

(12) 发明专利

(10) 授权公告号 CN 1934569 B

(45) 授权公告日 2012. 03. 14

(21) 申请号 200580008405. 9

(51) Int. Cl.

(22) 申请日 2005. 03. 15

G06F 17/30 (2006. 01)

(30) 优先权数据

60/553, 607 2004. 03. 15 US

60/614, 232 2004. 09. 28 US

(56) 对比文件

US 6181336 B1, 2001. 01. 30, 说明书第 13 栏第 9 行至第 48 行, 第 16 栏 46 行至 67 行.

US 5392428 A, 1995. 02. 21, 全文.

(85) PCT 申请进入国家阶段日

2006. 09. 15

审查员 王雪莲

(86) PCT 申请的申请数据

PCT/US2005/008635 2005. 03. 15

(87) PCT 申请的公布数据

W02005/091175 EN 2005. 09. 29

(73) 专利权人 雅虎公司

地址 美国加利福尼亚州

(72) 发明人 埃卡特·沃瑟尔 卢齐 大卫·库

凯文·李 仲·满·谭 阿利·迪比

(74) 专利代理机构 北京东方亿思知识产权代理

有限责任公司 11258

代理人 王怡

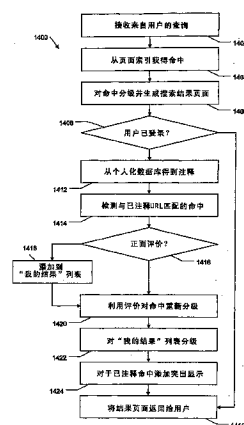
权利要求书 7 页 说明书 27 页 附图 15 页

(54) 发明名称

集成有用户注释的搜索系统和方法

(57) 摘要

计算机系统和方法允许用户对在诸如万维网之类的语料库中找到的内容项进行注释 (1412, 图 14)。注释可以包括与文档有关的任何描述性和 / 或评估性的元数据, 并且是从用户收集并与该用户关联存储的 (图 14)。用户能够注释并查看其在与语料库交互的同时遇到的任何文档的注释, 包括在语料库的搜索中返回的命中 (1414, 图 14)。用户还能够搜索其注释或者将搜索限制于其已注释的文档。也可对用户聚合来自注释的元数据, 并且聚合后的元数据被用于生成搜索结果 (1420, 图 14)。



1. 一种用于对用户查询作出响应的方法,所述方法包括:
接收由用户提交的查询;
搜索包括多个文档的语料库以识别一个或多个命中,其中每个命中是来自所述语料库的被确定与所述查询相关的文档;
访问由所述用户创建的注释库,每个注释与所述语料库中的一个文档相关联,并且包括与该文档有关的用户特定的元数据;
将与所述注释中的匹配注释相关联的每个命中识别为已注释命中;
生成包括命中列表的搜索报告,对于每个命中,所述搜索报告还指示所述命中是否是已注释命中;以及
将所述搜索报告发送给所述用户。
2. 如权利要求 1 所述的方法,其中所述搜索报告包括被应用于作为已注释命中的每个命中的可视突出显示元素。
3. 如权利要求 2 所述的方法,其中包括在所述注释中的用户特定的元数据包括评价,并且应用于每个已注释命中的可视突出显示元素依赖于包括在所述匹配注释中的评价。
4. 如权利要求 1 所述的方法,其中生成所述搜索报告还包括:
对于每个已注释命中,在所述搜索报告中提供控制元素,所述控制元素可由用户操作以请求显示所述匹配注释的用户特定的元数据。
5. 如权利要求 1 所述的方法,其中生成所述搜索报告还包括:
对于每个已注释命中,将来自所述匹配注释的至少某些用户特定的元数据结合到所述命中列表中。
6. 如权利要求 1 所述的方法,其中生成所述搜索报告包括:
对于作为未注释命中的每个命中,提供控制元素,所述控制元素可由用户操作以创建与所述命中相关联的新注释。
7. 如权利要求 6 所述的方法,还包括:
从所述用户接收所述新注释;以及
将所述新注释添加到所述注释库中。
8. 如权利要求 7 所述的方法,还包括:
重新生成所述搜索报告以进一步指示所述新注释的命中是已注释命中;以及
将重新生成的搜索报告发送给所述用户。
9. 如权利要求 1 所述的方法,其中生成所述搜索报告还包括:
生成只包括已注释命中的单独列表。
10. 如权利要求 9 所述的方法,其中包括在每个注释中的用户特定的元数据包括关联文档的评价,并且所述单独列表只包括所述匹配注释包括正面评价的已注释命中。
11. 如权利要求 1 所述的方法,其中包括在每个注释中的用户特定的元数据包括关联文档的评价,并且生成所述搜索报告包括至少部分基于所述已注释命中的评价确定所述命中列表中列表的顺序。
12. 如权利要求 11 所述的方法,其中生成所述搜索报告还包括在所述评价是负面评价的情况下从所述命中列表中去除已注释命中。
13. 如权利要求 1 所述的方法,还包括:

搜索所述注释库以识别一个或多个附加的已注释命中,其中每个附加的已注释命中对应于来自所述语料库的一个文档,对于该文档所述关联注释包括被确定为与所述查询相关的用户特定的元数据;以及

将所述附加的已注释命中结合到所述搜索结果页面中的命中列表中。

14. 如权利要求 13 所述的方法,其中搜索包括多个文档的语料库包括:

从所述用户查询中提取出搜索项;以及

将所述语料库中包含所述搜索项的每个文档识别为命中。

15. 如权利要求 14 所述的方法,其中搜索所述注释库包括将所述语料库中所述用户特定的元数据针对其包括所述搜索项的每个文档识别为附加的已注释命中。

16. 如权利要求 1 所述的方法,其中所述注释库还包括与所述语料库中的文档组相关联的至少一个注释,并且作为所述文档组之一的任何命中被识别为已注释命中。

17. 如权利要求 1 所述的方法,其中所述用户特定的元数据包括由所述用户明确输入的信息项。

18. 如权利要求 17 所述的方法,其中所述信息项是所述关联文档的评价。

19. 如权利要求 17 所述的方法,其中所述信息项是描述所述关联文档的关键字。

20. 如权利要求 17 所述的方法,其中所述信息项是从预定义的词表中选出的标签。

21. 如权利要求 17 所述的方法,其中所述信息项是所述关联文档的自由文本描述。

22. 如权利要求 1 所述的方法,其中所述语料库包括多个万维网页面。

23. 如权利要求 1 所述的方法,其中所述用户是人。

24. 如权利要求 1 所述的方法,其中所述用户是计算机。

25. 一种用于对用户查询作出响应的方法,所述方法包括:

接收来自用户的查询;

访问由所述用户创建的注释库,每个注释与属于一个语料库的多个文档之一相关联,每个注释还包括与所述关联文档有关的用户特定的元数据;

在与所述注释相关联的文档中识别一个或多个命中,其中每个命中是来自所述语料库的被确定与所述查询相关的文档;

生成包括命中列表的搜索报告,所述列表只包括针对其所述注释库具有关联注释的文档;以及

将所述搜索报告发送给所述用户。

26. 如权利要求 25 所述的方法,其中识别一个或多个命中包括将所述查询和与所述注释之一相关联的每个文档的内容相比较。

27. 如权利要求 26 所述的方法,其中所述比较动作包括:

从所述查询中提取出搜索项;以及

对于与所述多个注释之一相关联的每个文档,检测所述搜索项是否存在于所述文档中,

其中在所述搜索项存在于所述文档中的情况下,所述文档被识别为命中。

28. 如权利要求 25 所述的方法,其中识别一个或多个命中包括将所述查询与所述注释的用户特定的元数据相比较。

29. 如权利要求 28 所述的方法,其中识别一个或多个命中还包括:

从所述查询中提取出搜索项 ;以及
对于每个注释,检测所述搜索项是否存在于所述用户特定的元数据中,
其中在所述搜索项存在于所述用户特定的元数据中的情况下,所述关联文档被识别为命中。

30. 如权利要求 29 所述的方法,还包括 :

对于与所述多个注释之一相关联的每个文档,检测所述搜索项是否存在于所述文档中,

其中在所述搜索项存在于所述文档中的情况下,所述文档被识别为命中。

31. 如权利要求 29 所述的方法,其中所述用户特定的元数据包括多个字段,并且所述查询指定在检测动作期间要考虑哪些字段。

32. 如权利要求 25 所述的方法,还包括 :

从所述用户接收对要搜索的注释库的子集的选择,

其中,识别一个或多个命中的动作只对所述注释库的选定子集中的注释执行。

33. 如权利要求 25 所述的方法,其中对于每个命中,所述搜索报告还包括控制元素,所述控制元素可由用户操作以请求显示所述注释中的匹配注释的用户特定的元数据。

34. 如权利要求 25 所述的方法,其中对于每个命中,所述搜索报告还包括来自所述关联注释的至少某些用户特定的元数据。

35. 如权利要求 25 所述的方法,其中包括在每个注释中的用户特定的元数据包括关联文档的评价,并且所述列表中的文档以至少部分基于所述命中的评价确定的顺序放置。

36. 如权利要求 25 所述的方法,其中所述注释库还包括与所述语料库中的文档组相关联的至少一个注释,并且作为所述文档组之一的任何文档被识别为命中。

37. 如权利要求 25 所述的方法,其中所述语料库是万维网。

38. 如权利要求 25 所述的方法,其中所述用户是人。

39. 如权利要求 25 所述的方法,其中所述用户是计算机。

40. 一种用于对用户查询作出响应的系统,所述系统包括 :

用于接收由用户提交的查询的装置 ;

用于搜索包括多个文档的语料库以识别一个或多个命中的装置,其中每个命中是来自所述语料库的被确定为与所述查询相关的文档 ;

用于访问由所述用户创建的注释库的装置,每个注释与所述语料库中的一个文档相关联,并且包括与该文档有关的用户特定的元数据 ;

用于将与所述注释中的匹配注释相关联的每个命中识别为已注释命中的装置 ;

用于生成包括命中列表的搜索报告的装置,对于每个命中,所述搜索报告还指示所述命中是否是已注释命中 ;以及

用于将所述搜索报告发送给所述用户的装置。

41. 一种用于对用户查询作出响应的系统,所述系统包括 :

用于接收来自用户的查询的装置 ;

用于访问由所述用户创建的注释库的装置,每个注释与属于一个语料库的多个文档之一相关联,每个注释还包括与所述关联文档有关的用户特定的元数据 ;

用于在与所述注释相关联的文档中识别一个或多个命中的装置,其中每个命中是来自

所述语料库的被确定与所述查询相关的文档；

用于生成包括命中列表的搜索报告的装置，所述列表只包括针对其所述注释库具有关联注释的文档；以及

用于将所述搜索报告发送给所述用户的装置。

42. 一种用于对用户查询作出响应的方法，所述方法包括：

访问由多个注释用户创建的注释存储，每个注释与所述注释用户之一相关联，并且与属于一个语料库的多个文档之一相关联，每个注释包括与所述关联文档有关的用户特定的元数据；

对所述多个注释用户聚合与所述文档中的同一文档有关的用户特定的元数据中的一项或多项，从而生成所述文档的一个或多个聚合属性；

接收由查询用户提交的查询；

搜索所述语料库以识别一个或多个命中，其中每个命中是来自所述语料库的被确定与所述查询相关的文档；

生成包括命中列表的搜索报告，其中所述列表是至少部分基于所述命中的至少一个聚合属性的；以及

将所述搜索报告发送给所述查询用户。

43. 如权利要求 42 所述的方法，其中在所述聚合动作期间聚合的用户特定的元数据中的一项或多项包括所述关联文档的评价。

44. 如权利要求 43 所述的方法，其中所述聚合动作包括计算所述文档的平均评价，其中所述平均评价是所述聚合属性之一。

45. 如权利要求 44 所述的方法，其中生成所述搜索报告包括：

在所述搜索报告中包括每个命中的平均评价。

46. 如权利要求 44 所述的方法，其中生成所述搜索报告包括：

以至少部分基于每个命中的平均评价的顺序安排所述命中列表。

47. 如权利要求 42 所述的方法，其中所述用户特定的元数据中的一项或多项包括描述所述关联文档的用户提供的关键字。

48. 如权利要求 47 所述的方法，其中所述聚合动作包括识别最常提供的关键字的关键字集，其中所述关键字集是所述聚合属性之一。

49. 如权利要求 48 所述的方法，其中搜索所述语料库包括：

从所述查询中提取出搜索项；以及

将所述语料库中针对其所述关键字集包括所述搜索项的任何文档识别为命中。

50. 如权利要求 48 所述的方法，其中所述搜索报告包括每个命中的关键字集。

51. 如权利要求 42 所述的方法，其中所述用户特定的元数据中的一项或多项包括由所述注释用户从预定义的词表中选出的标签。

52. 如权利要求 51 所述的方法，其中所述预定义的词表包括多个类别标签。

53. 如权利要求 51 所述的方法，其中所述聚合动作包括识别最常选择的标签的标签集，其中所述标签集是所述聚合属性之一。

54. 如权利要求 53 所述的方法，其中搜索所述语料库包括：

从所述查询中提取出搜索项；以及

将所述语料库中针对其所述标签集包括所述搜索项的任何文档识别为命中。

55. 如权利要求 54 所述的方法,其中所述搜索报告包括每个命中的标签集。

56. 如权利要求 42 所述的方法,还包括:

将针对其所述注释存储包含由所述查询用户创建的并与该命中相关联的注释的每个命中识别为已注释命中;以及

对于所述注释存储包含由所述查询用户创建的注释的每个命中,在所述搜索报告中包括关于每个命中是否是已注释命中的指示。

57. 如权利要求 56 所述的方法,其中所述搜索报告还包括一个控件,所述控件可由所述查询用户操作以查看所述查询用户对于每个已注释命中的注释。

58. 如权利要求 56 所述的方法,其中所述搜索报告还包括从所述查询用户对于每个已注释命中的注释中提取出的元数据。

59. 如权利要求 42 所述的方法,其中所述语料库是万维网。

60. 如权利要求 42 所述的方法,其中所述查询用户是人。

61. 如权利要求 42 所述的方法,其中所述查询用户是计算机。

62. 一种用于对用户查询作出响应的系统,所述系统包括:

用于访问由多个注释用户创建的注释存储的装置,每个注释与所述注释用户之一相关联,并且与属于一个语料库的多个文档之一相关联,每个注释包括与所述关联文档有关的用户特定的元数据;

用于对所述多个注释用户聚合与所述文档中的同一文档有关的用户特定的元数据中的一项或多项,从而生成所述文档的一个或多个聚合属性的装置;

用于接收由查询用户提交的查询的装置;

用于搜索所述语料库以识别一个或多个命中的装置,其中每个命中是来自所述语料库的被确定为与所述查询相关的文档;

用于生成包括命中列表的搜索报告的装置,其中所述列表是至少部分基于所述命中的至少一个聚合属性的;以及

用于将所述搜索报告发送给所述查询用户的装置。

63. 一种用于收集与语料库中的文档有关的元数据的方法,所述方法包括:

提供具有第一控制元素的用户界面,所述第一控制元素可由用户操作以指示注释当前文档的期望;

经由网络从所述用户接收与所述当前文档有关的元数据;

将包括由所述用户提供的元数据的注释添加到数据存储,所述注释持久与所述用户相关联,并且与所述当前文档相关联;以及

其后响应于用户请求而经由所述网络将所述注释的元数据发送给所述用户。

64. 如权利要求 63 所述的方法,其中所述第一控制元素还可由所述用户操作以提交所述元数据。

65. 如权利要求 63 所述的方法,其中所述第一控制元素还可由所述用户操作以获得用于提交所述元数据的表格。

66. 如权利要求 63 所述的方法,其中提供所述用户界面包括将所述第一控制元素添加到响应于由所述用户提交的查询而生成的搜索报告。

67. 如权利要求 66 所述的方法,其中:

所述搜索报告包括命中列表,每个命中是来自所述语料库的文档;

所述搜索报告提供了对应于每个命中的第一控制元素的不同实例;并且

在所述用户操作所述第一控制元素的相应实例的情况下,所述命中被识别为所述当前文档。

68. 如权利要求 66 所述的方法,其中所述搜索报告包括命中列表,每个命中是来自所述语料库的文档,所述方法还包括:

对于每个命中,确定所述数据存储是否包括与所述用户和所述命中相关联的注释;

在所述数据存储包括与所述用户和所述命中相关联的注释的情况下,在所述搜索结果页面中包括第二控制元素,所述第二控制元素可由所述用户操作以查看所述注释的元数据;以及

在所述数据存储不包括与所述用户和所述命中相关联的注释的情况下,在所述搜索结果页面中包括对应于所述命中的第一控制元素的实例,其中在所述用户操作所述第一控制元素的相应实例的情况下,所述命中被识别为所述当前文档。

69. 如权利要求 63 所述的方法,其中所述第一控制元素被提供在文档浏览器的工具栏界面中,所述文档浏览器被配置为显示来自所述语料库的文档。

70. 如权利要求 69 所述的方法,其中当所述用户操作所述第一控制元素时被所述文档浏览器显示的文档被识别为所述当前文档。

71. 如权利要求 69 所述的方法,其中所述第一控制元素只在所述文档浏览器正在显示这样的文档时是活动的:对于该文档,所述数据存储不包括与所显示的文档和所述用户相关联的注释。

72. 如权利要求 69 所述的方法,还包括:

在所述工具栏界面中提供第二控制元素,所述第二控制元素可由所述用户操作以请求发送正由所述文档浏览器显示的文档的元数据,

其中所述第二控制元素只在所述文档浏览器正在显示这样的文档时是活动的:对于该文档,所述数据存储包括与所显示的文档和所述用户相关联的注释。

73. 如权利要求 63 所述的方法,其中所述元数据包括所述关联文档的评价。

74. 如权利要求 73 所述的方法,还包括:

当向所述用户显示所述关联文档时显示所述评价。

75. 如权利要求 73 所述的方法,还包括:

接收来自所述用户的查询;

响应于所述查询为所述用户生成搜索报告,所述搜索报告包括所述语料库中被确定为与所述查询相关的文档的列表;以及

在所述搜索报告中包括所述列表中每个文档的评价,对于该文档,所述数据存储包括与所述文档和所述用户相关联的注释。

76. 如权利要求 63 所述的方法,其中所述元数据包括描述所述关联文档的关键字。

77. 如权利要求 76 所述的方法,还包括:

接收来自所述用户的查询;以及

利用所述关键字来确定所述关联文档是否与所述查询相关。

78. 如权利要求 63 所述的方法,其中所述元数据包括描述所述关联文档的标签,所述标签由所述用户从预定义的词表中选择。

79. 如权利要求 63 所述的方法,其中所述元数据包括所述关联文档的自由文本描述。

80. 如权利要求 79 所述的方法,还包括:

接收来自所述用户的查询;以及

利用所述自由文本描述来确定所述关联文档是否与所述查询相关。

81. 如权利要求 63 所述的方法,其中发送所述元数据包括提供第二控制元素,所述第二控制元素可由所述用户操作以编辑所述元数据,所述方法还包括:

从所述用户接收已编辑的元数据;以及

利用所述已编辑的元数据来更新所述数据存储中的注释。

82. 如权利要求 63 所述的方法,还包括:

提供第二控制元素,所述第二控制元素可由所述用户操作以请求显示来自所述数据存储的注释;以及

响应于所述第二控制元素的操作,从所述数据存储中检索已请求的注释,并将检索到的注释的元数据发送给所述用户。

83. 如权利要求 82 所述的方法,其中所述第二控制元素被提供在文档浏览器的工具栏界面中,所述文档浏览器被配置为显示来自所述语料库的文档。

84. 如权利要求 63 所述的方法,其中所述语料库是万维网。

85. 如权利要求 63 所述的方法,其中所述用户是计算机。

86. 如权利要求 63 所述的方法,其中所述用户是人。

87. 一种用于收集与语料库中的文档有关的元数据的方法,所述方法包括:

在个人化数据存储中存储多个注释,每个注释与注释用户相关联,并且与所述语料库中的文档之一相关联,每个注释包括与所述文档之一有关的元数据;以及

由通信耦合到所述个人化数据存储的注释模块执行以下步骤:

输入控制步骤,接收来自所述注释用户的元数据并将所述接收的元数据与所述语料库中的文档之一相关联;

存储控制步骤,在所述个人化数据存储中存储所述接收的元数据,作为与所述注释用户的标识符和所述文档之一相关联的注释;以及

检索控制步骤,通过参考所述注释用户的标识符和所述文档中的关联文档来在所述个人化数据存储中定位注释。

集成有用户注释的搜索系统和方法

[0001] 相关申请的交叉引用

[0002] 本申请要求下面的两个美国临时专利申请的优先权：

[0003] 2004 年 3 月 15 日提交的申请 No. 60/553,607，题为“Search Systems and Methods with Integration of User Judgments”；以及 2004 年 9 月 28 日提交的申请 No. 60/614,232，题为“Search Systems and Methods with Integration of User Judgments”。

[0004] 这两个申请各自的内容通过引用结合于此以用于一切目的。

技术领域

[0005] 本发明一般地涉及搜索并导航文档的语料库 (corpus) 或其他内容项，更具体地，本发明涉及集成有与语料库中的个体项有关的用户提供的元数据（这里称为“注释” (annotation)）的搜索和导航系统和方法。

背景技术

[0006] 万维网 (Web) 提供了大量的与几乎每个可想象的主题有关的相互链接信息源（具有各种格式，包括文本、图像和媒体内容）。随着 Web 的发展，用户搜索该集合并且识别与特定主题相关的内容的能力变得越来越重要，并且目前出现了多个搜索服务提供者以满足该需求。通常，搜索服务提供者公布一个网页 (Web page)，用户可以经由该网页提交指示用户对什么感兴趣的查询。响应于查询，搜索服务提供者一般以“搜索结果”页面的形式生成到被认为与该查询相关的网页或站点的链接列表并将该列表发送给用户。

[0007] 查询响应通常包括以下步骤。首先，利用从查询中提取出的一个或多个搜索项 (search term) 搜索网页或站点的预先创建的索引或数据库，以生成命中 (hit)（通常是目标页面或站点，或者是到目标页面或站点的引用，其包含搜索项或者以其他方式被标识为与查询相关）的列表。随后，根据预定义的标准对命中分级 (rank)，并且最佳结果（根据这些标准）被给予最显著的放置，例如在列表顶部。将分级后的命中列表发送给用户，通常以“结果”页面（或一组互连页面）的形式，该页面包含到命中页面或站点的链接列表。在结果页面上也可以包括其他特征，如受赞助链接或广告。

[0008] 命中分级通常是用户的搜索是成功结束还是失败的一个重要因素。查询常常会返回如此大量的命中以致于用户不可能在合理时间内浏览所有命中。如果用户遵循的前几个链接未能引向相关内容，则用户常常会放弃搜索，甚至可能放弃搜索服务提供者，即使相关内容可能在列表下方就可得到也是如此。

[0009] 为了使相关内容被显著放置的似然性最大，搜索服务提供者已经开发了越来越复杂的页面分级标准和算法。在 Web 搜索的早期，分级通常是基于给定页面上搜索项的出现次数和 / 或邻近度的。这已经被证明是不适当的，今天使用的算法一般除了搜索项在页面上的存在情况以外还结合了其他信息，如 Web 上链接到给定目标页面的其他站点的数目（其反映了其他内容提供者认为目标页面有多大用处）。一种算法允许查询用户通过评价

(rate) 返回的命中来提供反馈。用户的评价 (rating) 与查询关联存储, 并且先前的肯定评价被用作下次由任何用户输入相同查询时对命中分级的一个因素。

[0010] 然而, 现有的算法通常不考虑个体用户之间的差异。例如, 输入相同查询的两个用户实际上可能对不同的东西感兴趣; 对于一个用户相关的页面或站点可能对于另一个用户不相关。另外, 用户可以具有个人偏好, 如内容是如何组织并显示的, 其信任哪些内容提供者等等, 这会影响其如何评估或评价给定站点。从而, 满足一个用户 (或许多用户) 需求的站点可能不满足输入相同查询的下一用户的需求, 并且该用户仍然可能以失败告终。

[0011] 另一种用于帮助个体用户找到其感兴趣的内容的工具是“书签 (bookmarking)”。传统上, 书签已实现在 Web 浏览器程序中, 并且在查看任何页面的同时, 用户可以选择保存该页面的书签。书签通常包括页面的 URL (统一资源定位符)、标题和 (可能) 其他信息, 如用户何时访问该页面或者用户何时创建书签。Web 浏览器程序维护书签的列表, 并且用户可以通过在其书签列表中找到书签标记的页面来导航到该页面。为了简化导航书签列表的任务, 大多数书签工具允许用户将其书签组织到文件夹中。更近一些, 某些基于因特网的信息服务具有实现的书签工具, 其允许注册用户从连接到因特网的任何计算机创建并访问个人的书签列表。

[0012] 尽管书签可能是有用的, 但是该工具也有其限制。例如, 将书签组织到文件夹中一般需要相当程度的用户努力, 并且即使利用文件夹, 用户也很难记住哪一个被书签标记的页面有用户在给定时刻可能正在查找的特定信息项。另外, 现有的书签工具通常不帮助用户识别其是否已经对给定页面加上了书签, 也不提供任何用于搜索加上书签的信息的工具。

[0013] 从而, 希望提供用于帮助个体用户收集并搜索其感兴趣的内容的改进的工具。

发明内容

[0014] 本发明的实施例提供了允许用户注释在文档的语料库中找到的内容项或其他内容项 (例如万维网) 的系统和方法。这里所用的术语“注释”通常指与文档 (例如网页或站点) 有关的任何描述性和 / 或评估性的元数据, 其中元数据是从用户收集的, 并随后与该用户关联存储。在本发明的实施例中, 注释可以包括元数据的各种字段, 如文档的评价 (其可以是正面的 (favorable) 或负面的 (unfavorable))、标识文档的 (一个或多个) 主题的关键字列表、文档的自由文本描述和 / 或其他字段。注释有利地被从语料库的用户收集, 并与创建注释的用户的标识符和与之有关的文档 (或其他内容项) 的标识符关联存储。

[0015] 在一个实施例中, 查看搜索结果页面的用户能够注释搜索命中并保存这些注释。在另一个实施例中, 用户可以对其访问的任何页面或站点创建并保存注释。在其他实施例中, 所存储的注释可按各种方式用于增强和个人化搜索和浏览操作。这种增强和个人化的示例包括: 在对应于已注释页面的搜索结果列表中突出显示 (highlight) 命中和 / 或允许用户查看来自搜索结果页面的注释数据; 从注释中提取出用户评价信息并利用该信息来影响搜索结果被表现的方式; 允许用户搜索其已注释页面 (或注释元数据) (除了页面内容外或替代页面内容); 或者允许用户在每次其访问已注释页面或站点时查看和 / 或编辑其注释。

[0016] 根据本发明的一个方面, 一种用于对用户查询作出响应的方法包括接收由用户提

交的查询并搜索包括多个文档的语料库以识别一个或多个命中,其中每个命中是来自语料库的被确定与查询相关的文档。访问由用户创建的注释库,每个注释与语料库中的一个文档相关联,并且包括与该文档有关的用户特定的元数据。将与注释中的匹配注释相关联的每个命中识别为已注释命中。生成搜索报告;搜索报告包括命中列表,并且对于每个命中,还指示该命中是否是已注释命中。搜索报告被发送给用户。语料库可以是例如万维网,并且用户可以是人或计算机(或操作计算机的人)。

[0017] 在某些实施例中,可视突出显示元素被应用于搜索报告中作为已注释命中的每个命中。如果包括在注释中的用户特定的元数据包括评价,则应用于每个已注释命中的可视突出显示元素可以依赖于包括在匹配注释中的评价。生成搜索报告还可以包括对于每个已注释命中,在搜索报告中提供控制元素,该控制元素可由用户操作以请求显示匹配注释的用户特定的元数据,或者对于每个已注释命中,将来自匹配注释的至少某些用户特定的元数据结合到命中列表中。

[0018] 生成搜索报告还可以包括生成只包括已注释命中的单独列表,并且在包括在每个注释中的用户特定的元数据包括关联文档的评价的实施例中,单独列表可以只包括针对其匹配注释包括正面评价的已注释命中。在包括在每个注释中的用户特定的元数据包括关联文档的评价的其他实施例中,生成搜索报告可以包括至少部分基于已注释命中的评价确定命中列表的顺序,并且在评价是负面评价的情况下可以从命中列表中去除已注释命中。

[0019] 在某些实施例中,生成搜索报告还包括对于作为未注释命中的每个命中,提供控制元素,该控制元素可由用户操作以创建与命中相关联的新注释。新注释可以从用户接收并被添加到注释库中。其后,可以重新生成搜索报告以进一步指示新注释的命中是已注释命中,并且重新生成的搜索报告可被发送给用户。

[0020] 在某些实施例中,该方法还包括搜索注释库以识别一个或多个附加的已注释命中,其中每个附加的已注释命中对应于来自语料库的一个文档,对于该文档关联注释包括被确定为与查询相关的用户特定的元数据,并且还包括将附加的已注释命中结合到搜索结果页面的命中列表中。例如,在搜索语料库包括从用户查询中提取出搜索项并将语料库中包含搜索项的每个文档识别为命中的情况下,搜索注释存储可以包括将语料库中用户特定的元数据针对其包括搜索项的每个文档识别为附加的已注释命中。

[0021] 在某些实施例中,注释库包括与语料库中的文档组相关联的至少一个注释,并且作为文档组之一的任何命中被识别为已注释命中。

[0022] 用户特定的元数据有利地包括由用户明确输入的信息项,如关联文档的评价、描述关联文档的关键字、从预定义的词表中选出的标签、关联文档的自由文本描述等等。

[0023] 根据本发明的另一个方面,一种用于对用户查询作出响应的方法包括接收来自用户的查询并访问由用户创建的注释库,其中每个注释与属于一个语料库的多个文档之一相关联,并且包括与关联文档有关的用户特定的元数据。在与注释相关联的文档中识别一个或多个命中;每个命中是来自语料库的被确定与查询相关的文档。生成包括命中列表的搜索报告。列表只包括针对其库具有关联注释的文档。搜索报告被发送给用户。语料库可以是例如万维网,并且用户可以是人或计算机。

[0024] 在某些实施例中,识别一个或多个命中包括将查询和与注释之一相关联的每个文档的内容相比较。例如,可以从查询中提取出搜索项,并且对于与多个注释之一相关联的每

个文档,可以检测搜索项是否存在于文档中;在搜索项存在于文档中的情况下,文档被识别为命中。

[0025] 在其他实施例中,除了将查询与文档内容相比较外(或者替代这种比较),识别一个或多个命中还包括将查询与注释的用户特定的元数据相比较。例如,可以从查询中提取出搜索项,并且对于每个注释,可以检测搜索项是否存在于用户特定的元数据中;在搜索项存在于用户特定的元数据中的情况下,关联文档被识别为命中。对于与多个注释之一相关联的每个文档,还可以检测搜索项是否存在于文档中,并且在搜索项存在于文档中的情况下,文档可以被识别为命中。另外,在用户特定的元数据包括多个字段的情况下,查询可指定在检测动作期间要考虑哪些字段,并且查询还可指定是否要考虑文档内容。

[0026] 在另一个实施例中,用户可以将搜索限制于库的一部分,并且该方法还可以包括从用户接收对要搜索的注释库的子集的选择,其中,识别一个或多个命中的动作只对库的选定子集中的注释执行。

[0027] 在某些实施例中,对于每个命中,搜索报告包括控制元素,该控制元素可由用户操作以请求显示匹配注释的用户特定的元数据。除此之外(或替代此),对于每个命中,搜索报告可包括来自关联注释的至少某些用户特定的元数据。在包括在每个注释中的用户特定的元数据包括关联文档的评价的情况下,列表中的文档可以按至少部分基于命中的评价确定的顺序放置。

[0028] 根据本发明的另一个方面,一种用于对用户查询作出响应的计算机系统包括索引数据存储、个人化数据存储以及通信耦合到索引数据存储和个人化数据存储的搜索服务器。索引数据存储被配置为存储来自一个语料库的多个文档的可搜索表示。个人化数据存储被配置为存储由用户创建的注释库,每个注释与语料库中的文档之一相关联,并且包括与该文档有关的用户特定的元数据。搜索服务器包括输入控制逻辑、搜索控制逻辑、个人化控制逻辑和报告控制逻辑。输入控制逻辑被配置为接收来自查询用户的查询。搜索控制逻辑被配置为搜索索引数据存储以识别一个或多个命中,其中每个命中是来自语料库的被确定为与接收的查询相关的文档。个人化控制逻辑被配置为将与由查询用户创建的注释中的匹配注释相关联的每个命中识别为已注释命中。报告控制逻辑被配置为生成包括命中列表的搜索报告,对于每个命中,搜索报告还指示该命中是否是已注释命中,报告控制逻辑还被配置为将搜索报告发送给用户。

[0029] 根据本发明的另一个方面,一种用于对用户查询作出响应的计算机系统包括索引数据存储、个人化数据存储以及通信耦合到索引数据存储和个人化数据存储的搜索服务器。索引数据存储被配置为存储来自一个语料库的多个文档的可搜索表示。个人化数据存储被配置为存储由用户创建的注释库,每个注释与语料库中的文档之一相关联,并且包括与该文档有关的用户特定的元数据。搜索服务器包括输入控制逻辑、搜索控制逻辑和报告逻辑。输入控制逻辑被配置为接收来自查询用户的查询。搜索控制逻辑被配置为将与由查询用户创建的注释相关联的一个或多个文档识别为命中。报告控制逻辑被配置为生成包括命中列表的搜索报告,列表只包括针对其注释库包括由查询用户作出的查询的文档;报告控制逻辑还被配置为将搜索报告发送给用户。

[0030] 根据本发明的又一方面,一种用于对用户查询作出响应的方法包括访问由多个注释用户创建的注释存储。每个注释与注释用户之一相关联,并且与属于一个语料库的多个

文档之一相关联,每个注释包括与关联文档有关的用户特定的元数据。对多个注释用户聚合与文档中的同一文档有关的用户特定的元数据中的一项或多项,从而生成文档的一个或多个聚合属性。接收由查询用户提交的查询,并搜索语料库以识别一个或多个命中,其中每个命中是来自语料库的被确定与查询相关的文档。生成包括命中列表的搜索报告,其中列表是至少部分基于命中的至少一个聚合属性的。搜索报告被发送给查询用户。语料库可以是例如万维网,并且用户可以是人或计算机(或使用计算机的人)。

[0031] 可以聚合元数据的各种项。例如,用户特定的元数据可包括关联文档的评价,并且聚合可包括计算文档的平均评价,其中平均评价是聚合属性之一。在某些实施例中,每个命中的平均评价可被包括在搜索报告中。在其他实施例中,命中列表以至少部分基于每个命中的平均评价的顺序来安排。

[0032] 作为另一个示例,用户特定的元数据可包括描述关联文档的用户提供的关键字,并且聚合可包括识别最常提供的关键字的关键字集,其中关键字集是聚合属性之一。在某些实施例中,搜索语料库包括从查询中提取出搜索项,并将语料库中针对其关键字集包括搜索项的任何文档识别为命中。在其他实施例中,搜索报告包括每个命中的关键字集。在其他实施例中,除了用户提供的关键字外(或替代用户提供的关键字),还可以使用由注释用户从预定义的词表中选出的标签;预定义的词表可包括例如各种类别标签。

[0033] 在某些实施例中,该方法还包括将针对其注释存储包含由查询用户创建的并与该命中相关联的注释的每个命中识别为已注释命中。对于注释存储包含由查询用户创建的注释的每个命中,在搜索报告中有利地包括关于每个命中是否是已注释命中的指示。搜索报告还可以包括一个控件,该控件可由用户操作以查看用户对于命中的注释和/或从用户对命中的注释中提取出的元数据。

[0034] 根据本发明的另一个方面,一种用于对用户查询作出响应的计算机系统包括索引数据存储、个人化数据存储、通信耦合到个人化数据存储和索引数据存储的聚合模块以及通信耦合到个人化数据存储和索引数据存储的搜索模块。索引数据存储被配置为存储包括多个文档的语料库的可搜索表示。个人化数据存储被配置为存储由多个注释用户创建的注释,每个注释与所述语料库中的文档之一相关联,每个注释包括与该文档有关的用户特定的元数据。聚合模块被配置为对多个注释用户聚合与文档之一有关的用户特定的元数据中的一项或多项,从而生成文档的一个或多个聚合属性,并且还被配置为在索引数据存储中存储文档的每个聚合属性。搜索模块包括输入控制逻辑、搜索控制逻辑和报告控制逻辑。输入控制逻辑被配置为接收由查询用户提交的查询。搜索控制逻辑被配置为搜索索引数据存储并识别一个或多个命中,其中每个命中是语料库中的被确定为与查询相关的文档。报告控制逻辑被配置为生成包括命中列表的搜索报告,列表是至少部分基于命中的聚合属性的,报告控制逻辑还被配置为将搜索报告发送给查询用户。

[0035] 根据本发明的又一方面,一种用于收集与语料库中的文档有关的元数据的方法包括提供具有第一控制元素的用户界面,该第一控制元素可由用户操作以指示注释当前文档的期望,并经由网络从用户接收与当前文档有关的元数据。将包括由用户提供的元数据的注释添加到数据存储,注释持久与用户相关联,并且与当前文档相关联。其后,响应于用户请求而经由网络将注释的元数据发送给用户。语料库可以是例如万维网,并且用户可以是人或计算机(或操作计算机的人)。

[0036] 在某些实施例中,第一控制元素还可由用户操作以提交元数据。在其他实施例中,第一控制元素还可由用户操作以获得用于提交元数据的表格。

[0037] 在某些实施例中,提供用户界面包括将第一控制元素添加到响应于由用户提交的查询而生成的搜索报告。例如,如果搜索报告包括命中列表,每个命中是来自语料库的文档,则搜索报告可以提供对应于每个命中的第一控制元素的不同实例,从而在用户操作第一控制元素的相应实例的情况下,该命中被识别为当前文档。

[0038] 在某些实施例中,如果搜索报告包括命中列表,每个命中是来自语料库的文档,则该方法还包括对于每个命中,确定数据存储是否包括与用户和命中相关联的注释。在数据存储包括与用户和命中相关联的注释的情况下,在搜索结果页面中包括第二控制元素,该第二控制元素可由用户操作以查看注释的元数据。在数据存储不包括与用户和命中相关联的注释的情况下,在搜索结果页面中包括对应于命中的第一控制元素的实例,从而在用户操作第一控制元素的相应实例的情况下,该命中被识别为当前文档。

[0039] 在其他实施例中,第一控制元素被提供在文档浏览器的工具栏界面中,该文档浏览器被配置为显示来自语料库的文档。并且当用户操作第一控制元素时被文档浏览器显示的文档被识别为当前文档。第一控制元素可以只在文档浏览器正在显示这样的文档时是活动的:对于该文档,数据存储不包括与所显示的文档和用户相关联的注释。在工具栏界面中也可以提供第二控制元素,该第二控制元素可由用户操作以请求发送正由文档浏览器显示的文档的元数据;第二控制元素有利地只在文档浏览器正在显示这样的文档时是活动的:对于该文档,数据存储包括与所显示的文档和用户相关联的注释。

[0040] 元数据的各种项可被包括在注释中。例如,元数据可以包括关联文档的评价,并且可以在向用户显示关联文档的任何时刻显示评价。在某些实施例中,该方法还包括接收来自用户的查询并响应于查询为用户生成搜索报告,其中搜索报告包括语料库中被确定为与查询相关的文档的列表。搜索报告还可以包括列表中每个文档的评价,对于该文档,数据存储包括与文档和用户相关联的注释。

[0041] 作为另一个示例,元数据可以包括描述关联文档的用户提供的关键字、由用户从预定义的词表中选出的标签和/或由用户提供的文档的自由文本描述。在某些实施例中,该方法还可以包括接收来自用户的查询,并利用关键字和/或标签和/或描述来确定关联文档是否与查询相关。

[0042] 在某些实施例中,发送元数据包括提供第二控制元素,该第二控制元素可由用户操作以编辑元数据。当从用户接收到已编辑的元数据时,可以利用已编辑的元数据来更新数据存储中的注释。

[0043] 在某些实施例中,该方法还包括提供第二控制元素,该第二控制元素可由用户操作以请求显示来自数据存储的注释。响应于第二控制元素的操作,从数据存储中检索已请求的注释,并将检索到的注释的元数据发送给用户。第二控制元素可以被提供在例如文档浏览器的工具栏界面中,该文档浏览器被配置为显示来自语料库的文档。

[0044] 根据本发明的另一个方面,一种用于收集与语料库中的文档有关的元数据的计算机系统包括个人化数据存储和通信耦合到个人化数据存储的注释模块。个人化数据存储被配置为存储多个注释,每个注释与注释用户相关联,并且与语料库中的文档之一相关联。每个注释包括与文档之一有关的元数据,该注释与该文档是相关联的。注释模块包括输入控

制逻辑、存储控制逻辑和检索控制逻辑。输入控制逻辑被配置为接收来自注释用户的元数据并将接收的元数据与语料库中的文档之一相关联。存储控制逻辑被配置为在个人化数据存储中存储接收的元数据,作为与注释用户的标识符和文档之一相关联的注释。检索控制逻辑被配置为通过参考注释用户的标识符和文档中的关联文档来在个人化数据存储中定位注释。

[0045] 下面的具体实施方式以及附图将提供对于本发明本质和优点的更好的理解。

附图说明

- [0046] 图 1 是根据本发明实施例的信息检索和通信网络的框图。
- [0047] 图 2 是根据本发明另一个实施例的信息检索和通信网络的框图。
- [0048] 图 3 是根据本发明实施例的注释记录的示意图。
- [0049] 图 4 是根据本发明实施例的用于组织注释的文件夹记录的示意图。
- [0050] 图 5 是根据本发明实施例的用于查看和组织注释的用户界面页面的示例。
- [0051] 图 6 是根据本发明实施例的搜索结果页面的示例。
- [0052] 图 7 是根据本发明实施例的 Web 浏览器的搜索工具栏的示例。
- [0053] 图 8 是根据本发明实施例的用于创建注释的过程的流程图。
- [0054] 图 9 是根据本发明实施例的用于创建注释的用户界面页面的示例。
- [0055] 图 10 是根据本发明另一个实施例的搜索结果页面的示例。
- [0056] 图 11 是根据本发明另一个实施例的 Web 浏览器的搜索工具栏的示例。
- [0057] 图 12 是根据本发明实施例的另一个搜索结果页面的示例。
- [0058] 图 13 是根据本发明实施例的显示注释的页面覆盖图的示例。
- [0059] 图 14 是根据本发明实施例的用于执行搜索的过程的流程图。
- [0060] 图 15 是根据本发明实施例的用于搜索注释库的搜索界面页面的示例。

具体实施方式

[0061] 本发明的实施例提供了允许用户注释其在文档语料库中找到的内容项或其他内容项(例如万维网)的系统和方法。这里所用的术语“注释”通常指与文档(例如网页或站点)有关的任何描述性和/或评估性的元数据,其中元数据是从用户收集的,并随后与该用户关联存储。在本发明的实施例中,注释可以包括元数据的各种字段,如文档的评价(其可以是正面的或负面的)、标识文档的(一个或多个)主题的关键字列表、文档的自由文本描述和/或其他字段。注释有利地被从语料库的用户收集,并与创建注释的用户的标识符和与之有关的文档(或其他内容项)的标识符关联存储。

[0062] 在一个实施例中,查看搜索结果页面的用户能够注释搜索命中并保存这些注释。在另一个实施例中,用户可以对其访问的任何页面或站点创建并保存注释。在其他实施例中,所存储的注释可按各种方式用于增强和个人化搜索和浏览操作。例如,当用户搜索语料库时,对应于用户已注释的页面的任何命中(这里称为“已注释命中”)可被突出显示,同时提供链接以允许用户查看其注释。在注释包括诸如数字评价之类的判断数据的情况下,已注释命中可被突出显示以指示用户的判断是正面的还是负面的。评价也可用于对响应于用户查询的搜索结果分级,其中正面的判断倾向于增大给定页面或站点的等级,而负面的

判断倾向于减小等级。在注释包括用户提供的自由文本和 / 或描述性关键字或标签的情况下,用户可以具有搜索其注释 (除了页面内容外或替代页面内容) 的选项。在其他实施例中,在用户访问其已注释的页面的任何时刻,都提供了允许用户查看和 / 或编辑其注释的控件。

[0063] 为了说明目的,本发明的描述和附图可能利用特定查询、搜索结果页面、URL 和 / 或网页。这种使用并不是要暗示对任何实际的网页或站点的任何意见、认可或者蔑视。另外,应当理解,本发明并不限于这里说明的具体示例。

[0064] I. 概述

[0065] A. 网络实现方式概述

[0066] 图 1 图示了根据本发明实施例包括客户端系统 20 的信息检索和通信网络 10 的概况。在计算机网络 10 中,客户端系统 20 通过因特网 40 或其他通信网络 (例如经由任何局域网 (LAN) 或广域网 (WAN) 连接) 耦合到任何数目的服务器系统 50_1 至 50_N 。如这里所描述的,根据本发明,客户端系统 20 被配置为与服务器系统 50_1 至 50_N 中的任何一个通信,以例如访问、接收、检索和显示媒体内容和其他信息 (如网页)。

[0067] 图 1 所示的系统中的若干元件包括这里不需要详细说明的传统的公知元件。例如,客户端系统 20 可包括桌面个人计算机、工作站、膝上型电脑、个人数字助理 (PDA)、手机或任何能够使用 WAP 的设备或任何其他能够直接或间接接口到因特网的计算设备。客户端系统 20 一般运行浏览程序,如微软的 Internet Explorer™ 浏览器、Netscape Navigator™ 浏览器、Mozilla™ 浏览器、Opera™ 浏览器、或在手机、PDA 或其他无线设备的情况下的能够使用 WAP 的浏览器等等,从而允许客户端系统 20 的用户访问、处理和查看可经由因特网 40 从服务器系统 50_1 至 50_N 获得的信息和页面。客户端系统 20 一般还包括一个或多个用户接口设备 22,如键盘、鼠标、触摸屏、笔等等,其用于结合由服务器系统 50_1 至 50_N 或其他服务器提供的页面、表格和其他信息,与由显示器 (例如,监视器屏幕、LCD 显示器等) 上的浏览器提供的图形用户界面 (GUI) 交互。本发明适合于用于因特网,因特网指网络的特定全局互连网络。然而,应当理解,也可以使用其他网络来替代因特网或与因特网一起使用,如内联网、外联网、虚拟专用网络 (VPN)、基于非 TCP/IP 的网络、任何 LAN 或 WAN 等。

[0068] 根据一个实施例,客户端系统 20 和其所有组件都可由操作者利用包括计算机代码的应用来配置,所述计算机代码是用中央处理单元 (如 IntelPentium™ 处理器、AMD Athlon™ 处理器等或多处理器) 来运行的。如这里所述的用于操作和配置客户端系统 20 传输、处理和显示数据和媒体内容的计算机代码优选地被下载并存储在硬盘上,但是全部程序代码或其一部分也可以存储在任何其他公知的易失性或非易失性存储器介质或设备 (如 ROM 或 RAM) 中,或者提供在能够存储程序代码的任何介质上,如致密盘 (CD) 介质、数字通用盘 (DVD) 介质、软盘等。另外,全部程序代码或其一部分可以从软件源发送和下载,例如从服务器系统 50_1 至 50_N 之一经由因特网发送到客户端系统 20,或者利用任何通信介质和协议 (例如 TCP/IP、HTTP、HTTPS、以太网或其他传统介质和协议) 经由任何其他网络连接 (例如外联网、VPN、LAN 或其他传统网络) 发送。

[0069] 应当意识到,用于实现本发明某些方面的计算机代码可以是 C、C++、HTML、XML、Java、JavaScript 等代码,或任何其他合适的脚本语言 (例如 VBScript)、或任何其他合适的编程语言,这些语言可以在客户端系统 20 上执行或可以被编译以在客户端系统 20 上执

行。在某些实施例中,不向客户端系统 20 下载任何代码,并且所需的代码由服务器执行,或者已经存在于客户端系统 20 上的代码被执行。

[0070] B. 搜索和注释系统概述

[0071] 图 2 图示了根据本发明实施例用于传输媒体内容的另一个信息检索和通信网络 110。如图所示,网络 110 包括客户端系统 120、一个或多个内容服务器系统 150 和搜索服务器系统 160。在网络 110 中,客户端系统 120 通过因特网 140 或其他通信网络通信耦合到服务器系统 150 和 160。如上所述,客户端系统 120 和其组件被配置为经由因特网 140 或其他通信网络与服务器系统 150 和 160 以及其他服务器系统通信。

[0072] 根据一个实施例,在客户端系统 120 上执行的客户端应用(表示为模块 125)包括用于控制客户端系统 120 和其组件与服务器系统 150 和 160 通信并处理和显示从其接收的数据内容的指令。客户端应用 125 优选地从诸如远程服务器系统(例如,服务器系统 150、服务器系统 160 或其他远程服务器系统)的软件源发送并下载到客户端系统 120,但客户端应用模块 125 也可以提供在如上所述的任何软件存储介质上,如软盘、CD、DVD 等。例如,在一个方面,客户端应用模块 125 可以经由因特网 140 以 HTML 包装(wrapper)的形式被提供到客户端系统 120,该 HTML 包装包括各种控件,例如嵌入式 JavaScript 或 Active X 控件,其用于操作数据并在各种对象、框架和窗口内表现数据。

[0073] 另外,客户端应用模块 125 包括各种用于处理数据和媒体内容的软件模块,如用于处理搜索请求和搜索结果数据的专用搜索模块 126,用于在文本和数据框架和活动窗口(例如浏览器窗口和对话框)内表现数据和媒体内容的用户接口模块 127 以及用于接口到客户端 120 上执行的各种应用并与之通信的应用接口模块 128。根据本发明的某些方面,应用接口模块 128 优选地被配置为与之接口的在客户端系统 120 上执行的应用的示例包括各种电子邮件应用、即时消息(IM)应用、浏览器应用、文档管理应用等等。另外,用户接口模块 127 可包括浏览器,如配置在客户端系统 120 上的默认浏览器或不同的浏览器。

[0074] 根据一个实施例,搜索服务器系统 160 被配置为向客户端系统 120 提供搜索结果数据集和媒体内容,内容服务器系统 150 被配置为例如响应于在由搜索服务器系统 160 提供的搜索结果页面中选定的链接而向客户端系统 120 提供诸如网页的数据和媒体内容。在某些变体中,搜索服务器系统 160 返回内容以及/或者链接和/或其他对内容的引用。搜索服务器系统包括被配置为接收来自用户的查询并为其生成搜索结果数据的查询响应模块 162 以及被配置为检测用户与搜索结果数据的交互并响应于该交互的用户注释模块 164,如下所述。

[0075] 在一个实施例中,查询响应模块 162 参考各种页面索引 170,这些页面索引 170 被填充以例如页面、到页面的链接、代表索引页面的内容的数据等等。页面索引可通过各种收集技术生成,包括自动 web 挖掘器(crawler) 172 和/或各种网蜘蛛(spider)等,以及用于对分级体系结构内的网页分类并分级的手工或半自动分类算法和界面。这些技术可以实现于搜索服务器系统 160 中或单独的系统(例如 web 挖掘器 172)内,这种单独的系统生成页面索引 170 并使其可用于搜索服务器系统 160。各种页面索引实现方式和格式是本领域中已知的,并且可用于页面索引 170。

[0076] 在一个实施例中,查询响应模块 162 还参考个人化数据库 166。可以利用传统数据库技术实现的个人化数据库 166 包括用户特定的信息,尤其是各种网页或站点的用户提供

的注释的具体记录。如下所述,注释可以包括任何类型的包括描述性和 / 或评估性信息的用户提供的元数据;元数据可以包括自由文本、关键字或标签、反映用户对于特定网页或站点的价值的判断(例如,正面的或负面的)的数字评价等等。注释的内容和格式以及用于收集存储在个人化数据库 166 中的注释的技术的示例将在下面描述。

[0077] 查询响应模块 162 被配置为提供响应于接收自客户端系统 120(尤其接收自搜索模块 126)的各种搜索请求(查询)的数据。这里所用的术语“查询”包含从用户到搜索服务器 160(例如经由客户端 120)的任何请求,其可通过搜索由页面索引 170 索引的 Web(或其他语料库)来满足。在一个实施例中,经由搜索模块 126 向用户呈现搜索界面。界面可包括文本框(用户可以向文本框中输入查询(例如通过打字))、复选框和 / 或单选按钮,单选按钮用于从预定义的查询中选择使用户能够将搜索限制在完全搜索语料库的预定义子集中(例如,限制在页面索引 170 内的某些网站或分类部门中)的目录或其他结构等。可以使用任何搜索界面。

[0078] 查询响应模块 162 有利地被配置具有与搜索相关的算法,其用于处理与给定查询相关的网页并对网页分级(例如,基于以下各项的组合:由查询中搜索项的出现模式测量的逻辑相关度;与搜索项和 / 或具体页面或站点相关联的上下文标识符;页面赞助者;从多个页面收集的连通性数据等)。例如,查询响应模块 162 可以分析接收到的查询以提取出一个或多个搜索项,然后利用这些搜索项访问页面索引 170,从而生成“命中”列表,即被确定与查询至少具有某些相关性的页面或站点(或对页面或站点的引用)。查询响应模块 162 随后可利用一种或多种分级算法对命中分级。用于识别命中并对命中分级的具体算法对本发明而言是不重要的,并且可以使用传统算法。

[0079] 在本发明的实施例中,查询响应模块 162 还被配置为从个人化数据库 166 中检索与输入当前查询的用户相关联的任何注释数据,并将这种注释数据结合到搜索结果中。例如,在至少某些注释包括评价(或反映用户对页面或站点的评估的其他数据)的情况下,查询响应模块 162 可以基于由该用户预先注释的特定页面或站点的正面用户评价来生成“正面”结果的单独列表;或者查询响应模块 162 可以在搜索结果的等级中结合特定页面或站点的用户评价;或者查询响应模块 162 可以使用特定页面或站点的负面用户评价来确定是否从结果列表中丢弃命中。在注释包括自由文本、关键字或标签的情况下,在搜索命中的识别和 / 或分级期间可以考虑这些元素中的任何一种中搜索项的出现情况。

[0080] 用户注释可以用各种方式提供给个人化数据库 166。在某些实施例中,搜索结果数据被呈现为包括命中列表的结果页面。对于每个命中,结果页面可包括例如页面或站点标题、到页面或站点的链接、来自该页面或站点的内容一个或多个摘录(例如,显示搜索项出现的上下文)以及其他选项,如到内容的缓存拷贝的链接。结果数据也可以包括按钮或允许用户注释任何命中页面或站点的其他界面元素。例如,用户可被邀请按预定义刻度(例如,拇指向上或拇指向下,0 到 4 星,从 1 到 10 的数字评价等)评价页面或站点,输入页面或站点的自由文本描述,从预定义列表中选择描述页面或站点的标签,或者输入描述页面或站点的一个或多个关键字。

[0081] 当用户选择注释命中页面或站点时,用户注释模块 164 接收来自用户(例如经由客户端系统 120)的新的注释数据并更新个人化数据库 166。

[0082] 在一个实施例中,用户注释模块 164 响应于新的注释数据还发起结果页面的自动

刷新。在该刷新操作期间,在结果页面上列出的命中可以利用新数据重新分级,并且反映新分级的更新后的结果页面被发送给用户。在新注释包括正面评价的情况下,新页面也可以在“正面”结果中列出新注释的站点。因此,用户的新注释可以对当前搜索的显示结果有直接影响,并且被存储以用于处理来自该用户的未来查询。

[0083] 为了启用诸如用户注释之类的搜索个人化特征,搜索服务器 160 有利地提供用户登录特征,其中“登录”通常指用于识别和 / 或验证计算机系统的用户的任何方案。大量示例在本领域中是已知的,并且可以与本发明的实施例结合使用。例如,在一个实施例中,每个用户有唯一的用户标识符 (ID) 和密码,并且搜索服务器 160 通过向客户端 120 传递登录页面 (用户可经由登录页面输入该信息) 来提示用户登录。在其他实施例中,除了用户 ID 和密码外 (或者替代这些技术),还可以使用生物度量、语音或其他识别和验证技术。一旦用户已例如通过登录识别其自身,则用户就可以通过与用户注释模块 164 交互来输入和 / 或更新其注释,如下所述。另外,登录用户输入的每个查询可以与该用户的唯一用户 ID 相关联;基于用户 ID,查询响应模块 162 可以访问个人化数据库 166 以将用户的存储注释结合到对该用户的查询的响应中。用户登录有利地是持久的,其意义是一旦用户已登录 (例如经由客户端应用 125),就可以在用户操作客户端应用 125 的同时,在任何适当的时刻将用户的身份传输到搜索服务器 160。从而,这里描述的个人化特征对用户来说可以是可连续访问的。

[0084] 除了使用用户自己的响应于查询的注释外,查询响应模块 162 还可以使用关于其他用户的注释的聚合信息。例如,在一个实施例中,从为页面或站点提供了具有一定评价的注释的每一用户的评价计算页面或站点的聚合评价 (例如,平均评价)。在另一个实施例中,可以例如通过识别由已对页面或站点作出注释的用户最常应用于该页面或站点的关键字或标签,来确定描述页面或站点的聚合关键字或标签。这种用于给定页面的聚合注释可例如存储在页面索引 170 中,并且被查询响应模块 162 用来响应于查询而对命中分级 (无论用户对于搜索服务器 160 来说是否是已知的)。

[0085] 在一个实施例中,用户注释模块 164 在接收到新的注释数据时将其转发到聚合器模块 (未在图 2 中示出),聚合器模块更新存储在页面索引 170 中的聚合注释数据。聚合注释数据可以以定期间隔 (例如每日或每小时) 更新,或者近似实时更新。聚合注释数据的收集和使用将在下面进一步描述。

[0086] 应当意识到,这里描述的搜索系统是示例性的,也可以有变化和修改。内容服务器和搜索服务器系统可以是单个组织的一部分,例如由 Yahoo! 公司提供给用户的分布式服务器系统,或者它们可以是不同组织的一部分。每个服务器系统通常包括至少一个服务器和关联的数据库系统,并且可以包括多个服务器和关联的数据库系统,并且尽管被示为单个块,但是也可以是地理上分布的。例如,搜索服务器系统的所有服务器可以彼此相邻 (例如在位于单个建筑物或校园内的服务器群中),也可以分布在彼此远离的位置处 (例如,一个或多个服务器位于 A 城市中,一个或多个服务器位于 B 城市中)。从而,这里所用的“服务器系统”一般包括分布在本地或分布在一个或多个地理位置处的一个或多个逻辑和 / 或物理上相连的服务器;术语“服务器”和“服务器系统”可互换使用。另外,这里描述的查询响应模块和用户注释模块可以实现在同一服务器上,也可以实现在不同服务器上。

[0087] 搜索服务器系统可被配置具有一个或多个页面索引和算法,该算法用于响应于接

收自客户端系统的搜索查询而访问（这一个或多个）页面索引并向用户提供搜索结果。搜索服务器系统可自己生成页面索引，从另一个源（例如单独的服务器系统）接收页面索引，或者从另一个源接收页面索引并对其执行进一步的处理（例如各种页面信息的添加或更新）。另外，尽管搜索服务器系统被描述为包括组分模块的特定组合，但是应当理解，对模块的划分仅仅是为了描述方便；可以定义更多的、更少的或不同的模块。

[0088] 另外，在某些实施例中，这里描述的由搜索服务器 160 维护的某些模块和 / 或元数据可以整体或部分驻留在客户端系统上。例如，用户注释的某些或全部可以存储在客户端系统 120 本地，并且由客户端应用 125 的组分模块管理。其他数据（包括页面索引 170 的某些部分或全部）可以周期性地从搜索服务器 160 下载并由客户端系统 120 存储以供后续使用。另外，客户端应用 125 可以创建并管理存储在客户端 120 本地的内容索引，并且还可以提供搜索本地存储的内容的能力，将包括本地存储的内容的搜索结果结合到 Web 搜索结果中等等。从而，搜索操作可包括由搜索服务器系统和 / 或客户端系统进行的操作的任何组合。

[0089] 根据本发明的一个实施例中，用户可以经由用户注释模块 164 与搜索服务器 160 交互以注释搜索结果和其他页面。注释可以包括肯定或否定的评价（或者反映用户对主题页面或站点的正面或负面意见的其他信息）、关键字、自由文本描述和 / 或提供关于页面或站点的描述性和 / 或评估性信息的其他元素。每个用户的注释被存储在个人化数据库 166 中，并且被有利地用于对由该用户提交的当前和未来的查询作出响应。在某些实施例中，用户注释也可用来提供各种定制化搜索服务，如下所述。

[0090] 在本发明的实施例中，可以用多种方式从用户收集注释，包括从搜索结果页面输入的注释、利用工具栏界面输入的注释等等。收集注释数据的示例如下所述。

[0091] II. 用户注释数据格式

[0092] 用户注释可以用多种格式收集，并且可以提供关于页面或站点的各种信息。尽管这里描述了示例性格式，但是应当理解，用户注释并不限于该示例的特定内容或格式，并且注释可以包括与页面或站点有关的用户提供的和 / 或自动生成的元数据的任何组合。

[0093] 这里所用的“页面”指可由唯一定位符（例如 URL）标识并且可由适当配置的浏览器程序显示的内容单元。“站点”指与共同主题有关并且位于同一服务器上的一个或多个页面的群组。在本发明的某些实施例中，创建注释的用户可以指示注释是应当应用于单个页面还是应用于一组相关页面。在后者的情况下，用户可以有利地定义站点的范围，如下所述。在某些实施例中，除了注释应用于的页面的数目不同以外，页面注释和站点注释之间没有其他区别。

[0094] A 注释内容

[0095] 在一个实施例中，每个注释在个人化数据库 166 中是结构化的条目。图 3 图示了注释 300 的内容字段。左列 302 中的字段可以由用户注释模块 164 自动生成和更新；右列 304 中的字段优选地是由用户提供的。

[0096] 自动生成的字段包括存储创建注释的用户的用户 ID 的“作者 ID” (Author ID) 字段 306 和标识注释所属的页面或站点的“URL”字段 308。“主机标志” (Host flag) 字段 310 指示注释是应用于页面还是应用于站点。如果主机标志被设为“页面” (page)，则注释只应用于其 URL 与字段 308 中的串精确匹配的页面，而如果主机标志被设为“站点” (site)，则

注释应用于其 URL 以字段 308 中所示的串开始的任何页面。从而,主机标志被设为“站点”的注释可以应用于任何数目的页面。主机标志字段 310 可自动设为默认值(例如“页面”),并且用户可被给予改变值的选项。

[0097] “标题”(Title) 字段 312 存储主题页面(或站点)的标题。该字段有利地被填充以从注释页面的源代码提取的页面标题作为默认值;在某些实施例中,允许用户改变标题。“摘要”(Abstract) 字段 314 存储主题页面或站点的文本摘要;该摘要可以自动生成(如下所述),或者用户可被邀请应用摘要。

[0098] 列 302 中的剩余字段提供关于注释的历史信息。例如,“介绍”(referral) 字段 316 提供关于用户如何到达主题页面或站点的上下文信息。在注释从搜索结果界面输入的情况下,介绍字段 316 有利地包括查询(响应于该查询产生了结果页面,如图 3 所示)。在用户经由工具栏界面提供当前显示的页面的注释的情况下,介绍字段 316 可包括关于用户在导航到主题页面并注释主题页面之前正在查看的内容的历史信息。例如,如果用户通过遵循来自另一页面的链接导航到主题页面,则介绍字段 316 可包含该另一页面的 URL。如果用户通过点击另一程序(例如,电子邮件客户端或字处理器)上的链接导航到主题页面,则介绍字段 316 可标识该程序。如果用户手工输入主题页面的 URL(例如,将其打字到浏览器窗口的地址栏中),则介绍字段 316 可指示手工输入。应当注意,介绍字段并不限于单个条目;例如,在某些实施例中,介绍字段可存储更长的导航踪迹(例如,引导向导航到主题页面并注释主题页面的全部浏览器会话,包括在该会话期间输入的任何搜索查询(无论主题页面是作为命中返回的还是由用户访问的))。

[0099] 在用户已注释页面并随后修改该注释的情况下,介绍字段 316 有利地被更新以标识引导致修改后的注释的介绍源。“老的介绍”(Oldreferral) 字段 318 可用于存储与前一注释有关的上下文信息;该信息类似于存储在介绍字段 316 中的信息。可以维护任何数目的老的介绍。

[0100] “最后更新”(Last updated) 字段 320 提供指示用户何时最后更新注释的时间戳。“最后访问”(Last Visited) 字段 322 提供指示用户何时最后访问注释页面的时间戳。尽管图 3 以年-月-日时:分:秒格式示出了这些时间戳,但是应当理解,可以换用其他格式和任何期望的精度。该信息可用来例如识别可能可靠性更低的更老的注释(尤其在注释页面的更新比用户最后访问该页面的时间更近的情况下)。

[0101] 列 304 中的字段由用户提供,并且有利地保持为空,直到并且除非用户提供数据。在优选实施例中,并不要求用户输入所有这些字段的数据,并且在注释数据被用于搜索处理时(如下所述)可以忽略任何空字段。

[0102] “关键字”(keywords) 字段 324 存储描述被注释页面的一个或多个用户提供的关键字或用户选择的标签。这里所用的“关键字”(在现有技术中有时也被称为 tag)指由可自由选择任何词或短语的用户提供的词或短语,而“标签”(label)指由用户从系统定义的词表(如类别标识符的分级列表)中选择的词或短语。在使用关键字的某些实施例中,各种提示或其他技术可被结合到系统中以鼓励用户采用“正规化”关键字(例如,标准拼写、同义词间的标准选择,如“bicycling”和“cycling”和“biking”,等等)。在页面的聚合关键字是由不同用户从注释中导出的情况下,关键字正规化尤其有用,如下所述。在某些实施例中,用户可被限制于某个最大数目(例如 10、20 或另一个数)的关键字或标签。

[0103] “描述”(Description) 字段 326 存储由用户提供的对页面或站点的自由文本描述。在填充该字段时,用户不被限制于词或短语,也不被限制于任何特定长度。在某些实施例中,描述字段 326 允许用户编写并存储相当长的内容论述(例如,长达 500 或 1000 个词),并且用户还可被允许将到其他内容的链接包括为描述的一部分。链接可被包括,以例如用来标识提供关于由注释页面提及的主题的更详细信息的其他站点。

[0104] “评价”(Rating) 字段 328 存储反映用户对页面的意见或判断的数字值或其他指示符。评价可以利用各种刻度提供,刻度优选地允许至少“正面”(favorable)、“负面”(unfavorable) 和“中立”(neutral) 评价。例如,在一个实施例中,在注释创建期间用户被提示对已注释页面给予正面(例如拇指向上)或负面(例如拇指向下)评价。正面和负面评价各自被分配以数字值(例如分别是 +2 和 -2);未被评价的页面被给予为 0 的默认(或中立)评价。也可以使用其他评价系统(例如 0-4 星、1-10 级等)。存储在字段 328 中的评价指示符不需要与用户使用的评价刻度匹配(例如,如果用户按 1-10 的刻度评价站点,则这可被转换为范围从 -4 到 5 的评价指示符)。用户加以注释但是不评价的任何站点都有利地被视为具有中立评价。

[0105] 应当理解,注释条目 300 是示例性的,也可以使用具有不同字段的其他注释格式。例如,在某些实施例中,注释可包括压缩或未压缩形式的主题页面的内容的一部分或全部的表示。在其他实施例中,用户可以将描述连接到主题页面的内容的特定部分,并且描述所连接到的部分可被存储在注释中。在另一实施例中,搜索服务器 160 还可以根据某种分类法对页面或站点分类,并且这种分类数据可被保存为注释的一部分。

[0106] 与已注释页面(或站点)有关的其他元数据也可以被收集在注释记录中,并且在用户继续浏览时被自动更新。例如,可以提供计数器来对用户访问已注释页面或站点的次数计数。计数器和/或最后访问的时间戳可以在每次用户访问页面或站点时自动更新。在某些实施例中,只有那些在用户登录到搜索服务器 160 中的同时发生的访问才导致自动更新。

[0107] 在某些实施例中,注释还可以包括非用户特定的元数据。例如,元数据还可能包括与主题页面或站点有关的真实位置(例如,纬度和经度坐标、街道地址等)或电话号码、与主题页面或站点有关的 UPC(通用产品码)或 ISBN(国际标准书号)或 ISSN(国际标准序列号)和/或其他类似信息。另外,与主题页面或站点的各种属性或行为有关的元数据(如其是否包括成人内容,其使用了哪种语言(例如英语、日语、西班牙语等),其或是否包括弹出式窗口等等)也可以与用户输入相独立地被结合到注释中。

[0108] 注释条目可以采取适合于存储在个人化数据库 166 中的任何格式(例如关系数据库方案、XML 记录等),并且可通过引用各种字段来访问。在一个实施例中,注释记录可通过至少作者 ID、URL、标题和关键字来访问。

[0109] B. 注释的组织

[0110] 在某些实施例中,用户可以利用文件夹来组织其注释。例如,每个用户可以具有“主”(Main)文件夹,用户的新注释默认放置到主文件夹中。用户可以按需要创建附加文件夹。在某些实施例中,用户还可以定义文件夹内的子文件夹。用于创建并管理文件夹的用户界面可以是传统的设计。

[0111] 在一个实施例中,每个文件夹在个人化数据库 166 中利用文件夹条目定义。图 4

图示了根据本发明实施例的文件夹条目 400。文件夹条目 400 包括提供对注释和 / 或属于文件夹 400 的子文件夹的引用（例如持久指针）的引用字段 404；链接列表或其他合适的数据结构可用于实现引用 404。

[0112] 文件夹条目 400 还有利地包括可用于文件夹管理的其他字段。在一个实施例中，这些字段包括存储文件夹所属的用户的用户 ID 的“作者 ID”字段 406 和存储用户提供的文件夹名（例如，上限为 80 个字符）的“名称”(Name) 字段 408。“名称”字段 408 可以默认是“新建文件夹”(New Folder) 或某些其他合适的串。“描述”(Description) 字段 410 存储用户可编辑的对文件夹目的或内容的自由文本描述；该字段可默认为空状态。“活动”(Active) 字段 412 存储指示该文件夹中的注释是否应当用于对查询作出响应的标志（例如布尔值）。

[0113] 应当意识到，文件夹格式可以变化，并且可以包括其他字段。除了“主”文件夹外，用户可以自由地创建、重命名和删除文件夹。在某些实施例中，多个文件夹可以存储对相同注释的引用；在其他实施例中，每个注释一次只分配给一个文件夹，并且用户可以将注释从一个文件夹移动到另一个，或者在不同文件夹中创建注释的拷贝。在某些实施例中，每个注释条目还可以包括存储返回到对注释被分配给的（一个或多个）文件夹的引用的“文件夹 ID”(folder ID) 字段。

[0114] 尽管文件夹是可选的，但是提供文件夹允许用户对搜索体验的控制度更大。例如，用户可以在多个文件夹中安排其注释，并且对于文件夹中的一个或多个将活动标志设为真(true)，而对于其他文件夹设为假(false)。当用户输入查询时，只有（一个或多个）活动文件夹中的注释会影响结果。用户还可以使用文件夹来以与各种 Web 浏览器程序或因特网入口服务所支持的“书签”或其他个人站点列表在某种程度上类似的方式收集并组织已注释页面。在优选实施例中，这里描述的文件夹和注释数据是由搜索服务器 160 为用户维护的，并且对于用户可用，而无论用户访问搜索服务器 160 的位置如何。

[0115] 在另一个实施例中，不使用文件夹，而是基于注释记录中的用户提供的关键字或标签来管理注释的使用。例如，活动标志和 / 或公布标志可以按关键字而不是按文件夹来定义。

[0116] C. 到注释的界面页面

[0117] 随着用户找到并注释各种页面或站点，每个用户将建立对于该用户有用的个人 Web 内容“库”。在某些实施例中，用户注释模块 164(图 2) 提供了允许用户管理其注释库并以多种方式与其注释数据交互的界面。例如，用户注释模块 164 可被配置为生成定制的“我的库”(My Library) 页面并向每个用户提供该页面。登录到搜索服务器 160 的用户可以例如通过以下方式请求该页面：循着来自搜索界面或搜索结果页面的链接，循着来自搜索服务器 160 所容宿的个人化或通用入口页面或站点的链接，或者激活浏览器工具栏或其他持久界面上的按钮。在某些实施例中，未登录的用户可以请求其“我的库”页面，然后在页面被显示前被提示登录。

[0118] 图 5 是根据本发明实施例用户的库界面页面 500 的一种实现方式的示例。页面 500 包括显示来自用户的库的注释的部分 501。可以显示来自每个注释的各种字段（例如，图 3 中所示的字段），并且有利地提供到每个已注释页面或站点的链接。“编辑”(Edit) 按钮 503 允许用户编辑注释。编辑界面通常类似于创建界面（将在下面描述），并且也可以允许用户将注释从一个文件夹移动或拷贝到另一个。

[0119] 部分 502 提供了用于选择和安排要显示在部分 501 中的界面控件。例如,按钮组 504 允许用户选择要查看的一个或多个文件夹。默认选择是“所有”(all),并且激活组 504 中的“选择文件夹”(Select Folders) 选项将用户带到文件夹选择界面(未示出),其可以是传统设计,并且可允许用户选择一个文件夹、一组文件夹或者文件夹和其子文件夹。当用户完成选择要查看的文件夹并返回到页面 500 时,页面被更新以反映当前的文件夹选择。

[0120] 复选框 506 允许用户请求只显示那些包括一个或多个用户特定的关键字的注释。在选中复选框 506 并在文本框 508 中输入一个或多个期望的关键字后,用户激活“刷新”(Refresh) 按钮 510 以刷新列表。响应于按钮 510 的激活,搜索服务器 160 搜索用户注释的关键字字段以查找来自文本框 508 的(一个或多个)关键字,并更新部分 501 中的列表以只包括具有匹配关键字的注释。

[0121] 单选按钮 512 允许用户指定是应当显示所有注释,还是应当只显示具有正面(或负面)评价的注释。单选按钮 512 有利地默认为选择所有注释,并且用户可以随意改变该设置。

[0122] 按钮组 514 允许用户指定应当如何安排部分 501 中的注释列表,例如按文件夹、按标题、按评价、按关键字或者按注释的最后更新或对已注释页面或站点的最后访问的时间戳。

[0123] 应当注意,可以一起使用部分 502 中的各种控件。例如,用户可以指定应当只显示包括特定关键字的特定文件夹中的注释,或者应当只显示特定文件夹(或文件夹组)中具有肯定评价的注释,等等。

[0124] 在其他实施例中,可以查看注释的其他安排。例如,某一数目(例如 5、10、20、50)的最常访问的已注释页面、最近访问的已注释页面、或最近注释的页面的列表可能是可查看的。在其他实施例中,用户也可以查看关于注释的上下文信息(例如,图 3 中所示的介绍字段 316 和 / 或老的介绍字段 318 的内容)。

[0125] 在各种实施例中,页面 500 可以支持用户与注释的其他交互。例如,利用文本框 530,用户可以提供要注释的新页面或站点的 URL,然后激活“转到”(Go) 按钮 532 来创建注释。在某些实施例中,如果在文本框 530 中输入的 URL 与现有注释的 URL 匹配,则向用户被显示以现有注释并被邀请对其进行更新。

[0126] “搜索我的 Web”(Search My Web) 按钮 540 发起利用用户的库(或其一部分)作为搜索语料库的搜索操作。该操作将在下面描述。

[0127] 应当意识到,这里描述的库界面页面是示例性的,也可以有变化和修改。可以使用允许用户查看、组织并搜索其注释的任何界面,并且该界面与所示的特定组合相比,可以具有更多的特征、更少的特征或不同的特征。

[0128] III. 注释数据的收集

[0129] 在优选实施例中,提供了各种界面,通过这些界面用户可以指示注释特定页面或站点的期望,从而发起注释的创建。图 6 是搜索结果页面 600(这里也称为搜索报告)的示例,其提供了用于注释命中页面或站点的界面。搜索结果页面 600 是响应于如文本框 602 中所示的用户查询而生成的;“提交”(submit) 按钮 604 被用于提交查询。每个结果 605 包括各种信息,如标题、自动生成的摘要、URL 等等。在每个结果旁边(或下面)是“保存”(Save This) 按钮 606,其启动了用户可经由其注释该特定命中的页面或对话框。合适页面的示例

将在下面描述。

[0130] 图 7 是用于注释用户恰好正在查看的任何页面的基于工具栏的界面的示例。Web 浏览器窗口 700 包括诸如查看区域 702 和文本框 704 之类的传统元素,查看区域 702 用于显示 Web 内容和提供导航按钮(后退、前进等)的默认工具栏,文本框 704 显示当前显示的页面的 URL,并且还允许用户输入要在查看区域 702 中显示的页面的 URL。浏览器窗口 700 还包括搜索工具栏 706,其可被提供作为传统浏览器程序的附加项或作为浏览器程序的标准特征。

[0131] 搜索工具栏 706 有利地包括用户可经由其向搜索服务器 160(图 2)提交查询的文本框 708 和“搜索”(Search)按钮 709,允许用户查看其保存的注释并导航到已注释页面的“已保存列表”(List Saved)按钮 710 以及打开允许用户注释当前显示的页面的页面或对话框的“保存”(SaveThis)按钮 712。在某些实施例中,搜索工具栏 706 还包括“显示我的意见”(Show My Comments)按钮 714,其出现在用户正在查看其先前已注释的页面的活动中;用户可以操作按钮 714 来查看其先前的注释。在注释包括评价的情况下,按钮 714 的外观可以部分依赖于评价(例如,评价可由包括在按钮中的图标反映)。搜索工具栏 706 还可以包括其他组件(除了所示的以外,或者替代所示的)。另外,可以换用任何其他持久界面(即,在用户正查看任何网页的同时可访问的界面);不需要搜索工具栏。

[0132] 在某些实施例中,搜索工具栏 706 可被配置使得其可在“通用”状态中由未登录到搜索服务器 160 的用户使用,并且可在“个人化”状态中由已登录的用户使用。在通用状态中,工具栏提供了对基本搜索服务的访问(例如,经由文本框 708 和“搜索”按钮 709),并且还提供了允许用户登录以访问个人化服务的按钮。在个人化状态中,通过工具栏可以支持个人化特征。例如,“保存”按钮 712 可以只在工具栏 706 的个人化状态中提供;或者,按钮 712 也可以在通用状态中提供,并且如果在工具栏处于通用状态的同时按钮 712 被激活则浏览器被重定向到登录页面。

[0133] 在某些实施例中,用户可被提供以用于发起注释创建多个选项,包括例如上述“我的库”页面 500(图 5)上的 URL 输入框 530 和/或图 6 和 7 中所示的各种“保存”按钮。无论如何发起注释的创建,用于创建注释的过程有利地都是相同的。

[0134] 图 8 是根据本发明实施例的过程 800 的流程图,用户通过过程 800 创建页面(或站点)的注释。在步骤 802,用户激活“保存”按钮,如搜索工具栏 706(图 7)上的按钮 710 或搜索结果页面 600(图 6)上的“保存”按钮 606 之一。在步骤 804,服务器识别要注释的页面(这里称为“主题页面”)的 URL。

[0135] 在步骤 806,显示主题页面的注释编辑界面。编辑界面有利地显示在当激活“保存”按钮时打开的单独窗口或对话框中。在某些实施例中,编辑界面以这样的方式安排在用户的显示器上,这种方式使得用户既可以看见主题页面也可以看见注释,或者可以容易地在两者间来回翻动。在步骤 808,用户将注释信息输入到界面中。

[0136] 可以使用各种编辑界面。图 9 是根据本发明实施例的编辑界面页面 900 的示例。页面 900 可以在过程 800 的步骤 802 中显示。页面 900 在区域 902 中显示了主题页面的 URL。唯一标识主题页面的 URL 有利地在页面 900 被创建时被预先填充并且不是用户可编辑的。页面标题被显示在文本框 904 中;框 904 可被默认填充以从主题页面的源代码提取的页面标题,并且用户可被允许通过修改框 904 的内容来编辑注释的标题。

[0137] 在区域 906 中,用户被邀请利用单选按钮 908、910 为注释分配范围(主机标志)。在一个实施例中,主机标志默认为“页面”,意味着注释只应用于显示在区域 902 中的 URL。用户可以将主机标志改为“站点”,并且也可以通过编辑显示在文本框 912 中的 URL 来调整站点的范围以反映站点的期望范围。例如,主机服务器(由在双斜线和下一斜线之间的 URL 部分标识)可由多个独立网站共享,这多个网站的内容可以完全无关。在范围被设为“站点”的情况下,用户可能能够例如通过修改文本框 912 来在该主机服务器上指定注释应当应用于的页面的子集,以在主机服务器上进一步指定包含期望的页面子集的目录。

[0138] 区域 914 接受用户输入。例如,用户可以在文本框 916 中输入一个或多个关键字。关键字例如可被输入为以逗点分隔的列表。在某些实施例中,关键字可以限于最大数目(例如 5、10、20),或限于每关键字最大数目的字符(例如 40、80),或者限于字符的最大总数。在另一个实施例中,用户可能能够从系统提供的列表中选择要应用于页面或站点的一个或多个标签;可以使用用于从列表中选择标签的传统界面元素。文本框 918 可用于自由形式的文本输入,并且有利地接受相当大量的字符(例如,最高约 5000 个)。用户还可以利用评价单选按钮 920 评价页面或站点。在一个实施例中,默认是没有评价,并且用户可以通过选择按钮 920 中的不同的一个来改变默认值。在优选实施例中,用户不需要完成部分 914 中的所有字段。

[0139] 再次参考图 8,在步骤 810,用户指示新的注释完成。例如,在图 9 的界面 900 中,用户通过激活“保存”(Save)按钮 922 或“取消”(Cancel)按钮 924 来指示完成。“保存”按钮 922 指示完成的注释应当被保存,“取消”按钮 924 指示注释应当被丢弃。

[0140] 在步骤 812,确定注释是否应当被保存,例如用户激活的是“保存”按钮 922 还是“取消”按钮 924。如果不保存注释,则过程 800 退出(步骤 814)。如果要保存注释,则在步骤 816,自动生成注释的摘要。在一个实施例中,摘要是通过利用用户提供的关键字和/或描述来识别页面内容的最相关部分而生成的。该算法可以类似于用于基于查询项生成搜索命中的摘要的传统算法。在步骤 818,新的注释被添加到个人化数据库 166。步骤 818 可包括创建数据库记录;利用经由编辑页面 900 提供的信息填充记录;压缩注释数据以高效地存储、搜索和检索等等。

[0141] 在步骤 820,当用户激活“保存”按钮时用户正在查看的页面被重新显示。在优选实施例中,重新显示的页面(或搜索工具栏)被更新以反映新注释的存在。例如,在用户激活搜索结果页面 600(图 6)上的“保存”按钮 606 的情况下,搜索结果页面 600 被基于注释更新,如下所述。在用户激活搜索工具栏 706(图 7)的“保存”按钮 712 的情况下,“显示我的意见”按钮 714 可以变为活动以指示新的注释存在。

[0142] 应当意识到,这里描述的用于创建注释的过程和编辑界面是示例性的,也可以有变化和修改。可以并行执行顺序描述的过程步骤,可以变化步骤的顺序,并且可以修改或组合步骤。在某些实施例中,用户可能能够在用户的身份对搜索服务器 160 已知之前激活“保存”按钮或类似的按钮。在该情况下,搜索服务器 160 可以在显示编辑界面之前提示用户登录。

[0143] 在一个实施例中,客户端应用 125(图 2)包括显示页面 900 或用于提供注释数据的其他用户可编辑形式的注释创建模块,并且搜索服务器 160 的用户注释模块 164 不参与过程 800,直到用户选择保存注释为止。此时,如果用户还未登录到搜索服务器 160,则其可

被提示登录从而使注释可以与其用户 ID 相关联。

[0144] 在给定实施例中,注释编辑界面有利地遵从注释数据结构的特定内容。例如,图 10 是允许用户输入搜索命中的评价的内嵌式注释编辑界面 1000 的示例。界面 1000 以内嵌方式集成到搜索结果页面 1002 中。用户被邀请利用按钮 1004、1006、1008 之一来评价每个命中。(如果用户不选择评价,则页面不被注释。)类似地,图 11 是具有基于工具栏的注释编辑界面 1020 的浏览器窗口 1100 的示例,界面 1020 允许用户输入被查看的页面的评价。界面 1020 是搜索工具栏 1022 的一部分,搜索工具栏 1022 通常可以类似于上述图 7 的搜索工具栏 706。界面 1020 出现在用户正在查看未注释页面时,并且包括用于输入正面、中立或负面评价的按钮 1024、1026、1028。应当理解,可以换用适合于不同评价方案的不同界面。也可以组合不同界面;例如,用户可以输入评价,然后被邀请提供附加的元数据。

[0145] IV. 利用注释数据进行搜索

[0146] 一旦被收集,注释数据就可以以各种方式用于增强用户的搜索和 Web 浏览的体验。现在将描述这种增强的示例;应当理解,注释的使用并不限于任何特定的特征或特征集。

[0147] A. 增强的 Web 搜索

[0148] 在一个实施例中,搜索服务器 160 访问用户的注释库以在响应于来自该用户的查询而生成的搜索结果页面中提供附加信息。例如,已注释命中(即,对应于用户的库中的已注释页面的命中)的单独列表可被包括在搜索结果中,或者已注释命中只要出现在结果列表中就可被突出显示。在注释包括评价的情况下,可以提供正面评价的命中的单独列表,并且可以以反映了查询用户的评价的方式来突出显示已被评价的命中,或者评价数据可用作对命中分级时的一个因素。

[0149] 图 12 是根据本发明实施例的利用注释信息增强的搜索结果页面 1200 的示例。结果页面 1200 可由查询响应模块 162 响应于用户查询而生成。在该实施例中,结果页面 1200 包括横幅部分 1202。除了页面标识信息外,横幅部分 1202 还包括以可编辑形式示出了当前查询(例如“chinesefood sunnyvale”)的搜索框 1204 以及使得用户能够改变查询并执行新的搜索的搜索按钮 1206。这些特征可以采用传统的设计。

[0150] 部分 1208 是个人化结果区域(“我的结果”(My Result)),其中显示了用户先前已注释的任何命中。在某些实施例中,部分 1208 可以只显示用户的注释包括正面评价的结果;在其他实施例中,在部分 1208 中可以列出所有已注释的页面。每个页面有利地伴随有“显示我的意见”按钮 1210,用户可以激活该按钮以查看其先前的注释。在某些实施例中,命中可以基于用户已分配给各种站点的评价(如果有的话)被突出显示。

[0151] 图 13 是当用户激活按钮 1210 时可以生成的注释查看页面 1300 的示例。注释查看页面 1300 有利地是显示在页面 1200 上的覆盖图(例如,弹出式窗口),从而页面 1200 至少部分是可见的。在某些实施例中,页面 1300 被显示为其已注释页面本身上的覆盖图,而不是页面 1200 上的覆盖图。在其他实施例中,注释可以内嵌方式显示在页面 1200 中,或者与已注释页面一同内嵌。在注释包括连接到页面内容的特定部分的元数据(例如描述)的情况下,这种元数据可以覆盖显示在其所连接到的页面部分上,或者显示在该部分附近(例如,紧挨着其前面或后面)。

[0152] 页面 1300 在部分 1302 中包括从存储在个人化数据库 166 中的注释检索的标题、

描述、关键字和评价。“编辑”按钮 1304 允许用户编辑注释；在一个实施例中，激活按钮 1304 打开了编辑界面 900（图 9），并且可以利用注释的当前内容预先填充编辑界面 900 的各种文本块和其他元素。“关闭”（Close）按钮 1306 关闭页面 1300。

[0153] 再次参考图 12，结果部分 1216 显示了命中的某些或全部，其等级由查询响应模块 162 确定。传统的分级算法可用于生成该等级。部分 1216 中的每个条目 1218 对应于一个命中，并且包括该页面（或站点）的标题以及来自该页面内容的简要摘录（或摘要）。摘录或摘要可以利用传统技术生成。还显示了站点的 URL（统一资源定位符）。对于用户还未注释的命中，可以显示“保存”按钮 1218，并且在查看页面 1200 时，用户可以通过激活按钮 1218 来选择对未注释命中进行注释。“保存”按钮 1218 有利地与上述图 6 中的按钮 606 相同。

[0154] 部分 1216 中用户已注释的任何命中都可以被可视地突出显示以指示注释的存在。可以使用用于突出显示的各种设计，例如包括边界、阴影、特殊字体、颜色等等。在注释包括评价的某些实施例中，突出显示的类型依赖于评价，并且评价可被显示在页面 1200 上。例如，命中 1220 具有正面评价，而命中 1222 具有负面评价。

[0155] 在一个实施例中，由用户在查看结果页面的同时输入的注释被实时用于更新当前查询的显示结果（除了将信息存储到个人化数据库 166 中以供将来使用外）。从而，如果查看页面 1200 的用户激活“保存”按钮 1218，从而发起上述的过程 800（图 8），则重新显示的页面 1200 可被修改以反映新的注释。例如，新注释的命中可被突出显示，或者可以利用新给予命中的评价对结果重新分级。

[0156] 图 14 是可以实现在查询处理模块 162（图 2）中的过程 1400 的流程图，该过程用于将用户先前已存储的注释结合到对来自该用户的当前查询的响应中。在步骤 1402，接收查询。在步骤 1404，例如从页面索引 170（图 2）获得对应于查询的命中列表。在步骤 1406，查询处理模块 162 例如利用传统算法对命中分级。

[0157] 在步骤 1408，查询处理模块 162 确定用户是否已登录。如果没有，则查询处理模块 162 可以在步骤 1410 将结果页面发送到用户而不进行个人化，从而使用户能够执行搜索并获得结果而无需登录到（或者甚至无需注册到）搜索服务器 160。如果用户已登录，则基于个人化数据库 166 中的信息针对该用户定制结果页面。

[0158] 更具体而言，在步骤 1412，查询处理模块 162 向个人化数据库 166 提供用户的 ID，并且检索由该用户创建的注释。在步骤 1414，将检索注释的 URL 与命中的 URL 相比较以检测与用户先前已为其创建注释的 URL 相匹配的任何命中。对于主机标志被设为“站点”的注释，如果命中 URL 的开始部分与存储在注释中（例如在图 3 的 URL 字段 308 中）的 URL（或部分 URL）相匹配，则检测到匹配（这里也称为“部分匹配”）。如果主机标志被设为“页面”，则要求注释的 URL 和命中 URL 之间有“精确”匹配。对于每个部分或精确匹配，在步骤 1416 例如通过参考注释 300 的“评价”（rating）字段 328 来确定注释是否包括正面评价。如果是，则命中被添加到正面结果（“我的结果”）列表。在其他实施例中，所有已注释命中（无论其评价如何）都可被添加到“我的结果”列表。

[0159] 在步骤 1420，在完成了命中和注释之间的比较后，可选地利用在注释中包含的评价对结果列表重新分级。例如，利用传统分级算法可以为每个命中（无论其是否有注释）生成基本分数。对于具有带有正面或负面评价的注释的命中，可以从评价中确定“奖

励”(bonus)。奖励有利地被定义为使得正面评价的站点趋向于在等级中向上移动,而负面评价的站点趋向于向下移动。例如,如果低分数对应于高等级,则正面评价的奖励可被定义为负数,而负面评价的奖励可被定义为正数。在某些实施例,部分 URL 匹配可能被给予比精确 URL 匹配少的奖励。未被评价的(或中立评价的)命中不接收奖励。该奖励可被添加到(以代数方式)基本分数以确定每个命中的最终分数,并且重新分级可以是基于最终分数的。

[0160] 在某些实施例,在步骤 1420 重新分级还可以包括从要显示的命中列表中丢弃与负面评价的页面或站点相匹配的任何命中。在这些实施例,传递给用户的搜索结果页面可包括由于负面评价而丢弃的命中数目的指示和/或“显示所有命中”(Show all hits)按钮(或其他控件),其允许用户查看包括负面评价的命中的被显示的搜索结果。在另一种变化中,用户可以点击链接以仅查看负面评价的命中。

[0161] 在步骤 1422,“我的结果”列表被分级并被添加到搜索结果页面。在某些实施例,该分级可以是基于上述基本分数或最终分数的。在其他实施例,“我的结果”列表中的命中按用户评价分类;具有相同评价的命中可进一步根据上述基本分数分类。在步骤 1424,搜索结果页面被修改以指示任何注释的存在;例如,突出显示和/或“显示我的意见”按钮可被添加到已注释命中。修改后的搜索结果页面(在这种情况下包括个人化的“我的结果”部分)在步骤 1410 被发送给用户。

[0162] 应当意识到,这里描述的过程是示例性的,也可以有变化和修改。可以并行执行顺序描述的步骤,可以变化步骤的顺序,并且可以修改或组合步骤。在某些实施例,在查看动作之前,注释的某些或全部内容可以以内嵌方式显示在搜索结果页面中。例如,可以显示指示正面或负面评价的可视突出显示元素,或者用户的关键字或描述可以出现在自动生成的摘要下方,等等。在注释内容以内嵌方式显示的情况下,可以忽略“显示我的意见”按钮(或允许用户查看注释的其他控制元素)。

[0163] 在其他实施例,用户的注释可用于在搜索操作期间识别命中。例如,除了搜索页面索引 170 外,查询响应模块 162 还可以利用与用来搜索页面索引 170 相同的搜索项中的某些或全部来搜索用户的注释的选定字段。在一个这样的实施例,搜索注释的关键字和/或描述字段,并且如果搜索项出现在这些字段之一中,则将已注释页面识别为命中,无论在页面索引 170 的搜索中是否已将该已注释页面识别为命中。

[0164] 在其他实施例,可以对不同注释(或不同已注释页面)对用户的相对重要性进行推断,并且可使用该推断来组织搜索命中的呈现方式。例如,可以推断用户更经常访问的已注释页面对于用户来说比不常访问的页面更有价值。从而,当用户最常访问的页面之一出现在搜索结果中时,该页面可被更明显地显示(例如,在等级中靠上),标记有与平常的已注释页面指示符不同的特殊指示,或者被包括在特殊的“收藏”(favorities)区域中。

[0165] 在某些实施例,用于对查询作出响应的注释可被限制为用户注释的子集。例如,只有活动标志被设为“真”(true)文件夹中的注释可用于识别命中和/或突出显示结果。

[0166] B. 用户库中的搜索

[0167] 在其他实施例,用户可以搜索其个人的已注释内容库,而不是搜索整个 Web。例如,图 5 的“我的库”页面 500 包括可用于发起用户的库搜索的“搜索我的库”(Search My Library)按钮 540。尽管未明示,但是在图 7 的工具栏 706 上也可以包括类似的按钮。激活

“搜索我的库”按钮有利地启动了库搜索界面页面,经由该页面查询可被引导到用户的库。

[0168] 图 15 是根据本发明实施例的库搜索界面页面 1500 的示例。页面 1500 提供了用于在用户的已注释页面或站点的库的全部内容或在用户的库的任何子集内进行字段特定的搜索的用户界面。范围部分 1502 支持要搜索的页面的选择。在这种情况下,选择“所有页面”(All Pages)选项 1504(其可以是默认的),并且将会搜索用户已注释的所有页面。如果选择“选定文件夹”(Selected Folders)选项 1506,则区域 1508 将会提供每个文件夹带有复选框或其他选择/去选择转换的文件夹列表。类似地,如果选择“选定页面”(Selected Pages)选项 1510,则区域 1508 将会提供每个页面带有复选框或其他独立的选择/去选择转换的已注释页面的列表(其最初可被显示为可扩展的文件夹列表)。页面可由来自用户注释的标题、URL 或任何其他(一个或多个)字段标识。在另一个实施例中,用户可以选择只搜索库中“活动”(active)标志被设为真的那些部分(文件夹和/或个体注释),或者选择不考虑活动标志而搜索所有的库内容。

[0169] 查询部分 1512 提供了各种文本框,用户可以在这些文本框中输入搜索项以搜索页面内容和/或搜索注释中的特定字段。在该示例中,用户可以单独指定用于页面内容(框 1514)、注释标题(框 1516)、关键字字段(框 1518)、描述(框 1520)和/或介绍(框 1521)的搜索项。单选按钮 1522 可用于约束命中的评价。在默认时,选择“任何评价”(Anyrating),从而评价(或没有评价)不限制搜索;用户可以选择将搜索限于例如具有正面评价的命中或具有负面评价的命中。“搜索”按钮 1526 提交查询以供处理,“重置”(Reset)按钮 1528 清除查询部分 1512 中的所有字段。

[0170] 搜索的处理依赖于查询部分 1512 中的哪些框提供搜索项。在要搜索页面内容的情况下,数据可以从页面索引 170 获得,或者从个人化数据库 166 中的注释获得(如果页面内容的表示存储在其中的话)。利用个人化数据库 166 中的用户注释搜索其他字段。应当理解,用户可以留下部分 1512 中的某些或全部框;在框为空的情况下,相应字段不用于约束搜索。例如,用户可以通过在框 1514 中输入搜索项并使其他框为空来搜索其注释的页面的页面内容;实际的搜索可以利用页面索引 170 执行,其中在将结果发送给用户之前不对应于已注释页面或站点的任何命中被丢弃。搜索结果有利地利用类似于上述页面 1200(图 12)的搜索结果页面传递,只不过在限于用户的库的搜索中,每一页面都有注释。

[0171] 查询界面可以变化。例如,在另一种界面中,提供了单个文本框,并且用户被提示选择文本框中的搜索项是否应当在页面内容中搜索和/或应当在注释记录的各种字段(例如,标题、关键字、描述和/或其他字段)中搜索。在另一个实施例中,默认提供了具有单个文本框的“基本”(basic)搜索界面,并且对页面内容和注释字段执行搜索。用户可以接受该基本搜索配置,或者选择查看查询部分 1512(或另一个查询界面)以执行更高级的搜索。其他查询界面和界面的组合也是可能的。

[0172] 在用户可以按关键字搜索其注释的实施例中,用户可以有利地开发出一种个人化的关键字方案,以用于对其在 Web 上发现的内容进行索引。利用上述搜索和/或基于工具栏的界面,用户可以通过创建注释来给其感兴趣的任何页面分配关键字并将页面添加到其个人库中。随后,用户可以利用关键字字段进行搜索以在其库中找到与特定主题有关的页面。从而,可搜索的注释为个体用户提供了一种有力的工具,用户可以利用该工具来以对其有用的方式组织 Web 内容并对 Web 内容分类。

[0173] 应当意识到,搜索页面 1500 是示例性的,也可以有变化和修改。在某些实施例中,搜索页面 1500 也可以经由工具栏上的按钮或持久用户界面的其他合适元素访问,某种从搜索提供者的主页面访问。如果未登录到搜索服务器 160 的用户试图访问页面 1500,则在页面 1500 被显示之前用户可被提示登录。

[0174] C. 搜索有关或类似页面

[0175] 在某些实施例中,用户还可以搜索与其库中的页面或站点类似或有关的其他文档(例如页面或站点)。“类似”文档是包含相对于已注释页面满足某种类似性标准的内容的文档。类似性标准的示例包括:具有某数目的公共的词、短语、或其他多词单元;具有词、短语、或其他多词单元的类似的出现模式;在系统定义的分类法中属于同一类别或紧密相关的类别;等等。用于确定两个页面之间的类似性的算法是本领域中已知的,并且可用于本发明。“有关”文档与被评价的页面共享 URL 的一部分(例如,至少域名);同样,可以使用用于确定有关度的已知算法。

[0176] 在一个实施例中,从搜索页面 1500(图 15)中,用户可以利用上述的范围部分 1502 选择其已注释页面的全部或任何子集,然后激活按钮 1536 以搜索类似文档或激活按钮 1538 以搜索有关文档。搜索类似或有关文档有利地并不限于已注释页面或站点,并且可以由搜索服务器 160 利用页面索引 170 执行。在某些实施例中,搜索有关或类似页面还可以包括搜索项限定词(qualifier),并且搜索项可以利用上述的查询部分 1512 或其他界面来指定。在某些实施例中,用户访问选定子集内的不同已注释页面的相对频率可用于给活动子集中的不同已注释页面分配不同权重,以确定哪些其他文档与选定文档最类似或最紧密有关;更经常被访问的页面相比于不常被访问的页面有利地被分配更高的权重。

[0177] D. 利用注释进行迭代搜索

[0178] 上述技术可被综合考虑以支持迭代搜索 workflow,其中用户可以过滤已注释页面或站点以创建子集并对该子集执行进一步的过滤。例如,用户可用的文件夹可以包括被自动创建并用来支持迭代搜索的“工作”(Working)文件夹。用户可以搜索在介绍框 1521 中有期望查询项的所有已注释页面(例如,利用页面 1500 的搜索界面 1512)。用户随后可以检查结果,并将任何感兴趣的结果保存到“工作”文件夹,并且填充或不填充各种注释字段。在一个实施例中,搜索结果页面(见图 12)还可以包括“添加到工作文件夹”(Add to Working Folder)按钮。

[0179] 在另一个实施例中,“工作”文件夹也可以在用户利用浏览器导航 Web 时填充。例如,“添加到工作文件夹”按钮可经由工具栏(例如,图 7 的工具栏 706)或其他持久界面提供。用户还可能能够经由合适的界面将注释记录从其他文件夹移动或拷贝到工作文件夹,并直接编辑工作文件夹的内容(例如,经由图 5 的页面 500)。在另一个实施例中,用户最初利用某种查询搜索整个 Web,然后例如通过将某些或全部命中添加到文件夹来从搜索结果页面填充“工作”文件夹。用于清空“工作”文件夹的界面元素可被包括在页面 500 中,或者按需要包括在工具栏或其他持久界面中。

[0180] 在填充了“工作”文件夹之后,用户可以执行进一步的查询,例如经由图 15 的页面 1500,从而只选择“工作”文件夹。在某些实施例中,用户可以通过激活适当的界面按钮将“工作”文件夹中的全部搜索结果保存到某个其他文件夹(其可以是例如由用户新创建的文件夹)中。这些结果可用于浏览或进一步的搜索。在某些实施例中,用户可以将对“工作”

文件夹进行的搜索保存为可以应用于其他页面或搜索结果的过滤器。

[0181] E. 工具栏增强

[0182] 在某些实施例中,浏览器工具栏(如上述的工具栏 700(图 7))可以基于用户注释加以增强。例如,如上所述,工具栏 700 有利地提供了“显示我的意见”按钮 714,其指示用户是否已经注释了当前页面并且允许用户查看其注释,例如在弹出式窗口或覆盖图(如上所述并如图 13 所示)中。注释也可以内嵌方式显示在已注释页面中。

[0183] 在其他实施例中,提供了进一步的工具栏增强。例如,在某些实施例中,注释集成有“书签”特征。如本领域中所知的,由某些因特网入口提供者提供的浏览器工具栏附加项包括“书签”特征,其允许入口服务的注册用户在其浏览时将书签保存到页面。每个书签通常包括 URL 并且在某些情况下还包括页面标题。用户可以利用安装有工具栏附加项的任何 Web 浏览器客户端访问其书签(假定用户首先登录到搜索服务器 160)。

[0184] 在某些方面中,注释可用作增强的书签。与仅仅给页面加上书签不同的是,用户可以保存关于已加上书签的页面的附加信息(元数据),如关键字、评价或其他描述性和/或评估性信息。该信息可例如提醒用户其认为正被讨论的页面或站点中哪些是有趣的或有价值的。另外,与试图记住数十或数百已加上书签的页面中的哪些包括特定信息项不同的是,用户可以搜索其注释库以找到期望页面。

[0185] 在某些实施例中,用户的传统书签列表是从其已注释页面或站点库(或者仅从那些具有正面评价的已注释页面或站点)自动生成的。用户可以选择在创建新的注释时不输入任何内容;这种情况下,注释实质上充当传统的书签。

[0186] F. 用户偏好

[0187] 在某些实施例中,注册用户可能能够控制其注释数据的使用。例如,“我的库”页面 500(图 5)可以包括“选项”(Options)按钮 550。按钮 550 有利地链接到偏好页面(没有明示),其使得用户能够指定其对于各种与注释有关的特征的偏好。示例包括在生成搜索结果时是否应当考虑用户的评价,在生成搜索结果时是否应当考虑“全局评价”(将在下面描述),是否应当公布任何的用户注释以及如何公布等等。

[0188] 用户对这些和任何其他选项的偏好有利地存储在个人化数据库 166 中,并且在用户登录时被应用。在某些实施例中,可从搜索结果页面(例如如上所述)控制的按钮和选项可以临时推翻经由用户偏好页面指定的用户偏好;对于下一查询或用户下次登录时,可以恢复原始偏好。

[0189] V. 注释数据的聚合

[0190] 在某些实施例中,搜索服务器 160 使用整个注册用户团体的聚合注释数据来增强所有用户的搜索经历。下面将描述示例。

[0191] A. 全局评价

[0192] 例如,由不同用户分配给页面或站点的评价可被取平均或者以其他方式聚合以生成该页面或站点的“全局”评价。在一个这样的实施例中,搜索服务器 160 或者与搜索服务器 160 在共同控制下操作的另一服务器周期性地(例如每天一次)执行一种算法,该算法搜索通过 URL 搜索个人化数据库 166 中的注释以收集任何用户可能已分配给具有该 URL 的页面或站点的所有评价。随后这些评价被取平均或以某种其他方式组合以确定页面或站点的全局评价。全局评价可例如缓存在页面索引 170 中。

[0193] 全局评价可以按各种方式来使用。在一个实施例中,在页面或站点的 URL 与给定搜索的命中相匹配时,页面或站点的全局评价可用作搜索结果分级算法中的一个因素。在另一个实施例中,全局评价与用户自己的评价(如果用户已经给予了一个评价)一起可以显示在搜索结果页面上每个命中的旁边,而无论在页面分级中是否使用了全局评价。

[0194] 在某些实施例中,用户可以具有显示或隐藏全局评价和/或在对搜索命中分级时使用或忽略全局评价的选项。用于这些选项的控件可以包括在搜索结果页面或用户偏好页面上,如上所述。

[0195] B. 群体分类 (Folksonomy)

[0196] 作为另一个示例,聚合注释数据可用于对 Web 内容分类或归类以创建“群体分类”。这里所用的“群体分类”指一种内容的分类系统,其基于并反映了用户团体所选择的描述内容并对内容分类的方式。在注释包括关键字(或标签)的实施例中,分配给特定页面的关键字模式往往将指示其内容。例如,如果注释了特定页面的用户中的 50% 使用了关键字“cycling”(自行车),则可以推断页面与自行车有关。

[0197] 在一个这样的实施例中,搜索服务器 160 或者与搜索服务器 160 在共同控制下操作的另一服务器周期性地(例如每天一次)执行一种算法,该算法搜索通过 URL 搜索个性化数据库 166 中的注释以找到已经由注册用户用来注释该 URL 的所有关键字。搜索服务器 160 分析关键字模式以确定页面的“群体分类关键字”集。该集合可包括例如最常使用的关键字(例如,达到 5 或 10 或某个其他最大数)、最近使用的关键字等等。在某些实施例中,使用频率和新近度的组合来选择群体分类关键字。特定页面的群体分类关键字可以与 URL 相关联地被添加到页面索引 170,并且用于对后续搜索作出响应。例如,如果 URL 的群体分类关键字之一与查询的搜索项相匹配,则无论页面内容是否实际包括搜索项中的任何一个,该 URL 都可被返回作为任何查询(来自任何用户)的命中。对命中分级可以部分依赖于搜索项是与页面内容匹配,与群体分类关键字匹配,还是与这两者匹配。

[0198] 在某些实施例中,用户(注册用户或所有用户)可以选择是否搜索群体分类关键字(除了搜索页面内容外,或替代搜索页面内容)。

[0199] C. 识别用户兴趣

[0200] 用户注释也可由服务提供者(例如,搜索服务器 160 的所有者)用来开发用户简档(user profile)。例如,已注释页面或站点可利用各种专有的或非专有的分类方案(包括上述的群体分类关键字)被分类到不同类别中,或者其可基于特定用户已分配的关键字被分类。通过分析用户的各种已评价的页面或站点所属的类别,搜索提供者可以识别用户尤其感兴趣的主体。例如,如果用户已注释了属于单个类别的许多页面,则可以推断出用户大体上对该类别中的内容感兴趣。搜索提供者还可以使用关于用户最经常访问的已注释页面的统计信息来进一步帮助识别用户的兴趣。

[0201] 搜索提供者可以使用以各种方式从注释收集的关于用户兴趣的信息来增强用户的体验。在一个实施例中,关于特定用户的兴趣的知识可用于解决从该用户接收的搜索查询中的不确定的地方,这种查询包括对整个 Web 进行搜索的查询和/或搜索(或搜索命中)限于用户的库的查询。例如,搜索查询中的项“美洲虎”(jaguar)可能指汽车或动物。如果用户注释了相对较多的与汽车有关的站点,而注释了相对较少的与动物有关的站点,则可以推断出该用户对汽车比对动物更感兴趣。该信息可用于增强与用户的实际兴趣相关的信

息将放置在搜索结果中的显著位置的似然性,其方式例如是通过基于页面是否与所识别的用户兴趣有关而调整页面分级。

[0202] 作为另一个示例,关于特定用户的兴趣的知识可用于定制 Web 内容以传递给该用户。在一个实施例中,用户的兴趣可用于确定如何在网页上安排内容;例如,由用户请求的新页面可被安排成与该用户的已知兴趣有关的内容首先出现。在另一个实施例中,在选择受赞助内容(例如广告)以显示在网页上时考虑用户的兴趣。也可以执行基于特定用户的已知兴趣的其他类型的定制。

[0203] 也可以对用户的子组(例如,所有用户、或者适合某种地理或人口简档的用户)聚合从注释获得的关于用户兴趣的信息。在用户的身份或特定兴趣还未知的情形中,该聚合信息可用于诸如以下目的:解决搜索查询中的不确定性,选择并安排网页的内容等等。在某些实施例中,对用户的不同子组进行的聚合是并行执行的,并且被用于在只有关于用户的部分信息(例如,用户的地理位置或人口特性)是已知的情形中。

[0204] 在其他实施例中,搜索提供者还可以使用关于用户兴趣的聚合信息来进行其他的策略性判决,如设置广告率,确定入口页面的内容或布局等。

[0205] D. 推断意见

[0206] 在对一个页面或站点的用户注释包括评价的某些实施例中,搜索服务器 160 可以从该评价推断出用户对其他(未注释)页面或站点的可能意见。这种推断可以基于 URL 的有关度和/或已注释页面和未注释页面之间的内容类似性。内容类似性可例如基于标题、摘要、内容中使用的词的模式、类别(基于群体分类关键字或其他分类方案)等来确定。未注释页面可例如基于被确定为类似的已注释页面或站点的用户平均评价而被分配以推断出的评价。

[0207] 在推断评价被用于对搜索命中分级的情况下,推断评价应当被给予比直接评价少的权重,从而反映了推断的不确定性。例如,如果用户给予页面或站点 X 否定评价,则可对与页面或站点 X 类似或有关的一页面或站点 Y 推断出相同的评价。直接否定评价可能导致页面或站点 X 被从搜索结果列表中移去,而页面或站点 Y 的推断出的否定评价可能导致页面或站点 Y 被包括在结果列表中,但是具有比其可能有的等级更低的等级。

[0208] 在某些实施例中,权重部分基于用户已注释了具有该属性的多少页面(或站点)以及对这些页面给予的评价的一致性来确定。例如,如果用户一直给予属于特定类别的大量页面以否定评价,则属于该类别的未注释页面的推断出的否定评价相比于用户只给予该类别中的一个或两个页面评价的情况可被给予更强的权重。

[0209] 在另一个实施例中,权重还部分基于已注释页面和未注释页面之间的类似度。例如,如果页面在内容上几乎是相同的,则类似评价的推断相比于如果类似性较弱的情况可被给予更高的权重。

[0210] VI. 另外的实施例

[0211] 尽管已经结合特定实施例描述了本发明,但是本领域技术人员将认识到可以有大量的修改。例如,各种搜索结果和用户界面页面的外观可以不同于这里所示的示例。界面元素并不限于按钮、页面的可点击区域、文本框或这里描述的其他的特定元素;可以使用任何界面实现方式。

[0212] 应当理解,在与评价相关时,本发明也不限于任何具体的评价方案,并且某些实施

例可能给用户选择替换评价方案的选项（例如，拇指向上 / 向下或者按某种刻度评价）。在某些实施例中，可能只支持正面或中立的评价。在其他实施例中，可能根本不收集评价。

[0213] 注释可以包括任何组合的任何数目的字段，并且可以包括与上述相比更多的字段、更少的字段或不同的字段。例如，在某些实施例中，用户可被邀请创建摘要或编辑自动生成的摘要。在另一个实施例中，用户可被提示指示被注释的页面或站点是否属于某种一般内容类别，例如“adult”（成人）或“foreign”（外国）或“spam”（垃圾广告）。用户随后可以选择在搜索期间包括或排除该类别中的内容。另外，关于在这些类别的一个或另一个中不同用户已经归类了哪些页面或站点的信息可被用于推断出正被讨论的页面或站点应当被这样对待。从而，例如，如果大量用户将特定页面识别为垃圾广告，则在未来的搜索结果中可以排除该页面。

[0214] 也可以提供用于查看注释和与注释交互的其他界面。例如，在某些实施例中，在每次已注释页面被显示在用户的浏览器中时，浏览用户的注释数据和 / 或聚合注释数据可被自动显示（例如，内嵌在页面内容中或者在覆盖图中）。在一个这样的实施例中，每个用户可能能够指示其对于应当自动显示其自己的注释、聚合元数据还是这两者的偏好。

[0215] 如上所述，某些实施例允许用户控制注释是应当应用于单个页面还是应当应用于页面组（或站点）。另外，在某些实施例中，用户还能够将注释应用于注册到与已评价的页面相同的域名注册者的所有页面。公共域名注册者的存在情况可以利用 WHOIS 或另一种类似服务确定。

[0216] 在其他实施例中，搜索服务器 160 的提供者还可以提供受赞助链接，其中内容提供者付费以使得其站点的链接在搜索结果中提供。受赞助链接通常显示在结果页面的指定部分中，与常规搜索结果相隔离。在本发明的一个实施例中，用户已注释的任何受赞助链接也可以被标记。例如，受赞助链接可能具有突出显示以指示用户具有该页面的注释，并且受赞助链接的用户评价（如果有的话）可用于确定突出显示，就好像对上述图 12 中所示的常规搜索结果一样。受赞助链接也可以伴随有“保存”按钮、“显示我的意见”按钮或类似的按钮或界面控件。

[0217] 上述实施例可能涉及网站、URL、链接和对于万维网（或其子集）用作搜索语料库的情形来说特定的其他术语。然而，应当理解，上述系统和方法也可以适用于不同的搜索语料库（如电子数据库或文档仓库），并且结果或注释可以包括内容以及链接或者对可以找到内容的位置的引用。

[0218] 结合了本发明各种特征的计算机程序可被编码在各种计算机可读介质上以供存储和 / 或发送；合适的介质包括磁盘或磁带、光存储介质（如 CD 或 DVD）、闪存和适于经由遵从多种协议的有线、光学和 / 或无线网络（包括因特网）发送的载波信号。编码有程序代码的计算机可读介质可以封装在兼容设备内，或者与其他设备分离提供（例如经由因特网下载）。

[0219] 尽管已经结合特定的硬件和软件组件描述了本发明，但是本领域技术人员将会意识到，也可以使用硬件和 / 或软件组件的不同组合，并且被描述为实现在硬件中的特定操作也可以实现在软件中，反之亦然。

[0220] 从而，尽管已经结合特定实施例描述了本发明，但是应当意识到，本发明试图覆盖在权利要求范围内的所有修改和等同物。

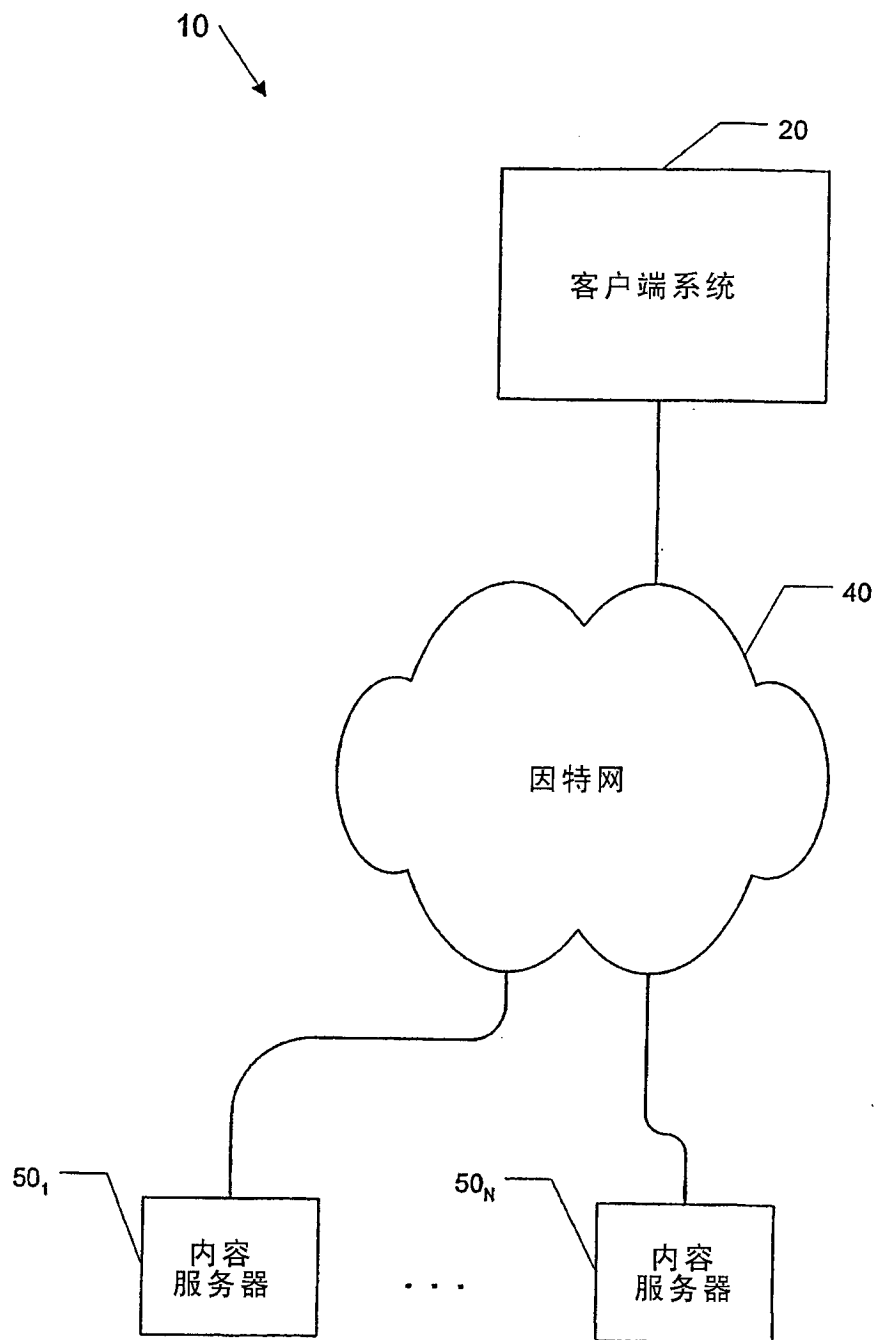


图 1

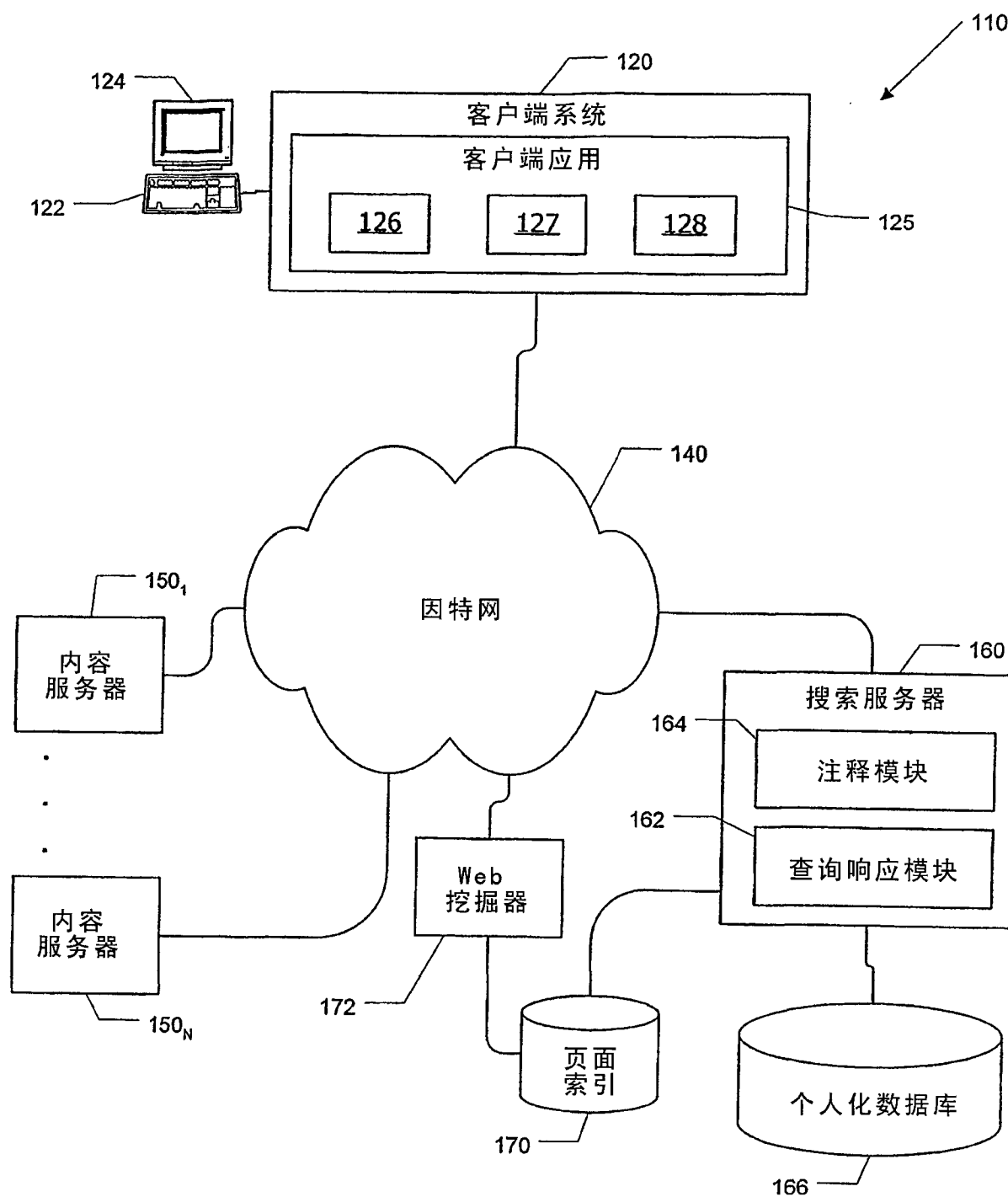


图 2

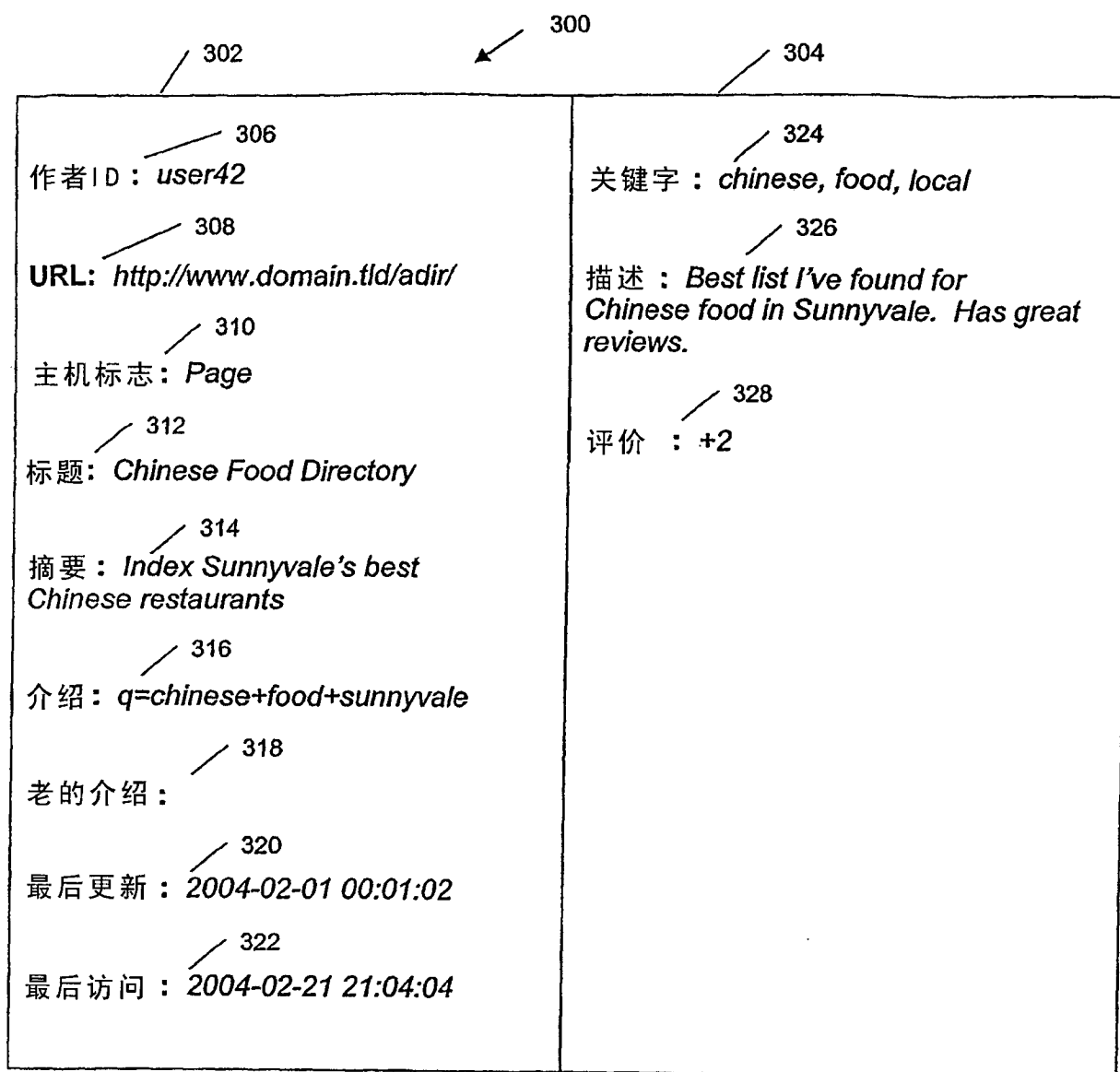


图 3

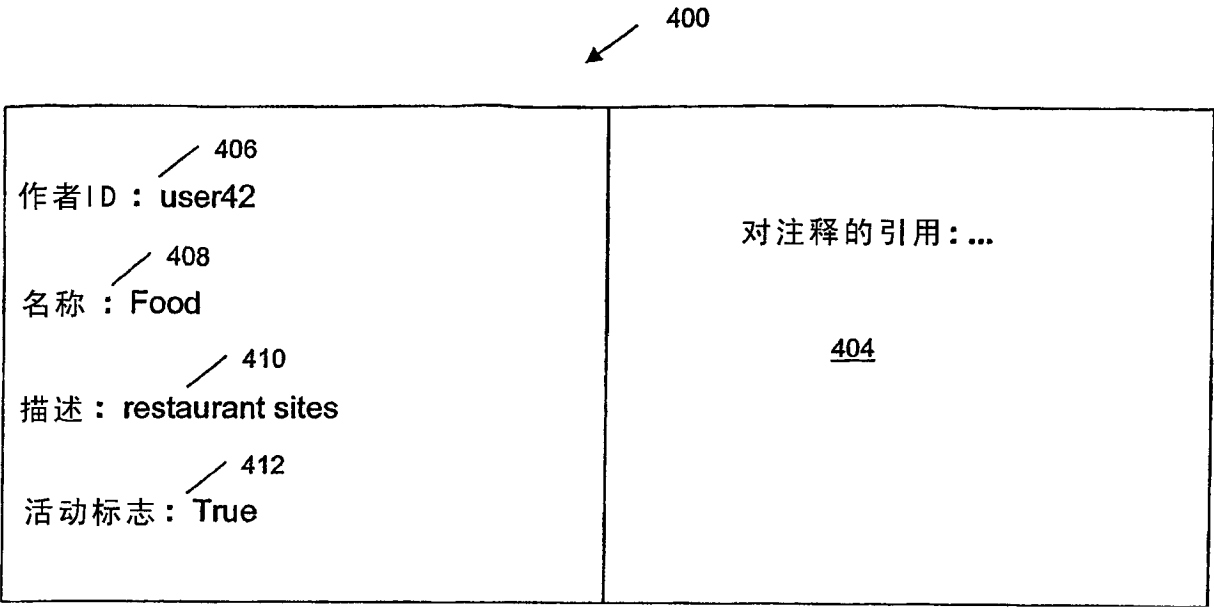


图 4

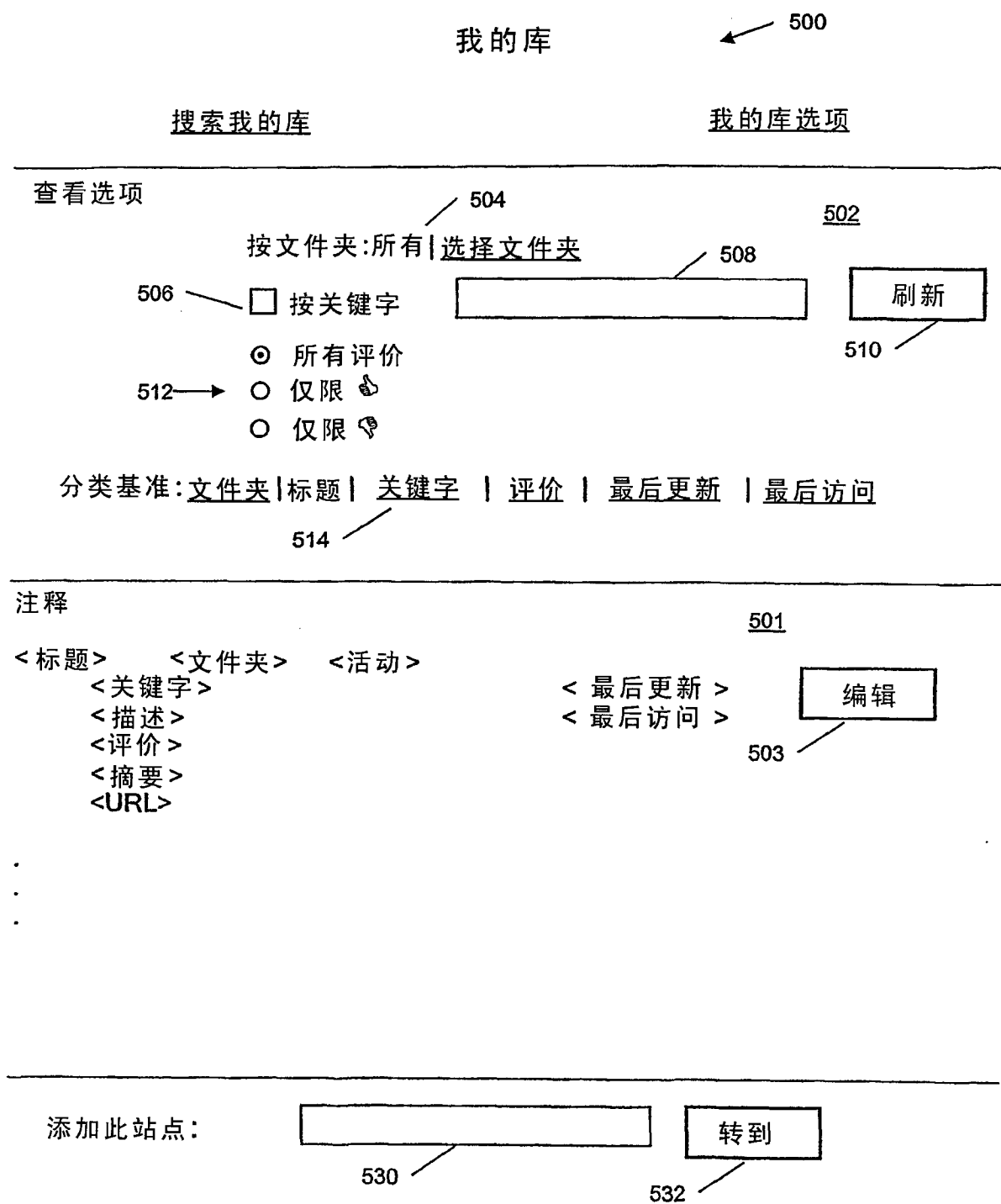


图 5

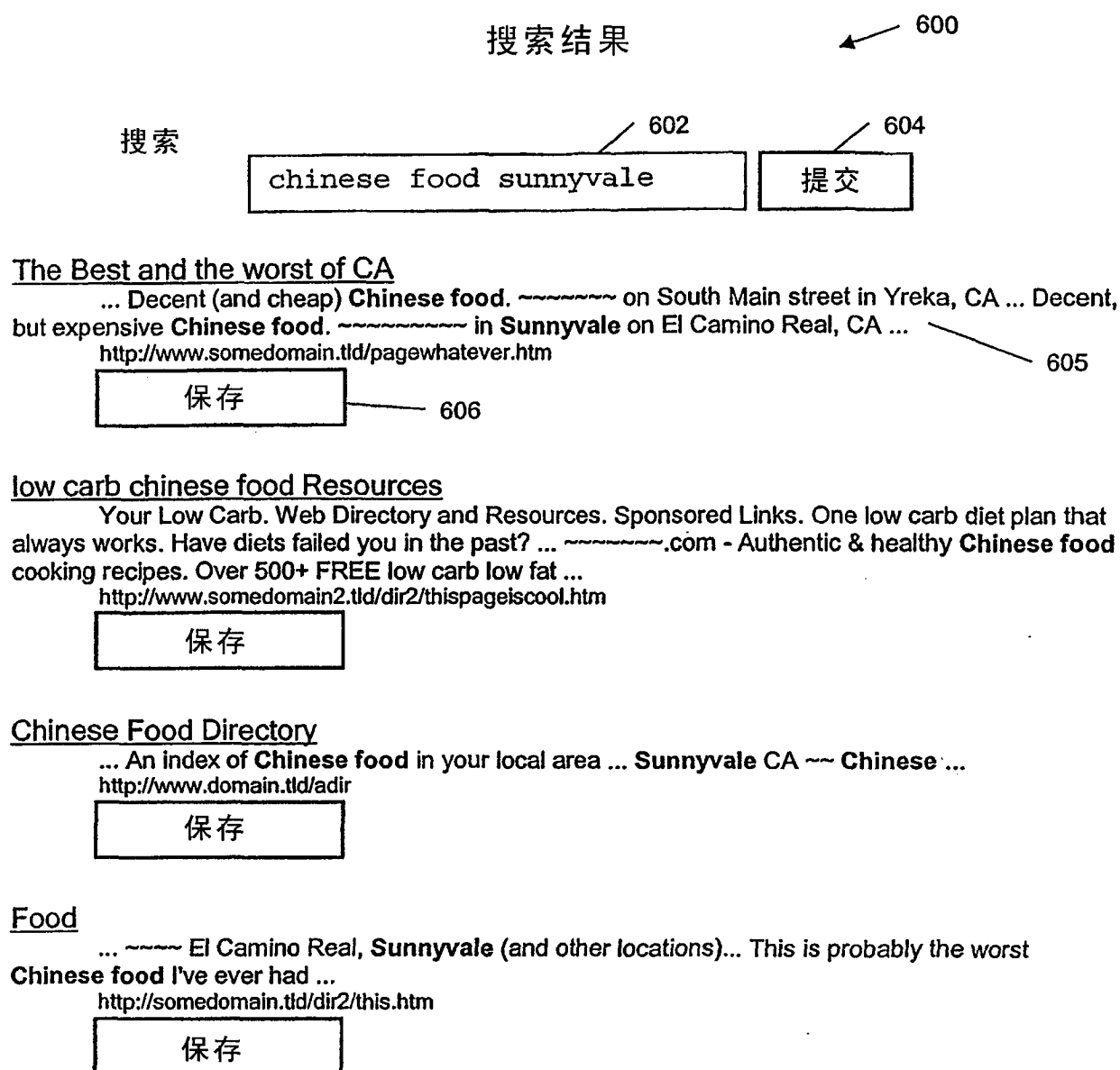


图 6

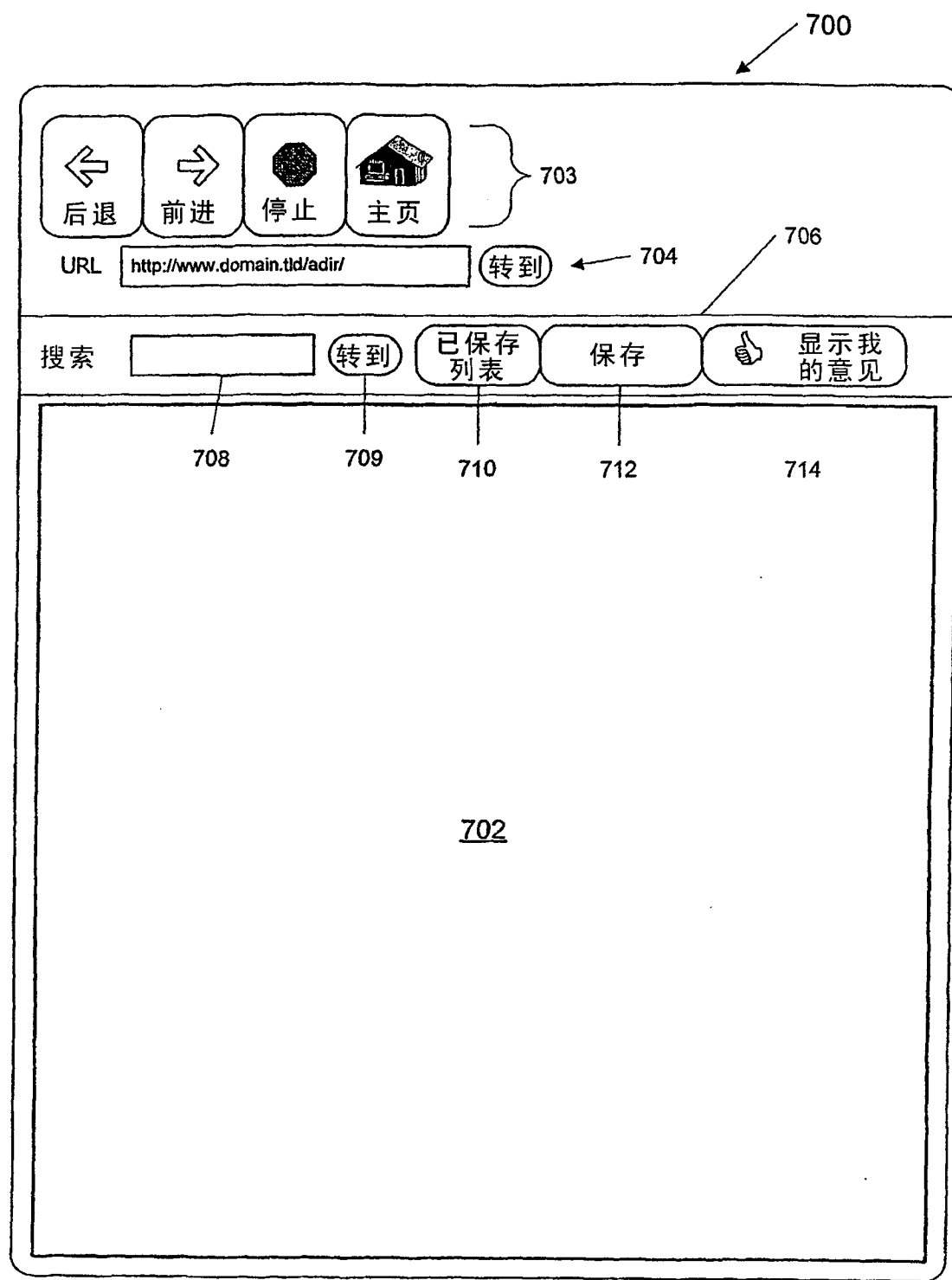


图 7

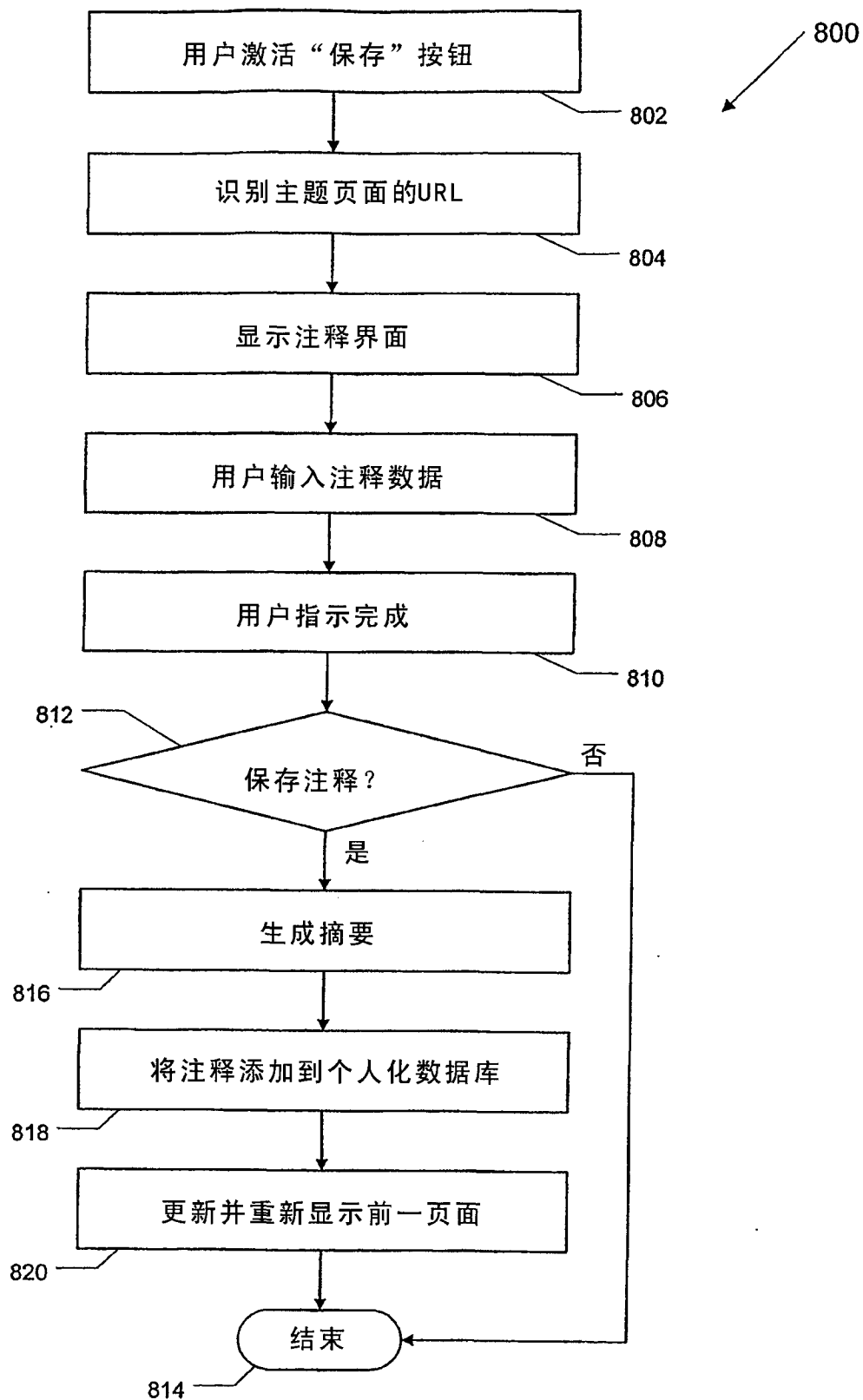


图 8

保存 900

902

URL: <http://www.domain.tld/some-page.html>

标题 904

Good Chinese Food? Here? YES!

保存什么? 908

☒ 仅限此页面

☐ 以下列内容开头的 所有URL 912

910 <http://www.domain.tld/>

关键字 916

描述 914

918

评价:

☐ 

☐  920

☒ 无

922 保存

924 取消

图 9

搜索结果

1002

搜索

The Best and the worst of CA
... Decent (and cheap) **Chinese food**. ~~~~~ on South Main street in Yreka, CA ... Decent, but expensive **Chinese food**. ~~~~~ in **Sunnyvale** on El Camino Real, CA ...
<http://www.somedomain.tld/pagewhatever.htm>

评价: 1000

1004 1006 1008

low carb chinese food Resources
Your Low Carb. Web Directory and Resources. Sponsored Links. One low carb diet plan that always works. Have diets failed you in the past? ... ~~~~~ .com - Authentic & healthy **Chinese food** cooking recipes. Over 500+ FREE low carb low fat ...
<http://www.somedomain2.tld/dir2/thispageiscool.htm>

评价:

Chinese Food Directory
... An index of **Chinese food** in your local area ... **Sunnyvale CA** ~~~ **Chinese** ...
<http://www.domain.tld/adir>

评价:

Food
... ~~~~ El Camino Real, **Sunnyvale** (and other locations)... This is probably the worst **Chinese food** I've ever had ...
<http://somedomain.tld/dir2/this.htm>

评价:

图 10

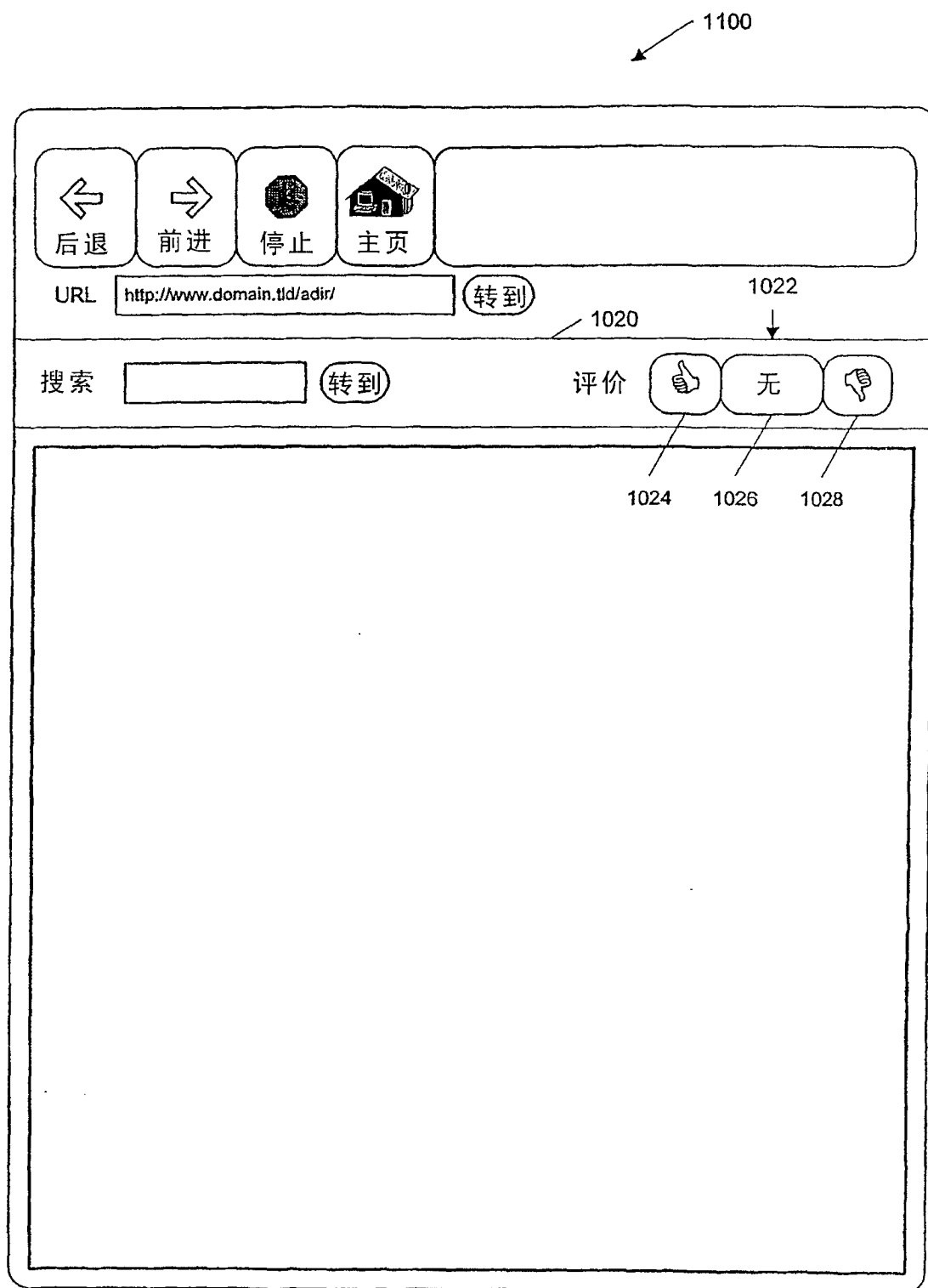


图 11

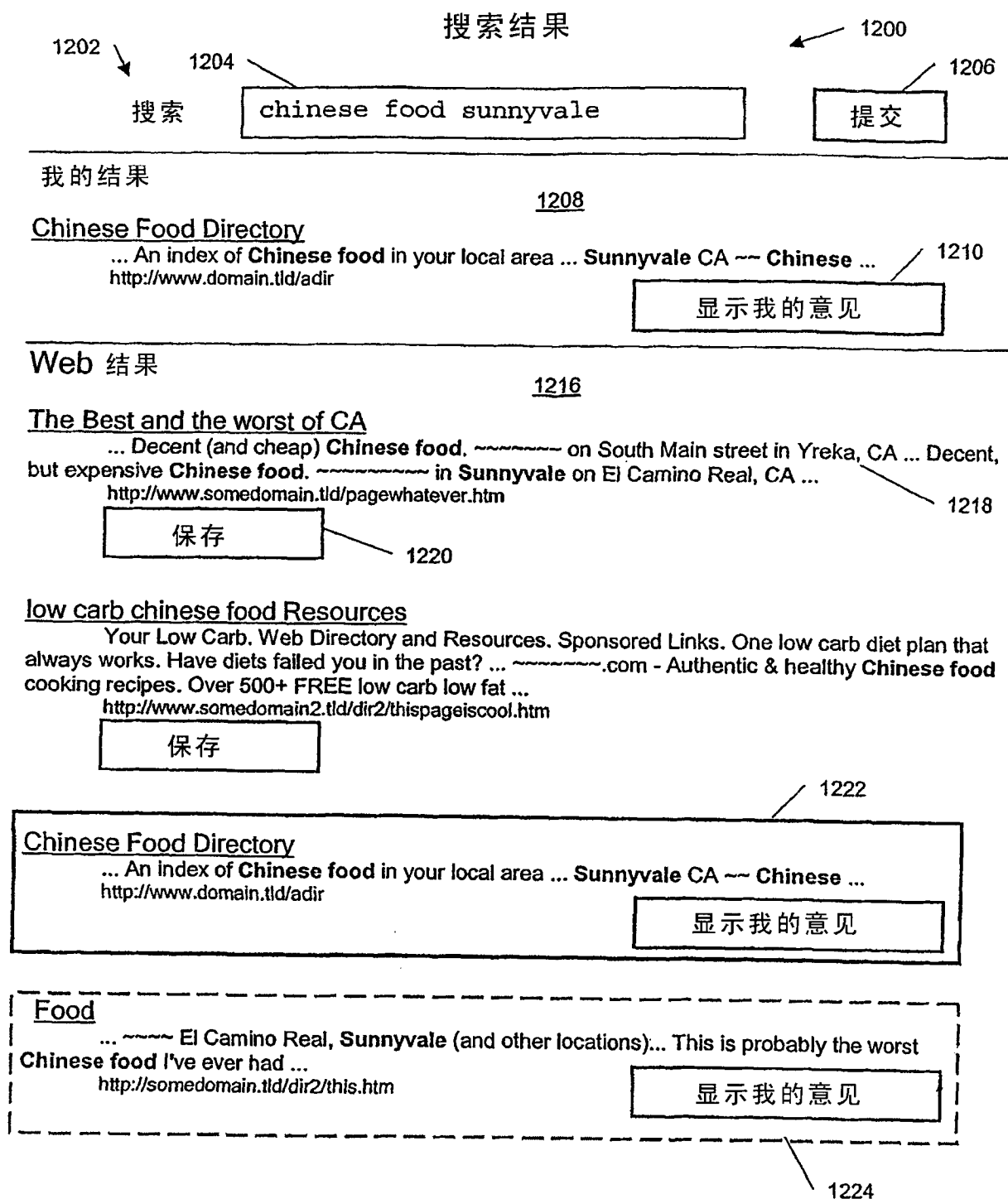


图 12

搜索结果

1200

搜索

chinese food sunnyvale

提交

我的结果

Chinese Food Directory
... An index of Chinese food in your local area ... Sunnyvale CA ~ Chinese ...
<http://www.domain.tld/adir>

显示我的意见

Web 结果

The Best and the worst of CA
... Decent (and cheap) Chinese food. ~~~~~ on S
but expensive Chinese food. ~~~~~ in Sunnyvale on E
<http://www.somedomain.tld/pagewhatever.htm>

保存

low carb chinese food Resources
Your Low Carb. Web Directory and Resources. Spon
always works. Have diets failed you in the past? ... ~~~~~
cooking recipes. Over 500+ FREE low carb low fat ...
<http://www.somedomain2.tld/dir2/thispageiscool.htm>

保存

标题: Chinese Food Directory
关键字: chinese, food, local
描述: Best list I've found
for Chinese food in Sunnyvale.
Has great reviews. ←1302
评价: 1304 1306
编辑 关闭

1300

Chinese Food Directory
... An index of Chinese food in your local area ... Sunnyvale CA ~ Chinese ...
<http://www.domain.tld/adir>

显示我的意见

Food
... ~~~~~ El Camino Real, Sunnyvale (and other locations)... This is probably the worst
Chinese food I've ever had ...
<http://somedomain.tld/dir2/this.htm>

显示我的意见

图 13

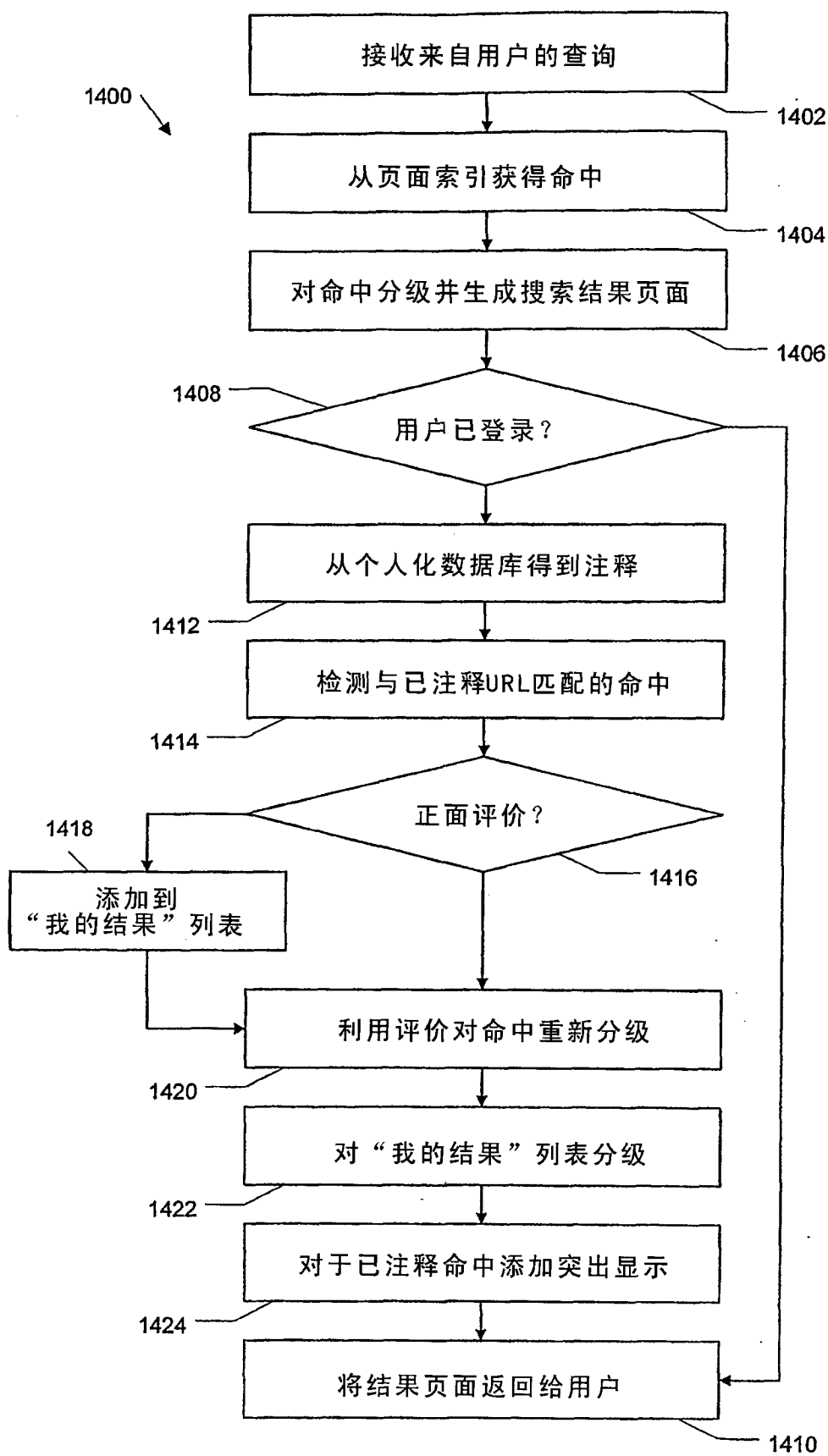


图 14

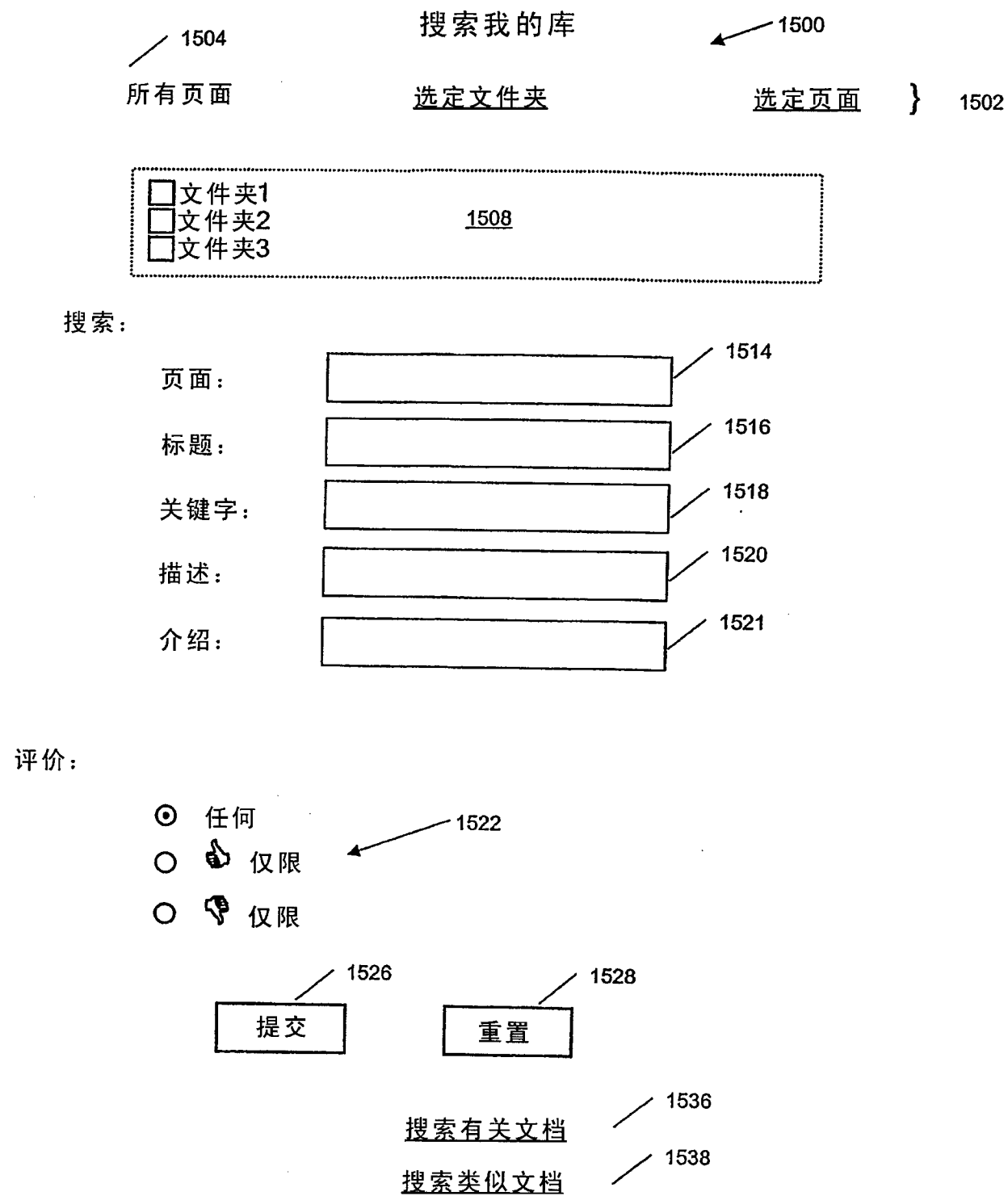


图 15