用户界面1200在被用户激活时可以呈现弹出显示1270。可以例如当用户悬停在替代性用户界面1200的特定区域(诸如,搜索结果1220中的一个搜索结果)上时激活弹出显示1270。激活弹出显示1270的其他方式可以被构想到且是可能的。弹出显示1270一般可以呈现关于搜索结果1220中的特定的一个搜索结果的附加信息而无需用户完整地打开文档以查看其内容。例如,如图14中所示,弹出显示1270可以提供关于呈现在搜索结果1220中的文档的概览部分的信息。在一些实施例中,可以在弹出显示1270的文本中突出显示各种项。

[0093] 作为对于弹出显示1270的替代物,在一些实施例中,替代性用户界面1200可以扩展以显示详细的结果部分1275,如图15中所示。详细结果部分1275可以与弹出显示1270(图14)和/或覆盖区域935(图9)相似。即,详细结果部分1275可以提供关于搜索结果1220中的

特定的一个搜索结果的附加信息而无需用户完整地打开文档以查看其内容。例如,如图15中所示,详细结果部分1275可以提供来自搜索结果1220的文档的文本的一个或多个经选择的部分,诸如,例如可被突出显示的包含了搜索项的部分,如在本文中更详细地描述的那样。另外,用户可以选择条1250、1255、1260中的一个条,以使详细结果部分1275显示包括了对应于条1250、1255、1260中的经选择的一个条的关键词的文档的文本。例如,如果用户选择对应于搜索项“internet”的第三条1260,则结果部分1275可以显示在其中包括了项“internet”的文本的一部分。另外,可以将项“internet”突出显示在被显示的文本内。

[0094] 在一些实施例中,替代性用户界面1200可以提供可浏览的扩展的详细结果部分1275,如图16中所示。更具体地,当扩展详细结果部分1275时,条1250、1255、1260可以各自扩展成多个条,由此允许用户选择多个扩展的条1280、1285、1290中的一个或多个条,以显示详细结果部分1275中的文档的文本的各种部分。这样的实施例可以允许用户通过在文档中的摘录和/或各种位置之间进行浏览来更高效地检阅来自搜索列表的文本,但仍不必打开整个文档以查看其内容。然而,应该理解到,这样的用户界面1200仅是说明性的,并且利用可视元素来向用户提供交互以便示出对应内容的其他方式位于本公开的范围之内。

[0095] 多个扩展的条1280、1285、1290仅是说明性的并且不受本公开限制。如此,允许用户可视化扩展的详细的结果部分1275内的文本的部分以及在该文本的部分之间进行浏览的其他方式被构想到并且是可能的。例如,在一些实施例中,可以使用呈现在集群中的一个或多个散列(与关于图9-11在以上描述的类似)。另外,替代性用户界面1200可以在无需用户到文档中浏览的情况下向用户传达其他类型的信息,包括命中的量、命中的比例、命中的位置等。

[0096] 在一些实施例中,替代性用户界面1200可以由用户可配置用于仅基于搜索项的相对于彼此的接近性来显示结果,如图17中所描绘的那样。这样的实施例可以确保将在文档内具有接近彼此的单元的集群的文档在搜索结果1220中显示得比在文档内具有散布排列的单元的文档要高。可替代地,可以完全地省略具有散布排列的文档。在一些实施例中,用户界面1200可以允许用户控制集群的灵敏度(例如,该灵敏度指示集群“散布”在文档的文本内的程度)。在非限制性示例中,用户可以通过指示文本中的100个单词内的5个命中是集群的最小阈值来调整集群的灵敏度。在另一个非限制性示例中,用户可以通过指示一个段落内的3个命中是集群的最小阈值来调整集群的灵敏度。在一些实施例中,可以由用户通过访问原型变量控制器弹出窗口1295并调整滑块1297来完成调整灵敏度。例如,将滑块1297移动至左侧可以“松开”阈值并示出在文本内较为散布的集群。类似地,将滑块1297移动至右侧可以“收窄”阈值并示出在文本内较为接近在一起的集群。应该理解到,这样的原型变量控制器弹出窗口1295仅是说明性的,并且调整灵敏度的其他方式被构想到并且是可能的。这样的特征可以表现为用户控制设定或表现为默认显示设定。

[0097] 在一些实施例中,本文中所描述的各种用户界面可以允许用户确定独立于由搜索引擎算法自动地生成的类别顺序的替代性类别顺序,这可以导致发现不一定由搜索算法预测到的对于用户而言更定制化的和更有用的搜索排名。因此,对于特定用户的需求的可视化的可定制性允许与搜索引擎和/或搜索算法相结合来使用的制衡方法。如此,可以在排名(例如,本文中所描述的过程)所确定的搜索结果与特定度量(例如,典型的搜索引擎过程)所确定的搜索结果之间达到平衡。

[0098] 应该理解到,各种用户界面仅是说明性的。如此,在不脱离本公开的范围的情况下,关于图9-17所描述的各种特征以及在本文中未具体描述的特征的任何组合被构想到并且是可能的。

[0099] 现在应该理解到,本文中所描述的系统和方法允许用户提交搜索查询,并且与实施搜索基本上同时地将搜索项分组为单元,并且将该单元的数目减少至低于阈值(如果必要的话)为了显示和呈现给用户的目的,这样的分组包括确定哪些项包含相似性以分组到一单元中。这样的减少包括确定哪些单元对于用户确定要查阅的文档是最有可能有用的。分组和减少一般与搜索基本上同时地被完成,以使得用户不在进行搜索中体验到延迟。一旦完成搜索、分组、以及减少,就向用户呈现显示。该显示允许用户通过嵌入微型图表来超越传统结果列表的有效性,从而在文档级别和集合级别上可视化搜索结果中的命中图案。如此,用户能够更快速地和更有效地确定来自结果列表的哪些文档是相关的和/或感兴趣的,而不必查看文档的整体。

[0100] 尽管本文示出和描述的特定实施例,但应理解可作出其它变化和改型而不偏离所要求保护主题的精神和范围。另外,虽然本文描述了所要求保护主题的各种方面,但不需要以组合的方式来使用这些方面。因此,所附的权利要求旨在覆盖在所要求保护主题范围内的全部变化和修改。

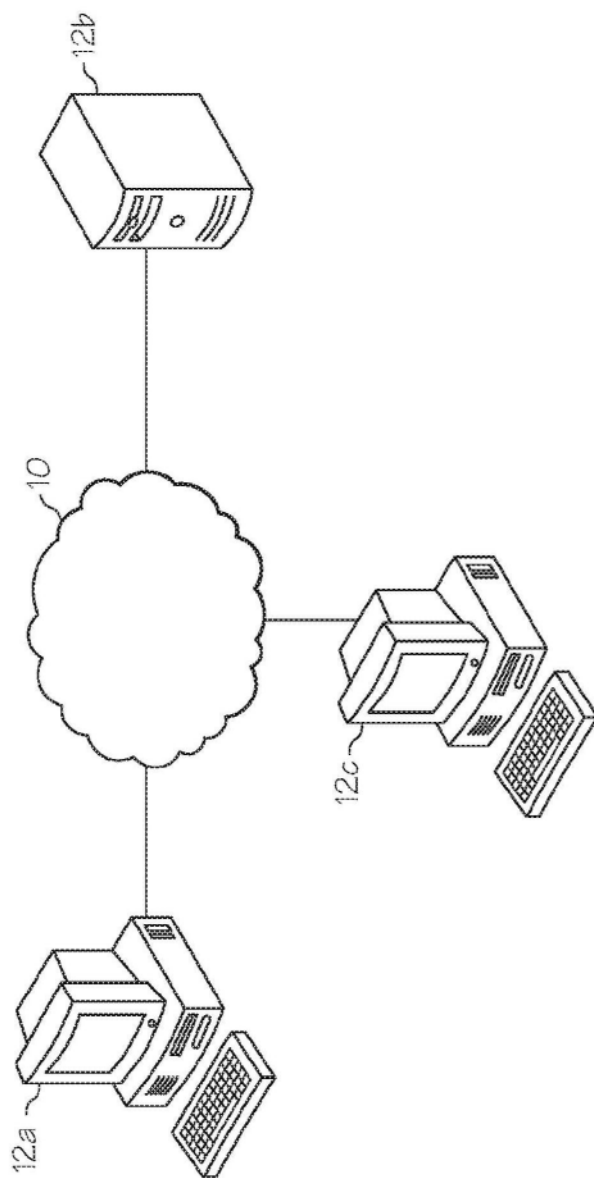


图1

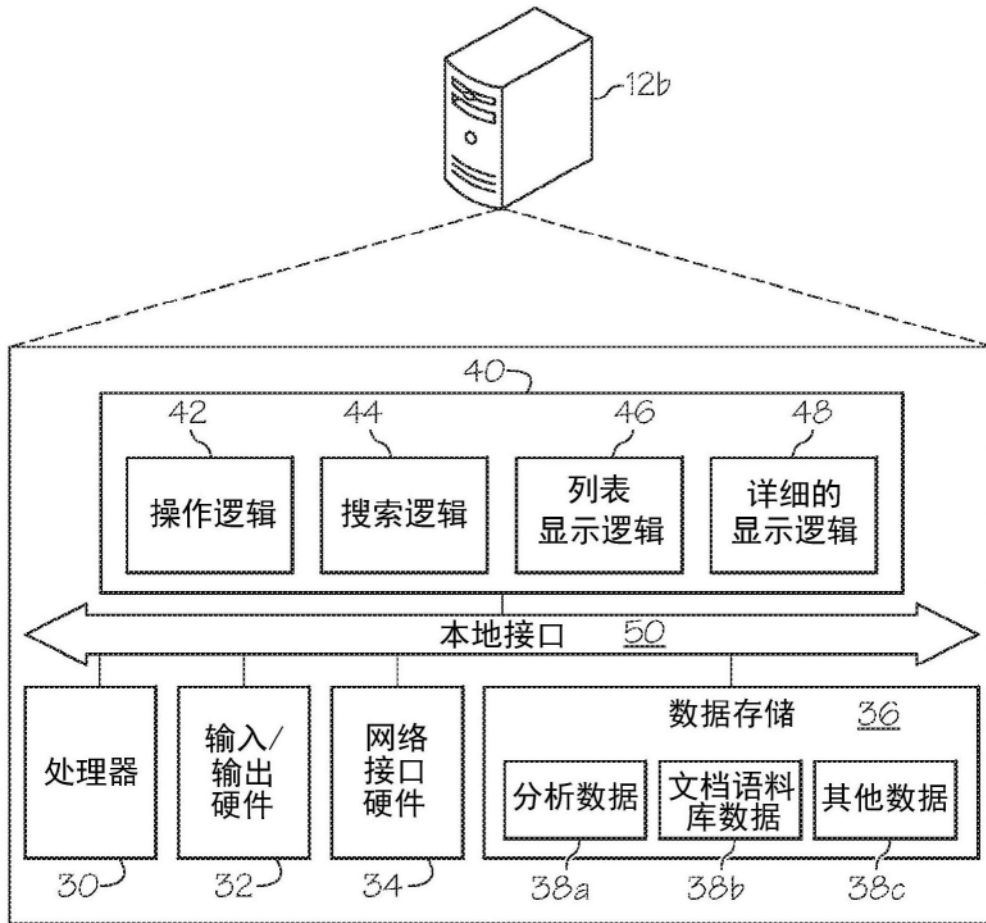


图2

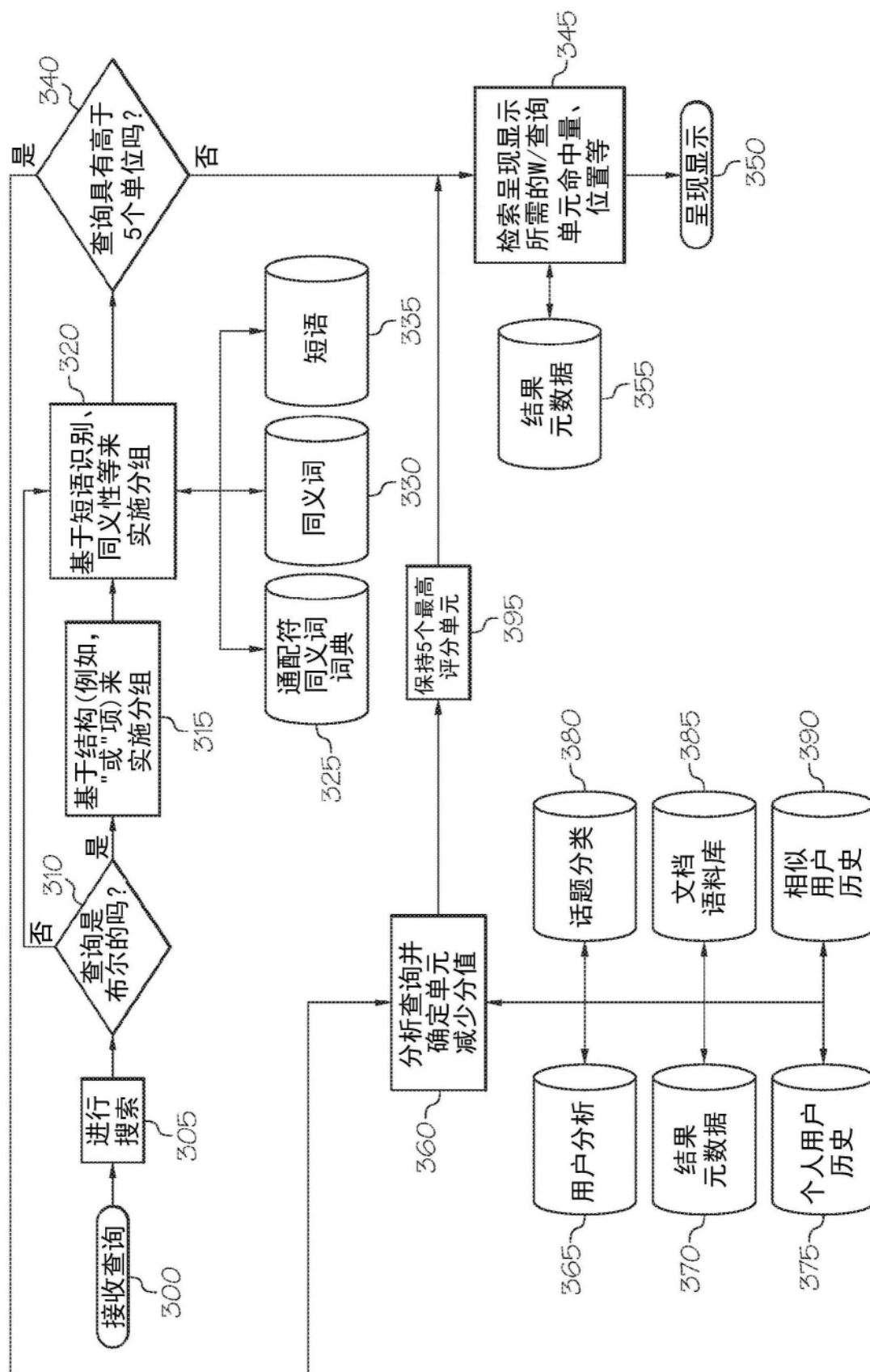


图3

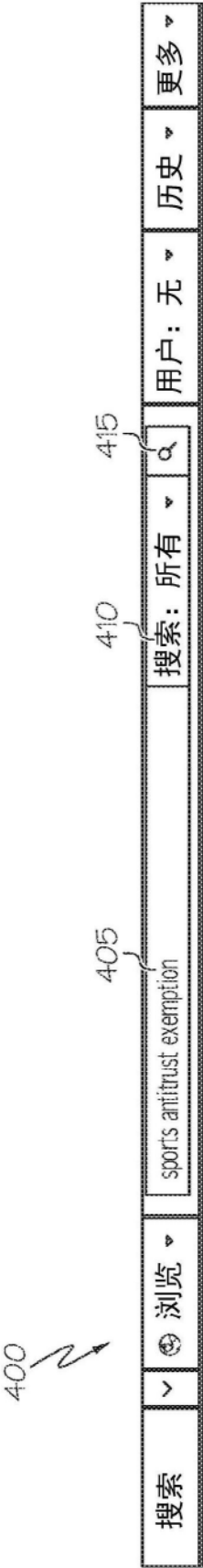


图4

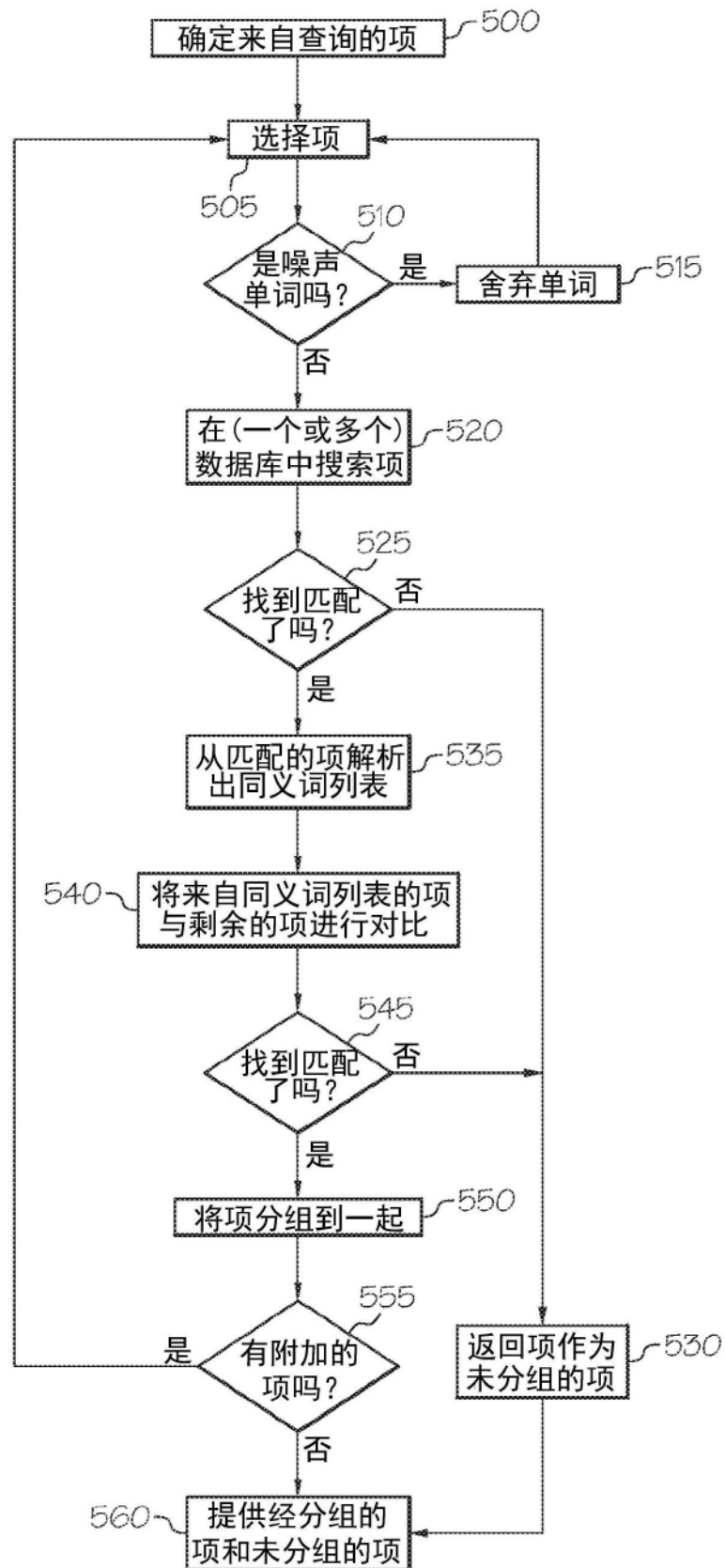


图5

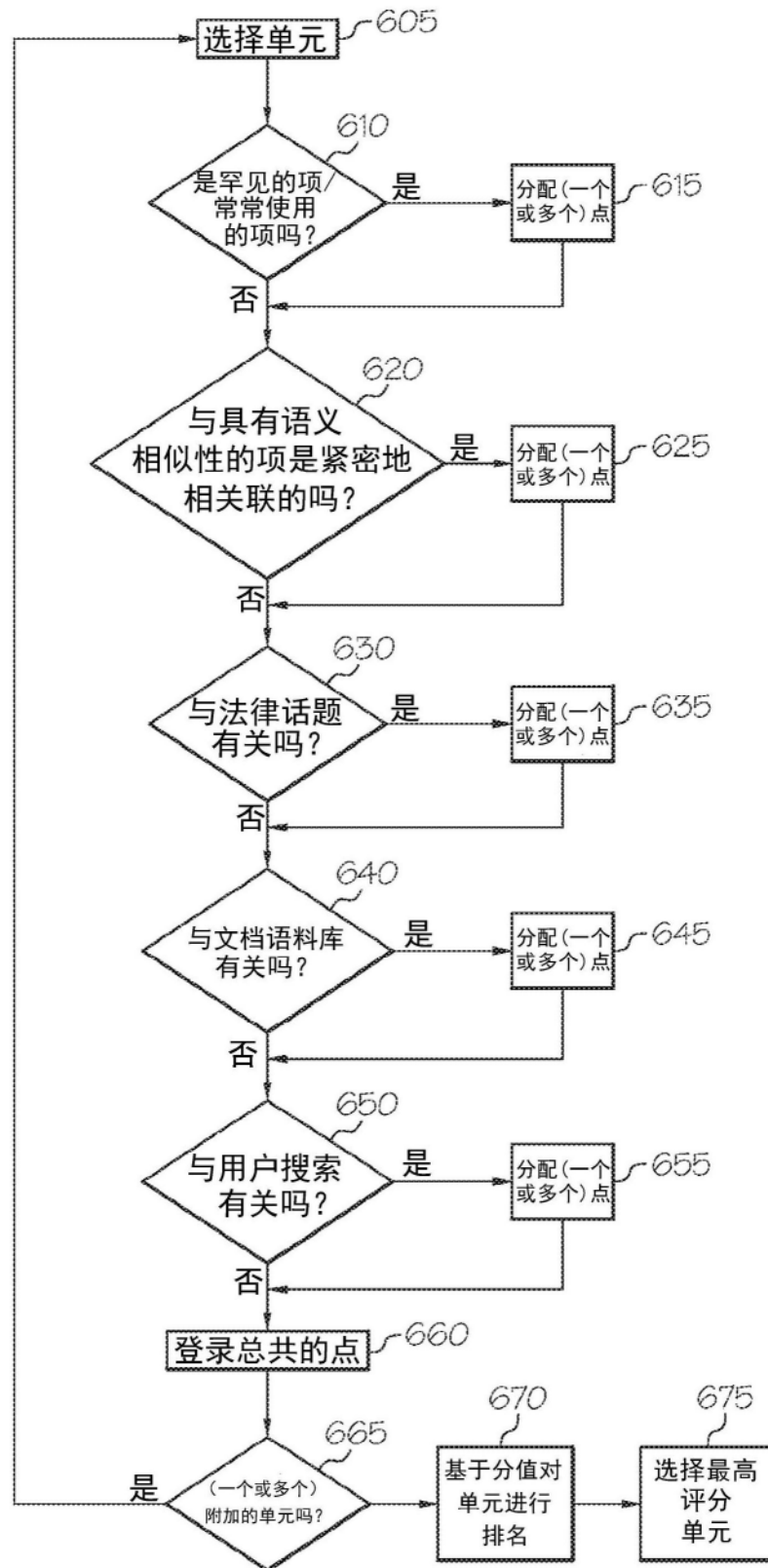


图6

项分值生成器

搜索	dac计数	罕见度百分比	常用度百分比
'toy-gun'	2389	100	0.0
'bank robbery'	24265	100	0.2
'weapon'	328285	98	2.5
'felony'	375228	96	4.4
'intent'	1765925	86	15.6
'threat'	847562	93	6.5
'toy gun' AND 'bank robbery'	243	100	0.0
'toy gun' AND 'weapon'	1259	100	0.0
'toy gun' AND 'felony'	789	100	0.0
'toy gun' AND 'intent'	829	100	0.0
'toy gun' AND 'threat'	1059	100	0.0
'bank robbery' AND 'weapon'	6578	100	0.0
'bank robbery' AND 'felony'	6559	100	0.0
'bank robbery' AND 'intent'	7229	100	0.1
'bank robbery' AND 'threat'	6719	100	0.1
'weapon' AND 'felony'	122556	99	0.9
'weapon' AND 'intent'	157469	99	1.1
'weapon' AND 'threat'	126505	99	1.0
'felony' AND 'intent'	222528	98	1.7
'felony' AND 'threat'	145092	99	1.1
'intent' AND 'threat'	320464	98	2.5

图7

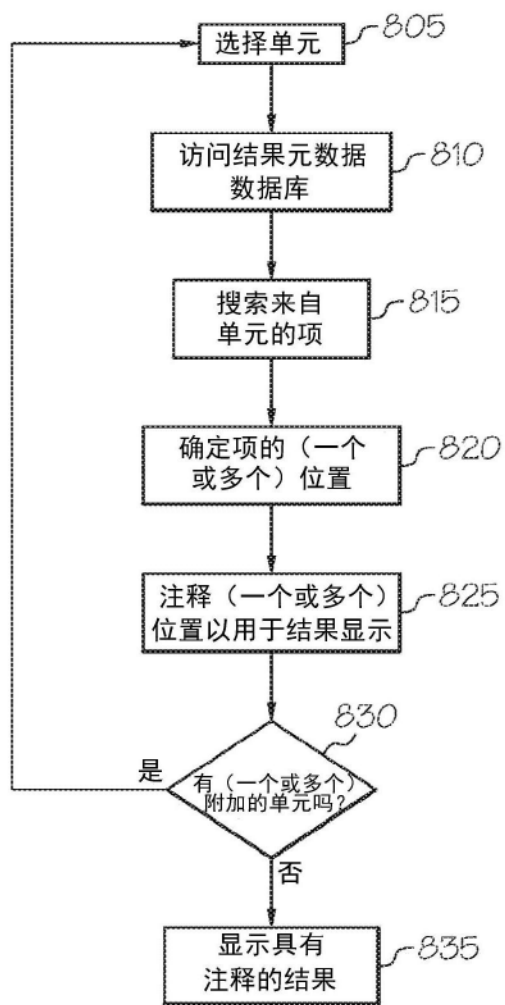


图8

900

905

搜索

浏览

sports antitrust exemption

搜索：所有

客户：无

历史

更多

结果针对：sports antitrust exemption

动作

930

Cases (10,000+)

多种颜色

sport

antitrust

exemption

展示/隐藏项高亮

915

分类依据：相关性

快照

Cases	10,000+
Statutes and Legislation	10,000+
Administration Codes and Regulations	10,000+
Agency Materials	10,000+
Secondary Materials	10,000+
Forms	1,530
Briefs, Pleadings and Motions	10,000+
Jury Instructions	3,138
Jury Verdicts and Settlements	8,099
Expert Witness Analysis	9,876
Dockets	Get
Directories	10,000+
News	Get
Legal News	10,000+

1. Robertson v. National Basketball Assn., 389 F. Supp. 867

920

912

910

914

927

926

927

926

927

926

927

925

928

923

924

935

Jurisdiction

US Federal Court

New York Southern District Court

Date

Feb 14, 1975

意见

摘要

项：摘要

... determined that it had jurisdiction even though such activities might have been unfair labor practices. The joinder of plaintiffs' union was not mandatory because the sports clubs did not demonstrate any potential harm from the non-joinder. A preliminary injunction preventing the merger of the sports clubs was not required to be dissolved by Norris La-Guardia Act 29 U.S.C.S. § 101 which was inapplicable to all ...

在完整文档中查看该段

923

924

935

图9

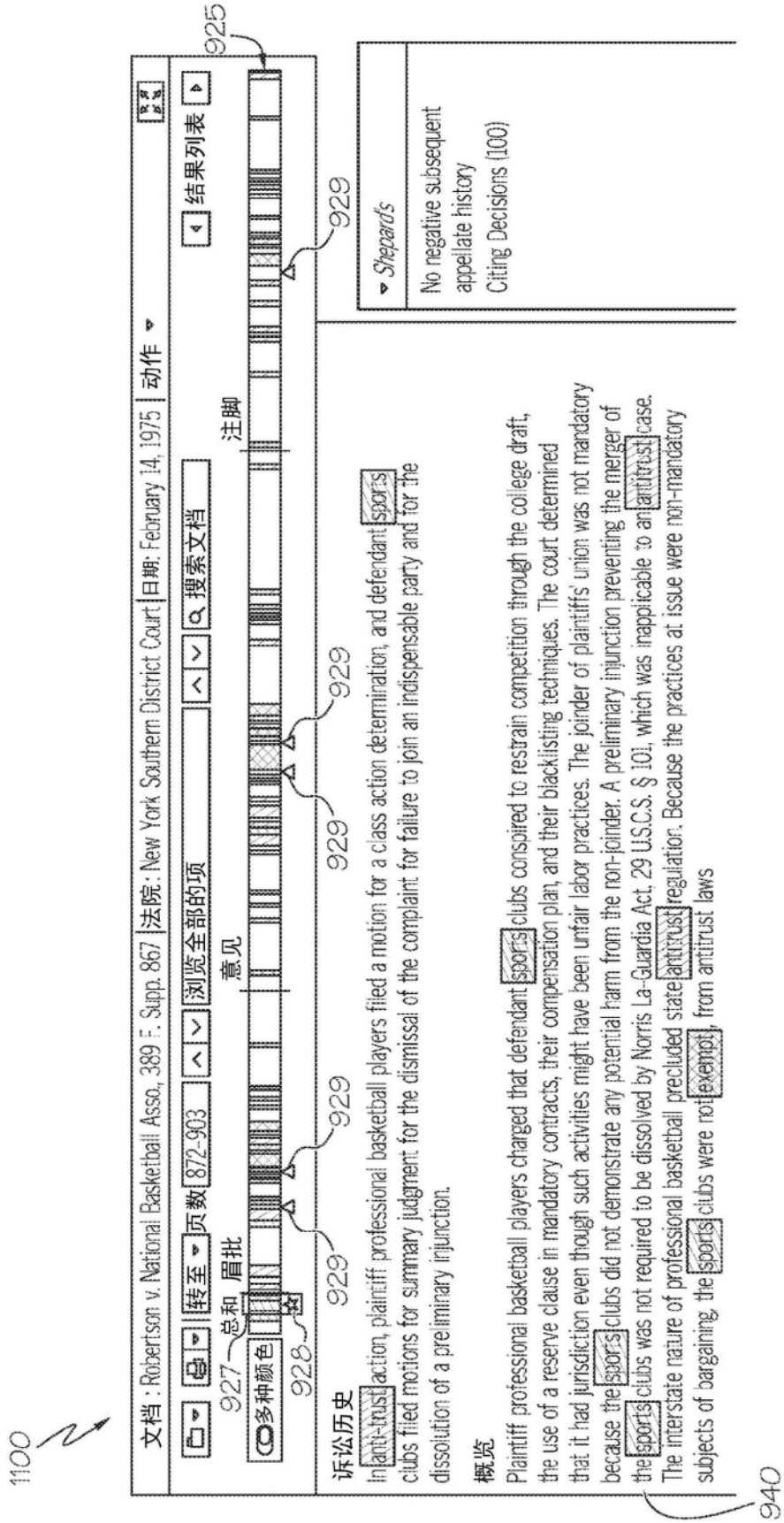


图11

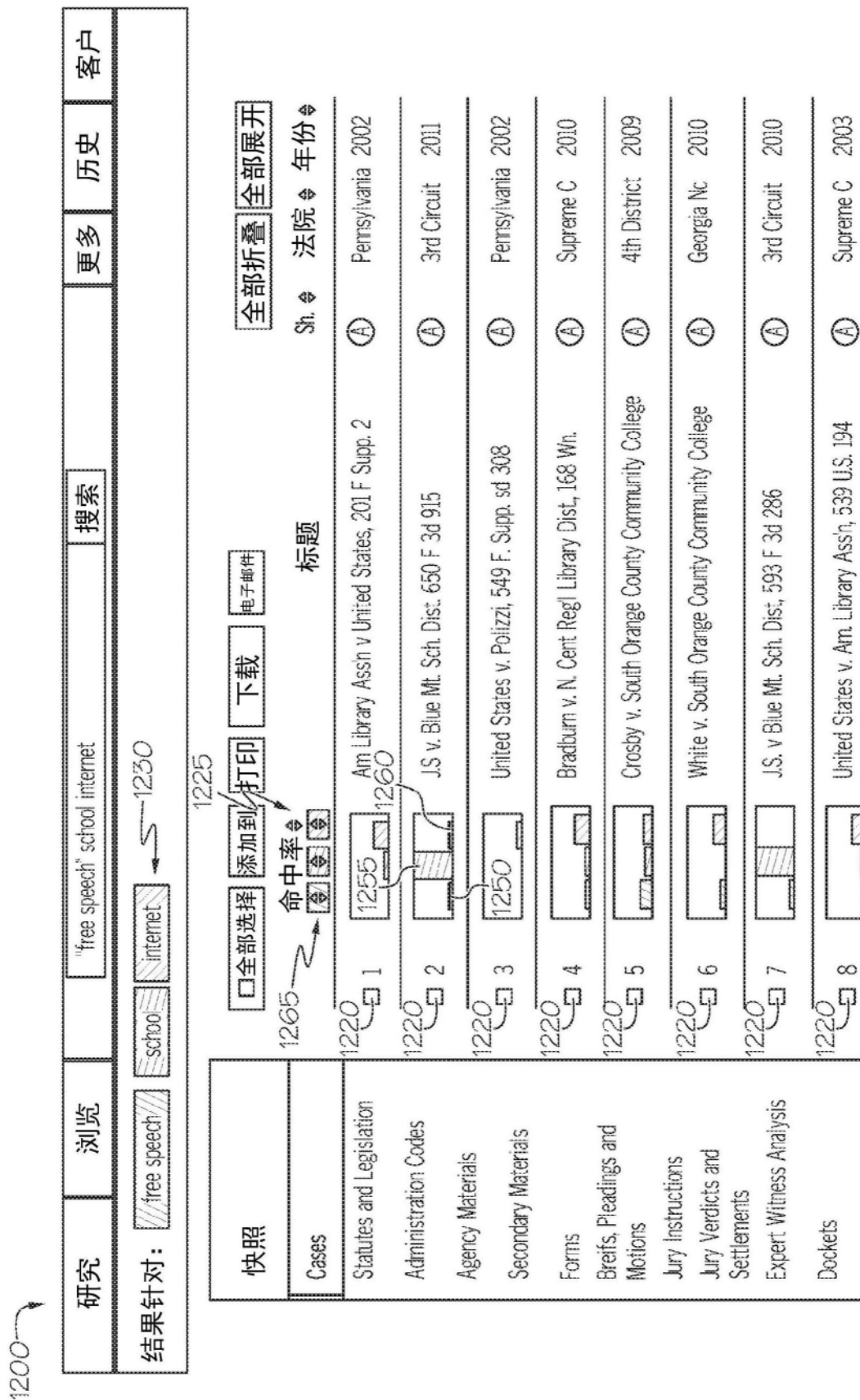


图12

1200

结果针对: ☐ free speech ☐ school ☐ internet

1225

1220

1265

1260

1250

1255

命中

全部选择

添加到

打印

下载

电子邮件

全部折叠

全部展开

Sh 法院 年份

Sh	法院	年份
1	3rd Circuit	2011
2	3rd Circuit	2010
3	4th District	2009
4	Supreme C	2010
	Supreme C	2003
	Pennsylvania	2002

快照

Cases

Statutes and Legislation

Administration Codes

Agency Materials

Secondary Materials

原型可变控制器

搜索变量

检索 10 个文档

从 10 个文档部分中提取摘录

JS v. Blue Mt. Sch. Dist. 650 F.3d 915

JS v. Blue Mt. Sch. Dist. 593 F.3d 286

Crosby v. South Orange County Community College

Bradburn v. N. Cent Reg'l Library Dist., 168 Wh.

United States v. Am. Library Ass'n, 539 U.S. 194

Am Library Ass'n v United States, 201 F Supp

图13

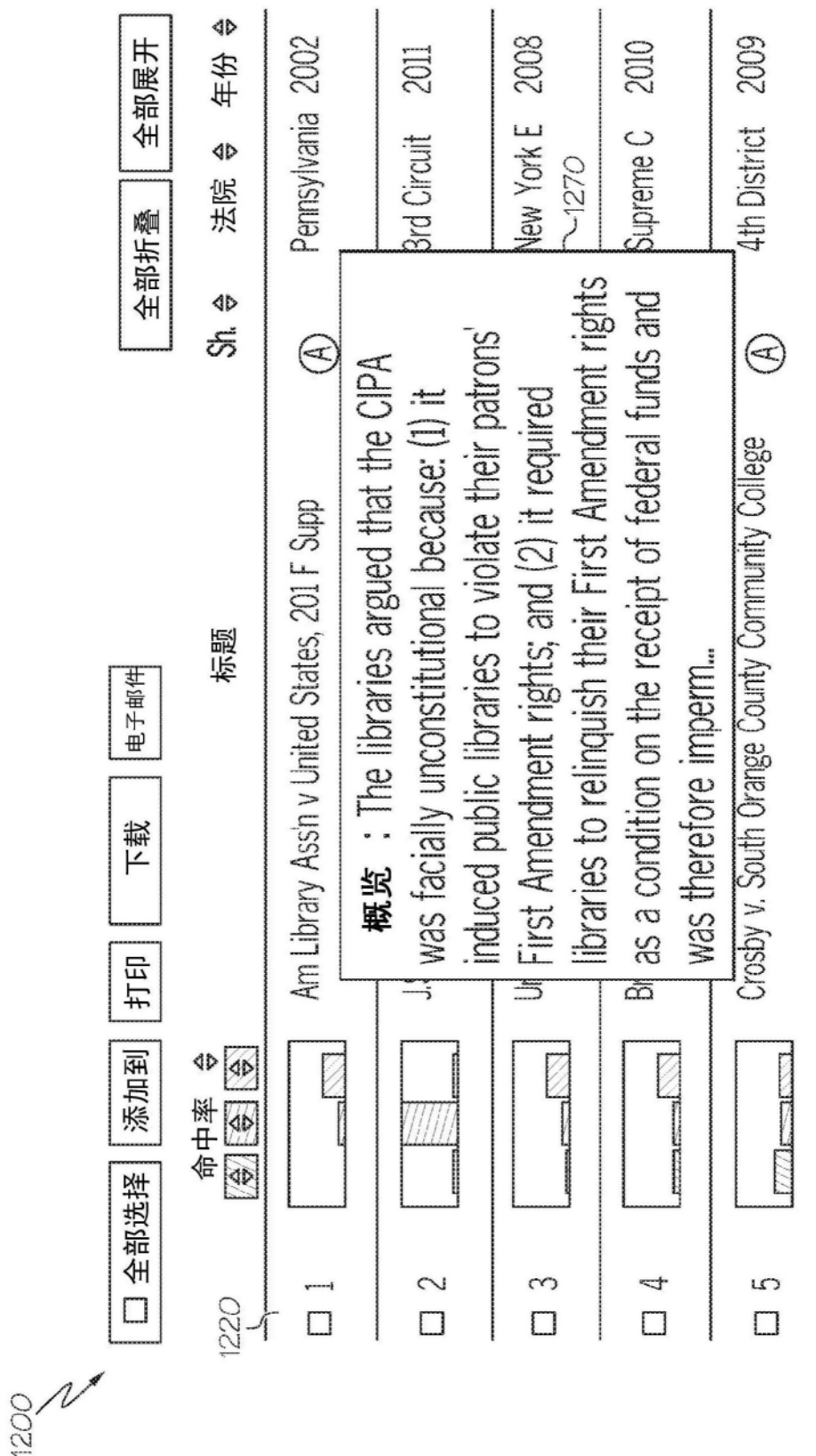


图14

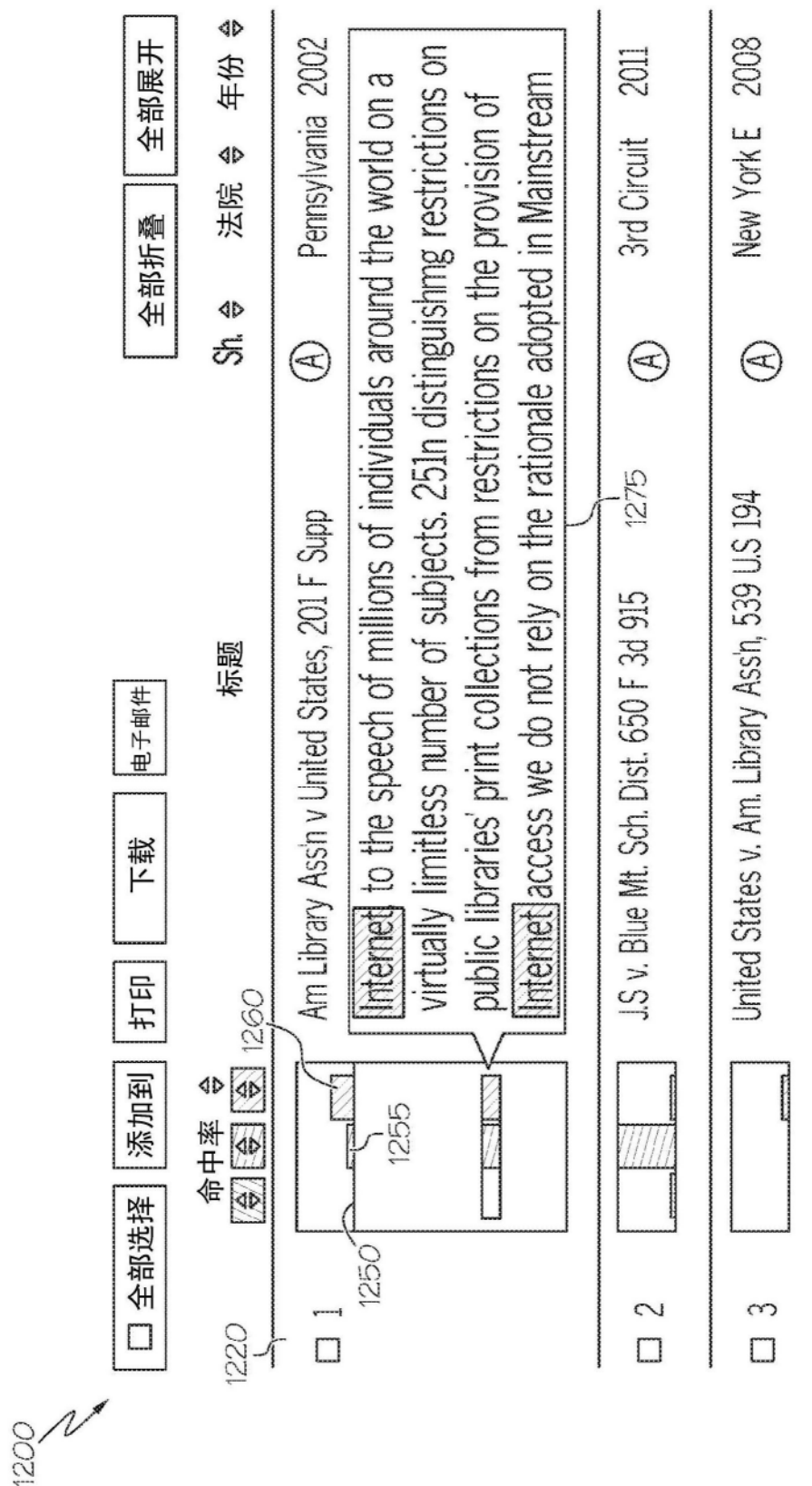


图15

结果针对:		<input type="checkbox"/> free speech	<input checked="" type="checkbox"/> school	<input type="checkbox"/> internet		
快照		<input type="checkbox"/> 全部选择 <input type="button" value="添加到"/> <input type="button" value="打印"/> <input type="button" value="下载"/> <input type="button" value="电子邮件"/>				
Cases	命中率 <input type="text"/> 命令 <input type="button" value="上"/> <input type="button" value="下"/> <input type="button" value="左"/> <input type="button" value="右"/>					
Statutes and Legislation	全部折叠 全部展开					
Administration Codes	Sh. 法院 年份					
Agency Materials	Pennsylvania 2002					
Secondary Materials	Am Library Assn v United States, 201 F Supp. 2					
Forms	United States v. Polizzi, 549 F. Supp. sd 308 Pennsylvania 2002					
Briefs, Pleadings and	Bradburn v. N. Cent Regl Library Dist., 168 Wn.					
	Crosby v. South Orange County Community College 4th District 2009					
	White v. South Orange County Community College Georgia Nc 2010					
	J.S. v Blue Mt. Sch. Dist., 593 F 3d 286 3rd Circuit 2010					
	United States v. Am. Library Ass'n, 539 U.S. 194 Supreme C 2003					

原型变量控制器

搜索变量 检索 个文档

从 个文档部分中提取摘录

查看

- ☒ 摘录位置
- ☐ 原始

阈值 (宽松<->严格)

———— 12

图16

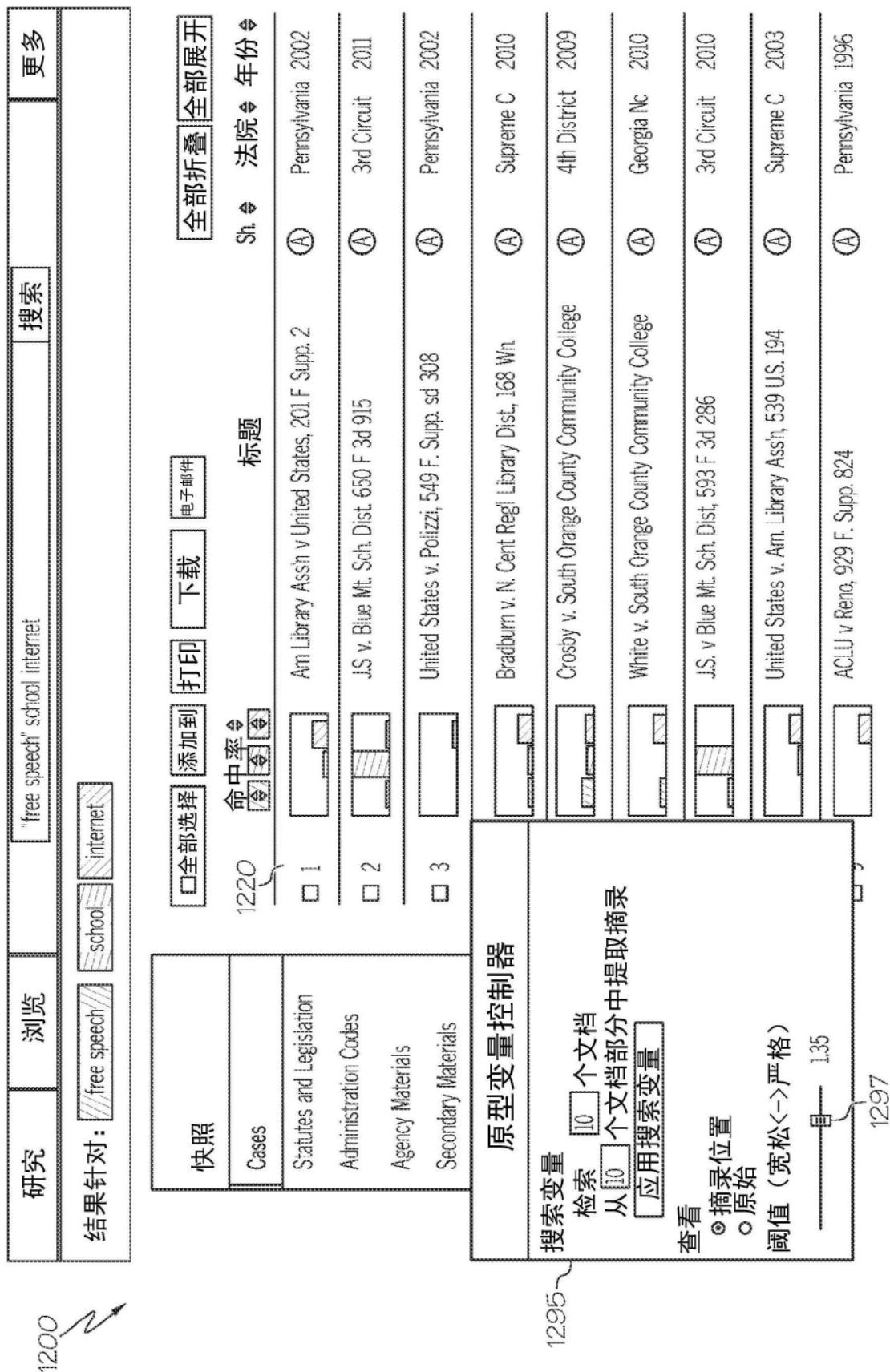


图17