



(12)发明专利

(10)授权公告号 CN 101595476 B

(45)授权公告日 2017.09.29

(21)申请号 200780032062.9

(22)申请日 2007.08.31

(65)同一申请的已公布的文献号

申请公布号 CN 101595476 A

(43)申请公布日 2009.12.02

(30)优先权数据

11/469,258 2006.08.31 US

11/550,457 2006.10.18 US

11/625,452 2007.01.22 US

(85)PCT国际申请进入国家阶段日

2009.02.27

(86)PCT国际申请的申请数据

PCT/CA2007/001546 2007.08.31

(87)PCT国际申请的公布数据

W02008/025167 EN 2008.03.06

(73)专利权人 启创互联公司

地址 加拿大安大略省

(72)发明人 彼得·斯威尼 罗伯特·古德

(74)专利代理机构 北京市金杜律师事务所
11256

代理人 王茂华 李峥宇

(51)Int.Cl.

G06F 17/30(2006.01)

(56)对比文件

US 6292792 B1,2001.09.18,全文.

CN 1395193 A,2003.02.05,说明书第1页第
5行—第4页第23行,图1—6.

审查员 李佳

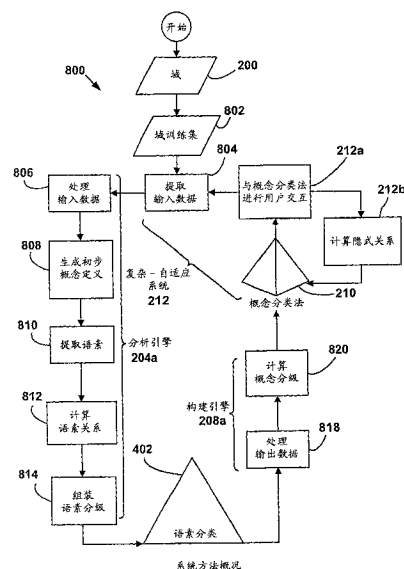
权利要求书7页 说明书49页 附图37页

(54)发明名称

用于客户定义信息架构的系统、方法和计算机程序

(57)摘要

提供一种用于基于来自反馈代理的输入来组织和管理数据结构的系统、计算机系统和方法,该方法包括:(a)适用于信息域的用于分面式分类的方法,所述分面式分类方法包括:(i)对所述域进行分面分析或者接收域的分面分析结果;以及(ii)应用所述域的分面式分类综合;以及(b)用于根据所述分面式分类综合来选择和返回关于一次或者多次迭代的信息的复杂-自适应方法,所述复杂-自适应方法响应于所述返回的信息来改变数据结构的组织和管理。



1. 一种用于信息域的分面式分类的计算机实现的方法,其特征在于:

a) 提供包括用来对信息进行分类的分面属性的分面式数据集,这种分面属性可选地包括用于所述分面属性的分面属性分级;

b) 提供维度概念分类法,其中所述分面属性根据将含义与待分类的域的对象进行关联的概念而分配给所述对象,所述概念由概念定义表示,所述概念定义在所述维度概念分类法中使用所述分面属性来定义并且与所述对象相关联,所述维度概念分类法根据所述分面式数据集来表达所述概念定义之间的维度概念关系;以及

c) 提供或者实现一种复杂-自适应系统,其用于选择和返回维度概念分类法信息,以响应于所述维度概念分类法信息来改变所述分面式数据集和维度概念分类法;

其中对输入信息的分面分析根据源数据结构选自信息域,所述方法的特征在于:

使用模式扩充和/或统计分析来发现所述输入信息的至少一个所述分面属性,以标识所述输入信息中的分面属性关系模式;以及

所述分面属性是语素,并且所述分面属性关系是语素关系。

2. 根据权利要求1所述的方法,其特征在于:

根据从所述输入信息获得的概念定义和概念关系来推断所述分面属性关系,由此实现分面属性分级的构建,所述概念定义包括用于定义所述分面属性的属性。

3. 根据权利要求2所述的方法,其特征在于:

根据所述输入信息中的概念定义和概念关系中的分面属性关系排列,来建立潜在分面属性关系。

4. 根据权利要求3所述的方法,其特征在于:

根据潜在分面属性关系来建立分面属性关系,以减少潜在分面属性关系的数目,以便扩充用于统计分析的所述模式。

5. 根据权利要求4所述的方法,其特征在于:

根据如下各项中的至少一项来建立潜在分面属性关系:(a) 存在跨过相关概念定义的配对而共享的分面属性;(b) 在所述分面属性分级中直接或者间接地存在所述分面属性关系;以及(c) 外部词典。

6. 根据权利要求5所述的方法,其特征在于:

对于每个潜在分面属性关系,评估相应潜在分面属性关系保持基本上适用于其中存在所述相应潜在分面属性关系的所述概念关系的可能性;以及响应于所述评估,使所述相应潜在分面属性关系构成用于所述分面属性分级的候选分面属性关系。

7. 根据权利要求6所述的方法,其特征在于:

所述评估包括确定所述相应潜在分面属性关系在所有概念关系的集合体背景中的普及度,以及其中所述构成响应于所述普及度。

8. 根据权利要求6所述的方法,其特征在于:

所述评估包括标识所述潜在分面属性关系之中存在循环关系,以及标识与分面属性关系之间的分级假设相抵触的相应潜在分面属性关系,并且其中所述构成响应于所述标识。

9. 根据权利要求6所述的方法,其特征在于:

将所述候选分面属性关系组装成分面属性多分级,从而使所述多分级中的候选分面属性关系集在集合体中在逻辑上基本一致。

10. 根据权利要求2所述的方法,其特征在于:

通过将具有多个父属性的分面属性的级别重新考虑为所述多个父属性的祖先的属性,来将所述分面属性分级定义为严格分级。

11. 根据权利要求10所述的方法,其特征在于:

根据所述分面属性分级内的根节点来确定所述分面属性。

12. 根据权利要求1所述的方法,其特征在于:

使用所述分面属性以及可选地使用所述分面属性分级,来对从第二信息域中选择的更多输入信息执行分面分析。

13. 根据权利要求1所述的方法,其特征在于:

提供包括所述分面属性以及可选地包括分面属性分级的分面式数据集。

14. 根据权利要求13所述的方法,其特征在于:

将所述分面式数据集用于分面分类综合。

15. 根据权利要求1所述的方法,其特征在于:

访问包括多个统计分析的数据储存设备,以及通过所述复杂-自适应系统的操作将所述统计分析用于通过聚集所选维度概念分类法信息,来改变所述分面式数据集和维度概念分类法。

16. 根据权利要求1所述的方法,其特征在于:

执行分面式分类综合以将概念集相关,所述概念集由根据包括分面属性以及可选地包括分面属性分级的分面式数据集所定义的概念定义来表示,所述分面式分类综合包括:表达所述概念定义之间的维度概念关系,其中通过检查显式关系和隐式关系中的至少一个关系在两个概念定义的相应分面属性之间在所述分面式数据集中是否存在,来确定所述两个概念定义在特定维度概念关系中相关。

17. 根据权利要求16所述的方法,其特征在于:

在维度概念分级中组装所述维度概念关系。

18. 根据权利要求16所述的方法,其特征在于:

根据一个或者多个分面属性来定义维度轴,所述维度轴用来根据所述维度概念关系来生成维度概念分级。

19. 根据权利要求18所述的方法,其特征在于:

定义以下一个或者多个:(i) 将要分类的域,以及(ii) 来自所述概念集之中的活跃概念,所述活跃概念用作用于生成所述维度概念分级的祖先概念或者后代概念。

20. 根据权利要求16所述的方法,其特征在于:

所述两个概念定义的所述相应分面属性定义所述两个概念关系之间的祖先或者后代关系。

21. 根据权利要求19所述的方法,其特征在于:

定义来自所述域中的一个或者多个概念的所述活跃概念。

22. 根据权利要求19所述的方法,其特征在于:

如果定义所述活跃概念,则根据所述活跃概念的一个或者多个分面属性来定义所述维度轴。

23. 根据权利要求19所述的方法,其特征在于:

借助人类输入来定义所述域、所述活跃概念和所述维度轴中的一个或者多个。

24. 根据权利要求19所述的方法,其特征在于:

借助机器输入来定义所述域、所述活跃概念和所述维度轴中的一个或者多个。

25. 根据权利要求18所述的方法,其特征在于:

定义所述维度概念分级中的分级阶跃的数目限制。

26. 根据权利要求18所述的方法,其特征在于:

定义将要相关的概念的数目限制。

27. 根据权利要求25或者26所述的方法,其特征在于:

借助人类输入来定义所述限制。

28. 根据权利要求25或者26所述的方法,其特征在于:

借助机器输入来定义所述限制。

29. 根据权利要求16所述的方法,其特征在于:

定义多个维度轴,各轴根据一个或者多个分面属性的相应集合来定义,以便定义多个维度概念分级。

30. 根据权利要求18所述的方法,其特征在于:

在特定维度概念关系中,如果两个概念定义之一的所述分面属性与所述两个概念定义中另一概念定义的所有分面属性或者分面属性子集相关,则在特定维度轴上对所述两个概念定义进行关联。

31. 根据权利要求18所述的方法,其特征在于:

在特定维度概念关系中,如果两个概念定义之一的所述分面属性的子集与所述两个概念定义中另一概念定义的所有分面属性或者分面属性子集相关联,则在特定维度轴上对所述两个概念定义进行关联。

32. 根据权利要求18所述的方法,其特征在于:

选择用于定义所述维度轴的一个或者多个分面属性,根据以下各项中的至少一项进行所述选择:(i) 所述分面属性分级中的所述分面属性的相应优先级;以及(ii) 用以将特定含义关联到所述维度轴的、包括所述一个或者多个分面属性的特定概念定义。

33. 根据权利要求16所述的方法,其特征在于:

针对间接关系的存在来评估维度概念关系集,以及组装无所述间接关系的所述维度概念分级。

34. 根据权利要求16所述的方法,其特征在于:

基于以下各项中的至少一项为所述维度概念分级中的所述概念定义建立优先级:(i) 所述分面属性分级中的分面属性优先级;以及(ii) 相关概念的所述相应概念定义中所述分面属性的分析。

35. 根据权利要求16所述的方法,其特征在于:

通过根据与将要分类的域的内容节点相关联的概念,将分面属性分配给所述内容节点来定义维度概念分类法,所述概念由在所述维度概念分级中使用所述分面属性而定义的概念定义来表示。

36. 根据权利要求16所述的方法,其特征在于:

推断两个或者多个概念定义之间的概念关系,其中在所述两个概念定义之间没有显式

概念关系。

37. 根据权利要求35所述的方法,其特征在于:

将概念推断为由一个或者多个分面属性定义的概念定义,其中在所述域中没有与所述两个概念定义中的另一概念定义相关联的关联内容节点。

38. 根据权利要求16所述的方法,其特征在于:

针对将要分类的域来定义所述维度概念关系,所述域包括各自与所述概念定义表示的一个或者多个概念相关联的一个或者多个内容节点。

39. 根据权利要求38所述的方法,其特征在于:

针对概念节点的相应一个或者多个概念定义来检查所述概念节点。

40. 根据权利要求38所述的方法,其特征在于:

针对将要分类的域来定义所述维度概念关系,从而所述定义限于使用与所选内容节点邻近的内容节点来处理所述域的局部化区域。

41. 根据权利要求1所述的方法,其特征在于:

通过定义候选概念定义的集合使所述概念相关,所述候选概念定义通过分析与所述概念相关联的所述分面属性或者所述分面属性的子集而被确定为相互邻近。

42. 根据权利要求16所述的方法,其特征在于:

基于与所述两个概念相关联的分面属性,通过定义所述分面属性分级来建立相互邻近的概念定义,从而派生与所述两个概念相关联的候选概念定义集。

43. 一种用于组织和管理数据结构的计算机系统,包括基于来自反馈代理的输入,其特征在于:

所述系统包括或者链接到复杂-自适应系统,所述复杂-自适应系统用于选择和返回维度概念分类法信息,以响应于维度概念分类法信息来改变分面式数据集和维度概念分类法,其中:

所述复杂-自适应系统可操作用以处理分面式数据集,所述分面式数据集包括分面、分面属性以及可选地包括用来对信息进行分类的用于所述分面属性的分面属性分级;以及

所述复杂-自适应系统还可操作用以定义所述维度概念分类法,其中所述分面属性根据将含义与待分类的域的对象进行关联的概念而分配给所述对象,所述概念由概念定义表示,所述概念定义在所述维度概念分类法中使用所述分面属性来定义并且与所述对象相关联,所述维度概念分类法根据所述分面式数据集来表达所述概念定义之间的维度概念关系;

所述复杂-自适应系统可操作用以 (i) 根据源数据结构来处理从信息域中选择的输入信息;以及 (ii) 使用模式扩充和/或统计分析 来发现所述输入信息的至少一个所述分面属性,或者可选地发现所述输入信息的所述分面属性分级之一,以标识所述输入信息中的分面属性关系模式;以及

所述分面属性是语素,并且所述分面属性关系是语素关系。

44. 根据权利要求43所述的系统,其特征在于:

所述复杂-自适应系统可操作用以返回用于加强所述分面属性或者可选地加强所述分面属性分级的更多输入信息,所述分面属性和所述分面属性分级构成用于所述返回的更多输入信息的基础,所述更多输入信息根据基于机器或者基于用户的返回路径中的一个或者

多个路径而派生。

45. 根据权利要求44所述的系统,其特征在于:

(a) 所述更多输入信息借助对所述复杂-自适应系统生成的更多输入信息的请求来获得;

(b) 这种请求与用于响应于这样的请求的时间限制相关联。

46. 根据权利要求43所述的系统,其特征在于:

所述复杂-自适应系统包括或者链接到统计分析数据储存设备;以及

所述复杂-自适应系统可操作用以通过聚集所选维度概念分类法信息基于所述统计分析,来改变所述分面式数据集和维度概念分类法。

47. 根据权利要求45所述的系统,其特征在于:

所述复杂-自适应系统可操作用以响应于所述维度概念分类法信息,来控制所述分面式数据集和维度概念分类法的变化。

48. 根据权利要求46所述的系统,其特征在于:

所述复杂-自适应系统可操作用以针对根据所述维度概念分类法信息派生的所述分面属性,对于所述分面、所述分面属性和所述分面属性分级来应用以下一个或者多个:(i) 统计障碍;以及(ii) 模式匹配约束。

49. 根据权利要求43所述的系统,其特征在于:

所述复杂-自适应系统包括基于机器的复杂自适应系统,所述基于机器的复杂自适应系统可操作用以使用统计分析以分析所述维度概念分类法以及选择将要返回的维度概念分类法信息。

50. 一种计算机系统,用于根据源数据结构来执行从信息域中选择的输入信息的分面分析,其特征在于所述系统包括:

处理器;

与所述处理器进行数据通信的计算机可读介质,其中所述计算机可读介质上包括处理器可执行的指令;

当由所述处理器执行所述指令时,使得所述计算机系统可操作用以使用模式扩充和统计分析来派生所述输入信息的分面属性,以及可选地派生所述输入信息的分面属性分级,以标识所述输入信息中的分面属性关系模式;以及

所述分面属性是语素,并且所述分面属性关系是语素关系。

51. 根据权利要求50所述的计算机系统,其特征在于:

所述计算机系统可操作用以根据从所述输入信息获得的概念定义和概念关系来推断所述分面属性关系,以由此构建所述分面属性分级,所述概念定义包括用于定义所述分面属性的属性。

52. 根据权利要求51所述的计算机系统,其特征在于:

所述计算机系统可操作用以通过递增地添加分面属性关系,来构建所述分面属性分级。

53. 根据权利要求51所述的计算机系统,其特征在于:

所述计算机系统可操作用以根据所述输入信息中的概念定义和概念关系中的分面属性关系排列来建立潜在分面属性关系。

54. 根据权利要求53所述的计算机系统,其特征在于:

所述计算机系统可操作用以根据概念关系配对的所述概念定义中的分面属性的计数,以对所述潜在分面属性关系的处理进行排序。

55. 根据权利要求53所述的计算机系统,其特征在于:

所述计算机系统可操作用以校准来自所述潜在分面属性关系的分面属性关系,以减少潜在分面属性关系的数目,从而扩充用于统计分析的所述模式。

56. 根据权利要求53所述的计算机系统,其特征在于:

所述计算机系统可操作用以根据以下各项中的至少一项来校准所述潜在分面属性关系:(a) 存在跨过相关概念定义配对而共享的分面属性;(b) 在所述分面属性分级中直接或者间接地存在所述分面属性关系;以及(c) 外部词典。

57. 根据权利要求53所述的计算机系统,其特征在于:

所述计算机系统可操作用以:对于各潜在分面属性关系,评估相应潜在分面属性关系保持基本上适用于存在所述相应潜在分面属性关系的所述概念关系的可能性;以及响应于所述评估使所述相应潜在分面属性关系构成用于所述分面属性分级的候选分面属性关系。

58. 根据权利要求57所述的计算机系统,其特征在于:

所述评估包括确定所述相应潜在分面属性关系在所有概念关系的集合体背景中的普及度,以及其中所述构成响应于所述普及度。

59. 根据权利要求58所述的计算机系统,其特征在于:

所述评估包括确定所述潜在分面属性关系之中存在循环关系,以及标识与分面属性关系之间的分级假设相抵触的相应潜在分面属性关系,以及其中所述构成响应于所述标识。

60. 根据权利要求58所述的计算机系统,其特征在于:

所述计算机系统可操作用以:如果可以使用其他候选分面属性关系的集合来逻辑地构造所述候选分面属性关系之一,则删减所述一个候选分面属性关系。

61. 根据权利要求58所述的计算机系统,其特征在于:

所述计算机系统可操作用以将所述候选分面属性关系组装成分面属性多分级,从而所述多分级中的候选分面属性关系集在所述集合体中在逻辑上基本一致。

62. 根据权利要求61所述的计算机系统,其特征在于:

所述计算机系统可操作使得所述组装使用为了创建所述逻辑一致多分级而必需的最少数目的多分级关系。

63. 根据权利要求61所述的计算机系统,其特征在于:

所述计算机系统可操作使得所述组装响应于各分面属性的相对通用性的度量来对所述多分级中的相应分面属性进行排序。

64. 根据权利要求63所述的计算机系统,其特征在于:

所述计算机系统可操作使得使用与相应分面属性相关联的一个或者多个概念定义的通用性,来确定所述相应分面属性的通用性。

65. 根据权利要求50所述的计算机系统,其特征在于:

所述计算机系统可操作用以通过将具有多个父属性的分面属性的顺序重新考虑为所述多个父属性的祖先的属性,来将所述分面属性分级定义为严格分级。

66. 根据权利要求50所述的计算机系统,其特征在于:

所述计算机系统可操作用以根据所述分面属性分级内的根节点来确定所述分面属性。

67. 根据权利要求51所述的计算机系统,其特征在于:

所述计算机系统可操作用以根据所述输入信息来派生概念定义和所述概念关系,其中所述概念关系定义为分面属性。

68. 根据权利要求50所述的计算机系统,其特征在于:

所述计算机系统可操作用以标识所述输入信息中的结构标记器,以及响应于所述结构标记器从所述输入信息提取属性。

69. 根据权利要求50所述的计算机系统,其特征在于:

所述计算机系统可使用所述分面属性来操作,用以对从第二信息域中选择的更多输入信息执行分面分析。

70. 根据权利要求50所述的计算机系统,其特征在于:

所述计算机系统可操作用以提供包括所述分面属性以及可选地包括所述分面属性分级的分面式数据集。

71. 根据权利要求70所述的计算机系统,其特征在于:

所述计算机系统可操作用以提供用于在多个域之间共享的所述分面式数据集,所述分面式数据集用来针对每个域派生相应维度概念分类法。

72. 根据权利要求70所述的计算机系统,其特征在于:

所述计算机系统可操作用以将所述分面式数据集用于分面分类综合。

用于客户定义信息架构的系统、方法和计算机程序

[0001] 优先权

[0002] 本申请要求以下专利申请的优先权：于2006年8月31日提交的美国专利申请第11/469,258号，于2006年10月18日提交的美国专利申请第11/550,457号和于2007年1月22日提交的美国专利申请第11/625,452号。

技术领域

[0003] 本发明总体上涉及分类系统。具体而言，本发明涉及一种用以对信息进行分类的系统、方法和计算机程序。本发明还涉及一种用于综合用于特定信息域的分类结构的系统、方法和计算机程序。

背景技术

[0004] 分面式分类是基于信息具有多维性质并且可以用许多不同方式来分类这一原理。信息域的对象被细分成分面以代表这一多维性。域的属性按分面分级而相关。然后基于这些属性对域内的素材进行标识和分类。

[0005] 图1图示了现有技术中的分面式分类的大体方式，其中该方式（例如）应用于酒的分类。

[0006] 分面式分类称为分析-综合方法，因为它涉及到分析和综合两种过程。为了构思用于分面式分类的方案，分析信息域以确定它们的基本分面。然后可以通过将这些分面的属性应用于域来对分类进行综合（或者构建）。

[0007] 许多学者已经将分面式分类认同为用于组织海量信息储存器（例如因特网上的信息储存器）的理想方法。分面式分类易受迅速变化和动态信息的影响。另外，通过将对象细分成分面，它提供用以访问信息的多个各种方式。

[0008] 尽管分面式分类有可能解决分类需要，但对其的使用一直是缓慢的。相对于因特网上的海量信息，只有很少的域使用分面式分类。实际上，其使用已经在具体纵向应用（比如电子商务储存器和图书馆）中进行划分。它一般停留在学者、专业分类和信息设计者的范围中。

[0009] 采用分面式分类的障碍在于它的复杂性。分面式分类是劳动密集型并且在理解力上颇具挑战性的工作。这一复杂性随着信息的规模而增加。随着规模增加，维度（或者分面）的数目在域内更复杂，这使其越来越难以组织。

[0010] 为了帮助解决这一复杂性，学者已经构思了用于分面式分类的规则和指南。这一学术群体可追溯至早在现代计算和数据分析问世之前的几十年。

[0011] 最近，已经在分面式分类的服务中谋求技术。大体上，已经在历史分类方法和组织原理内应用这一技术。受传统方法束缚，对提供全自动化分面式分类方法的尝试一般受到了挫折。

[0012] 为了表明现有技术，自动化分类和分面式导航系统的一个例子是ENDECA™。ENDECA被认为是信息分类和访问系统业内产品的优良领军者 (<http://www.usatoday.com/tech/>)

products/cnet/2007-06-29-endece-google_N.htm)。

[0013] ENDECA的技术使用了指导式导航和如下元关系索引,该元关系索引包容数据和文档的维度以及维度之间的关系:例如参见“Hierarchical data-driven search and navigation system and method for information retrieval”;美国专利No.7,035,864,2006年4月25日:“Hierarchical data-driven search and navigation system and method for information retrieval”。

[0014] ENDECA的系统包括由公司描述为分类法定义和分类的分类方式,参见:美国专利No.7,062,483,2006年6月13日:Hierarchical data-driven search and navigation system and method for information retrieval。

[0015] 现有自动化分类技术得到最为主流地使用,并且对于以下事物是有用的,业界专家将这些事物称为“结构化数据仓库”和“受管理的内容仓库”。

[0016] 现有自动化分类技术的另一限制在于:在分类的认知要求方面,它缺乏基于人类的反馈。例如,尽管ENDECA具有分面式导航的反馈环——包括用以驱动搜索结果呈现的使用普及度以及包括优先级——但是它没有用以改进内容的语义定义和语义关系的基于使用的反馈环。

[0017] 另一大类混合分类系统可以描述为大规模协作式分类。这一方式尝试对人工分类的认知优点与自动化系统的处理能力进行结合。这一新兴领域中的协作式分类系统冠有各种称谓:“Web 2.0”、“协作式分类”、“公众分类法(folksonomy)”、“社会索引”、“社会标签”,“群体智慧”和其他称谓。FLICKR™(照片共享群体)、DEL.ICIO.US™(社会书签管理器)和WIKIPEDIA™(基于wiki的协作式百科全书)是这一类新兴协作式分类的例子。

[0018] 在比例可变的情况下,这些系统使用用以在允许协作者针对类别、概念和关系进行决策之时,为广范围和分布式协作提供框架的技术。这一方式的一个挑战在于,它在主题和分类专家的指导与外行个人终端用户的输入之间产生抵触,这些用户通常对于内容持有大相径庭的观点和分类方式。这些系统可以通过标识不明确和不一致的范围以及通过突出协作者之间的竞争性主张,来帮助人们协作。但是最终,就协作式系统而言,人们应当优选地和解他们的分歧并且就最难以应对的条款达成广泛共识。因此这一过程难以在大型和各种信息域内展开和扩展。

[0019] 协作式分类方式的首要例子是Metaweb Technologies公司,该公司着眼于通过使用协作式分类方式以在Web以及其他复杂和各种信息环境内创建可搜索的数据库,来对广范围的开放信息域进行分类。

[0020] Metaweb Technologies因其用以创建语义网的先驱协作式方式而倍受关注。Metaweb Technologies已经向美国专利商标局提交2个专利申请(美国专利申请20050086188,“Knowledge web”,2005年4月21日;美国专利申请20030196094,“Method and apparatus for authenticating the content of a distributed database”,2003年10月16日)。

[0021] Metaweb Technologies的协作式本体构建依赖于用于其协作式分类的“群众智慧”。终端用户利用它来定义和扩展每个人都可以使用的多个方案。根据著名的业界观察员Esther Dyson的观点:“Metaweb的创建者已经‘聪明地设计’了如何指定关系的语法,但是他们依赖于群众的智慧(或者具体知识)和努力,以创建实际内容——不仅仅是具体数据,

而是在具体事物之间具体种类的关系”(第0.9版:Metaweb-Emergent Structure vs. Intelligent Design, 2007年3月11日, http://www.huffingtonpost.com/esther-dyson/release-09-met_b_43167.html)。这一方式的限制在于数据库范围和质量受制于其用户输入的语义有关内容。它也依赖于专家和外行人士用以就具体数据元素达成共识并且指定内容之间的关系以消除冗余度,从而数据库包含明确信息的能力。

[0022] 因此,现有技术自动化分面式分类、自动化分类和大规模协作式分类方面存在许多弊端。在传统方法内或者基于传统方法来应用技术。需要如下增强的分类方法,这些方法实现对信息结构的基本改变。

[0023] 对于分面分析,一般需要人类认知的输入,因为不存在对所有信息域均起作用的用于分面分析的通用模式或者探索方式。目前,只有人类才拥有全广度的模式识别技能。遗憾的是,一般需要在将要分类的整个信息域内标识结构模式(比如语义或者语法结构),并且存在标识分面和属性的许多不同模式。尽管可以训练人们识别小型(局部)数据集上的这些模式,但是随着域的大小增加,该任务的难度变得令人望而却步。

[0024] 当分析和综合过程的计算需要超过人类认知能力时,还由于人类介入而引入限制。人类熟练于小规模评估信息元素之间的关系、但是无法管理在集合体中的整个域内的复杂性。需要能够跨整个信息域来聚集小型局部化人类输入的系统。

[0025] 分面式分类方案支持多角度,这是一个常常提到的益处。遗憾的是,当这些角度跨多个分级而分开时,它们并不直观。这引起可视化、整合和整体角度的严重问题。随着结构中的分面数目(或者维度)增加,可视化变得越来越困难。因而,分面式分类方案的可视化常常缩减为“平面式”的一维结果集;其一次仅跨一个分面来导航结构。这一类缩减掩盖了下层结构的丰富复杂性。

[0026] 需要在整合的和描述丰富的分级内组合分面式方案的表达性和灵活性的方法和技术。另外,这一灵活性在将分面动态构造为组织基础时最优地向下扩展至分类方案本身的基本级。

[0027] 一旦被选择,分面本身就是静态的并且难以修正。这在分面式方案的开发中表示相当大的风险。分类常常缺乏信息域的完整知识,因此这些组织基础的选择易于出错。在动态分类系统之下,将通过易于添加或者更改下层分面的能力来减轻这些风险。传统分类方法和派生技术在这一基础级中缺乏灵活性。

[0028] 任何分类系统还可以考虑动态环境中的维护要求。随着域中的素材改变,分类可以相应地调整。相对于分面式分类方案的初始开发,维护常常是甚至更令人沮丧的挑战。在术语学(terminology)显现和改变时必须对其加以更新;一般需要评估和用符号表示域中的新素材;一般需要调整分面和属性的布置以包含演变的结构。许多时候,为了整个新分类而简单地放弃现有分面式分类。

[0029] 混合系统在分析、综合和维护的关键阶段涉及到人类。如在该过程中的早期涉及到的那样,人类常常成为分类工作的瓶颈。这样,该过程一直处于缓慢和高成本。需要这样一种系统,该系统以无需中央化控制和权限的更为分散、自制的方式接受来自人们的分类数据。这些系统可以支持隐式反馈机制,其中信息访问和信息消费的真实活动为分类方案的维护和成长提供积极维护。

[0030] 为了指导该过程,混合系统常常基于现有的通用分面式分类方案。然而,这些通用

方案并不总是适用于海量的和迅速演变的现代信息世界。需要针对单独域的需求而定制的专门化方案。

[0031] 由于不能通用地应用通用分面式分类方案,所以也需要将不同信息域连接在一起。然而,尽管提供用以整合域的机会,但是解决方案应当考虑单独域所有者的私密和安全。

[0032] 分类需要的最重要之处要求系统可以在涉及到大型协作者组的广泛分散环境中管理的系统。然而,在意味深长和含义不明的影响下,分类应对复杂概念。解决这些含义不明确和冲突常常涉及到甚至在小型组中仍脱离协作的密集协商和个人冲突。

发明内容

[0033] 在本发明的第一方面中,提供一种用于组织和管理数据结构的方法,其中包括基于来自反馈代理的输入,该方法包括:(a) 适用于信息域的用于分面式分类的方法,所述分面式分类方法包括(i) 对所述域进行分面分析或者接收域的分面分析结果,以及(ii) 应用所述域的分面式分类综合;以及(b) 用于根据所述分面式分类综合来选择和返回关于一次或者多次迭代的信息的复杂-自适应方法,所述复杂-自适应方法响应于所述返回的信息来改变数据结构的组织和管理。

[0034] 在本发明的另一方面中,一种用于信息域的分面式分类的方法包括:(a) 提供包括用来对信息进行分类的分面属性的分面式数据集,这样的分面属性可选地包括用于分面属性的分面属性分级;(b) 提供维度概念分类法,在该维度概念分类法中,根据将含义与将要分类的域的对象相关联的概念来将分面属性分配给对象,所述概念由概念定义表示,这些概念定义在维度概念分类法中使用所述分面属性来定义并且与对象关联,所述维度概念分类法根据分面式数据集来表达概念定义之间的维度概念关系;以及(c) 提供或者实现一种复杂-自适应系统,用于选择和返回维度概念分类法信息,以响应于维度概念分类法信息来改变分面式数据集和维度概念分类法。

[0035] 在本发明的又一方面中,用于信息域的分面式分类的方法还包括:执行分面式分类综合,以将概念定义所表示的概念集相关,所述概念定义是根据包括分面属性以及可选地包括分面属性分级的分面式数据集来定义的,所述分面式分类综合包括:表达概念定义之间的维度概念关系,其中通过检查两个概念定义的相应分面属性之间在分面式数据集中是否存在显式关系和隐式关系中的至少一个关系,来确定两个概念定义在特定维度概念关系中相关。

[0036] 在本发明的又一方面中,提供一种用于根据源数据结构来执行对从信息域中选择的输入信息的分面分析的计算机系统,该计算机系统:(a) 可操作用以使用模式扩充和统计分析,来派生输入信息的分面属性以及可选地派生输入信息的分面属性分级,以标识输入信息中的分面属性关系模式。

[0037] 在本发明的另一方面中,提供一种用于使用户能够操控维度概念关系的计算机系统,该计算机系统包括:(a) 处理器;(b) 与处理器进行数据通信的计算机可读介质,其中该计算机可读介质上包括处理器可执行的指令和被确定为在特定维度概念关系中相关的多个数据元素;(c) 输入工具,配置用以允许外界实体与处理器进行接口通信;(d) 显示器,操作用以提供至少所选数据元素的可视描绘;以及(e) 编辑器,允许外界实体修改数据元素和

特定维度概念关系。

[0038] 在本发明的又一方面中,提供一种用于组织和管理数据结构的系统,包括基于来自反馈代理的输入,其中:(a) 该系统包括或者链接到复杂-自适应系统,所述复杂-自适应系统用于选择和返回维度概念分类法信息,以响应于维度概念分类法信息来改变分面式数据集和维度概念分类法;(b) 该系统可操作用以处理分面式数据集,该分面式数据集包括分面、分面属性以及可选地包括用来对信息进行分类的用于分面属性的分面属性分级;以及(c) 系统还可操作用以定义维度概念分类法,在该维度概念分类法中,根据将含义与将要被分类的域的对象关联的概念来将分面属性分配给对象,所述概念由概念定义表示,这些概念定义在维度概念分类法中使用分面属性来定义并且与对象进行关联,所述维度概念分类法根据分面式数据集来表达概念定义之间的维度概念关系。

附图说明

[0039] 参照附图将更好地理解本发明。注意,对于这里包含的说明,三角形状用来代表相对简单的数据结构,而锥形形状用来代表将较高维度具体化的相对复杂的数据结构。三角形和锥形的可变大小代表压缩和扩展变换、但是决不表明或者显示压缩或者扩展的精确比例。

[0040] 在说明书中并入的并且构成说明书一部分的附图图示了对本发明各方面的各种例子实施例进行说明的各种例子系统、方法等。应该认识到图中所示元素边界(例如方框、方框组或者其他形状)代表边界的一个例子。本领域普通技术人员将认识到,一个元素可以设计为多个元素或者多个元素可以设计为一个元素。作为另一元素的内部部件而示出的元件可以实施为外部部件,并且反之亦然。另外,元素可能不是按比例绘制的。

[0041] 图1是图示了现有技术的分面式分类方法的示意图;

[0042] 图2图示了操作概况,这些操作示出了用以创建用于域的维度概念分类法的数据结构变换;

[0043] 图3图示了用于图2的操作的知识表示模型;

[0044] 图4更具体地图示了图2中操作概况的细节;

[0045] 图5图示了提取输入数据的方法;

[0046] 图6图示了源结构分析的方法;

[0047] 图7图示了提取初步概念-关键词定义的过程;

[0048] 图8图示了提取语素的方法;

[0049] 图9-图10图示了根据概念关系来计算潜在语素关系的过程;

[0050] 图11A-图11B、图12和图13图示了根据潜在语素关系集来组装多分级语素关系的过程;

[0051] 图14A、图14B和图15图示了使用归因方法将语素多分级重新排序成严格分级;

[0052] 图16A和图16B图示了来自语素分级和关键词分级的样本片段;

[0053] 图17图示了预备用于在构造维度概念分类法时使用的输出数据的方法;

[0054] 图18图示了操作根据元素构造来生成维度概念的方式;

[0055] 图19图示了操作如何组合维度概念关系以生成维度概念分类法;

[0056] 图20、图21和图22图示了如何使用分面式输出数据来构造维度概念分类法;

- [0057] 图23图示了为局部化域集而构建的维度概念分类法；
- [0058] 图24图示了动态综合的模式；
- [0059] 图25图示了用于动态综合的候选集组装的方法；
- [0060] 图26图示了在维度概念分类法内编辑内容容器的用户交互的过程；
- [0061] 图27图示了复杂-自适应系统中的一系列用户交互和反馈环；
- [0062] 图28图示了个性化操作；
- [0063] 图29图示了基于机器的复杂-自适应系统的操作；
- [0064] 图30图示了根据一个实施例的用于执行操作的系统的计算环境和架构部件；
- [0065] 图31图示了一个实施例中的简化数据模式；
- [0066] 图32图示了根据一个实施例的用以执行数据结构变换操作的系统概况；
- [0067] 图33图示了一个实施例中使用的分面式数据结构和支持这些结构的多级架构；
- [0068] 图34图示了基于浏览器的用户界面中的维度概念分类法的视图；
- [0069] 图35图示了有助于动态综合模式的基于浏览器的用户界面；
- [0070] 图36图示了基于概括器(outliner)的用户界面中用于用户交互的环境；以及
- [0071] 图37图示了根据本发明的允许操控分面式分类信息各方面的计算机系统的具代表性的实施。

具体实施方式

[0072] 系统操作

[0073] 具体实施方式具体描述本发明一些方面的一个或者多个实施例。

[0074] 具体实施方式的描述分成下述标题和子标题。

[0075] (1) “本发明概况”：该部分大体地描述与本领域有关的包括本发明的信息分类领域，并且还大体地描述本发明的目的和一些优点。

[0076] (2) “系统操作”：该部分大体地描述在实施本发明时涉及到的步骤。小标题“操作概况”大体地描述包括系统的一些部件。小标题“分面式分析方法”大体地描述本发明的分面分析部件。小标题“分面式分类综合方法”大体地描述本发明的分面综合部件、包括本发明的静态和动态综合部件。小标题“复杂-自适应反馈机制”大体地描述本发明对各种用户交互的响应。

[0077] (3) “实施”：该部分大体地描述本发明可操作的具代表性的实施例。小标题“系统架构部件”大体地描述本发明的可能实施例。小标题“数据模型和模式”大体地描述本发明变换数据的方法。小标题“维度变换系统”大体地描述将在本发明的仅一个可能实施例中出现的本发明系统的操作。以下小标题涉及本发明的代表性实施：“多级数据结构”、“分布式计算环境”、“XML模式和客户端侧变换”和“用户界面”。

[0078] 本发明概况

[0079] 鉴于现有技术中的限制和不足，可以认识到对信息架构的构造性和协作式系统的具体要求以解决这里提到的挑战和问题。因而，本发明的若干目的和优点总结为如下几点：这些目的或者优点并非穷尽而仅用以说明本发明的一些方面及其可能的优点和益处。

[0080] 在本发明的一个方面中，本发明的系统在构造最优信息结构的基本级上操作。大多数现有分类、搜索和可视化解决方案是在具有缺陷的结构基础之上进行的修补，因此先

天性地受到限制。本发明的系统提供一种用于复杂信息结构的本体和分类框架,这是用于实现的实施途径。本发明的系统在其一方面中支持复杂结构,该结构与如今主导信息领域的现有技术中的简单平面式结构不同。

[0081] 本发明的系统支持概念分级作为用于使信息相关的最熟悉和鲁棒模型(术语“多分级”描述一种将维度性和概念分级的核心要求进行组合的结构模型)。然而,本发明的系统在以一个方面中减少使概念分级、分类法和本体构造变得麻烦的个人和协作式协商。还应当提供一种用于对来自不同信息域的分级进行链接的可靠机制。

[0082] 本发明的系统在其一方面中在维度空间内的各种相交处提供结构完整性。这可以通过消除在节点中以及在节点之间的链接和连接中存在的信息空白问题来解决。

[0083] 本发明的系统在其一方面中涉及用以提供重要的上下文认知部件的人类。虽然机器提供用于发现和协作的有用工具,但是机器并不拥有为了“理解”复杂知识而必需的人工智能。这样,本发明的系统在其一方面中以人类熟悉和可介入方式涉及人类。

[0084] 本发明的系统在其一方面中涉及到用以管理巨大信息域中维度结构和概念多分级的极度复杂性、并且用以在概念描述和关系中促成协作者之间共识的机器。

[0085] 本发明的系统在其一方面中在协作中纳入非技术性外行人士。专业设计者的缺乏和问题的范围要求解决方案的通用易访问性。本发明可以使人们避开维度结构的复杂性而无损于他们的技术优势。

[0086] 本发明的系统可操作用以支持海量分布式并行处理(“人多力量大”)。信息领域的规模和复杂性一般对处理具有物理限制,这种物理限制目前看来在实践中不可变的。在许多情况下,大规模和分散式并行性对于挑战这些限制而言是优选的。

[0087] 本发明的系统在其一方面中可操作用以支持能够如下综合操作,这些操作能够避免无界信息和知识的物理限制。本发明的系统在其一方面中提供如下能力:对用于实质上无限数目的数据连接的可能性进行编码,而无需在信息的客户请求这些数据连接之前实际地生成它们。另外,本发明的系统在其一方面中提供各种综合模式,使得仅呈现与客户的设定兴趣和角度相匹配的数据连接。

[0088] 本发明的系统在其一方面中支持和涵盖信息领域的动态性。它提供可以适应信息以及随信息而演变的结构,而不是如在某一时间点处信息的静态快照。

[0089] 本发明的系统是成本有效的。虽然搜索成本提供一种对发现针对信息充斥和信息蔓延的解决方案的巨大激励,但是组织项目没有实行空白检查。对更为结构化的因特网的一个阻碍是,使用现有技术和方法来组织它的庞大成本。这些组织成本不仅是财务性的、而且在人类方面和计算机处理限制上也是与生俱来的。

[0090] 本发明的系统在其一方面中为域所有者和系统终端用户提供一种机会,以便在共享集体才智和汇集式知识资产的益处的同时,维持不同的、私人的和高度个性化知识仓库。

[0091] 本发明在其一方面中提供如下方法和系统,该方法和系统能够管理多个信息形式、包括结构关系、数字媒体如文本和多媒体、消息接发和电子邮件、电子商务以及许多形式的人类交互和协作,并且为终端用户提供一种用以跨各种介质(包括网站和软件客户端)来输出结构信息的分散式系统。

[0092] 更多目的和优点将从后续描述和附图的考虑中变得更清楚。

[0093] 系统操作

[0094] 操作概况

[0095] 图2、图3、图18、图19、图32、图33和图4提供用于构造和管理复杂的维度信息结构以便例如创建用于域的维度概念分类法的操作和系统的概况。具体而言,图2、图3、图18、图19、图32、图33和图4示出了用于此类操作的知识表示模型以及某些维度数据结构和构造。还示出了包括复杂-自适应系统和增强型分面式分类方法的数据结构变换方法。这一描述从特别适用于知识表示的复杂维度结构的简要概况开始。

[0096] 复杂维度结构中的知识表示

[0097] 存在可以用来表示信息和知识的抽象分级的级别。“维度”概念常常用来表达复杂性的程度。简单列表(如购物列表或者好友列表)可以描述为一维数组。表和电子数据表(二维数组)比简单列表更复杂。某些图标可以在三维空间等中描述信息。

[0098] 结构内的每个维度可以建立用于所含信息的组织基础。维度性因此可以建立用于信息结构的复杂性规模。复杂结构可以涉及这些基础中的许多基础,并且常常标识为n维结构。

[0099] 还需要重点注意:维度本身的技术属性可以在结构之间提供大量多样性。例如,维度可以作为变量存在,结构因此建立多变量空间。在这些类型的模型之下,节点可以采用各维度所代表的变量内的具体值或者数据点。备选的是,节点可以不那么严格,仅为除了离散变量之外的信息提供容器。节点之间的距离可以是相对的,而不是严格地加以量化。通过改变这些类型的技术属性,相关联的结构可以在组织刚性与描述灵活性之间达到某种平衡。

[0100] 一些信息结构可以包含每个相交处的节点:其他信息结构可以是不完整的、遗漏一些维度之间相交的节点。这在人工构造信息结构时特别地相关。当结构的复杂性超过人类设计者的认知能力时,可能造成信息结构的错误和空白。

[0101] 作为一个例子,当人们创建网络结构(如万维网)中的超链接时,他们提供的链接在给定域内很少是全面的。如果对于域中的链接而言存在适当目标,但是缺乏该链接,则可以认为这是信息结构中的空白。另外,如果信息结构提供用于信息的分类,但是该信息当前并不存在,则这也是结构中的空白。

[0102] 结构的完整性可以部分地通过信息结构中的空白来描述。如果没有用以管理关系的下层分类系统或者显式的本体,则结构可能开始随着节点和维度的数目增加而退化。信息空白是这一退化的一个标志。

[0103] 复杂结构比简单结构具有远远更多的承载信息容量。正如添加楼层使建筑物的体积增加一样,添加维度使可以包含在结构中的信息量增加。在没有多个维度支持的情况下,随着充斥的信息超过容量,结构最终将在负载之下崩溃。

[0104] 复杂维度结构的另一吸引人的特点在于它们的易访问性。平面式结构将随着信息增加而蔓延,就好像小型建筑物的郊区造成城市蔓延一样。

[0105] 显然,复杂结构的维度性指向对信息充斥和信息蔓延的强制补救。由于其固有优点,人们将预期其增生扩展。遗憾的是,情况并非如此。复杂结构的采用(特别是在最需要它们的普通大众之中采用)一直步履艰难。

[0106] 采用复杂结构受限的原因不言而喻:它们的固有复杂性。尽管存在这些突出的基础和结构问题,仍然已经提出一种足以创建和管理复杂结构而对于大规模市场采用而言足够简单的解决方案。

[0107] 系统方法概况

[0108] 分析和压缩。

[0109] 图2图示了用以构造用于域200的维度概念分类法210的操作,其中域200包括作为分类主题的信息全集。域200可以通过用于向分析和压缩过程204输入的如下源数据结构202来表示,该源数据结构202包括源结构模式和从域200派生的源数据实体集。分析和压缩过程204可以派生语素词典206,该语素词典206是包括用于为新的分面式分类方案提供基础的初等构造集的初等数据结构。

[0110] 域200中的信息可以涉及虚拟或者物理对象、过程以及此类信息之间的关系。作为一个例子,这里描述的操作可以涉及可通过Web页面来访问的内容分类。域200的替代实施例可以包括:文档仓库、用于音乐的推荐系统、软件代码仓库、 workflow模型和商务过程等。

[0111] 语素词典206内的初等构造可以是基本信息构建块和信息关系的最小集,其中该集合体提供用来对源数据结构202进行分类的承载信息的容量。

[0112] 综合和扩展

[0113] 语素词典206可以是对综合和扩展方法208的输入。综合和扩展操作可以将源数据结构202变换成这里称为维度概念分类法210的第三数据结构。术语“分类法”是指如下结构,该结构将类别组织成为分级树,并且将类别与相关对象(如文档或者其他数字内容)进行关联。维度概念分类法210可以在从源数据结构202派生的复杂维度结构中对来自域200的源数据实体进行分类。这样,源数据实体(对象)可以跨过许多不同组织基础来相关,这允许从许多不同角度找到它们。

[0114] 复杂-自适应系统

[0115] 有利的是,分类系统和操作适应于动态环境中的变化。在一个实施例中,通过复杂-自适应系统212来满足这一要求。可以通过与维度概念分类法210进行用户交互,来建立回到源数据结构202的反馈环。变换过程(204和208)可以重复,而所得结构206和210可以随时间来细化。

[0116] 在一个实施例中,复杂-自适应系统212可以管理如下终端用户的交互,这些终端用户使用输出结构(即维度概念分类法210),以便在分类过程中利用人类认知能力。

[0117] 这里描述的操作寻求相对简单地将源数据结构变换成更复杂的维度结构,以便可以用各种方式组织和访问源数据对象。可以通过扩展许多类型的信息系统的下层数据结构的维度性和复杂性来增强这些系统。正如更高分辨率提高图像质量一样,更高维度性可以提高数据结构的分辨率和细节特征性。这一增加的维度性又可以增强数据结构的效用。可以通过改进的、更灵活的内容发现(例如通过搜索)、信息检索的改进和内容聚集来实现增强效用。

[0118] 由于可以通过复杂系统来实现变换,所以维度性的增加并非必然是线性或者可预计的。变换也可以部分地依赖于源数据结构中所含信息量。

[0119] 为了实施针对大型因特网规模的系统,关键区别在于,对于节点和连接的集合呈指数增加,维度信息结构最优地提供如下可能性:直到并且除非需要这些连接,否则不会造成实际地构建这些连接的令人望而却步的成本。

[0120] 维度知识表示模型

[0121] 图3图示了可以在图2的操作中使用的包括知识表示实体、关系和变换方法的知识

表示模型的一个实施例。在下文的描述中将参照图3、图18、图19、图32、图33和图4来描述了知识表示模型及其变换方法的更多细节。

[0122] 在本发明一个实施例中,知识表示实体是内容节点302的集合、内容容器304的集合、概念306的集合(为了简化说明,在图3中呈现仅一个概念)、关键词308的集合和语素310的集合。

[0123] 将要分类的域的对象称为内容节点302。内容节点可以包括服从于分类的任何对象。例如,内容节点302可以是文件、文档、文档块(如注释)、图像或者存储的字符串。内容节点302可以引用物理对象或者虚拟对象。

[0124] 内容节点302可以包含于内容容器304的集合中。内容容器304可以提供可以用来取回内容节点302的可寻址(或者可定位)信息。例如,可通过URL来访问的Web页面的内容容器304可以包含以本和图像形式的内容节点302。内容容器304可以包含一个或者多个内容节点302。

[0125] 概念306可以与内容节点302进行关联,以抽象出一些含义(比如内容节点302的描述、目的、使用或者意图)。单独内容节点302可以被赋予以许多概念306;可以跨许多内容节点302共享单独概念306。

[0126] 可以通过概念与其他实体的关系来按照复合的抽象级、以及在结构上按照其他更基本的知识表示实体(例如关键词308和语素310)来定义概念306。这样的结构在这里称为概念定义。

[0127] 语素310代表:系统已知的(即,为了构造语素词典206而已经对其分析的)跨域存在的最小的有意义的知识表示实体。单个语素310可以与许多关键词308相关联;单个关键词308可以包括一个或者多个语素310。

[0128] 另外,在术语“语素”在本说明书的上下文中的含义与它在语言学领域中的传统定义有所区别。在语言学中,语素是“有意义的最小语言单位”。在本说明书的上下文中,语素是指“在系统已知的任何域中存在的最小的有意义的知识表示实体”。

[0129] 关键词308包括语素310的集合(或者组)。单个关键词308可以与许多概念306关联;单个概念306可以包括一个或者多个关键词308。关键词308因此可以代表在概念306与语素310之间的附加数据结构级。它们有助于“原子概念”作为对用户可识别的最低知识表示级。

[0130] 由于可以从内容节点302抽象出概念306,所以概念签名305可以用来表示概念节点302内的概念306。概念签名305是内容节点302的如下特征,这些特征代表了存在于内容中的组织主题。

[0131] 在本发明的一个实施例中,与初等构造一样,内容节点302趋近它们的最不可缩减形式。内容容器304可以缩减至与实际一样多的内容节点302。当在本发明中与极精细的分类模式组合时,这些初等内容节点302可以扩展用于内容聚集和过滤的选项。因此可以在维度概念分类法中沿着任何维度来识别和重新组合内容节点302。

[0132] 特殊类别的内容节点302、也就是标签(在分类领域中常称为“项”)可以加入到各知识表示实体。与内容节点302一样,可以从相应实体抽象出它们在知识表示模型中描述的标签。因此在图3中标识以下类型的标签:用以描述内容容器304的内容容器标签304a;用以描述内容节点302的内容节点标签302a;用以描述概念306的概念标签306a;用以描述关键

词308的集合的关键词标签308a的集合;以及用以描述语素310的集合的语素标签310a的集合。

[0133] 在图18中呈现了语素310的样本。语素310可以在从源数据派生的初等构造之中。其他初等构造集可以包括语素关系集。正如语素代表概念定义的初等构建块并且是从概念派生的一样,语素关系代表概念之间关系的初等构建块并且是从这样的概念关系派生的。下文将更具体地讨论图9-10中图示的语素关系。

[0134] 标记提供人类可辨别的知识表示实体。在一个实施例中,从源域的唯一词汇表派生每个标签。换言之,从域中呈现的语言和项汲取向各数据元素分配的标签。

[0135] 下文描述并在图7-8中图示了概念、关键词和语素抽象。下文参照输入数据抽象出更具体地讨论概念签名和内容节点以及标签抽象(图5)。

[0136] 本发明的一个实施例跨实体及其关系来使用多级知识表示模型。这使它有别于如图1(现有技术)中所示传统分面式分类中的概念-原子概念二级模型及其平面式的(单级)关系结构。

[0137] 虽然参照一种知识表示模型来描述操作和系统的某些方面,但是本领域普通技术人员将认识到可以通过相应地改造操作和系统来使用其他模型。例如,可以将概念组合在一起以创建更高阶的知识表示实体(比如meme,作为用以包括观念的概念汇集(collection))。也可以收缩表示模型的结构。例如,可以移除关键词抽象层,从而仅针对语素310定义概念。

[0138] 系统变换方法概况

[0139] 图4图示了图2中介绍的变换操作800的一个实施例的广义概况。

[0140] 输入数据提取

[0141] 操作800可以开始于从将要分类的域200的域所有者的人工标识。可以根据域训练集802来定义源数据结构202。训练集802可以是更大域200的代表性子集并且可以用作替代集。也就是说,训练集可以包括用于整个域200的源数据结构202或者代表性部分。训练集在本领域中是公知的。

[0142] 可以从域训练集802提取(804)输入数据集。可以分析输入数据以发现和提取初等构造(下文更具体地讨论图5中所示这一过程)。

[0143] 域分面分析和数据压缩

[0144] 在本实施例中,上文介绍的和在图33中描述的分析引擎204a可以如图4中的括号所示遵循方法806至814。对输入数据进行分析 and 处理(806),以提供元结构分析集。源数据分析可以提供与源数据结构202的结构特征有关的信息。下文更具体地讨论图6中所示这一过程。

[0145] 可以生成(808)初步概念定义集(下文更具体地讨论图7中所示这一过程)。初步概念定义可以在结构上表示为关键词308的集合。

[0146] 可以从初步概念定义中的关键词308提取(810)语素310,由此将概念定义的结构扩展至另一抽象级(下文更具体地讨论图8中所示这一过程)。

[0147] 为了开始构造语素分级402的过程,可以计算(812)潜在语素关系集。可以根据输入数据中的概念关系的分析来派生潜在语素关系。

[0148] 语素结构分析可以适用于潜在语素关系以标识将用来创建语义分级的语素关系。

[0149] 可以组装被选择 (814) 用于包含在语素分级中的语素关系,以形成语素分级402 (下文更具体地讨论图9-15中所示这一过程)。

[0150] 维度结构综合和数据扩展

[0151] 在本实施例中,上文介绍的和在图32中描述的构建引擎可以如图4中的括号所示遵循方法818至820。增强的分面式分类方法可以用来对复杂维度结构210a和维度概念分类法210进行综合 (下文更具体地讨论图20-22中所示这一过程)。

[0152] 可以准备 (818) 用于新维度结构的输出数据210a。输出数据是用于域的分类方案的结构表示。其可以用作分面式数据,以创建维度概念分类法210。如上文所述,输出数据可以包括与内容节点302和关键词分级710相关联的概念定义708。具体而言,分面式数据可以包括关键词分级710的结构和概念定义中的关键词308,其中按照语素词典206的语素310来定义关键词308 (下文更具体地讨论图17中所示这一过程)。

[0153] 可以构造 (820) (聚集形式的多分级中的) 维度概念关系集。维度概念关系表示维度概念分类法210中的概念关系。可以基于增强的分面式分类方法的组织原理来计算维度概念关系。维度概念关系可以进行合并,并且在概念306的分类内 (如概念定义中编码的那样) 可以形成维度概念分类法210 (下文更具体地讨论图20-22中所示这一过程)。

[0154] 对于增强的分面式分类方法而言,可以有各种综合操作模式。在一个实施例中公开一种“范围有限”分面式分类综合操作的系统,在该系统中根据分析引擎方法尚未完全处理或者完全没有处理的域来综合概念关系。在另一实施例中公开一种“动态”分面式分类综合的系统,在该系统中直接地基于为信息的终端用户而提供的综合参数,来近实时地处理维度概念分级 (下文更具体地讨论综合操作模式)。

[0155] 复杂-自适应系统和用户交互

[0156] 在本实施例中,上文介绍的和在图2中描述的复杂-自适应系统212的操作可以如图4中的括号所示遵循与概念分类法210相关联的方法212a、212b和804。

[0157] 如讨论的那样,可以通过表示层608向用户表达维度概念分类法210。在一个实施例中,表示层608是网站 (下文更具体地讨论图23-27和34-36中所示表示层)。经由表示层608,可以将域200中的内容节点302呈现为在与每个内容节点302相关联的概念定义内进行分类。

[0158] 这一表示层608可以提供用于收集用户交互212a的集合作为维度概念分类法信息的环境。用户交互212a可以包括终端用户和域所有者可以与域维度概念分类法210交互的各种方式。用户交互212a可以通过步骤804经由反馈环耦合到分析引擎,以提取用以实现复杂-自适应系统的输入数据 (下文更具体地讨论图27中所示这一过程)。

[0159] 在一个实施例中,可以对显式反馈环中返回的用户交互212a进行排队,以便在资源变得可用时处理。因而,可以提供隐式反馈环。隐式反馈环可以基于增强型分面式分类方法的组织原理的子集,以计算隐式概念关系212b。通过隐式反馈环,可以近实时地处理与维度概念分类法210的用户交互212a。

[0160] 通过复杂-自适应系统212,可以持续地磨练和扩展用于派生维度概念分类法210的分类方案。

[0161] 分面分析方法

[0162] 提取输入数据

[0163] 图5图示了在本发明的一个特定方面中的如下操作900,这些操作包括如参照图4概要讨论的用以提取(804)输入数据的操作及其一些预备步骤。

[0164] 标识结构标记器

[0165] 可以在训练集802内标识(902)结构标记器,以表明可以从训练集何处提取输入数据。结构标记器可以包括源结构模式。结构标记器可以存在于内容容器304中,并且可以包括但不限于文档的标题、与内容相关联的描述元标记、超链接、在数据库中的表之间的关系或者存在于内容容器中的关键词308的普及度。标记器可以由域所有者或者其他标识。

[0166] 可以利用跨域而适用的默认结构标记器来配置操作900。例如,Web页面的URL可以是用于内容节点302的共用结构标记器。这样,可以用如下多个默认结构模式来配置操作902,这些默认结构模式将在源结构模式中的那些区域中不存在任何显式引用时适用。

[0167] 结构标记器可以显式地位于输入数据中,或者可以位于用于输入数据的替代数据中。例如,可以使用在内容节点302中间的关系作为用于概念关系的替代结构标记器。

[0168] 在一个实施例中,可以组合结构标记器以生成关于源结构模式的逻辑推断。如果概念关系在源结构模式中不是显式的,则可以根据结构标记器(例如与内容节点302相关联的概念签名)以及内容节点关系集来推断它们。例如,如进一步所述,概念签名可以是如下文档中的标题,该文档被映射到用于将要定义的概念的替代概念。可以根据内容节点302之间的结构链接(例如连接网页的超链接)来派生内容节点关系。

[0169] 概念签名到内容节点302的连接以及内容节点302到其他内容节点302的连接可以推断相交概念之间的概念关系。这些关系可以形成附加的(显式)输入数据。

[0170] 存在本领域普通技术人员所知的用以标识结构标记器的许多不同方式。

[0171] 将源结构模式映射到系统输入模式

[0172] 可以将源结构模式映射(904)到输入模式。在一个实施例中,输入模式可以包括概念签名906的集合、概念关系908的集合和概念节点302的集合。

[0173] 这一模式设计是变换过程的代表而并非旨在限制。输入操作并不需要跨系统输入模式中每个数据元素的源输入数据以便涵盖很简单的结构。

[0174] 系统输入模式也可以扩展成映射到系统数据变换模式中的每个元素。系统数据变换模式可以对应于变换过程中存在的每个数据实体。也就是说,系统输入模式可以扩展成映射到系统中的每个数据实体。换言之,源结构模式可以包括系统输入模式的子集。

[0175] 此外,域所有者可以映射来自非常复杂结构的源数据模式。作为一个例子,关系数据库的表和属性可以在各种抽象级建模为分面分级,并且映射到系统数据变换模式的多级结构。

[0176] 同样,分析引擎204a和构建引擎208a的操作提供数据结构变换引擎,并且可以在将一类复杂数据结构(比如在关系数据库中建模的复杂数据结构)变换成另一类复杂数据结构(通过这里描述的方法和系统所产生的复杂维度结构)时实现显著的新效用。产品目录提供了从这一类复杂数据结构到复杂数据结构的变换中受益的复杂数据结构的一个例子。下文提供关于图30中所示例子数据变换模式的更多信息。

[0177] 提取输入数据

[0178] 可以对训练集应用输入数据映射,以将其源结构模式映射到输入模式,从而提取(804)输入数据。如本领域中所知,本发明的一个实施例使用XSLT对用来从源XML文件提取

数据的数据映射进行编码。

[0179] 提取方法因许多因素而变化,这些因素包括源结构模式的参数和结构标记器的位置。例如,如果概念签名是精确的(如文档标题、基于关键词的元标记或者数据库关键词段那样),则可以直接使用签名来表示概念标签。对于较为复杂的签名(比如文档本身中的关键词的普及度),则可以使用普通文本挖掘方法。一种简单方法使关键词提取基于文档中的最普及关键词的简单计数。在信息提取和文本挖掘的广泛领域内,存在普通技术人员已知的许多其他提取方法。

[0180] 一旦被提取,则输入数据可以存储于耦合到分析引擎204a的一个或者多个存储装置中。为求便利,这里包含的图和描述引用数据存储器910作为存储装置,但是可以使用其他存储器。

[0181] 例如,可以使用域数据储存设备706,特别是在计算环境是受控环境的情况下。

[0182] 系统输入数据可以拆分成它们的组成集,并且传递到变换引擎中的后续过程。

[0183] 概念关系是用于下文描述的和在图6中图示的源结构分析的输入(A)。

[0184] 可以处理概念签名,以提取下文描述的和在图7中图示的初步概念定义(B)。

[0185] 可以将内容节点作为下文描述的和在图17中图示的系统输出数据来处理(C)。

[0186] 如上文所述从源数据结构提取输入数据是可以用于提取输入数据的许多实施例之一。在一个实施例中,通向分析引擎204a的其他初步输入通道包括复杂-自适应系统的反馈环。这样,返回用户交互212a (O) 以提供更多输入数据。下文描述图27中所示这一输入数据通道和包括复杂-自适应系统的反馈环的细节。

[0187] 处理源数据结构

[0188] 图6示出了在本发明的一个特定方面中用以提取源结构分析的源数据结构处理。源数据结构分析可以提供涉及源数据结构的拓扑的数据。源数据的拓扑是指对其形状进行描述的源数据结构的技术特征集(特征例如结构中所含节点的数目以及源数据结构中的节点之间关系的分散模式)。

[0189] 这一分析方法的一个主要目的在于度量概念306的通用或者专用(关于训练集802中的其他概念306)的程度。这里,将概念的相对通用性或者专用性的度量称为“通用性”。下文描述在一个实施例中分析的源数据特征。关于分析和特征的细节随源数据结构而变化。

[0190] 可以组装概念关系908以供分析。可以标识(表明存在非分级关系)并且求解概念306之间的循环关系1002。

[0191] 可以从集合1004删减(prune)被系统标识为非分级的所有概念关系。在后续处理中没有涉及到删减的概念关系,但是可以使这些概念关系可用于基于不同变换规则的处理。

[0192] 可以将未删减的概念关系作为分级关系来处理。系统可以将这些概念关系1006组装成被排序为扩展间接关系集的所有分级概念关系的输入概念分级1008。组装输入概念分级1008可以包括到对集合体中的节点进行排序,并且移除可以从其他关系集推断的任何冗余关系。输入概念分级1008可以包括多分级结构,其中实体可以具有多个直接父实体。

[0193] 一旦被组装,则如在以下步骤中所述,输入概念分级1008可以包括用于度量概念关系集中概念306的通用性的结构,并且可以用于变换过程中的其他方法。如下文所述和图9-10中所示,输入概念分级1008中的概念关系可以用来计算潜在语素关系(D)。如下文所述

和图17中所示,输入概念分级中的概念关系也可以用来处理用于系统的数据(E)。

[0194] 对输入概念分级的分析可以进行至度量各概念1010的通用性。同样,通用性是指任何给定节点相对于分级1008中的其他节点如何通用或者专用。各概念3006可以基于它在输入概念分级1008中的位置来进行通用性度量评估。

[0195] 可以计算每个概念308同与概念306相交的树中的每个根的加权平均分离度。加权平均分离度是指各概念306与在根节点处的概念306的距离。明确地为根节点的概念306被赋予唯一的通用性度量。通用性度量针对更具体的概念306而增加,这反映它们与驻留于根节点的最通用概念306的增加分离度。本领域技术人员将认识到许多其他通用性度量是可能的。

[0196] 用于各概念306的通用性度量可以存储于概念通用性指数1012中(例如数据储存器910中)。如下文所述和图12-图13中所示,概念通用性指数1012可以用来推断用于语素的一般性度量集(F)。

[0197] 在一个实施例中描述的方法可以适用于也称为父-子关系的分级型关系。父-子关系涵盖其可以支持的大量的各种关系类型。例子包括:整体-部分、基因-物种、类型-实例和类-子类。换言之,通过支持分级型关系,本发明适用于广泛的分类任务。

[0198] 处理初步概念定义

[0199] 图7图示了用以生成初步概念定义的关键词提取方法。这一过程的一个主要目的在于,按照关键词308来生成用于概念306的结构定义。在一个实施例中,在这一阶段可以将概念定义描述为“初步”是因为它们将在以后阶段中受到修正。

[0200] 本领域普通技术人员将认识到,对于涉及提取关键词308作为概念306的结构表示这一目标,存在许多方法和技术。

[0201] 在一个实施例中,可以限制适用于关键词提取的抽象级。可以设计这些限制以派生具有以下性质的关键词:关键词是响应于直接关系集内词的独立性、使用(基于原子概念提取的)原子概念(其中概念存在于训练集的其他区域中)来定义的。

[0202] 可以汇集概念签名906和概念关系908以供分析。在一个实施例中,这一过程是基于文本实体的提取。这样,在以下描述中,可以假设概念签名906直接地映射到向概念306分配的概念标签。

[0203] 当在概念标签906中标识标签时,可以提取文本串的相关部分,并且使用该相关部分作为概念标签306a。在后续方法中,当在概念306中标识关键词308和语素310时,可以从概念标签306a的相关部分提取用于关键词308a和语素310a的标签。

[0204] 最终,这些域特有的标签可以写入到输出数据。如果操作800变换先前已经分析和分类的数据结构,则实体标签可以直接在源数据结构中可用。

[0205] 注意,在概念签名与概念标签提取之间的这一接合点表示用于各类实体提取工具的整合点,其涉及许多类型的内容节点302,如图像、多媒体的广泛各种实体提取工具与物理对象分类。

[0206] 可以在概念标签中标识一系列关键词描绘符。可以基于关键词308的共用结构描绘符(比如括号、引号和逗号)、根据概念标签306a来解析(1102)初步关键词范围。然后再次使用普通字描绘符(比如空格和语法符号)根据初步关键词范围来解析(1104)全部的字。用于文本实体解析的这些基于模式的方式在本领域中是公知的。

[0207] 来自初步关键词范围1102的解析词可以包括用于关键词提取过程中的下一阶段的一个输入集。其他输入集可以是直接概念关系集1106。可以根据概念关系908的集合来派生直接概念关系集1106。直接概念关系集1106可以包括用于每个概念306的所有直接关系(所有直接父关系和所有直接子关系)。

[0208] 这些输入用来检查初步关键词范围1108中词的独立性。直接关系集1106内单个词的独立性可以包括用于关键词308的描绘符。在描绘关键词范围之后,可以执行校验以保证所得出的关键词308的所有部分有效。具体而言,被描绘为关键词308的概念标签306a的所有部分最优地通过词独立性测试。

[0209] 在一个实施例中,可以基于下文称为“词干化(stemming)”的词干(或称词根)匹配方法来执行针对词独立性的校验。本领域中公知许多词干化方法。如下文在图8中所示语素提取方法中所述,词干化为分类提供极为精细的基础。

[0210] 基于初步关键词范围中的词独立性,可以标识潜在关键词描绘符1110的附加集合。简言之,如果一个词在一个概念标签306a中与其他词一起存在、而在相关概念标签306a中缺乏这些相同的词,则该词可以描绘关键词。

[0211] 然而,在基于这些关键词描绘符将概念标签306a解析成关键词标签308a之前,可以验证(1112)候选关键词标签。一般要求所有候选关键词标签通过上述词独立性测试。这一校验防止关键词提取过程分裂在目标抽象级以外的概念(即原子概念)。

[0212] 一旦生成关键词标签的初步集合,系统可以检查集合中的所有初步关键词标签。这里的意图在于标识复合关键词1114。符合关键词可以作为单个概念标签306a内的多个有效关键词标签而存在。这一测试可以直接基于作为概念-关键词抽象范围的原子概念目标。

[0213] 在一个实施例中,可以使用递归来将复合关键词的集合穷尽地拆分成由训练集802支持的关键词308的最初等集合。

[0214] 如果复合关键词保留于关键词标签的演变汇集,则可以生成(1110)潜在关键词描绘符的附加集合,其中使用匹配关键词来对描绘符进行定位。同样,可以将描绘的关键词范围作为有效关键词来校验,提取关键词,并且该过程重复直至无法找到更多复合关键词。

[0215] 该方法的最后一轮联合可以用以跨过整个域来消除关键词标签的不明确性。消除不明确性在本领域中是一个公知的要求,并且它有许多方式。一般而言,消除不明确性用来解决在实体共享相同标签时显现的不明确性。

[0216] 在一个实施例中,可以通过将关键词联合为共享同一标签的单个结构实体来提供一种消除不明确性的方法。具体而言,如果关键词共享标签和相交的直接概念关系集,则可以存在用于关键词标签、将其与单个关键词实体进行关联的基础。

[0217] 备选地,可以放宽这一消除不明确性的方法的限制。具体而言,通过移除相交的直接概念关系集的标准,域中的所有共享关键词标签可以联合到相同关键词实体。在域相对较小或者非常汇集于它的主题内容时,这是一种有用的方式。备选地,可以通过直接和间接概念关系的更广泛谱系来改变在这一消除不明确性方法中使用的概念关系集。许多消除不明确性的方法在本领域中是已知的。

[0218] 这一关键词提取方法的结果可以是抽象至“原子概念”级的关键词集合1118。将关键词与派生它们的概念306相关联(1120),作为初步概念定义708a。以后可以扩展这些初步

概念定义708a以在它们的结构中包括语素实体,从而扩展为更深和更基本的抽象级。如下文进一步所述,可以进一步扩展这些初步概念定义,以利用由输入数据中的概念关系来表明的关键词和语素的隐式属性。

[0219] 从这一过程派生的实体708a可以传递到本公开内容中描述的变换引擎中的后续过程。初步概念定义708a是向下文描述和在图8中图示的语素提取过程(G)以及下文描述和在图17中图示的输出数据过程(H)的输入。

[0220] 提取语素

[0221] 在传统分面式分类中,用于分面的属性一般可以限于可以使用人类认知来标识、并且与其他概念相关联的概念。结果,属性可以视为原子概念,因为属性构成概念而无需更深的上下文。

[0222] 这里描述的方法可以跨过大型数据集使用统计工具,以标识初等的(语素性的)不可缩减的概念属性及其关系。在这一抽象级,许多属性将无法由人类分类者识别为概念。

[0223] 图8图示了可以用来解析语素310并且将语素与关键词308进行关联以扩展初步概念定义708a的方法。该语素提取方法可以从上文描述和在图7中图示的生成初步概念定义的方法继续。

[0224] 注意,在一个实施例中,语素提取方法可以具有与关键词提取方法公共的元素。这里,为对话素提取进行较为粗略的处理而提供描述,其中这些方法于此重叠。

[0225] 关键词的池库1118和直接概念关系1116的集合可以是对这一方法的输入。

[0226] 可以定义模式以用作用于标识语素候选1202的标准。如本领域公知的,这些模式可以建立用于词干化的参数,并且可以包括用于整个词以及部分词匹配的模式。

[0227] 与关键词提取一样,直接概念关系1106的集合可以提供用于模式匹配的上下文。可以在出现关键词的直接概念关系集内针对关键词的池库1118来应用模式(1024)。可以标识(1206)基于词干化模式的共享根集。共享根集可以包括用于每个关键词的候选语素根1208的集合。

[0228] 可以对用于每个关键词的候选语素根以保证它们相互一致(1210)。可以假设,在同一关键词的上下文和出现该关键词的直接概念关系集内驻留的根具有重叠根。另外,假设从这些重叠根的相交派生的初等根将保留在用来标识有效语素的参数内。

[0229] 这一验证校验可以提供一种用于在应用模式匹配以标识潜在语素时校正存在的错误(词干化方法的一个通病)的方法。更重要的是,验证可以约束过度的语素拆分并且可以提供上下文有意义的而仍然是基本的抽象级。

[0230] 在一个实施例中设计的对话素和关键词提取的一系列约束还可以在复杂-自适应系统的背景内提供负反馈机制。具体而言,这些约束可以起到消除复杂性以及在用于分类的设置参数内对其进行管理的作用。

[0231] 通过这一语素验证过程,可以从关键词集合移除任何不一致的候选语素根(1212)。可以重复模式匹配以标识语素候选的过程直至移除所有不一致的候选。

[0232] 一致语素候选集可以用来派生与关键词相关联的语素。与关键词提取方法一样,描绘符可以用来提取语素(1214)。通过检查潜在根的群组,可以针对每个关键词来标识一个或者多个语素描绘符。

[0233] 可以基于描绘符在各关键词标签内的位置来提取语素(810)。更重要的是派生一

个或者多个语素实体以向关键词提供结构定义的过程。可以通过将语素与派生它们的关键词相关(或者映射)来构造关键词定义(1216)。这些关键词定义可以存储于域数据储存设备706中。

[0234] 可以基于语素类型(例如自由、受限、屈折或者派生)(1218)。在构造过程的以后阶段中,用于构建概念的规则可以基于涉及到的语素类型以及这些语素是否受限于其他语素而变化。

[0235] 一旦被确定类型,则提取的语素可以包括域1220中所有语素的池库。这些实体可以存储于系统的语素词典206中。

[0236] 可以维护每个语素标签的持久详细目录以用来通知将来多轮语素解析(关于更多信息,参见上文图33中所示数据结构变换的概述)。

[0237] 如下文所述和在图9-图10中所示,从这一过程派生的语素可以传递到变换引擎中的后续过程,以处理语素关系(I)。

[0238] 本领域技术人员将认识到,存在可以用来发现和提取包括语素的关键词定义的许多算法。

[0239] 计算语素关系

[0240] 语素可以提供对系统的多级分面式数据结构进行锚定的一个初等构造集。其他初等构造可以是语素关系。如上文讨论的和图3、图18-图19中图示的那样。语素关系提供用于创建维度概念关系的有力基础。

[0241] 然而,挑战在于在分类数据中存在的不明确性噪声中真实地标识语素性的语素关系。本发明的多级结构提供一种针对这一挑战的解决方式。通过跨多个提取集来验证关系,相继地削减不明确性。

[0242] 以下部分解决了发现语素关系。具体而言,在本发明的这一特定方面中,模式扩充方法用来去掉噪声以增强初等构造的统计标识。

[0243] 潜在语素关系概述

[0244] 图9图示了用于根据训练汇集的概念关系来推断潜在语素关系的方法。

[0245] 可以计算潜在语素关系以检查所有概念关系的集合体中的独立潜在语素关系的普及度。基于这一检查,可以应用统计测试来标识如下候选语素关系,这些候选语素关系在其中存在它们的所有概念关系的背景中成立的可能性高。

[0246] 在本发明的系统的一个实施例中,可以将潜在语素关系构造为可以在相关概念中的语素之间存在的所有关系排列,其中保留关系的父-子方向性。

[0247] 在图9中的例子中,输入概念分级1008的一部分示出了两个概念之间的关系。父概念及其有关子概念可以分别包含语素{A,B}和{C,D}。

[0248] 同样,可以按照(在一个实施例中经由关键词来分组的)一个或者多个语素来定义概念。结果,在两个概念之间的任何关系将意味着在定义这些概念的语素之间的至少一个(并且常常是多个)关系。

[0249] 在这一例子中说明了计算潜在语素关系的过程。可以根据单个概念关系来推断四个潜在语素关系812a。维持通过概念关系来建立的父-子方向性并且不允许任何重复,可以派生四个潜在语素关系:A.C、A.D、B.C、B.D。

[0250] 一般而言,如果父概念包含x个元素而子概念包含y个元素,则将存在 $x \times y$ 个潜在

语素关系:潜在语素关系的数目是父概念和子概念中的语素数目的乘积。

[0251] 在一个实施例中,可以对计算语素关系的这一简单说明进行细化,以改进所生成的统计指示符。下文在对图10所示潜在语素关系计算方法的描述中指出了这些细化(即,排列语素)。

[0252] 对标识潜在语素关系的基本方法的这些细化可以用以减少潜在语素关系的数目。这一减少又可以减少噪声数量,由此扩充对语素关系进行标识的模式,并且使语素关系的统计标识更可靠。

[0253] 同样,本领域普通技术人员将认识到,存在可以用来根据给定的概念关系集来派生潜在语素关系的许多算法。

[0254] 计算潜在语素关系的方法

[0255] 图10更具体地给出了计算潜在语素关系的过程的一个实施例。

[0256] 这里的意图在于生成如下潜在语素关系集,以后可以分析该潜在语素关系集,以评估它们本质上真正是语素性(也就是说,它们在其出现的每个上下文中成立)的可能性。

[0257] 计算潜在语素关系的当前方法从上文描述和在图6中图示的资源结构分析方法D继续。

[0258] 该方法也从如上文描述和在图8中图示的语素提取方法I扩展。

[0259] 对确定潜在语素关系的这一方法的输入可以是域提取的语素池库1220和包含来自域的概念关系验证集的输入概念分级1008。

[0260] 可以校准每个概念关系对内的语素(1404),以减少可以推断的潜在语素关系的数目。具体而言,如果校准两个数据元素,则这些元素可以与同一概念关系对内的任何其他元素组合。通过校准,可以减少候选语素关系的数目。

[0261] 在一个实施例中,轴可以基于共享的语素来校准,并且包括限于共享语素的所有语素。例如,如果一个概念是“加拿大政治”而另一概念是“国际政治”,则可以使用关键词“政治”中的共享语素作为用于校准的基础。

[0262] 也可以基于语素词典内的现有语素关系来校准轴。具体而言,如果可以通过语素词典中的语素关系来表示(使用语素关系集来直接或者间接地构造)任何给定的潜在语素关系,则可以在这一基础上校准潜在语素关系。

[0263] 外部词典(图10中未示出)也可以用来指导潜在语素关系的校准。例如WORDNET™是可以适用于校准的词典。可以使用外部词典内包含的各种信息作为用于指导的基础。在一个实施例中,可以先通过言语部分对关键词进行分组;约束潜在语素关系以便仅在这些语法分组内进行组合。换言之,如外部词典所指导的那样,校准可以基于言语的语法部分。还可以使用可以根据外部词典来推断的直接语素关系作为校准的基础。

[0264] 可以计算潜在语素关系(812)作为已校准集合中未涉及到语素的所有组合。上文描述并在图9中图示了这一计算。

[0265] 潜在语素关系1406的所得集合可以保持于域数据储存设备910中。这里,潜在语素关系的详细目录可以在它们存在于训练汇集时加以跟踪,并且通过后续分析阶段来删减。

[0266] 根据这一过程来派生的潜在语素关系可以传递到如下文描述的和在图11-图13中图示的用于删减和语素关系组装(J)的过程。

[0267] 删减潜在语素关系

[0268] 可以将通过上文描述并在图9-图10中图示的方法所生成的潜在语素关系池库删减成候选语素关系集。

[0269] 可以基于潜在语素关系在训练汇集的普及度的评估来删减潜在语素关系。普及度高的那些潜在语素关系真正作为语素性(也就是关系在每个上下文中成立)的可能性更大。

[0270] 此外,可以假设语素关系在它们与更多通用(更广义)的相关语素的关系上是明确的。用于这一不明确性的结构标记器可以是多分级。语素关系可以将更少属性具体化并且提供用于相关语素的更多确定基础。这样,还可以在潜在语素关系存在于多分级中时对其进行删减。

[0271] 可以通过也是分级的语素关系对的集合来构造语素关系分级。这样,可以在集合体中分析潜在语素关系池库以标识与这一分级假设矛盾的关系。

[0272] 在这一删减过程之后留存的候选语素关系可以组装成语素分级。虽然候选语素关系是父-子配对,但是语素分级可以扩展到多代父-子关系。

[0273] 图11A和图11B图示了在潜在语素关系与减除后的候选语素关系集之间的差异。

[0274] 在图11A中存在分级的四个潜在语素关系对(父-子)。这些关系中的前三个关系在域中相对地普及,但是第四个关系相对地少见。因而,从潜在语素关系集中减除第四对。

[0275] 潜在语素关系1406的集合中的前三个关系对也与分级假设一致。然而,双向的第五关系1502与这一假设冲突。关系D.C的方向与关系C.D的方向冲突。这一语素对被重新归类为通过关联关系而相关,并且从候选语素关系1504的集合中被移除。图11B示出了减除后的候选语素关系集。

[0276] 组装语义关系

[0277] 合并语义关系

[0278] 图12图示了将候选语素关系联合为整体语素多分级。所有候选语素关系对可以并入到一个全集合中,该集合连接在逻辑上一致的世代树(generational tree)(如下文更具体地所述)。

[0279] 这一数据结构可以描述为“多分级”,因为它可以造成在与更多通用语素(多个父节点)的一个以上的直接关系中涉及到的奇异语素(singular morpheme)。这一多分级可以在该过程的后续阶段中变换成严格分级(仅单个父节点)。

[0280] 在(上文描述的和在图11B中图示)冲突删减过程之后留存的潜在语素关系可以汇集到候选语义关系1504的集合中。候选语素关系集可以合并成整体语素多分级1602。

[0281] 在一个实施例中,对构造整体多分级的过程的约束可以是:1)多分级中的候选语素关系集在集合体中逻辑上是一致的;2)多分级使用创建逻辑上一致的结构所需的最少数目的多分级关系。

[0282] 递归排序算法可以用来组装树并且突出冲突和提出的解决途径。适用于下例的推理说明了这一算法的逻辑。

[0283] 基于关系分级#1,A比C高级(也就是更通用)。基于分级#2,B比C高级。基于分级#3,A比D高级。四个语素可以与比C高级的A和C、以及比D高级的A在逻辑上组合。

[0284] 当可能不止一个逻辑排序,概念通用性指数1012可以用来解决不明确性(通过上文描述的和在图6中图示的源结构分析方法来创建概念通用性指数)。这一指数可以用来比较语素以评估语素比其他语素相对更通用还是更专用(按照从跟节点起的分离度来度量

通用性)。

[0285] 在该例中,A和B基于候选语素关系均为逻辑上一致的最高节点。A和B也均为C的父节点。因此,可以在C处生成多分级的关系集。由于在样本集中没有与多分级的关系集冲突的信息,所以可以假设关系有效。处理可以继续后续阶段中分解多分级。

[0286] 如果存在如下新数据,该数据表明A和B代之以是经过间接关系的相关节点,则系统可以立即分解多分级并且将A和B排序于同一树中。可以通过通用性指数来确定A和B优先级。这里,A具有低于B的通用性排名。因此在所得多分级1602中对其给予更高的(或者更通用的)位置。

[0287] 语素多分级组装

[0288] 图13图示了可以用于从候选语素关系组装语素多分级的方法。

[0289] 可以通过分析集合体中的候选语素关系来组装语素多分级。与在输入概念分级组装中一样,目的在于将独立的关系对联合成统一的整体。

[0290] 语素关系组装方法可以从上文描述和在图9-图10中图示的计算潜在语素关系的方法J继续。

[0291] 潜在语素关系的集合1406可以是对这一方法的输入。可以基于包含语素的概念关系的分析来对候选语素关系进行排列(1702)。可以基于各概念关系对中的语素的总数来对概念关系进行排列(最低到最高)。

[0292] 随着概念关系对中包含的语素的数目减少,语素关系的可能性可以增加(因为按照该对中的潜在候选的数目将针对任何给定语素关系候选的概率因数化)。因此在一个实施例中,操作可以以更低语素计数的分析来区分概念关系的优先次序。降低该对中的语素的数目,则可以增加发现真正语素性的语素关系的机会。

[0293] 可以设置用以限定语素关系的统计相关边界的参数(1704)。这些参数可以是基于集合体中的语素关系的普及度。目的在于标识域中流行度高的语素关系。对语素关系的这些约束也可以对复杂-自适应系统的负反馈机制起作用。可以对集合体中的关系集进行分析(1706)以确定各关系的总普及度。这一分析可以组合在由系统管理员控制的灵敏度参数内进行的统计工具。确切的参数可以针对各域来定制,并且可以由域所有者和系统管理员来改变。

[0294] 与概念关系分析一样,可以使用循环关系(1708)作为结构标记器来否定分级关系的假设。如果潜在语素关系没有通过普及度和分级的过滤器则可以将其删减(1710)。

[0295] 删减后的潜在语素关系集可以包括候选语素关系的集合1504。如在概念通用性指数1012中具体化的那样,可以根据源结构概念的通用性来推断语素的通用性(1010a)。

[0296] 可以使用将最低数目的语素具体化的概念作为用于各语素的通用性的替代概念。为了说明这一假设的基础,假设概念包括仅一个语素。在概念与包括它的单个语素之间的高相关程度的给定情况下,有可能的是语素的通用性将与概念的通用性紧密地相关。

[0297] 这一推理在一个实施例中指导语素通用性的计算。具体而言,系统可以收集将集合体中的最低数目的语素具体化的概念集。也就是说,系统可以选择代表集合中所有语素的概念集。

[0298] 概念通用性指数1012可以用来区分维度概念关系的优先次序,并且可以存储(未示出)于域数据储存设备706中。

[0299] 使用如上文描述的和在图12中图示的方法,语素分级可以组装到整个多分级结构(1712)中。这可以包括对集合体中的节点进行排序,并且移除可以根据其他间接关系集来推断的任何冗余关系。创建的概念通用性指数可以从最通用到最专用,来对语素进行排序。

[0300] 本领域普通技术人员将认识到,存在本领域中已知的可以用来将分级语素关系汇集合并成多分级的许多算法。

[0301] 组装语素分级

[0302] 图14-图16图示了将语素多分级变换成语素分级。

[0303] 语素多分级归因(attribution)

[0304] 图14A-图14B图示了语素归因过程和例子结果。在此上下文中的归因是指对分面属性进行排序并且将其分配给数据元素的方式。正如操作对实体提取(比如关键词和语素提取)设置约束一样,可以使用对话素关系的显式约束来构建语素分级。

[0305] 按照定义,将语素链接成为分级的语素关系是语素性的。语素实体是初等和明确的。一般要求语素仅涉及一个父节点。在语素关系集(语素分级)中,语素可以仅在一个位置中存在。

[0306] 基于一个知识表示模型中的这些定义,可以在语素数据分面分级内将语素呈现为属性。知识表示模型因此可以提供分面式数据和多级增强型分面式分类方法。

[0307] 在前述方法中,候选语素关系的聚集可以呈现语素多分级的集合1802。因此,归因可以用来估算知识表示模型中的这些冲突并且给出解1804。

[0308] 在一个实施例中的归因方法可以包括为分级中与语素分级要求没有冲突的每个语素找到位置。

[0309] 多分级中的语素可以在它们的原有树内上升到新位置或者移动到全新的树。这一归因过程最终定义分面分级中的最高根语素节点。因此,可以将语素分级中的根语素节点定义为语素分面,其中各语素包含于语素分面属性树内。

[0310] 以下讨论说明了用于使用属性概念来移除多个父节点的方法。

[0311] 同样,用于冲突的结构标记器可以是语素多分级1802中出现的多个父节点的存在。为了移除冲突,可以将具有多个父节点的语素重新视为共享父节点的祖先的属性。

[0312] 可以创建属性类以维持由重新组织的语素原来共享的父节点的分组,并且将语素保持于与这些父节点分离的属性类中(在没有唯一祖先的情况下,该方法将语素升级为分级的根级作为新的语素分面)。

[0313] 可以将关系重新组织成从根节点到叶节点的属性类。可以先将多个父节点重新组织成属性,从而可以标识奇异父属性。也就是说,语素关系的自上而下的遍历提供了可以分解至解集1804的归因。

[0314] 一般而言,如果两个语素共享至少一个父属性,则它们在该共享父节点的上下文中是兄弟(关联关系)。可以在单个属性类之下对兄弟子节点进行分组(注意子节点只需共享一个父节点;它们无需共享所有父节点)。如果语素没有共享至少一个父节点,则可以将它们分组为共享祖先的单独属性。

[0315] 为了在两种备选方式之间选择,可以对源关系的相关度进行加权。上文在图6中所示源结构分析的讨论中介绍了关系相关度的度量。

[0316] 在上而下地开始,变换步骤可以分解如下:

[0317] 1. 兄弟组 {B,C,D,F,H} 共享单个父节点。校验各独立节点,以查看是否存在多个父节点。在这一情况下,这些节点无一具有多个父节点,因而无需重新组织这些关系。

[0318] 2. 语素E具有多个父节点。E的最近的单亲祖先是A。需要将E重新组织成A的属性。

[0319] 3. 在属性类A1之下对 {B,C,D,F,H}、E的父属性进行分组。E然后变成A1的兄弟,作为A的属性。

[0320] 4. 语素G也具有多个父节点。与在步骤(2-3)中一样,需要将它重新组织成A的属性。此外,由于E和G共享至少一个父节点,所以可以将其分组在单个属性类A2之下。

[0321] 5. 语素J具有唯一父节点H。无需重新组织这一父-子关系。

[0322] 6. 语素K具有多个父节点E和G。E和G的唯一祖先现在为A2。需要将K重新组织成A2的属性。

[0323] 7. 在属性类A2-1之下对 {E,G}、K的父属性进行分组。K然后变成A2-1的兄弟,作为A2的属性。

[0324] 最终结果是与通过本发明的知识标识模型来定义的真正语素属性和语素关系的假设相符的语素分级。

[0325] 语素分级重新组织

[0326] 图15给出了在一个实施例中可以提供归因方法的递归算法。这一语素分级重新组织的核心逻辑可以是上文描述和在图14A和图14B中图示的归因方法。

[0327] 用于这一方法的输入可以是上文描述并在图11-图13中图示的语素多分级(K)。向本方法的输入可以是语素多分级1602。可以从根节点到叶节点对关系进行分类(1902)。可以针对多个父节点来校验语素多分级中的每个语素。这里,将作为分析焦点的语素称为活跃语素。

[0328] 如果存在任何多个父节点,则可以将用于活跃约束的多个父节点的集合分组成下文称为语素属性类的集合(1906)。可以使用语素属性类来指导应当如何对重新组织的树中的语素进行排序。

[0329] 对于每个语素属性类,可以定位没有多个父节点的唯一祖先(1908)。祖先可以仅与属性类(由语素共享的父节点组)唯一地关联。

[0330] 如果存在祖先,则系统可以创建一个或者多个虚拟属性(1910)以包含语素属性内的所有语素。树中的这一节点称为“虚拟属性”是因为它没有直接地与任何语素相关联,因此不会包含在任何概念定义中。它是虚拟属性而不是实际属性。

[0331] 如果存在祖先并且创建一个或者多个属性,则可以将活跃语素重新组织成祖先的属性(1912),其或者与祖先直接相关,或者与语素属性类中的其他语素分组在一起。

[0332] 如果不存在唯一祖先,则可以将语素重新定位成树中的根节点(分面)(1914)。

[0333] 系统也可以允许管理员人工更改(1916)语素关系池库和所得语素分级以细化或者替换自动生成的结果。

[0334] 这一过程的最终结果可以是包括初等语素分级布置的语素分级402。作为系统的数据结构的初等构造之一,语素分级可以用来将实体分类和布置到越来越复杂的抽象级中。

[0335] 可以在语素词典206中录入语素分级中的语素关系。可以基于系统中存储的标签普及度将语素标签分配给语素。可以使用系统中最普及的语素标签作为用于该语素的单个

代表性标签。

[0336] 如下文所述和在图17中所示,可以对该方法的输出进行处理作为系统输出数据(L)。

[0337] 可以使用将多分级变换成严格分级的替代方式。可以基于多个加权因子中的任何加权因子来选择单个父节点,以移除多个父节点的情形。在一个简单解决方案中,可以删除多个父节点的关系。

[0338] 图16A图示了来自组装的语素分级的样本树片段。树(例如2002a)中的各节点可以代表语素分级中的语素。文件夹图标用来表明如下语素,这些语素是下面嵌套的相关语素的父节点(语素关系)。在各节点旁边的文本(例如2002b)是相关联语素标签(在许多情况下为部分的词)。

[0339] 分面式分类综合的方法

[0340] 这里开始基于增强的分面式分类方法来构建(或称综合)维度概念分类法210的过程。这一分类可以通过利用概念定义集(具体而言按照语素来定义,其中零个或者更多语素作为语素分级内的语素属性)检查语素分级来生成维度概念关系。

[0341] 可以在多个数据提取级应用本发明的分面式分类方法。以此方式,多个域可以在维持对域特有的边界之时共享用于分类的相同初等构造。

[0342] 处理分面式数据集

[0343] 以下几点总结了根据分析操作来准备输出数据以用于对分面式分类数据结构进行综合(如下文进一步所述)的一个方面中所包括的步骤:

[0344] 对于将要分类的各域,可以输出数据结构作为域特有的关键词分级和域特有的概念定义集合(具体而言,按照域特有的关键词来定义,零个或者更多域特有的关键词作为域特有的关键词分级内的关键词属性)。

[0345] 可以根据跨域共享的初等构造来派生上述域特有的分面式数据。可以用新信息来修正和明显地扩展初步概念定义。这是通过将语素分级中的信息与训练集中的原有概念关系做比较来实现的。

[0346] 具体而言,综合操作可以不仅基于由域所有者提供的显式定义的分析而且通过集合体中所有相交概念和概念关系的分析,来向内容节点分配概念定义。可以分配“显式”属性的初步定义,以后用与内容节点相交的概念关系“暗示”的丰富的多的属性集来补充该初步定义。

[0347] 候选语素关系可以组装成整个语素分级,该语素分级将用作用于分面式分类的数据内核。可以根据各域中的关键词及其语素的唯一相交来创建用于各域的单独分面分级。这一数据结构可以表达限于域边界的语素分级。

[0348] 分面分解可以在域的词汇表(它的唯一关键词集)中表达,并且可以仅包括被因子化到域中的那些语素关系。可以输出用于各域的分面式分类作为用于该域和分面分级的概念定义集。

[0349] 因此在一个实施例中,可以根据汇集式语素分级来推断域特有的分面分级。它可以针对更小的域提供更丰富的分面集。它可以建立在多个域的共享经验(可以纠正在更小的域中存在的错误)上,并且它可以有助于域的较快处理。

[0350] 在另一实施例中,系统可以直接地基于图14-图15中所示上述方法为域创建唯一

分面分级。在这一实施例中,属性分级组装过程可以直接地适用于从各域提取的域特有的关键词。

[0351] 在又一实施例中,综合操作可以基于从其他传统分类手段汇集的数据。这样的分类手段可以包括为传统分面式分类综合而准备的分面式数据以及如形式概念分析中使用的严格属性集来限定的概念。这些和其他补充分类方法是本领域技术人员公知的。

[0352] 图16A-图16B图示了如在一个实施例中派生的来自组装的语素分级2002(如上所述)的树的片段和来自域特有的关键词分级2004的树的片段。注意到,在用于关键词分级2004的树片段中,在代表关联关键词标签的每个节点(例如2004b)旁边的文本是完全的词,因为它们将存在于域中。另外,用于关键词分级2004的树分段可以是用于语素分级2002的树片段的如下子集,该子集被收缩成仅包括与派生关键词分级的域相关的那些节点。

[0353] 图17图示了为增强的分面式分类方法准备输出数据的操作。

[0354] 输出数据可以包括用于域的修正概念定义和关键词分级。关键词分级可以基于语素分级。

[0355] 该过程的输入可以是将要被分类的内容节点302的集合、输入概念分级1008、语素分级402和初步概念定义708a。上文描述了用以生成或者以别的方式获得这些输入的相应操作C、E、L和H。

[0356] 输入概念关系以及第一概念定义708a内的语素属性的交集可以用来(2102)将第一概念定义708a修正为第二概念定义708b。具体而言,如果无法根据语义分级来推断源数据中的概念关系,则可以扩展概念定义以提供由概念关系“暗示”的属性。结果是修正的概念定义708b的集合。

[0357] 可以标识来自参与域的所有语素的集合的语素分级中的相关语素关系的集合(2106)。

[0358] 可以使用来自域的关键词来标记语素分级的缩减的、域特有的版本中的语素(2108)。对于各语素,可以选择使用该语素的次数最多的签名关键词。可以分配用于各关键词的最普及关键词标签。独立关键词可以限于在分面分级中出现一次。一旦使用关键词作为签名关键词,则它可能无法用作其他语素的替代语素。

[0359] 可以将语素分级联合成如下语素关系集,该语素关系集仅包括参与域的语素,并且根据联合的语素分级来推断关键词分级2112(2110)。

[0360] 代表分面式分类的输出数据210a可以包括修正的概念定义708b、关键词分级2112和内容节点302。输出数据可以传送到域数据储存设备706。

[0361] 输入概念分级中的概念关系还可以直接地影响域数据储存设备706中的输出数据。具体而言,输入概念分级可以用来对根据操作的综合部分而推断的关系进行优先级化。直接从源数据汲取的概念关系池库可以代表“显式”数据,这不同于推断的维度概念关系。在输入概念分级中(直接地或者间接地)为显式的推断关系可以优先于在源数据中不存在的关系。也就是说,可以认为显式关系比从该过程推断的附加关系更有意义。

[0362] 输出数据现在可以用作复杂的维度数据结构以表现维度概念分类法(M)。

[0363] 应用分面式分类方法

[0364] 在上文的图3、图18-图19中首先图示了增强的分面式分类方法的组织原理,并且在下文中进行更具体地描述,在图20-图22进行示出,通过这些组织原理可以对初等构造进

行综合,以创建复杂的维度结构。

[0365] 如通过单式(未分裂)的复杂概念分级所提供的那样,这一增强的分面式分类方法使分面式分类方案的灵活性益处与简单性、可视化和整体透视的益处联姻。

[0366] 将分面式分级与简单(单式)分级形成对照阐明了这些益处。简单分级是直观的,并且易于可视化。它们常常同时整合许多组织基础(或者分面),这提供了对所有相关属性的较为整体的透视。跨分面边界来耦合属性并且可以并行导航属性。通过整合属性而不是分裂它们,这些属性提供一种更为经济和鲁棒的解释性框架。

[0367] 本领域技术人员将认识到,许多其他更简单的和传统的分类方法也可以如下文概括的那样从本发明的各种部件和模式中受益。分面式分类和基于集合的分类构造(如形式概念分析)的这些传统过程说明了将从这里描述的系统受益的两个此类可选分类方法。

[0368] 维度概念综合

[0369] 参照图18,包括概念定义的语素310可以在语素分级402中相关。语素分级402可以是删减冗余语素关系的、语素词典206中已知的所有语素关系的全集合。如果可以使用其他语素关系的集合(即通过间接关系)在逻辑上构造语素关系,则可以认为这些语素关系是冗余的。

[0370] 可以按关键词对独立语素310a和310b进行分组,以定义具体概念306b。注意,这些语素310a和310b因此可以与概念306b进行关联(经由关键词分组)并且与语素分级402中的其他语素310进行关联。

[0371] 通过这些互连,语素分级402可以用来创建概念关系的新的和扩展的集合。具体而言,包含通过语素关系来相关的语素310的任何两个概念306本身可以是相关概念。

[0372] 可以使用在概念定义内语素的相伴出现作为用于创建概念关系分级的基础。在概念306b的每个交线406a和406b(图18)表示将概念306b连接到其他相关概念(未示出)的维度轴。如下维度轴的集合可以是复杂维度结构的结构基础,各维度轴代表由定义该轴的语素(或者分面属性)集过滤的概念关系的单独分级。构造方法的简化概况在图19中继续。

[0373] 维度概念分类法

[0374] 图19图示了用于基于维度轴的相交来定义维度概念分类法210的复杂维度结构的构造。

[0375] 利用分别由语素310c、310d和310e定义的概念306c、306d和306e以及由语素310c、310d和310e的集合定义的概念306f来说明四个概念306c、306d、306e和306f的集合。借助语素310c、310d和310e的交集,概念306c、306d、306e和306f可以共享概念关系。综合操作(下文描述)可以基于概念定义中的语素310c、310d和310e来将维度轴406c、406d和406e创建为概念关系的不同分级。

[0376] 可以向域200中的所有或者一部分内容节点302处理这一综合维度概念关系的操作(下文描述图22-图23中所示的范围有限的和动态的处理操作模式)。因此可以将内容节点302分类到完全重新设计的复杂维度结构中,如维度概念分类法210。

[0377] 如上所述,可以向单个内容容器或者内容节点(比如web页面)分配多个概念。因而,单个内容容器或者内容节点在维度概念分类法中可以驻留于许多离散分级上。

[0378] 同样,包含通过语素关系来相关的语素310的任何两个概念306本身可以是相关概念。在一个实施例中,显式和隐式语素关系可以与域的上下文调查组合,以推断维度概念分

类法中的复杂维度关系。

[0379] 可以使用语素作为分面属性来描述概念定义。如上所述,分面属性(语素)在词典中是显式的(“已登记”或者“已知”)还是隐式的(“未登记”或者“未知”)可以无关紧要。应当简单地存在与概念定义相关联的有效描述,以在维度概念分类法中承载其定义。有效概念定义可以提供用以描述内容节点在维度概念分类法中的含义的原始素材。以这一方式,无论域中的对象先前是否作为训练集的部分来分析,都可以将它们放在维度概念分类法中加以分类。如本领域中公知的,存在可用以向将要分类的对象分配概念定义的许多方法和技术。

[0380] 在本发明的一个实施例中,知识表示模型的结构实体的相互影响(上文描述)可以在语素、语素关系、概念定义、概念节点和概念关系之间建立如下逻辑链接:

[0381] 如果活跃内容节点内的概念包含与其他内容节点(下文称为“相关节点”)中的概念具有相同谱系的分面属性(下文称为语素),则可以在活跃节点与相关节点的概念之间存在关系。换言之,各概念可以继承通过如它们存在于内容节点中的语素之间的关系来推断的所有关系。

[0382] 根据分面分级来直接推断的维度概念关系在这里称为显式关系。在这里如下维度概念关系称为隐式关系,该维度概念关系是根据向将要分类的内容节点分配的概念定义内的分面属性的交集来推断的。

[0383] 综合(构建)规则

[0384] 可以通过检查概念的概念关系中的属性之间的关系来计算概念之间的显式关系。如果该概念定义包含与正在分类的内容节点(下文称为“活跃节点”)的属性在分面分级中直接或者间接地相关(下文称为“谱系”相同)的属性,则沿着由涉及到的属性代表的维度轴在概念之间可以存在显式关系。

[0385] 受制于有限制的约束(下文描述),可以在如下任何概念之间推断隐式关系,这些概念共享其概念定义中的属性子集。属性的交集建立父-子关系。

[0386] 可以按照分面属性集来定义轴。在一个实施例中,可以通过分面分级中的分面(根节点)集来定义轴。这些属性集然后可以用来将概率过滤成维度概念关系的联合分级。备选的是,对于根据复杂维度结构派生的动态构造(定制)的分级,可以使用任何属性集作为维度轴的基础。

[0387] 如果可以针对父概念定义中的所有轴得到显式和/或隐式关系,则存在维度概念关系。因此,维度概念关系跨过属性定义的所有维度而在结构上完整无缺。

[0388] 优先级和方向性

[0389] (如语素分级所表达的)分面分级可以用来对内容节点进行优先级化。具体而言,各内容节点可以具体化在分面分级中存在于至多一个位置的属性。在分级中的属性优先级可以确定节点的优先级。

[0390] 可以首先通过检查所讨论的集合内的任何已登记语素的全部优先级,来确定概念关系内的优先级。最高的已登记语素可以建立用于集合的优先级。

[0391] 例如,如果第一集合包括优先级数为{3,37,303}的三个已登记语素,第二集合包括优先级为{5,490}的两个已登记语素,而第三集合包括优先级为{5,296,1002}的三个已登记优先级,则可以对集合进行排序:{3,37,303}、{5,296,1002}、{5,490}。可以基于在语素的集合中包含优先级3的语素的最高总排行来将第一排序集合优先级化。后两个集合可

以均具有最高语素优先级{5}。因此,可以检查各集合中的下一最高语素优先级,以揭示包含优先级为{296}的语素的集合应当是优先级更高的集合。

[0392] 当已登记语素没有区分概念关系中的内容节点时,系统可以使用隐式语素的数目作为优先级化的基础。可以假设语素数目最少的集合在分级中优先级更高。当内容节点包含相同的显式语素和相同数目的未登记隐式语素时,可以认为内容节点彼此处于同等地位。当内容节点处于同等地位时,可以按照系统发现这些内容节点中各内容节点的顺序来建立优先级。

[0393] 图20提供了一个实施例的简单图示,该实施例用于构造隐式关系和确定节点在所得分级中的优先级。

[0394] 在本例中,在语素词典中登记语素“商务”2201。假设通过用户交互,利用如下概念定义来构造内容节点,该内容定义包含这一语素加上在语素词典中没有识别的新语素“模型”2202。

[0395] 继续上例,语素“商务”具有最高优先级2203。集合“商务、模型”是“商务”的暗示子语素2204。向这一集合添加的任何附加语素如“广告”2205将创建分级中的附加层2206。

[0396] 任何语素,无论其在系统中是显式的还是隐式的,都可以用作内容分级的基础(或者轴)。继续上例,隐式语素“广告”2207是基于这一语素的分级的父节点2208。集合“商务、模型、广告”2205是这一分级中的子节点2209。包括“广告”的任何附加集合也将是这一分级的成员。在该例中,集合“广告、方法”2210也是“广告”2211的子节点。由于登记了语素“商务”,所以集合“商务、模型、广告”在广告分级中被给予比仅包含隐式语素的集合“广告、方法”更高的优先级。

[0397] 节点进行优先级化的一个备选实施例涉及了“签名”节点。这些被定义为最好地描述(或者给出含义)它们的关联概念的内容节点。例如,域所有者可以将与具体概念相关联的照片作为用于该概念的签名标识符。因此可以区分签名节点的优先次序。

[0398] 存在用以实施签名节点的许多方式。例如,一种方式是作为特殊一类内容节点的标签。可以对签名节点分配特殊属性,并且该属性可以在分面分级中被给予最高优先级。或者可以在内容节点的表中使用字段以规定这一属性。

[0399] 可以通过自动基础(如按字母顺序、数字和按事件顺序的排列)来补充基于分面分级的优先级化。在传统分面式分类中,优先次序区分和排列是符号表示和引证顺序的问题。系统通常提供用于优先次序区分和排列的动态的属性重新排序。此外,在此不对这些操作进行进一步讨论。

[0400] 轴定义和结构完整性

[0401] 在系统的一个实施例中,用于构建维度概念分类法的另一规则涉及维度轴的结构完整性。作为概念定义(轴定义)的每个语素(属性)集可以建立维度轴。根据这些语素来推断的维度概念关系必须跨如由父节点确定的所有维度而在结构上完整无缺。换言之,与父概念相交的所有维度也必须与节点的所有子概念相交。下例将说明:

[0402] 考虑内容定义为{A,B,C}的活跃内容节点,

[0403] 其中A、B、C是概念定义中的三个语素,而语素E、F、G分别是A、B、C在语素分级中的子语素;

[0404] {A,B,C}是指用语素A和B和C描述的概念定义;

[0405] $\{A, *\}$ 是指用以建立作为A的隐式子语素的节点的显式语素A和一个或者多个隐式语素 $\{*\}$ 的组合;

[0406] $\{A|B\}$ 是指语素 $\{A\}$ 或者 $\{B\}$ 。

[0407] 活跃节点中的三个语素A、B、C在这一例子中可以用来在维度概念分级中建立三个维度(或者相交轴)。对于将作为这一节点的子节点的任何其他内容节点,候选必须是相对于所有三个轴的子节点。以下符号表示是如本发明一个实施例定义的显式关系和隐式关系的解集:

[0408] $\{(A|E|A, *|E, *) , (B|F|B, *|F, *) , (C|G|C, *|G, *) \}$,

[0409] 其中第一维度的语素为A或者E、或者A的隐式语素、或者E的隐式语素;

[0410] 其中第二维度的语素为B或者F、或者B的隐式语素、或者F的隐式语素;

[0411] 其中第三维度的语素为C或者G、或者C的隐式语素、或者G的隐式语素。

[0412] 可以通过约束维度轴的概念定义来进一步限制处理范围。可以通过引用来自父节点的语素子集来建立独立轴(下文称为“活跃轴”),由此约束可以链接到活跃节点的父节点(祖先)的集合。有效的是,与活跃轴相关联的概念定义可以建立约束如下多分级的虚拟父节点,该多分级从活跃节点仅延伸到在由活跃轴的概念定义所定义的分级上驻留的那些内容节点。

[0413] 下例利用概念定义 $\{A, B, C\}$ 使用上文介绍的例子来说明这一约束。在这一例子中,派生的维度概念关系被约束至具有概念定义 $\{A, B\}$ 的活跃轴。在这一约束之下,对活跃节点的可能父节点(祖先)的集合限于集合 $\{(A, B) | A|B\}$ 。换言之,匹配的概念定义将仅包括A或者B的组合但是没有C(同样在本例中假设在语素分级中没有A或者B的父节点)。

[0414] 语素中的显式关系和隐式关系的组合因此可以建立用于在概念之间构建分级关系的规则。

[0415] 如本领域中已知,存在用以优化这些类型的过滤和排序功能的许多方式。它们包括数据管理工具如索引和高速缓存。这些细化在本领域中是公知的,并且在这里将不进一步加以讨论。

[0416] 综合操作模式

[0417] 各种综合操作模式对于本发明的分面式分类方法而言是可能的。可以改变综合以适应不同域的独立要求和终端用户要求。如下文所述,这些模式可以定义如下:

[0418] 静态综合与动态综合

[0419] 在一个实施例中,提供如下“静态”分面式分类综合,在该综合中可以预先定义对维度概念分级进行定义的轴。然后将所得维度概念分类法作为静态结构来利用。

[0420] 分面式分类综合的静态模式的优点在于,域所有者可以按照它们的准确规范来组织维度概念分类法。访问和消费这些静态结构内所含信息的终端用户因此可以从域所有者的组织知识中受益。静态综合因此例如在信息的终端用户对域内所含信息几乎一无所知时尤其有用。

[0421] 在另一实施例中,提供如下“动态”分面式分类综合系统,在该系统中可以直接地基于为信息的终端用户而提供的综合参数,来近实时地处理维度概念分级。这一动态操作模式有助于信息结构的递增和完全“按需”组装。

[0422] 动态处理可以提供巨大的信息经济和存储益处,这消除了对预先创建和存储终端

用户结构的需要。更重要的是,动态处理可以允许终端用户按照他们的要求来精确地定制输出,这提供了个性化益处(下文更具体地讨论综合操作模式)。

[0423] 又一实施例组合了上文介绍的静态和动态综合模式。在这一混合的综合模式之下,域所有者可以提供用于为维度概念分类法提供静态“全局”结构的轴定义选集。在该全局结构内,继而可以使用动态综合来使独立终端用户能够按照他们的需要来进一步定制结构。这一混合模式因此组合了静态综合和动态综合这二者的优点。

[0424] 对概念分级和内容节点的限制

[0425] 随着域和分面分级的规模增加,可以推断的维度概念关系的数目也迅速地增长。可以对生成的关系的数目设定限制。

[0426] 可以由用户输入限制以设置所得输出分级中的有关概念或者关联内容节点的最大数目。例如,管理员可以将综合操作配置成在系统将十个最紧密相关的概念组装成分级之后停止处理。

[0427] 可变抽象级

[0428] 如上文在知识表示模型和分析操作的描述中所述,可以将包括概念定义的属性定义成可变的抽象级。这里描述的一个实施例提供处于概念、关键词和语素这些抽象级的实体。在综合中所用概念定义的属性的抽象级变化可以实现综合操作的明显不同的输出。

[0429] 具体而言,由于属性趋向于域内更基本的语素实体,所以在使用这些属性定义的复杂概念之间可能存在更多连接。定义这些语素项中的属性因此可以提供用以组织所得综合输出的更大连接和更多样方式。

[0430] 反之,由于属性趋向于更抽象的复杂实体(如关键词或者复杂概念),所以所得综合结构可以更精确,通常具有更少的连接、但是整体质量更高。因此,改变综合操作中的抽象级可以允许管理员、域所有者或者终端用户根据他们的独立要求来定制信息。

[0431] 域处理的范围

[0432] 在一个实施例中,可以在生成维度概念分类法的完整视图之前检查和比较域中的所有内容节点。换言之,系统可以发现域中的所有如下内容节点,这些内容节点可以在针对这些相关节点之间的直接分级关系进行任何推断之前是相关的。

[0433] 域中所有节点的完整检查的益处在于它可以提供域内信息的穷尽探测和发现。对于高精确度和取回(recall)要求,这一综合模式可能是适合的。它对于相对更小的界定清晰的域而言也常常是优选的。

[0434] 在另一实施例中,取代分析整个域,可以基于用户的活跃焦点来分析域的局部化区域。这一局部化分析可以应用于素材,无论其先前是否作为训练集的部分被加以分析。参数可以由管理员设置以平衡分析深度与处理时间(延时)。

[0435] 对于没有作为训练集的部分来分析的素材,系统可以使用局部化分析的操作以在根据训练集素材所派生的增强型分面式分类方按之下对素材进行分类。

[0436] 注意到,如下文更具体描述的那样,对来自域的素材的局部子集进行分类的操作也可以用来对新的域进行分类。换言之,可以使用来自一个域的训练集作为用于构造方案的基础,以对来自新的域的素材进行分类,由此支持多域分类环境。

[0437] 图21更具体地图示了各种综合模式。在不限制本发明的范围情况下,这些例子示范了通过各种模式来提供的综合选项的广大范围。这一综合灵活性的益处在于提供一种可

以适应为数众多的域和用户要求的系统。

[0438] 静态(预索引)综合

[0439] 图21图示了本发明在其一个实施例中的如下方法,通过该方法,用于增强型分面式分类方法的输出数据可以产生用以重新组织域的维度概念分类词210。可以生成输出数据(M)(如上文所述和在图17中所示)。此方法的输入可以是来自域的修正概念定义2104、关键词分级2112和内容节点302。

[0440] 各概念定义708b可以映射到关键词分级2112中的关键词2302。可以通过如上文所述和在图3、图18-图20中所示的增强型分面式分类方法的规则来生成用于概念的新维度概念关系(820)。

[0441] 信息结构的管理人员可以偏好于人工调整(2304)自动生成的维度概念分类法构造的结果。操作可以支持这些类型的人工干预,但是针对全自动化操作无需用户交互。

[0442] 分析(2306)可以用来评估所得维度概念分类法的参数。同样,管理人员可以将统计参数设置(2308)为用于维度概念分类法的缩放因子。它们也可以通过减少处理范围将复杂性限制为复杂-自适应系统中的负反馈,因此往回缩放所并入的分级的数目。

[0443] 如下文所述和在图27中所示,维度概念分类法210可以用于用户交互(N)。

[0444] 域子集(范围有限的)综合

[0445] 图22图示了从域中选择内容节点,并且将这些内容节点排序成维度概念分级。可以采用域相对于活跃节点2402的受约束视图。取代处理整个域的是,操作可以执行活跃节点2402的紧接邻近区2404中所有内容节点(例如2406)的定向调查。

[0446] 递归概念分级组装

[0447] 在一个实施例中,递归算法可以用于将未区分的相关内容节点组细分成具体结构组。“候选集”描述了与活跃概念定义相关的概念集和关联内容节点,而不考虑它们如何精确地相关。可以相对于活跃概念或者内容节点将组描述为父节点和子节点(分级关系)以及兄弟节点(关联关系)。这些组描述的结构关系在本领域中是公知的。然后可以基于涉及到的下层语素关系和语素相对于活跃概念,来将这些邻近概念和关联内容节点排序成分级关系。

[0448] 在图22中将这一分级图示为内容节点候选集2404内的内容节点(例如2406)之间的关系子集。在分级2408中,与活跃节点2402直接相关的那些内容节点(直接子节点)在候选集2404内没有其他父节点。候选集中的其余内容节点可以在分级中定位得更深,其作为间接子节点(后代)。

[0449] 将一个域的分类方案应用于第二个域

[0450] 图23图示了对不是用来开发分面式分类方案的训练集的部分、来自域的素材的局部子集进行分类的操作。

[0451] 可以从域200中选择域素材2404a的局部子集以供处理。可以基于由域所有者建立的选择标准(2502)来选择素材。可以相对于作为用于局部化区域的基础的活跃节点2504来进行选择。选择过程可以生成局部子集2506的参数,比如对局部子集的边界进行描述的搜索项列表。

[0452] 存在用于局部集的许多可能选择标准。在一个实施例中,可以通过将与活跃节点相关联的概念定义传递到全文本信息检索(搜索)部件,以返回相关素材集来选择素材。这

样的全文本信息检索工具在本领域中是公知的。在一个替代实施例中,可以通过检查关键词分级根据活跃节点中的概念定义来派生扩展式搜索查询,以派生相关关键词的集合。这些相关关键词又可以用来扩展搜索查询以包括与活跃节点的概念定义相关的项。

[0453] 从选择过程派生的域的局部子集2404a可以包括将要分类的候选内容节点。对于局部子集中的各候选内容节点,可以提取概念签名(2508)。概念签名可以由域所有者标识并且可以用来映射(2302)为域特有的关键词分级2112中的关键词,以针对各候选内容节点提供概念定义。同样,构建部件并不要求根据概念签名来派生的所有关键词都为系统所知(在关键词分级中登记)。

[0454] 可以使用上述隐式和显式关系的构建规则针对候选内容节点来计算概念分级(820)。最终结果可以是局部概念分类法210c,其中在根据训练集针对该域而派生的构造方案之下,组织来自域的局部子集的内容节点。局部概念分类法然后可以用作用户交互的环境以进一步对分类进行细化。

[0455] 动态(实时)综合

[0456] 本发明的一个备选实施例使用将用户偏好实时并入综合操作中的动态综合模式。图24-图25和以下描述提供域在这一动态综合模式内的操作有关的更具体细节。

[0457] 在图24中,在广义的概况中图示了动态综合模式的一个实施例。动态综合过程可以遵循请求-响应的操作模型。动态综合操作由用户请求启动(2402)。用户可以指定他们的要求(例如他们的兴趣域、通过活跃概念定义来编码的他们的兴趣主题、通过轴定义来编码的他们对主题的看法以及受限制综合参数集约束的他们的兴趣范围)。在图24中,以简化形式将这些用户参数示意地表示为内部包括更多初等属性(四个点)2404的活跃概念定义(方框)。

[0458] 使用来自用户的这一动态输入,系统然后可以返回关联的概念分级(输出概念分级)2406。这一输出概念分级然后可以是用户进一步探测的焦点,或者它可以充当通向又一轮综合操作的桥梁。

[0459] 为了处理这一请求,与活跃概念定义关联的属性集可以是用于来自指定域2408内的如下概念集进行定位的基础,该概念集将用作用于所综合的概念分级的候选集2410。下文描述用以将这些概念与活跃概念定义相关的“派生”方法2412。派生可以对派生动态地进行排列,并且用作用以构造相关概念分级的参考。

[0460] 接着,提供域动态综合模式的主要步骤和部件有关的更多细节。

[0461] 用户启动的综合请求

[0462] 动态综合操作由用户请求启动(3502)。为了启动动态综合过程,用户可以提供域、活跃概念定义和轴定义。用户也可以经由下文讨论的其他输入综合参数来约束的概念分级的规模和形状。如下文在用户界面系统实施的讨论中所述,存在用以获取这一类用户输入的许多技术手段。

[0463] 动态综合输入和综合参数

[0464] 因此,对动态综合模式的输入可以包括用户特有的综合参数和域特有的分面式数据集。这些输入可以将综合操作约束于狭窄磨练的范围或者对象区域、约束于用户的精确要求。上文提供了与域特有的分面式数据集有关的细节。

[0465] 运行时间综合参数

[0466] 如上文讨论的那样,动态综合的一个实施例可以提供活跃域、活跃概念定义和活跃轴定义的用户输入。此外,用户可以通过提供对分离度进行规定的参数和按照概念和内容节点对综合操作的输出进行限制的参数来进一步描述他们的要求。

[0467] 分离度参数指定在输出概念分级中从活跃概念定义到相关概念定义的直接分级步骤的最大数目。

[0468] 例如,基于增强的分面式分类方法的构建规则,并且在给定有代表性的活跃属性集{A,B,C}的情况下,以下属性集将是被移除的一个分离度:

[0469] {A,B,C,?}:具有一个附加元素的所有超集,其中“?”代表一个其他属性;

[0470] {A,B},{A,C},{B,C}:基于隐式属性关系的所有子集;

[0471] {D,B,C},在给定 $A \rightarrow D$ 的情况下是显式属性关系。

[0472] 延时

[0473] 延时是终端用户可以操控的另一综合参数。在一个实施中,“上限”响应时间可以适用于系统,使得综合操作限于在用户的综合请求与用以满足该请求的构建引擎响应和输出之间的最大时间。这一延时控制的另一实施例将允许终端用户增加或者减少请求-响应的的时间以调节性能,以匹配其独立信息访问和发现要求。

[0474] 用于动态综合的候选集

[0475] 在图25中图示了用于动态综合的候选集组装的一个实施例。

[0476] 在动态综合中,可以针对属性分级来检查活跃概念的属性集,以便发现显式的相关祖先和后代属性集。上文在综合(构建)规则的描述之下提供了关于这些检查的更多信息。同样,无需在这一实时动态综合模式之下完全地检查整个域。系统仅检查由候选集定义的域的子集。发现候选集如下:

[0477] 可以考虑是子集或者具有如下元素或者兼而有之的属性集,这些元素是活跃属性集中的元素的显式祖先(这些代表可能的祖先概念)。在这些相关属性集2502a、2502b和2502c中的每个属性集内,每个属性可以具有它子集的匹配概念定义集。用于给定的活跃概念定义属性集的这些概念集2504a、2504b和2504c的交集可以包含该属性集的匹配概念(将匹配概念图示为实心点;将非匹配概念图示为空心点)。

[0478] 单独使用如下相关属性集来进行类似过程,这些相关属性集可以是超集或者具有如下元素或者兼而有之,这些元素是活跃属性集中的元素的显式后代(代表候选后代概念)。这里同样,用于相关属性集的概念集的交集可以包含该属性的匹配概念。

[0479] 来自所有相关属性集的交集的并集可以是候选集。可以将相关属性集约束于指定的轴定义。它们的数目也可以受制于指定的最大限制和分离距离程度。

[0480] 用于概念分级组装的派生

[0481] 在实时的动态综合模式之下,延时可能是一个主要限制因素。具体而言,几乎没有时间来以穷尽方式处理甚至相对较小的候选集。如上文讨论的那样,使用概念分级综合递归方法的静态综合手段常常由于它可能对更大的域引入的延时而在这一动态环境中错用。

[0482] 这样,动态综合的一个实施例使用派生方法以动态地实时组装概念分级。派生是对候选概念如何与活跃概念相关进行描述的操作集。

[0483] 除了上文介绍的性能和减少延时的益处之外,派生还引入概念综合的新颖益处、也就是将新概念定义推断为下文讨论的“虚拟概念”。这些虚拟概念通过推断新概念来在很

大程度上扩展系统的发现益处,即使这些新概念还没有与内容节点关联。这些派生也提供强大的排列和过滤手段作为用户可配置的群集机制。

[0484] 可以从与活跃概念的属性集相关的属性集中发现候选集。可以从分面式数据集中的属性分级中发现显式相关的元素。隐式相关的属性集可以由集合相交(也就是这些属性集的子集和超集)暗示。而在域中,用来发现隐式后代属性的附加属性可以为系统所知或者可以不为系统所知。

[0485] 活跃属性集可以与候选集中的概念所关联的各属性集配对。对于各对,可以派生将活跃属性集变换成它的配对集的集合操作序列。

[0486] 在试图发现相关属性集的过程中,可以对属性集执行四个派生操作。操作类型可以简写为如表1中所示。

[0487] 表1-派生操作类型

[0488]

	为了派生隐式关系	为了派生显式关系
对祖先	d:删除属性	p:用父属性取代属性
对后代	a:添加属性	c:用子属性取代属性

[0489] 注意,所有属性关系的方向性必须与潜在概念关系对一致。属性集对可以在它们元素之间具有祖先关系或者后代关系,但是并不兼有这二者。

[0490] 综合过程通过仅应用祖先操作(p,d)或者后代操作(c,a)但是并不兼用这二者,以建立概念之间的关系来保留这一方向性。这防止概念利用对应于无关概念的那些属性来替代让其所有属性。

[0491] 例如,在给定属性为{A,B,C}的活跃概念以及属性为{D,B,G,F}的候选概念的情况下,存在贯穿活跃概念与其三个属性对应的定义三个轴。为了确定在概念之间是否存在关系,可以首先使用显式关系,例如从A到D的显式关系和从C到G的另一显式关系。(这些均为c操作:用子属性取代属性)。最后,使用添加后代属性(即F)的隐式a操作获得与候选后代的属性集匹配活跃概念的属性集。因此可以认为该候选是活跃概念的后代。

[0492] 为了进行说明,当将活跃属性集和候选属性集配对时,存在三个可能的属性组:

[0493] 仅与候选集相关联的属性组(“仅有候选”属性);

[0494] 与候选集和活跃集二者相关联的属性组(“兼有”属性);

[0495] 仅与活跃集相关联的属性组(“仅有活跃”属性)。

[0496] 如果将活跃集转换成候选集需要删除“仅有活跃”属性,则候选集是活跃集的祖先。

[0497] 如果活跃集与候选集相同,则候选集是活跃集的兄弟。

[0498] 如果将活跃集转换成候选集需要添加“仅有候选”属性,则候选集是活跃集的后代。

[0499] 通过删除“仅有活跃”属性并且添加“仅有候选”属性来将活跃集转换成候选集是无效的,无论两个原有集合是否已经具有共同属性。认为这样的一对集合无关。对此仅有例外是,当“仅有”集合中的属性在属性分级中相关时。在这样的情况下,可以执行两个操作之一:

[0500] 用活跃集属性的父属性取代该活跃集(其中候选集为活跃集的祖先);

[0501] 用活跃集属性的子属性取代该活跃集(其中候选集为活跃集的后代)。

[0502] 所得属性然后是“兼有”集合的成员。

[0503] 在给定级,呈现兄弟的顺序可能至关重要。对于用户而言可能更重要的那些概念应当具有更高优先级。

[0504] 候选集中的各概念可以具有将它连接到活跃概念的唯一派生系列。通过综合对派生进行排列和处置的顺序影响结果分级中的概念排序。根据表2来确定分级中的候选概念的优先级。

[0505] 表2-派生在确定结果分级时的优先级

[0506]

	候选集中的普及度	域中的普及度
显式操作 (p,c)	1	2
隐式操作 (a,d)	3	4

[0507] 响应

[0508] 响应于在用户的请求中指定的要求,应用可以返回根据与域内的对象相关联的概念而构建的、与活跃概念相关的、并且沿着轴的概念分级。用户参考这一概念分级以发现与它们指定的活跃概念相关的概念。

[0509] 可以将派生构建成分级结果集。该分级中的各节点代表如下概念,该概念具有作为其概念定义的属性集。分级中的各边代表单个派生操作。

[0510] 虚拟概念

[0511] 在一些情况下,在概念分级节点的属性集没有匹配概念。可以使用虚拟概念作为用以表明这一点的占位符。

[0512] 例如,在给定属性集 {A,B,C} 的情况下,如果存在:

[0513] 显式关系 $A \rightarrow D$,

[0514] 显式关系 $D \rightarrow F$,

[0515] 没有具有 {D,B,C} 属性集的概念,

[0516] 则 {F,B,C} 将在具有来自 {A,B,C} 的一个分离度的候选集中。如果 {D,B,C} 属性集没有对应概念,则在分级中的这一节点处是虚拟概念。

[0517] 从活跃域内,动态综合过程可以隔离和返回与活跃概念相关的概念分级。相关概念可以沿着指定轴并且按照规定的那样,从活跃概念在祖先(更广义)和后代(更专用)的方向上产生分支。

[0518] 注意,出于多个目的,可以用多种方式来表示对维度概念分类法210进行派生的数据结构。在以下描述中图示了终端用户交互的目的。然而,这些结构也可以在其他数据操控技术的服务中使用,例如用作对另一信息检索或者数据挖掘工具(未示出)的输入。

[0519] 复杂-自适应反馈机制

[0520] 图27图示了用于在复杂-自适应系统中处理用户交互的方法。该方法在上述维度概念分类法过程N上构建。用户交互可以向系统建立一系列反馈。可以通过由终端用户启动的反馈来实现向复杂维度结构的自适应细化过程。

[0521] 图37图示了如下计算机系统4000的一种可能实施,该计算机系统4000允许以一个或者多个维度概念分类法4010的形式来操控分面式分类信息的各方面。系统4000可以包括

计算机可读介质4020,如盘驱动或者其他形式的计算机存储器,该计算机可读介质4020包含用于执行该实施的计算机程序、软件或者固件4080以及维度概念分类法的各方面,如例如概念定义4090、分级数据4100、内容节点4110、与内容节点对应的定义4120或者维度概念分类法4010的每个方面的分类4130或者分类4130中的一个。系统4000也可以包括处理器4030、用户接口4040(如键盘或者鼠标以及显示器4050)。在这一实施中,计算机处理器4030可以访问计算机可读介质4020并且取回根据源数据生成的维度概念分类法4010的至少一部分,而且在显示器4050上呈现分类法4010的该部分。处理器4030也可以从外界实体(用户或者机器)从如下界面4040(可选为用户界面)输入,该界面反映维度概念分类法4010的各方面的用户操控。处理器4030可以将第一维度概念分类法4010中发现的多个可能关系中的任一可能关系的所接收的外界实体操控并入到第二维度概念分类法中。例如,外界实体操控的形式可以是对第一维度概念分类法4010进行更改或者添加数据、编辑概念定义、分级数据、改变与概念相关联的内容节点相对于与概念相关联的其他内容节点的位置、更改对内容节点的主题内容进行描述的定义或者对分面式分类进行的其他改变。第二维度概念分类法可以完全取代第一维度概念分类法4010、与第一维度概念分类法4010完全并行或者分开存在、作为第一维度分类法4010的特例表而存在等。另外,对第二维度概念分类法的访问可以限于诸如域所有者和管理员、用户、专用远程计算机设备等某些类的外界实体。

[0522] 显示器4050可以呈现维度概念分类法4010的各方面,其形式是可以响应于界面4040的、由处理器控制的显示窗口或者编辑器4070。编辑器4070也可以采用web页面的形式,并且可以呈现根据维度概念分类法4010或者其变型而派生的内容节点和分面式分类。由编辑器示出的内容节点和分面式分类可以对应于由外界实体选择的活跃节点,并且例如可以采用树片段的形式。编辑器4070也可以呈现如下编辑功能,外界实体可以利用该功能来操控维度概念分类法4010的各方面或者引入新元素、关系和内容。编辑功能还可以包括回顾界面,该回顾界面允许外界实体更改与节点内容相关联的一个或者多个语素组、以及节点在维度概念分类法中的位置,以使它们与节点的内容一致。

[0523] 因此,可以将复杂-自适应过程的方法总结如下:

[0524] 提供维度概念分类法作为用于用户交互212a的环境。一旦已经将维度概念分类法210呈现给用户,则它可以变成用于修正现有数据的环境以及用于新数据(维度概念分类法信息)的源。输入数据804a包括用户对现有数据的编辑和新数据的输入。它还提供用于将分类进行演变并使其适应于动态域。

[0525] 用户交互可以包括对系统的反馈。可以基于集中式系统中存储的语素元素使用符号表示系统,来唯一地标识维度概念分类法信息中数据元素的唯一标识符。因此,可以以如下方式标识由系统产生的维度概念分类法中的各数据元素,该方式使得该数据元素可以合并回到集中式(共享)语素词典中。

[0526] 因此,当用户操控这些元素时,可以跟踪对相关语素元素的临时效果。这些改变可以反映系统中的新显式数据,以对由系统自动生成的任何推断数据进行细化。换言之,可以通过终端用户的显式交互来加强或者拒绝由系统原先推断的数据。

[0527] 用户交互可以包括新数据源和对已知数据源的修正。可以将对已知元素的操控转换回到其语素祖先。系统没有识别的任何数据元素可以代表新数据。然而,由于在由系统产生的现有维度概念分类法的上下文中进行改变,所以此新数据可以放置于已知数据的上下

文中。因此,可以在已知数据的上下文中提供由用户添加的任何新数据元素。在已知与未知之间的关系可以对可以根据用户的交互来推断的维度概念分类信息的数量进行很大程度的扩展。

[0528] 系统中的“快捷”反馈212c可以提供用于终端用户的实时交互环境。由用户启动的分类法和容器编辑2902可以在系统中排队等候,并且在系统资源变得可用时正式加以处理。然而,用户可能需要(或者偏好)他们对维度概念分类法的改变的实时反馈。为了通过系统的正式反馈来处理变化而需要的时间可能延迟向用户的这一实时反馈。结果,系统的一个实施例提供快捷反馈。

[0529] 该快捷反馈可以开始于针对当时存在的域数据储存设备706来处理用户编辑。由于用户的改变可以包括在域数据储存器中当前不存在的维度概念分类法信息,所以系统必须使用与这些改变的效果逼近的过程。

[0530] 可以将用于创建隐含关系212b的规则(上文描述)应用于新数据作为用于完全处理的短期替代规则。这一方式允许用户立即插入新数据并且与该新数据交互。

[0531] 与通过系统的正式过程来计算的维度概念关系不同,这一逼近过程可以使用在已知语素集中存在的系统未知的语素,以限定和调整集合中已知语素的维度概念关系。将这些调整的关系描述为上文更具体描述的“隐式关系”216。

[0532] 对于新数据元素,可以基于隐式关系(上文描述)来分配短期概念定义,从而有助于实时交互处理。在完成对域的下一完全处理循环时,可以用由系统修正的完整概念定义取代短期的暗示概念定义。

[0533] 本领域技术人员将认识到,可以存在许多算法用来逼近未知语素对系统中已知语素的关系的影响。

[0534] 提供用户交互

[0535] 维度概念分类法提供用户交互的环境。在本发明的一个实施例中,可以提供两个主要用户界面。导航“查看器”界面可以提供浏览分面式分类。这一界面是称为“分面式导航”的类。其他界面可以称为“概括器”,该概括器可以允许终端用户改变关系结构、概念定义和内容节点分配。

[0536] 分面式导航和概括器界面的一般特征在本领域中公知的。本领域技术人员将清楚下文描述的新颖方面,特别是所涉及的复杂-自适应系统212。

[0537] 查看概念分类法

[0538] 可以通过表示层来表达维度概念分类法。在一个实施例中,表示层是网站。网站可以包括表现维度概念分类法视图集的web页面。视图是活跃节点范围内的维度概念分类法的一部分(例如由一个或者多个轴过滤的多分级的子集)。在本文中的活跃节点是维度概念分类法内当前作为终端用户或者域所有者焦点的节点。在一个实施例中,“树片段”用来代表这些关系。

[0539] 用户可以向系统提供文本查询,以直接地移动到它们的搜索和信息检索的通用区域。正如本领域中公知的,可以通过与各概念相交的分面和属性对视图进行过滤和排列。

[0540] 可以按照各概念对内容节点进行分类。也就是说,对于任何给定的活跃概念,可以呈现与由用户过滤的该概念的属性相匹配的所有内容节点。

[0541] 可以围绕各节点来改变各视图的“分辨率”。这是指所显示关系的广度和总览的穷

尽度。也可以在已分析域部分的尺寸和选择背景下考虑视图的分辨率问题。同样,在分析深度与处理所需时间量(延时)之间存在权衡。表示层可以操作用以基于活跃节点的位置、视图的分辨率和由管理员配置的参数,来选择将要分析的域的一部分。

[0542] 在一个实施例中,查看维度概念分类法、操作动态综合模式(如上文讨论的那样)这些交互可以生成用于本发明的复杂-自适应系统的反馈。在这些条件之下,从终端用户的角度来看,通过查看交互而生成的隐式反馈实质上将是透明的。换言之,终端用户将仅通过查看维度概念分类法这些交互来为系统创建有价值的反馈。

[0543] 这一透明的由用户生成的反馈存在许多益处。终端用户将无需花费对维度概念分类法进行直接编辑(如下文具体讨论的那样)所需要的精力。另外,由于在这一动态综合模式之下,只有由用户请求的维度概念分级包括作为用于后续分析操作的反馈而返回的维度概念分类法。将反馈集约束至只有由终端用户实际上请求的信息的这一更窄范围,这具有改进由系统生成的反馈数据的质量的效果。

[0544] 编辑概念分类法

[0545] 表示层将维度结构精炼成人类交互所必需的简化视图(比如以下web页面,其中包括指向维度概念分类法中有关页面的链接)。这样,表示层也可以倍增作为用于如下信息结构的编辑环境,它是根据这些信息结构而派生的。在一个实施例中,用户能够从表示层内切换到编辑模式以立即编辑结构。

[0546] 概括器为用户提供用以操控分级数据的手段。概括器也允许用户操控与结构中的每个概念相关联的内容节点。

[0547] 用户交互可以更改向维度概念分类法中的节点分配的上下文和/或概念。上下文是指节点相对于结构中其他节点而言的位置(也就是建立结构的维度概念关系)。概念关系描述了表达为语素汇集的节点的内容或者主题。

[0548] 在一个实施例中,可以向用户呈现回顾过程,以使用户能够确认这样的用户编辑的参数。可以向用户展示以下维度概念分类法信息用于这一回顾:1)节点的内容;2)与内容相关联的语素组(表达为关键词);以及3)节点在分类法结构中的位置。用户可以更改后两个(语素和相对位置)的参数以使信息与前一个(在该节点的内容)一致。

[0549] 因此,在本发明的一个实施例中的交互可以总结为两个广义类型的某一组合:a)容器编辑;以及b)分类法编辑。

[0550] 容器编辑是对内容容器(比如URL地址)改变对在维度概念分类法内分类的内容节点的分配。容器编辑也改变对维度概念分类法内的内容节点的描述。

[0551] 分类法编辑是对维度概念分类法中节点位置的上下文改变。这些改变包括将新节点添加到结构中和对现有节点重新定位。这一维度概念分类法信息可以作为对与受用户交互影响的概念相关联的语素关系的改变而反馈到系统中。

[0552] 利用分类法编辑,可以创建在分类法中的概念之间的新关系。可以通过用户交互来构造这些概念关系。由于这些概念是基于语素的,所以新概念关系可以与新语素关系集关联。这一维度概念分类法信息可以反馈到系统中以重新计算这些暗示的语素关系。

[0553] 也可以在更初等的抽象级(如关键词和语素)提供用户交互。

[0554] 图26图示了容器编辑过程的一个实施例。容器编辑改变对描述各内容节点的概念定义和下层语素。利用这些改变,用户更改内容节点的下层概念描述。在这样做时,他们可

以更改被映射到这些内容节点处的概念定义的语素。

[0555] 用户交互可以构造表达为关键词汇集的向内容节点分配的概念定义。在这一构造中,用户可以与系统的语素词典和域数据储存设备交互。这里创建的任何新关键词可以发送到如上文所述的系统的语素提取过程。

[0556] 在这一例子中,文档2801是活跃容器。在用户界面中,描述内容的关键词2802的集合可以连同文档一起呈现给用户(为简化例子,这里没有示出这一节点在维度概念分类法中的相对位置)。

[0557] 在该例中,当用户回顾内容时,用户可以确定与页面相关联的关键词不是最优。可以由用户选择新关键词以取代载有页面的集合(2803)。用户将关键词列表2804更新为与文档相关联的新概念定义。

[0558] 这些改变被传送至域数据存储设备706。可以对数据存储设备进行搜索以便标识在系统中登记的所有关键字。

[0559] 在这一例子中,列表包括除了“狗”之外的由用户标识的所有关键词。结果,“狗”将作为如下隐式关键词来处理,该隐式关键词修改在系统2806中登记的显式关键词。

[0560] 当通过集中式变换引擎来回顾域时,可以完全地分析隐式关键词。它然后可以由显式关键词(作为现有关键词或者新关键词)取代并且与一个或者多个语素相关联。

[0561] 个性化

[0562] 图28图示了提供个性化特征的本发明的一个可选实施例,其中可以针对域的各单独用户维护维度概念分类法的个性化版本。

[0563] 个性化的一个实施例提供用以将共同概念分类法210e连同用于各单独用户的个性化概念分类法210f一起进行个性化的手段。在终端用户与系统第一次交互时,各终端用户可以参与共同概念分类法210e。后继交互可以参与分类法210f的用户的个性化看法。

[0564] 响应于代表各终端用户的偏好的用户交互212a,通过核对数据结构的唯一表示将数据结构“个性化”。可以将编辑的结果存储为来自用户交互的个性化数据(3004)。在一个实施例中,将这些编辑存储为共同概念分类法210e的“特例”。当处理个人概念分类法210f时,系统可以替换它在用户的特例表中发现的任何改变。

[0565] 所示元素可以标识系统的复杂-自适应过程中的协作者。它提供一种用以将唯一标识符与每个用户相关联并且存储他们的交互的手段。

[0566] 在另一实施例中,系统可以将唯一标识符分配给通过表示层来与维度概念分类法210e交互的每个用户。这些标识符可以视为语素。每个用户分配有全球唯一标识符(GUID)、优选为可以跨过所有计算机和网络来使用的128位整数(16字节)。用户GUID作为系统中的语素而存在。

[0567] 与系统中的任何其他元素相似,用户标识符可以在语素分级中加以登记(显式语素)或者是系统未知的(显式语素)。

[0568] 按照本领域公知的术语,两类标识符之间的区别类似于登记的拜访者和匿名的拜访者之间的区别。可以用来生成标识符并且将标识符(“跟踪符”)与用户相关联的各种方式在本领域中也是公知的并且将不在这里加以讨论。

[0569] 当用户与系统交互(例如通过编辑内容容器)时,系统可以将该用户的标识符添加到描述概念定义的语素集。系统还可以添加与系统所支持的各种交互性相关联的一个或者

多个语素。例如,用户“鲍勃”可能希望编辑具有内容定义“记录、工作室”的容器以包括地理参考。系统因此可以创建鲍勃特有的用于该容器的以下概念定义: {鲍勃, 华盛顿, (记录, 工作室)}。

[0570] 利用此维度概念分类法信息,通过在上述增强的分面式分类方法中应用相同的显式和隐式关系规则,系统可以以用户鲍勃所特有的方式呈现容器。容器可以出现在用于鲍勃的个人Web页面上。在他的个人概念分类法中,页面将与华盛顿的资源相关。

[0571] 维度概念分类法信息还将为其他用户全球可用,并且作为负反馈机制受制于由管理员建立的统计分析和障碍率。例如,如果足够的用户标识了具有记录工作室的华盛顿现场,则它将最终地作为有效关系呈现给所有用户。

[0572] 对与内容容器相关联的概念定义的这一类修改实质上新的维度层添加到代表各层用户交互性的维度概念分类法信息。它提供一种用于使用适用于其他信息和内容形式的现有构造过程来进行个性化的多功能机制。

[0573] 正如本领域中公知的,存在可用于添加个性化和定制表示层的许多技术和架构。这里讨论的方法利用系统的核心结构逻辑以组织协作者。它实质上用户交互仅视为另一类信元,这说明了系统的灵活性和可扩展性。然而,它在用于向系统添加定制和个性化的各种方法方面没有限制本发明的范围。

[0574] 基于机器的复杂-自适应系统

[0575] 图29图示了提供用于提供复杂-自适应系统的基于机器的手段的一个可选实施例,其中包括维度概念分类法210的维度概念关系作为系统输入数据804b直接返回到变换引擎过程(3102)。

[0576] 注意,就此而言,本发明提供了如在本公开内容中描述的终端用户用以创建和管理数据结构的能力。在本发明的某些方面中,终端用户提供如下反馈,该反馈进一步通知如这里说明的数据结构的创建和管理。这一反馈不仅可以由终端用户提供,而且还可以,例如,由例如从终端用户汇集反馈的计算机这样的机器、或者甚至是比如完全无人类介入的计算机这样的机器提供。在本文中,终端用户或者机器的角色在本公开内容中称为“反馈代理”。也应当注意到,出于说明的目的,在本公开内容中提供的许多例子涉及终端用户,但是应当理解,在这些情况中的许多情况下(即使不是在所有情况下),比如计算机这样的机器可以取代终端用户的角色。这一子标题说明了这样的实施。因而,本公开内容应当这样来解读:在许多情况下(即使不是在所有情况下),提及的“终端用户”可以解读为指代“反馈代理”。

[0577] 注意,在从源数据结构派生的原有概念关系与从系统构建引擎的过程中表现出来的维度概念关系之间有重要区别。前者在源数据结构中是显式的;后者是根据针对语素词典内的初等构造而应用的构造方法来派生的(或者通过这些构造方法来显现)。因此,与基于用户交互的复杂-自适应系统相似,基于机器的方式可以提供如下手段,该手段通过由初等构造来综合(复杂的)维度概念关系在系统操作800中引入变化,然后在源结构分析部件中该改变中选择。

[0578] 在这一基于机器的操作模式之下,针对复杂-自适应系统的选择要求可以由源结构分析部件(上文所述和在图6中所示)承担。具体而言,可以基于循环关系的标识(1002)以及可以用来分解这些循环关系的各种模式和参数来选择维度概念关系。如本领域中公知

的,存在用以提供基于机器的复杂-自适应系统的许多可选手段、选择标准和分析工具。

[0579] 可以从数据集删减与分级假设(该分级假设是通过循环关系的存在而在集合体中标识的)相抵触的维度概念关系(1004)。这一删减后的数据集可以重新组装(1006)到如下输入概念分类法1008中,根据该输入概念分类法,操作800可以通过分析引擎的其余操作来派生新的初等构造集。

[0580] 这一类基于机器的复杂-自适应系统可以与比如上文参照图4和27描述的基于用户交互的系统212这样的其他复杂-自适应系统结合使用。例如,图30的基于机器的复杂-自适应系统可以用来通过数次迭代该过程来对维度概念分类法进行细化。随后,可以在基于用户的复杂-自适应系统中向用户介绍所得维度概念分类法以求进一步细化和演变。

[0581] 实施

[0582] 如在系统架构的这一描述中通篇强调的那样,在包括数据储存设备的用于设计本发明许多实施例的方法和技术中,存在大量可变性。可以通过在现有技术中同样公知的多种形式的架构设计来解释和改变本发明的许多应用。

[0583] 系统架构部件

[0584] 计算环境

[0585] 图30图示了用于本发明的计算环境的一个实施例。

[0586] 在一个实施例中,可以将本发明实施为在四级架构之下操作的计算机软件程序。服务器应用软件和数据库可以在集中式计算机和分布式的分散系统上执行。可以使用因特网作为用以在集中式服务器和与它交互的各种计算设备和分布式系统之间通信的网络。

[0587] 用于建立这一类计算环境的可变性和方法在本领域中是公知的。这样,这里没有包含计算环境的进一步讨论。所有适用环境的共同之处在于,用户通过他的或者她的计算机或者计算设备来访问公共或者私有网络(如因特网或者公司的内部网),由此访问将本发明具体化的计算机软件。

[0588] 服务级

[0589] 各个级可以负责提供服务。一级3202和二级3204在集中式处理模型之下操作。三级3206和四级3208在分布式(分散式)处理的模型之下操作。

[0590] 此四级模型根据由系统用来分析域的共享集中式数据来实现专用域数据的分散化。下文讨论图33中示出了在共享数据与私有数据之间的这一描绘。

[0591] 在第一级,集中式数据储存器代表由系统管理的各种数据和内容源。在一个实施例中,数据库服务器3210可以提供数据服务以及访问和维护数据的手段。

[0592] 虽然在此将分布式内容描述为包含于“数据库”内,但是数据可以存储于多个链接的物理位置或者数据源中。

[0593] 元数据也可以是分散式的、并且存储于系统数据库外部。例如,HTML代码段包含可以由系统操作的元数据。来自外部模式的元素可以映射到在本系统的模式中使用的元素。用于呈现元数据的其他格式在本领域中公知的。信息领域因此可以提供分布式内容源的财富和用于终端用户以分散方式管理信息的手段。

[0594] 用于跨多个链接的物理位置或者数据源来管理数据的技术和方法在本领域中公知的,并且将不在这里进一步穷尽地加以讨论。

[0595] XML数据馈送和应用编程接口(API) 3212可以用来将数据储存器3210连接到应用

服务器3214。

[0596] 同样,本领域技术人员可以理解,XML可以符合广泛的专用和开放式模式。一定范围的数据交换技术提供用以将各种分布式内容格式并入到系统中的基础结构。对一个实施例中所用连接器的这一讨论和所有以下讨论并不限制本发明的范围。

[0597] 在第二级3204,在集中式服务器3214上驻留的应用可以包含用于本发明的核心编程逻辑。应用服务器可以提供用于实施本发明方法的各种方面的处理规则、以及对于数据库服务器的连通性。上文中具体描述了图4-图17和图20-图23中所示这一编程逻辑。

[0598] 在一个实施例中,可以将应用服务器所处理的结构信息输出为XML 3216。XML可以用来将外部数据储存设备和网站与应用服务器进行连接。

[0599] 同样,XML 3216可以用来将这一交互性传达回到应用服务器,以供在进行优化和细化的过程中进一步处理。

[0600] 在第三级,分布式数据储存器3218可以用来存储域数据。在一个实施例中,此数据可以用XML文件的形式存储于web服务器上。存在存储域数据的许多可选模式,比如外部数据库。分布式数据储存设备可以用来将输出数据分布到终端用户的呈现设备。

[0601] 在一个实施例中,可以将输出数据作为使用XSL变换文件(XSLT) 3220来表现的XML数据馈送进行分布。这些技术可以在第四级通过表示层来表现输出数据。

[0602] 表示层可以是任何分散式网站、客户端软件或者如下其他介质,该其他介质可以用可以由人类或者机器利用的形式来呈现分类法。表示层可以代表由终端用户用与分类法交互的环境和分类法的对外展现。在一个实施例中,可以将数据表现为网站,并且显示于浏览器中。

[0603] 这一结构化信息可以提供用于用户协作和输入的平台。本领域技术人员将认识到,XML和XSLT可以用来跨不同范围的计算平台和介质来表现信息。这一灵活性允许在广泛的信息处理任务内将系统作为过程来使用。

[0604] 例如,可以使用数据馈送中的关键词来表达语素。通过在数据馈送中包括语素引用,系统可以响应于具体语素标识符在表示层上提供附加处理。上文在对个性化的讨论中描述了这一灵活性的应用(图28)。

[0605] 使用基于web的表单和控件3224,用户可以添加和修改系统中的信息。这一输入继而可以作为XML数据馈送3226和3216经由分布式数据储存设备而返回到集中式处理系统。

[0606] 此外,也可以从因特网并入开放式XML格式如RSS作为向系统的输入。

[0607] 可以由应用服务器3214处理对结构信息的修改。来自这一处理的共享语素数据可以经由XML和API连接器3212返回,并且存储于集中式数据储存设备3210中。

[0608] 在广泛的系统架构领域内,存在许多公知的可能设计、模式和产品。这些包括集中式、分散式和开放式系统架构访问模型。在这里将不进一步讨论本发明所覆盖的这些实施和各种可选实施的技术运转。

[0609] 数据模型和模式

[0610] 图31提供了本发明一个实施例中的系统内的核心数据结构的简化概况。这一简化模式说明了可以通过系统的应用编程逻辑对数据进行变换的方式。它也说明了如何可以析构和存储语素数据。

[0611] 将系统的数据架构设计成在提供用于处理域特有的实体的临时数据储存器之时

使语素词典集中。

[0612] 注意,域数据可以流过系统;可以不将其存储于系统中。映射到域实体的表可以是临时数据储存设备,该临时数据储存设备继而被变换成用于域的输出数据和数据储存器。域数据储存器设备可以连同其他集中式资产一起存储或者分布到由域所有者维护的存储资源。

[0613] 在一个实施例中,应用和数据服务器(上文所述和图30中所示)可以主要地操控数据。可以在系统中的三个广义数据抽象区内组织数据:

[0614] 实体抽象层3302:其中实体是系统中的主要知识表示构建块。实体可以包括:语素3304、关键词3306、概念3308、内容节点3310和内容容器3312(由URL代表)。

[0615] 关系抽象层3314:其中实体定义由系统中所用各种实体之间的关系表示。实体关系可以包括:语素关系3316、概念关系3318、关键词-语素关系3320、概念-关键词关系3322、节点-概念关系3324和节点内容容器(URL)关系3326。

[0616] 标记抽象层3328是如下层,其中用来描述实体的项与实体本身的结构定义分离。标签3330可以包括:语素标签3332、关键词标签3334、概念标签3336和节点标签3338。可以跨过各种实体来共享标签。备选的是,可以按实体类型来划分标签。

[0617] 注意,这一简化模式决不限制一个实施例中使用的数据库模式。主要地考虑系统性能、存储和优化的问题。本领域技术人员已知,存在用以设计对这里描述的设计元素进行反映的数据库系统的许多方式。这样,在这里将不进一步讨论可以在本发明中用作实施例的各种方法、技术和设计。

[0618] 维度变换系统

[0619] 图32图示了根据一个实施例的用以执行上文描述和下文进一步描述的数据结构变换操作的系统概况。

[0620] 可以更具体地重新叙述在一个实施例中存在的上文介绍的三个广义变换过程:1)域分析和压缩,其按照复杂维度结构中的初等构造来定义的用以发现域200的结构分面;2)综合和扩展,其将域的复杂维度结构综合和扩展成维度概念分类法210,这是通过增强的分面式分类方法来提供的;以及3)管理,其通过分面式导航和编辑环境来管理维度概念分类法210内的用户交互,从而实现随时对结构(例如206和210)进行细化的复杂-自适应系统。

[0621] 分析初等构造

[0622] 在一个实施例中示意地示出了分布式计算环境600。用于集中式处理的一个计算系统601可以作为用于数据结构的变换引擎602进行操作。变换引擎可以获得来自一个或者多个域200的源数据结构202来作为其输入。变换引擎602可以包括:分析引擎204a、语素词典206和构建引擎208a。这些系统部件可以提供上文介绍和在图2中图示的分析和综合功能。

[0623] 在一个很具体的实施例中,可以将复杂维度结构编码成如下XML文件604,这些XML文件可以通过因特网606经由web服务(或者API或者其他分布信道)分布到一个或者多个计算系统以供分散式处理(例如603)。通过这一和/或其他分布和分散模式,广泛的开发者和发布者可以使用变换引擎602以创建复杂的维度结构。应用包括网站、知识库、电子商务商店、搜索服务、客户端软件、关联信息系统、分析等。

[0624] 这里注意,不应将集中式和分散式处理的这些描述与可以用来提供这些处理模式的各种集中式和分散式物理系统相混淆。这里,“集中式处理”是指用于变换处理的共享的、公共的、和/或汇集式数据和服务。“分散式处理”是指域特有的数据和服务。如本领域中公知的,存在可以被实施成实现集中式和分散式处理这一混合的多个物理系统和架构。

[0625] 利用增强型分面式分类实现的综合

[0626] 在XML文件604中具体化的复杂维度结构可以用作重新组织域内容的基础。在一个实施例中,一种增强的分面式分类方法可以用来重新组织域中的素材,其使用在XML文件604中具体化的复杂维度结构在第二计算系统603处派生维度概念分类法210。通常,第二计算系统(如系统603)可以由如下域所有者维护,这些域所有者也负责通过概念分类法210来重新组织域。下文提供并在图33中示出了由系统使用的多级数据结构相关的信息。

[0627] 在系统603的一个实施例中,可以提供用于维度概念分级210的表示层608或者图形用户界面(GUI)。客户端侧工具610(如浏览器、基于web的表单和软件部件)可以允许域终端用户和域所有者/管理员与维度概念分类法210交互。

[0628] 经由用户交互的复杂-自适应处理

[0629] 维度概念分类法210可以由各独立终端用户和域所有者定制和划界。这些用户交互可以由第二计算系统(例如603)用来向分类系统提供人类认知和附加处理资源。

[0630] 例如在XML 212a中编码的将用户交互具体化的维度分类法信息可以通过经由web服务或者其他装置进行分布来返回到变换引擎602。这允许数据结构(例如206和210)随时间而演变和改进。

[0631] 从第二系统603到变换引擎602的反馈建立了复杂-自适应的处理系统。尽管终端用户和域所有者通过维度概念分类法210在高的抽象级交互,但是用户交互可以转译成作为维度概念分类法信息的基础的初等构造(例如语素和语素关系)。通过将终端用户和域所有者交互耦合到初等构造以及将其反馈回变换引擎602,系统可以评价集合体中的交互。

[0632] 使用这一机制,可以移除在协作式分类中历史上出现的不明确性和冲突。因此,这一协作式分类方式寻求在概念级上避免其他这样的系统可能出现的个人和协作式协商。

[0633] 通过允许用户通过他们的交互来共享内容节点302和分类数据(维度概念分类法信息),用户交互也扩展可用的源数据202,这增强了分类的整体质量并且增加可用的处理资源。

[0634] 多级数据结构

[0635] 图33图示了如下装置,借助这些装置,通过连续多级抽象和维度性来复合从各源数据结构202收获的初等构造,从而创建用于各域200的维度概念分类法210。还图示了在每个域200中具体化的分散式私有数据(708、710和302)与由集中式系统用来通知为每个域而生成的分类模式的共享初等构造(语素词典)206之间的描绘。

[0636] 初等构造

[0637] 语素310和语素关系这些初等构造可以作为集中式数据存储在语素词典206中。可以跨过分布式计算机环境600(例如经由变换引擎系统601)来对集中式数据进行集中,并且使其可用于所有域所有者和终端用户,以便辅助域分类。由于集中式数据是初等的(语素性的)并且与由概念306和概念关系所代表的任何特有和私有知识的背景无关联,所以它可以在第二分散式计算系统603之间进行共享。系统601无需持久地存储包括各域中所含唯一信

息的这些初等构造的唯一表达和组合。

[0638] 语素词典206可以在语素属性702的表集中存储每个语素310的属性。语素属性702可以引用由变换引擎602的分析过程(如下文进一步所述)使用的结构参数和统计数据。可以在集合体中将语素关系排序成语素分级402。

[0639] 维度分面式输出数据

[0640] 域数据储存设备706可以存储由变换引擎系统601根据源数据结构202,并且使用语素词典206而派生的域特有的数据(复杂维度结构210a)。在一个实施例中,可以用XML形式来存储域特有的数据。

[0641] 每个域数据储存设备706中基于XML的复杂维度结构210a可以包括:域特有的关键词分级710、内容节点302的集合和概念定义708的集合。关键词分级710可以包括关键词关系的分级集合。XML输出本身可以编码为分面式数据。分面式数据代表作为其结构分面的源数据结构202的维度性和就分面属性而言的源数据结构202的内容节点302。这一方式允许域特有的资源(例如系统603)来将复杂的维度结构210a处理成更高的抽象级(如维度概念分类法210)。

[0642] 可以使用复杂维度结构210a作为用以管理内容节点302之间关系的组织基础。新的组织原则集然后可以适用于初等构造进行分类。组织原则可以包括如下文详述并在图20-22中所示的增强的分面式分类方法。

[0643] 增强的分面式分类方法可以适用于复杂维度结构210a。还可以应用其他更简单的分类方法,并且可以按照需要根据复杂维度结构210a来创建其他数据结构(简单或者复杂)。在一个实施例中,可以使用显式地表示分面式分类的输出模式。可以使用其他输出模式。可以使用各种数据模型来代表为各域生成的分面式分类。可用的分类方法与正在被分类的数据结构类型紧密关联。因此,用于分类的这些可选实施例可以与上文讨论的可选维度性实施例直接地相联系。

[0644] 域数据储存设备706中包含的数据实体(例如708、710)包括对存储于语素词典206中的初等构造的引用。以这一方式,在创建用于各域200的维度概念分类法210之后,可以对其进行重新分析,以便适应改变。当域所有者希望更新其分类时,可以将域特有的数据重新装载至分析引擎204a中用于处理。可以对域200进行实时(例如,通过经由XML 212a来通过终端用户交互)分析,或者通过(以队列方式)周期性更新对域200进行分析。

[0645] 共享数据与私有数据

[0646] 维度知识表示模型的一个优点在于,由系统用来将域处理成复杂维度结构210a的私有域数据和共享数据的清晰分离。数据分离提供了分布式计算的益处,比如掌控的应用服务提供商(ASP)处理模型、利用如上文所述环境这种效用计算环境的机会、或者作为服务的软件(SaaS)应用递送模型。在这些模型之下,第三方可以向域所有者赋予变化引擎服务。域所有者因此可以利用这些类型的模型所提供的规模经济。

[0647] 域所有者的域特有的数据可以在各种存储模型之下安全地加以掌控(例如经由ASP),因为它可与共享数据(即语素词典206)和其他域所有者的私有数据分离。备选地,域特有的数据可以由域所有者掌控、从共享数据在物理上加以移除。

[0648] 在这一分布式知识表示模型之下,域所有者可以受益于集中式知识变换服务的经济优点和专门化中受益以及从集中式分类数据的“汇集智慧”。然而,通过保持必需的域特

有的数据与这些集中式服务和数据资产分离,域所有者可以在整个用户群体的共享知识(例如语素词典)上构建而无需有损于他们的独有知识。

[0649] 企业设置内的知识仓库和内部网在专用知识域的背景内提供了这一共享汇集知识应用的一个例子。目前,在需要维护私有知识以求竞争优势的情况下,公司面临着汇集知识的经济优点与开放式协作之间的权衡。这里描述的系统允许这一类封闭式信息域,其不但从这里描述的集中式知识表示和变换服务以及如这里描述的语素词典中的群体数据资产中受益,同时又保持它们的综合知识和域特有的数据资产私有。

[0650] 分布式计算环境

[0651] 在一个实施例中,可以将构建引擎分布为在开源平台上运行的软件应用。一个这样的开源平台是包括Linux™、Apache™、MySQL™以及可以包括Perl、PHP、Python和其他语言的编程技术在内的“LAMP”技术栈。通过这样的应用,可以在域所有者的分布式物理系统上直接读取构建引擎的综合规则的多个副本。在这一模型之下,得到了运行集中式处理规则的分布式物理系统(因为构建引擎的各副本具有相同指令)。

[0652] 使用这一方式,用于对每个域的复杂维度结构进行综合的升级成本分布在了每个域所有者的资源中。以相似的方式,可以将构建引擎分布为轻量级客户端侧应用,该应用按照这些应用的终端用户的需要来综合复杂维度结构。

[0653] 除了在域所有者和终端用户的系统上直接运行这些分散式系统的机会之外,诸如Amazon Web Services™(AWS)的效用计算系统提供了一种用于集中式构建引擎规则的经济分布机制(运行构建引擎的可视化实例的直接成本可以多于以下,即跨域所有者的异型环境来分布和支持构建引擎的间接成本的偏移量)。取代物理地分布构建引擎的副本,可以在效用计算环境内提供虚拟化构建引擎应用。

[0654] 例如在AWS内,将要创建用于构建引擎的映像,并且将其上传到AWS弹性计算云服务(EC2)的虚拟化环境。EC2可以提供一或者多个虚拟服务器环境。AWS“映像”实质上是虚拟服务器的盘映像;“实例”是基于该盘映像的操作虚拟服务器。将提供在虚拟服务器上运行的构建引擎的新实例以按需处理域和适应用户活动。

[0655] 在这一分散式环境(以及许多其他环境)中,可以对域特有的数据与构建引擎进行解耦合。在AWS内,EC2可以用于处理、简单存储服务(S3)可以用于数据存储,而简单队列服务(SQS)可以用来协调上文介绍的并在下文更具体讨论的跨分析和复杂-自适应反馈的EC2、S3和其他集中式服务的消息传送。

[0656] AWS S3服务可以存储和分发对用于域的维度复杂结构进行编码的分面式数据集。这些域特有的分面式数据集可以在处理构建引擎规则的多个虚拟服务器之间共享。

[0657] 综合的概念关系可以存储于这一分散式环境中。可以综合构建请求,并且将构建请求并行发送到终端用户系统和S3。随后,可以从S3中的概念关系高速缓存满足域先前请求的参数匹配的综合请求,或者如果需要更新则可以由构建引擎直接地生成这些综合请求。同样重要的是,综合的关系将可用作如上所述用于集中式分析引擎服务中的下一分析循环的反馈。

[0658] 本领域技术人员将认识到,在分布式计算机领域中可以进行许多架构改进和发展。这种类型改进的示例例如:跨多个虚拟机的并行化、以及跨域的负载平衡、以及用户活动性。

[0659] XML模式和客户端侧变换

[0660] 分面式输出数据可以编码为XML并且通过XSLT来表现。可以用许多不同方式重新组织和代表分面式输出(例如参考公布的XFML模式)。用于代表分级的可选输出是可用的。

[0661] 一个实施例中,XML变换码(XSLT)在用来呈现表示层。由系统管理的所有信元(如果通过系统用信道发送分布式内容,则包括该分布式内容)可以通过XSLT来表现。

[0662] 客户端侧处理是一个实施例用以将数据馈送连接到系统表示层的过程。这些类型的连接器可以用来将来自应用服务器的信息输出到使用结构信息的各种介质。可以通过XSLT来处理来自应用服务器的XML数据以便在web页面上呈现。

[0663] 本领域技术人员将认识到,XML技术和类似呈现技术将在本发明的服务中提供的当前和将来功能。除了基本的发布和数据呈现之外,XSLT和类似技术还可以提供一定范围的编程机会。复杂的信息结构(例如由系统创建的信息结构)可以提供可作用信息(如数据模型)。软件程序和代理可以对关于表示层的信息起作用,以提供复杂化交互性和自动化。这样,由系统的核心结构优点提供的本发明范围可以远远超出简单发布以外扩展。

[0664] 本领域技术人员也将认识到,可能用于设计这些XML和XSLT位置的可变性。例如,文件可以局部地存储于终端用户的计算机上或者使用web服务来生成。ASP代码(或者类似技术)可以用来插入由系统管理的关于分布式表示层的信息(比如第三方发布者或者软件客户的web页面)。

[0665] 作为另一例子,包含来自系统的核心结构信息的XML数据馈送可以与系统组织的分布式内容组合。本领域技术人员将认识到,用以将这两类数据解耦综合单独数据馈送的机会。

[0666] 用于存储和分布这些呈现文件和数据馈送的这些和其他架构机会在本领域中公知的,因此在这里将不进一步讨论。

[0667] 用户界面

[0668] 以下部分提供与用于上文讨论的系统操作的各种用户界面有关的实施细节。这些操作是:查看维度概念分类法;提供在动态综合模式下的综合参数;以及编辑维度概念分类法。本领域技术人员将认识到,可以在上文讨论的系统操作服务中实施的各种可能用户界面。这样,用户界面实施的图示和描述决不限制本发明的范围。

[0669] 维度概念分类法查看器

[0670] 图34提供了用于终端用户查看和浏览的维度概念分类法呈现UI的主要部件的实例的截屏。

[0671] 内容容器2600可以保持域中的各类内容以及形成用于维度概念分类法的表示层的结构链接和概念定义。一个或者多个概念定义可以与容器中的内容节点相关联。如这里所述,系统可以管理系统中登记的任一类信元以及URI和用来计算维度概念关系的概念定义。

[0672] 在一个实施例中,可以将通常与传统线性(或者平面)信息结构相关联的用户界面设备进行复合或者堆叠,以表示复杂维度结构中的维度性。

[0673] 复合传统Web UI器件(如导航条、目录树2604和面包屑路径2602)可以用来示出信息架构中各种节点处的维度交集。与活跃内容节点2606相交的各维度轴(或者分级)可以表示为单独分级,一个分级用于一个相交轴。

[0674] 结构关系可以由从活跃内容容器到域中相关内容容器的指针(或者链接)来定义。这可以提供如维度概念分类法规定的在活跃容器与相关容器之间的多个结构链接。结构链接可以通过各种方式来呈现,这些方式包括概念的完全上下文呈现、仅显示活跃轴上关键词的概念的过滤呈现、内容节点标签的呈现等。

[0675] 结构链接可以提供用于维度概念分类法内如下内容节点2608的背景,这些内容节点被组织在一个或者多个关系类型(例如父、子或者兄弟)内优先级化的内容节点分组。

[0676] XSLT可以用来将结构信息呈现为网站上的导航路径,该路径允许用户将结构分级导航到与活跃容器相关的容器。作为网站上的导航器件的这一类结构信息呈现可以存在于系统的多数基本应用之中。

[0677] 这些和其他导航惯例在本领域中是公知的。

[0678] 动态综合用户界面

[0679] 在图35中示出了并入用户界面控制以提供动态综合操作(如上所述)的用户界面。

[0680] 用户界面可以包括如下用户界面控件,用户可以利用这些用户界面控件来指定:活跃概念定义3602、活跃轴定义3604和活跃域3606。用于指定活跃概念定义和活跃轴定义的控件可以包括:用于将概念定义规定为关键词并且启动编辑操作和基于文本的搜索(未示出)的链接(已示出)。

[0681] 在一个实施例中,用户可以从现有概念分级3608内布置的概念定义集中选择活跃概念定义。活跃概念定义的这一选择可以是基于先前执行的动态综合操作,以提供用于维度概念分类法的全局导航结构。

[0682] 在另一实施例中,为了指定活跃概念定义,用户可以将查询键入到文本框(未示出)中。可以针对与域关联的实体标签集来处理查询。在键入它们时,可以基于与域中的概念、关键词和语素这些其他实体所关联的标签之间的比较来赋予建议列表(上文更具体讨论了提取方法)。使用这些工具,用户可以基于域特有的标签的定制词汇表,来从赋予的建议中选择概念定义。

[0683] 可以使用活跃概念定义的一个或者多个属性的列表或者用户可能希望组装的属性的任何组合(如上文在综合操作的讨论之下所述)来指定轴定义。以来自用于动态综合操作的候选集内的属性的分析为基础的“标签云”3610可以是一种用于提供可能轴定义的总览的手段。例如,可以使用候选集中的最普遍关键词的计数作为基础,该基础既用于选择用于呈现的关键词子集,又用于基于关键词总计数来改变关键词标签的字体大小。

[0684] 在这一实施中,用户可以通过从跨屏幕的顶部来定位的标记集中选择来选择活跃域。

[0685] 为了控制该处理和所得综合输出的范围,用以限定如上所述综合参数的控制可以包括:作为滑块3610的分离程度和对作为链接3612而返回的内容节点的数目的限制(在这一实施例中,对于显示的内容节点的数目的限制与对返回的概念的限制存在联系。备选的是,可以移除对概念和内容节点的限制的联系,以在呈现中提供更多灵活性)。将可以用来显示或者隐藏虚拟概念的手段图示为复选框转换控件3614。

[0686] 维度概念分类法概括器

[0687] 维度概念分类法的视图可以通过上述用户界面呈现给用户。出于说明的目的,假设在回顾分类之后,用户希望重新组织它。从系统的角度来看,这些交互将在复杂-自适应

系统内生成显式用户反馈。

[0688] 图36图示了在一个实施例中可以提供这些交互的概括器用户界面。它指示设备改变节点2702在结构2704中的位置,并且编辑在各节点2076的容器和概念定义分配。

[0689] 在一个实施例中,使用客户端侧控制,用户可以移动分级中的节点,以重新组织维度概念分类法。在这样做时,用户可以在节点之间建立新的父-子关系。

[0690] 在编辑节点的位置时,可以使下层语素之间新的成组关系相关。这由此可能需要重新计算以确定所推断维度概念关系的新集合。这些变化可以排队等候以计算通过概念关系来推断的新语素关系。

[0691] 可以将变化存储为共享维度概念分类法(下文称为群体概念分类法)的特例,以满足用户的个性化需要(见下文关于个性化的更多细节)。

[0692] 本领域技术人员将认识到,存在可以向终端用户呈现多维度信息结构并且提供交互性的许多方法和技术。例如,多变量形式可以用来允许用户同时沿着许多不同维度查询信息架构。比如“数据透视表(pivot table)”这样的技术可以用来在信息结构中保持一个维度(或者变量)恒定而改变其他变量。软件部件如基于ActiveX和Ajax的部件可以嵌入于web页面中以提供与下层结构的交互性。可视化技术可以提供数据的三维视图。这些和其他变化对于本领域技术人员而言将是清楚的并且没有限制本发明的范围。

[0693] 本领域技术人员将认识到,本发明可以采用许多形式并且这样的形式在要求保护的本发明范围内。因此,所附权利要求的精神和范围不应限于这里所含特定版本的描述。

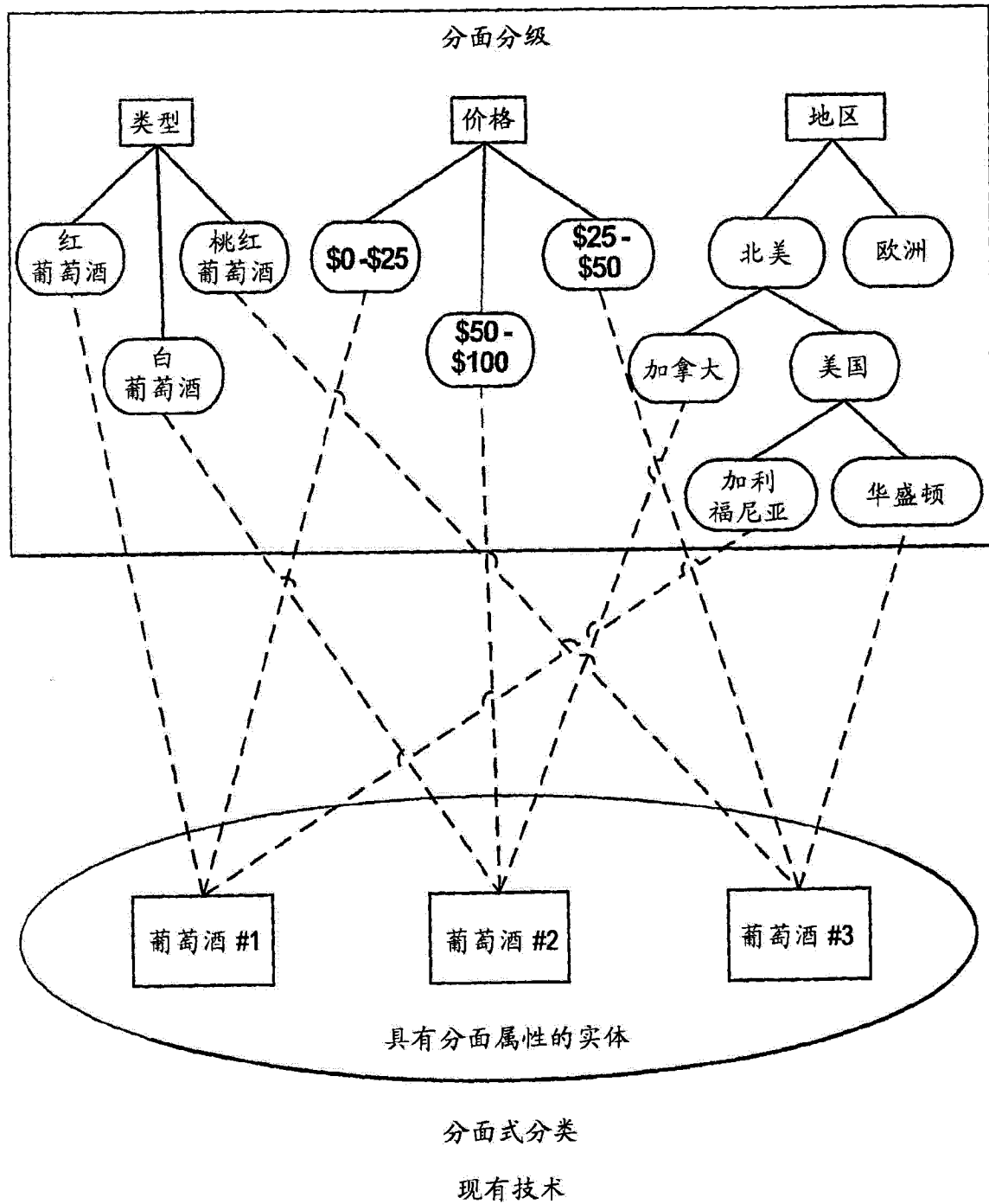


图1

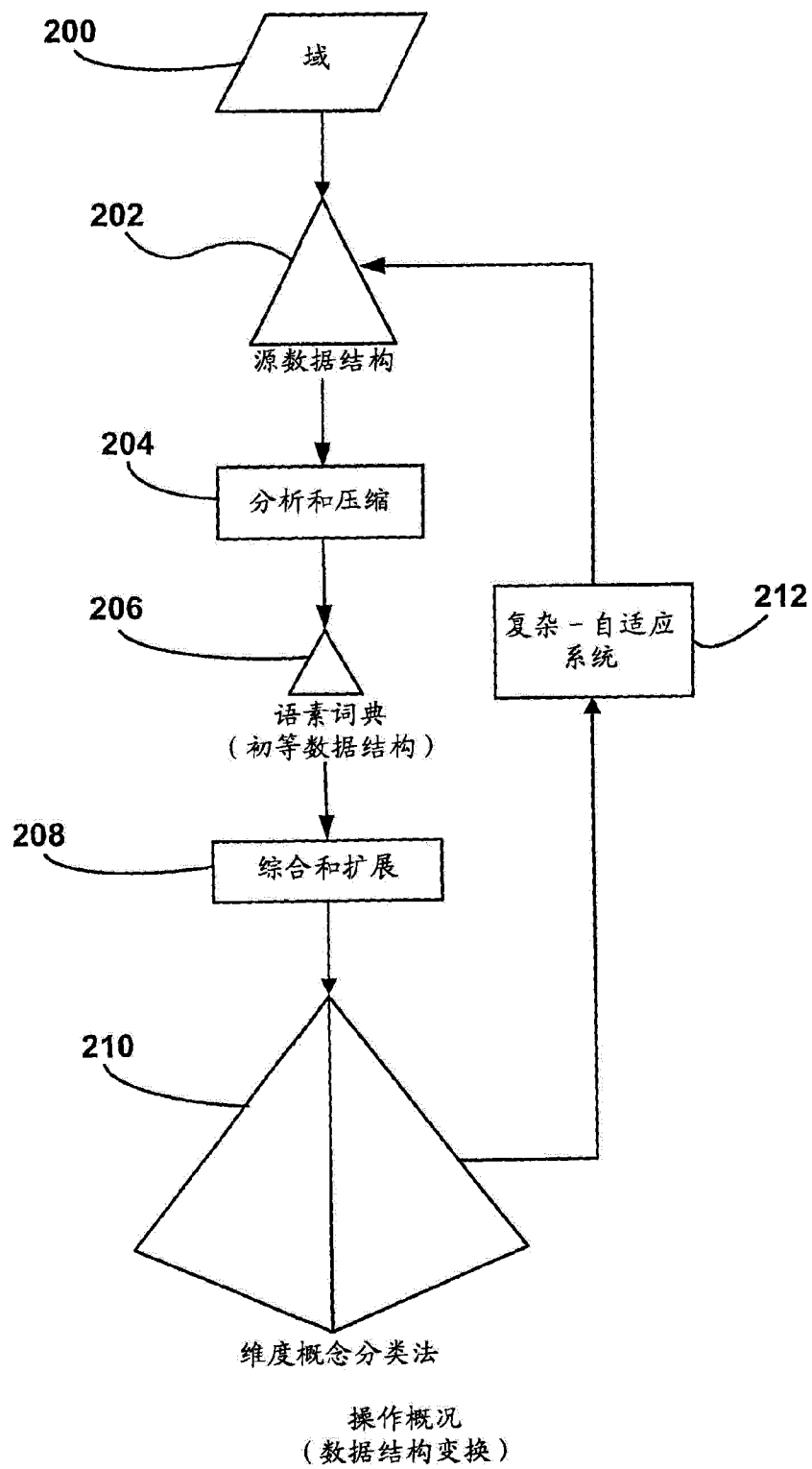
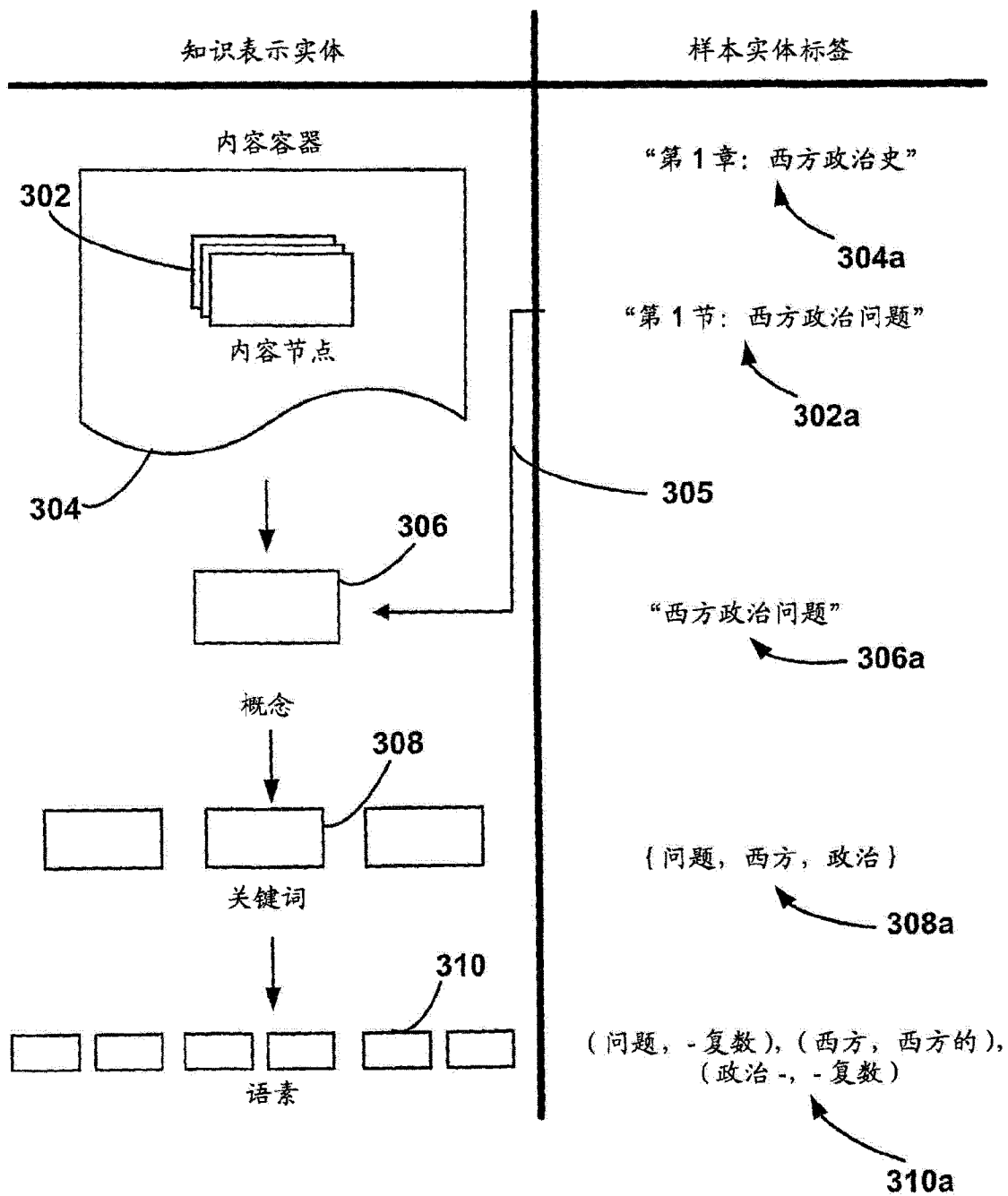


图2



知识表示模型

图3

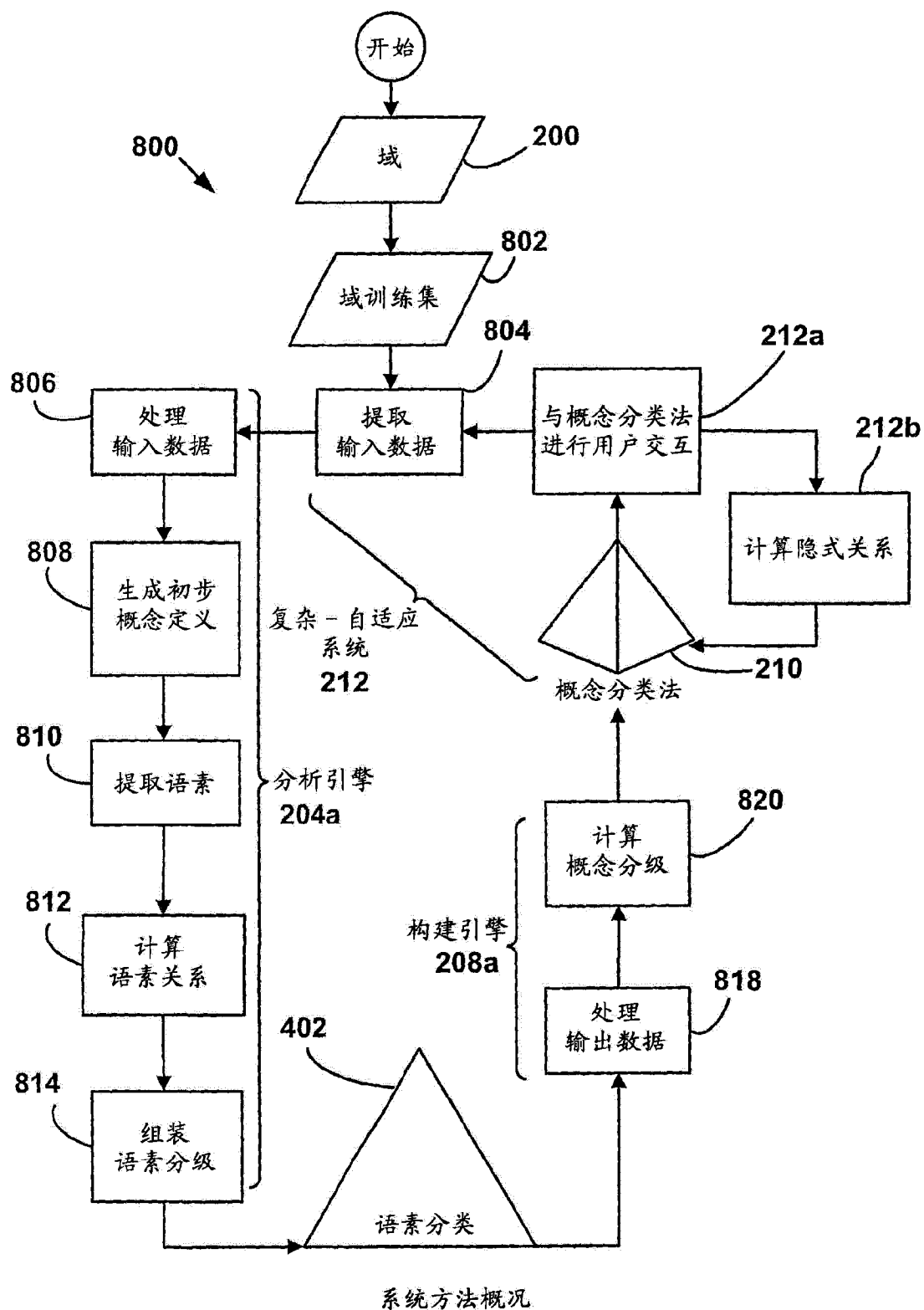


图4

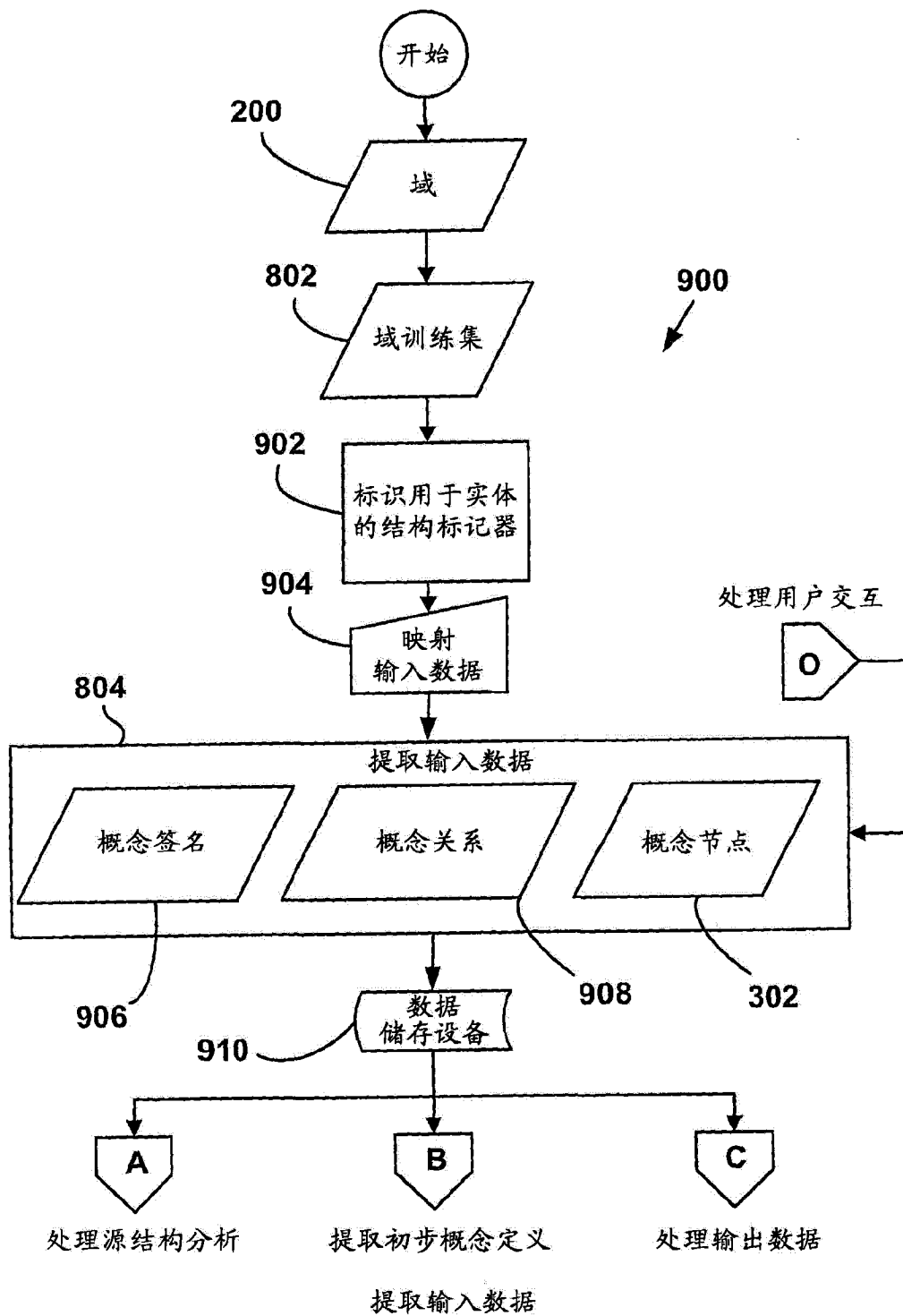


图5

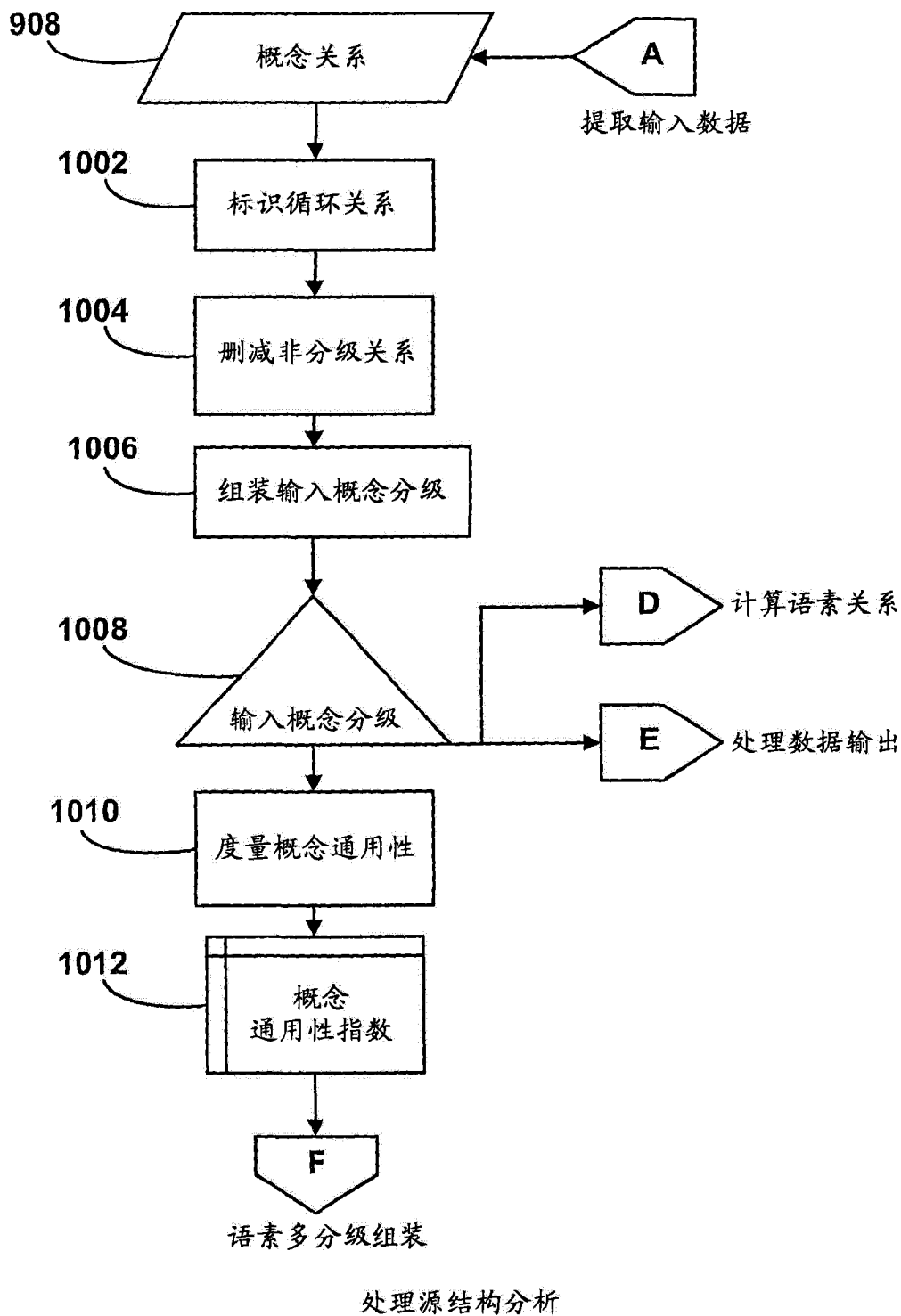


图6

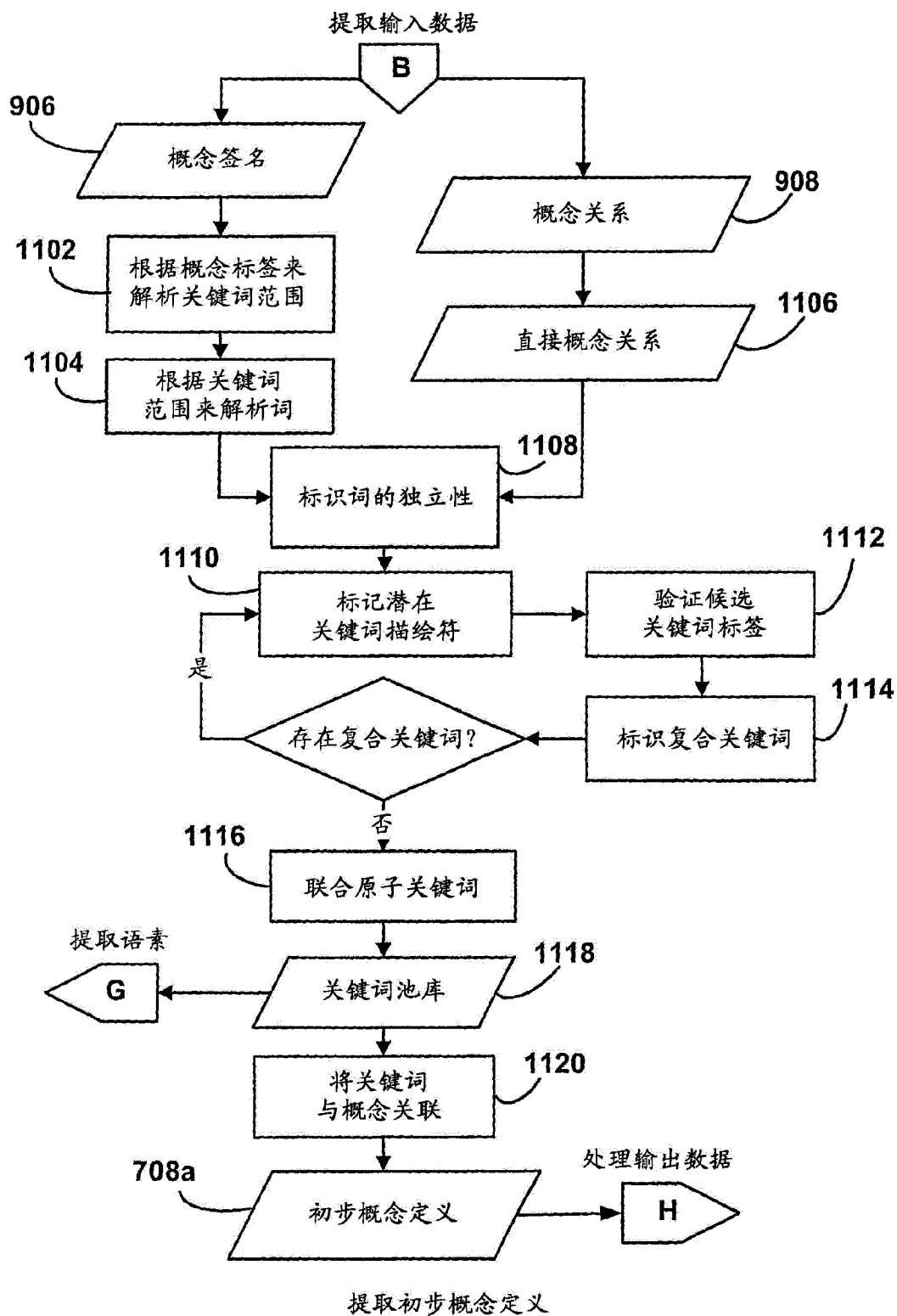


图7

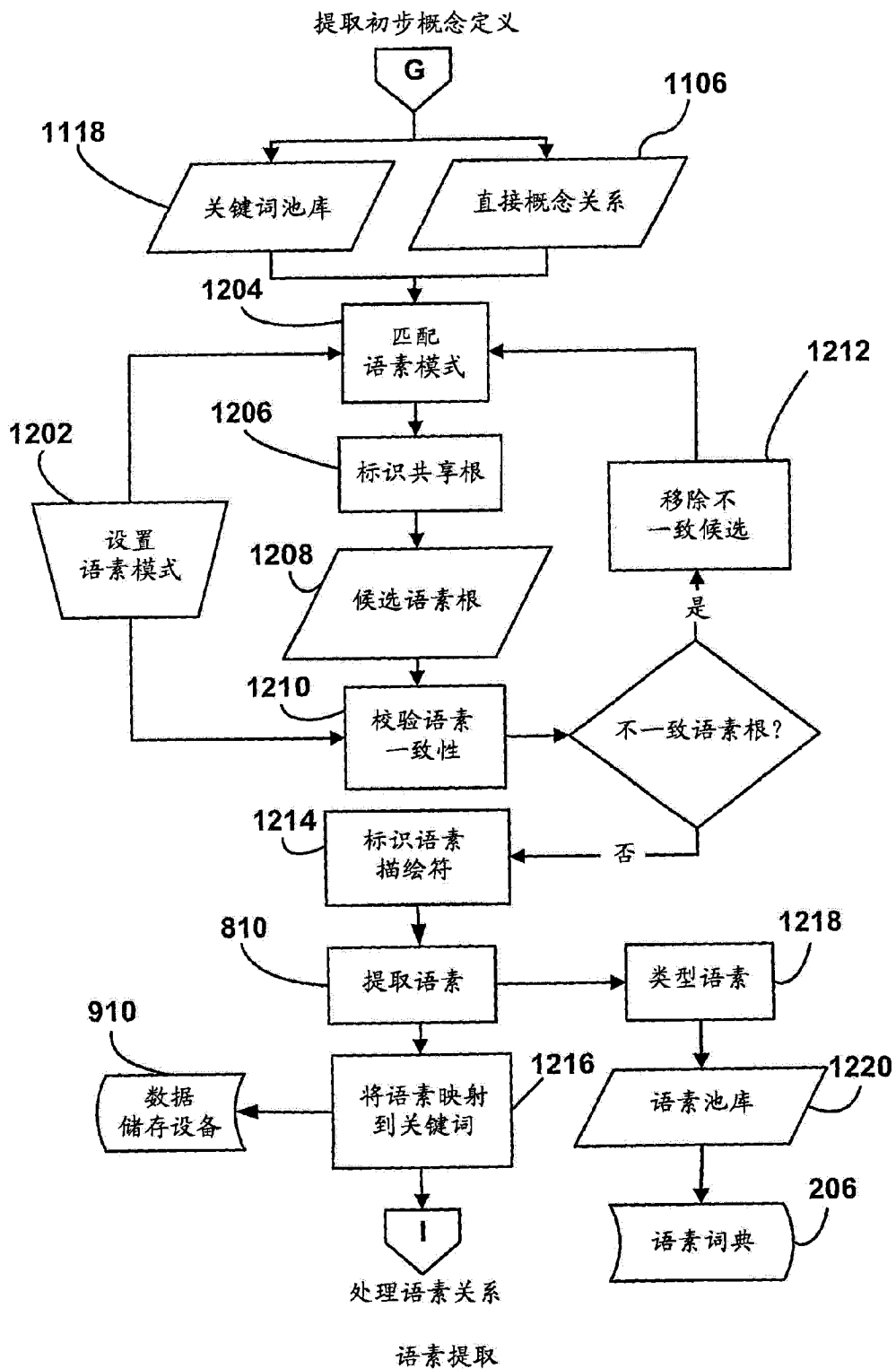


图8

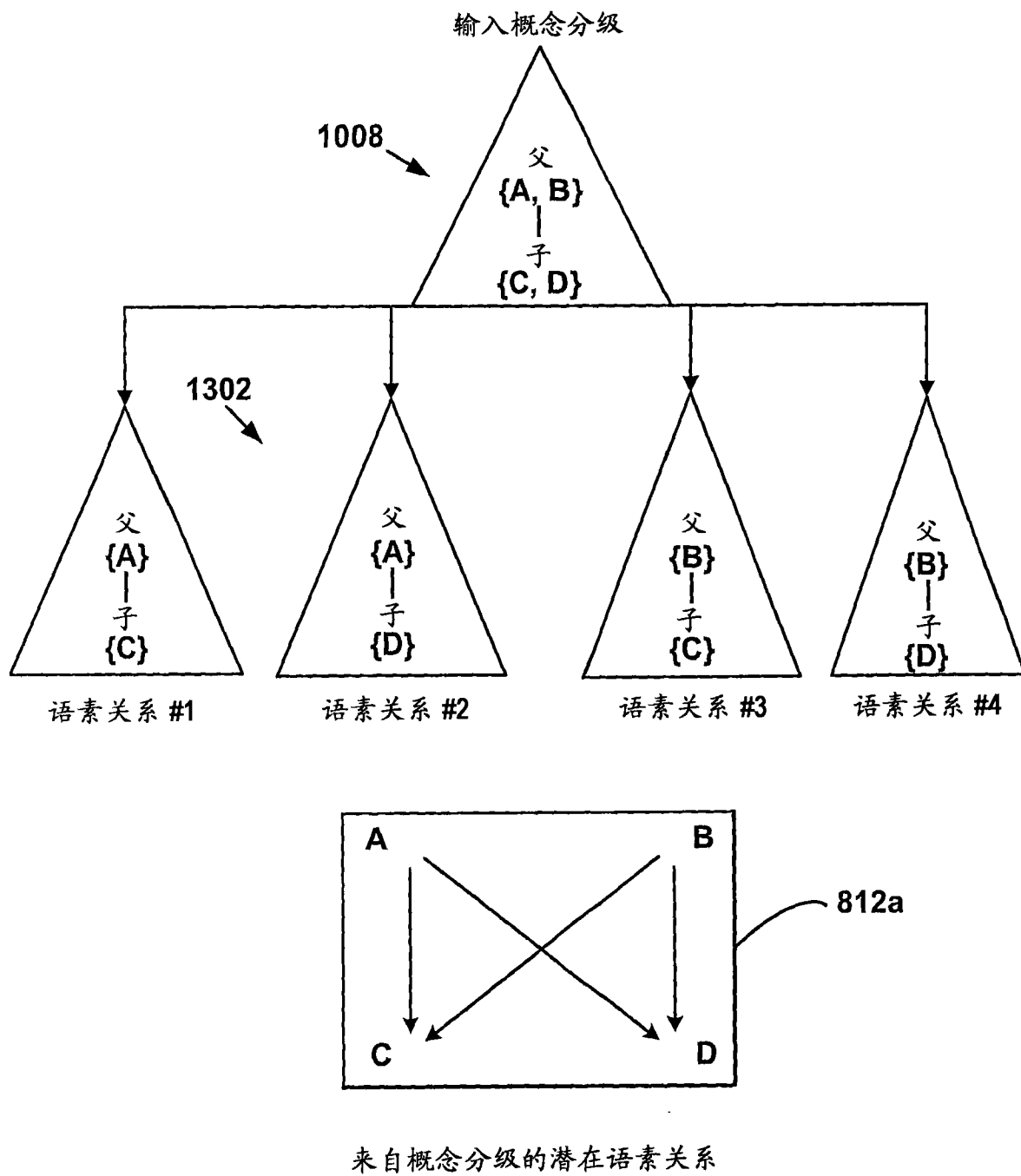
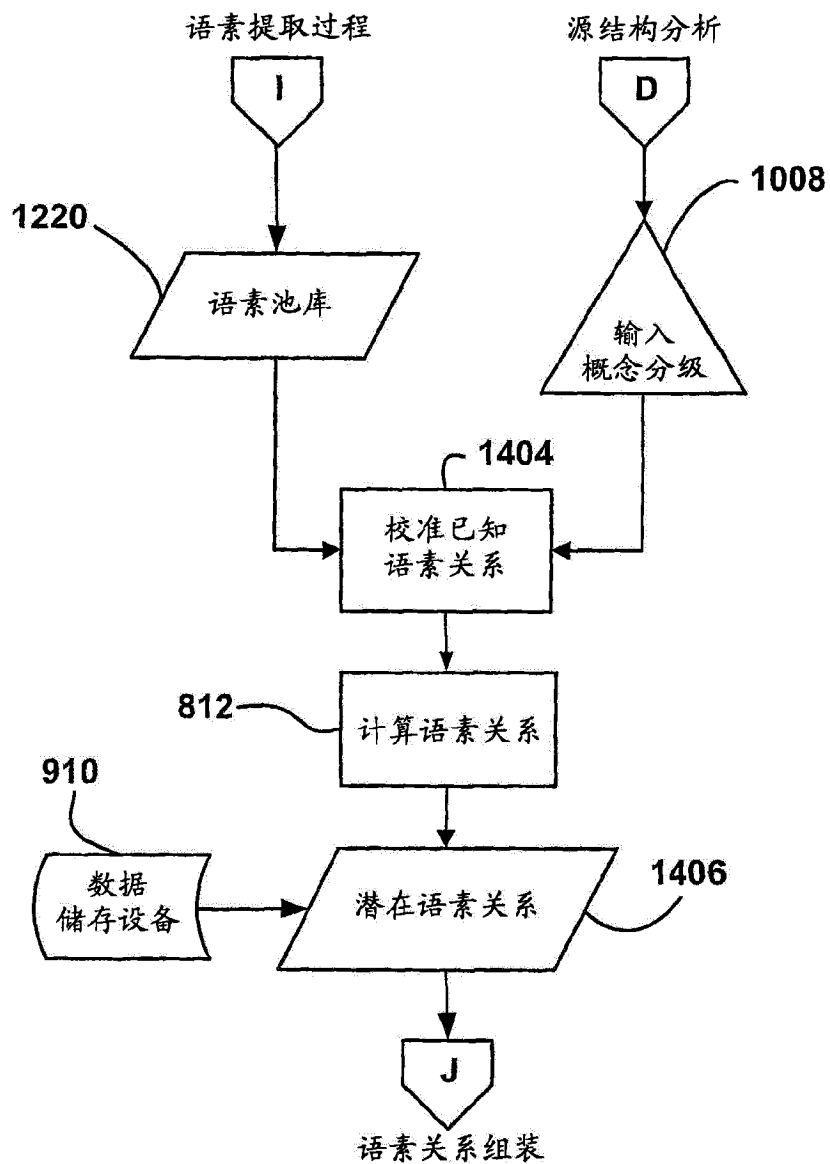


图9



计算潜在语素关系

图10

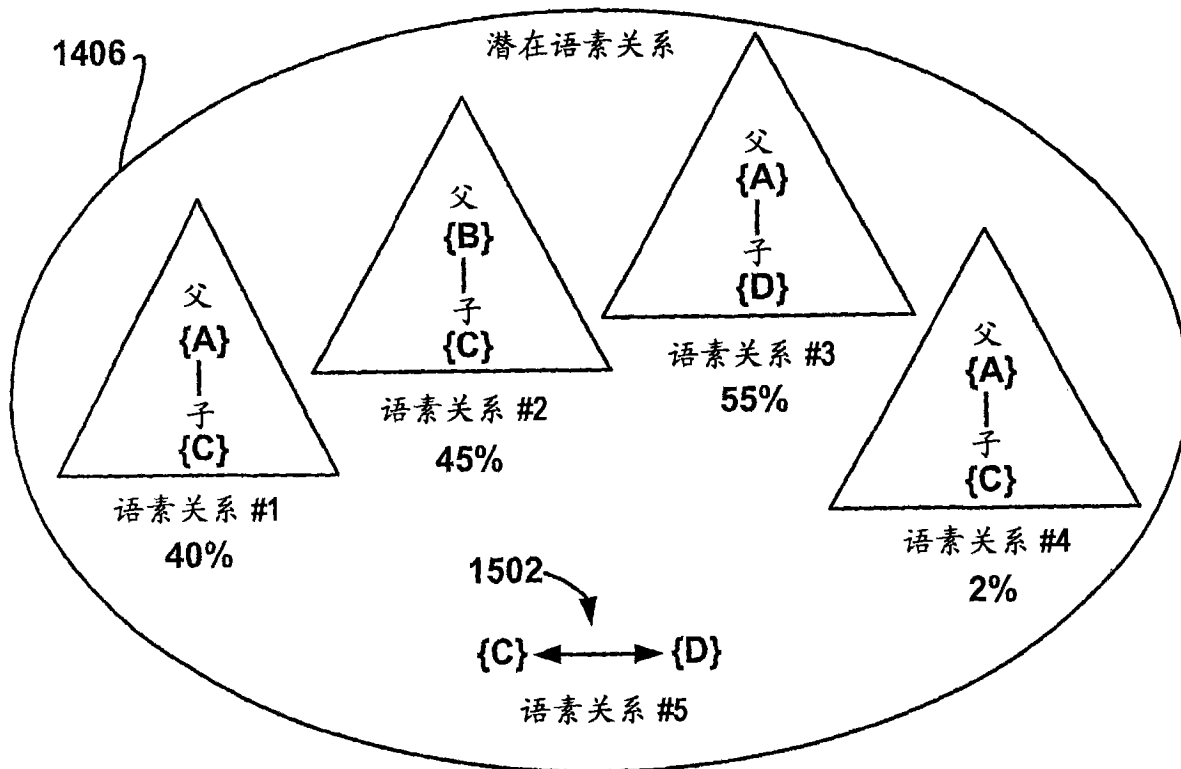


图11A

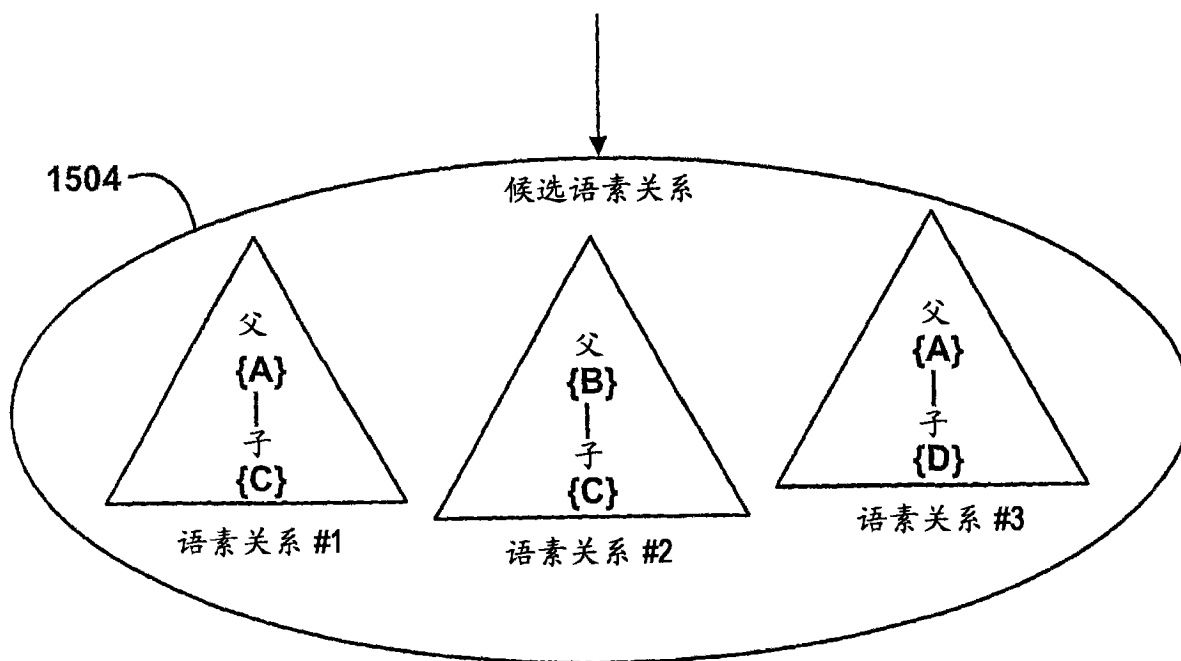


图11B

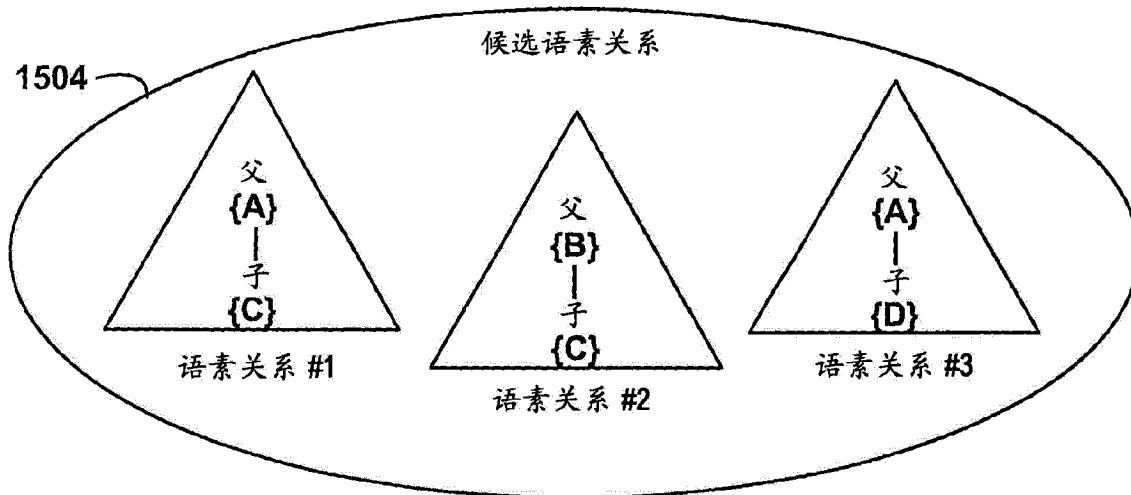


图 15B

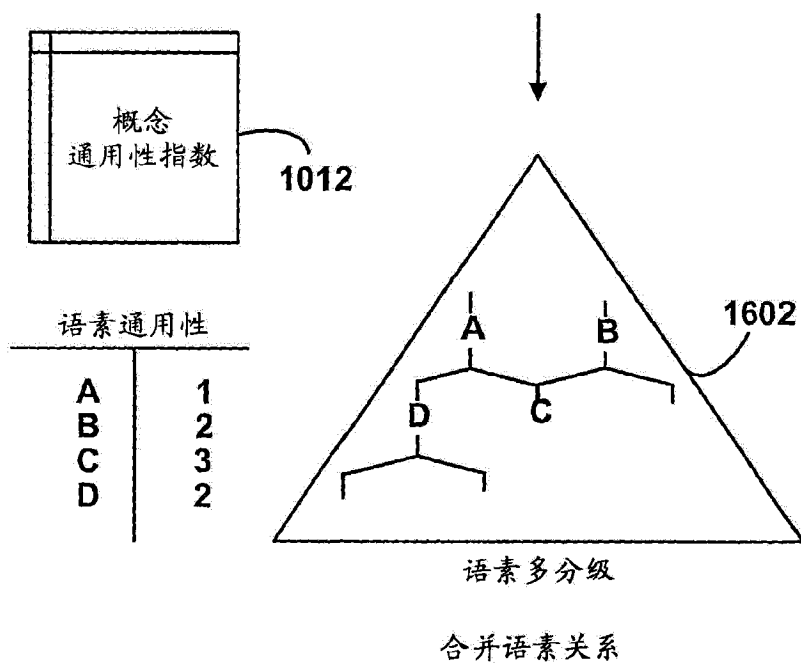


图12

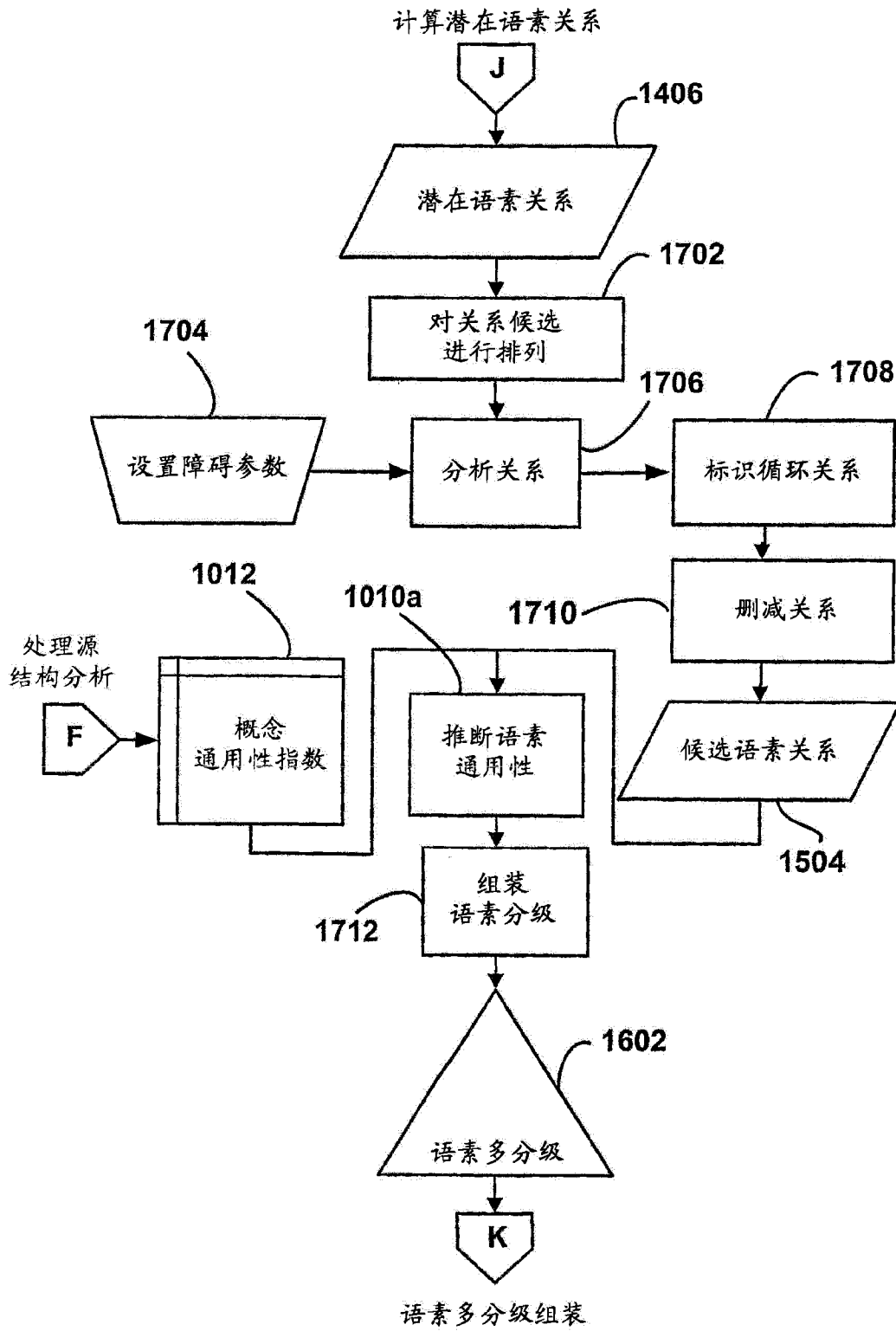
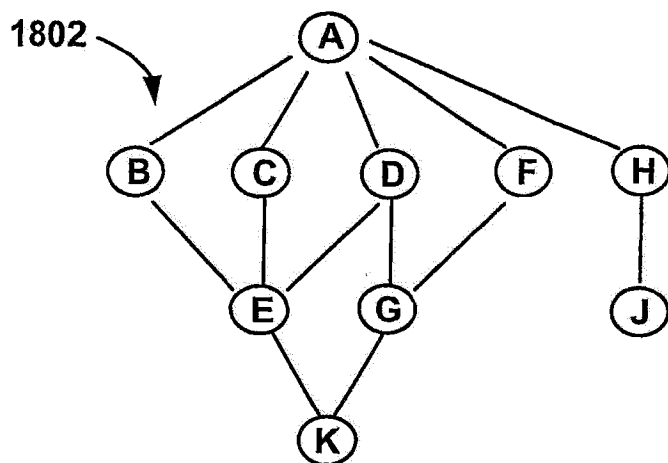
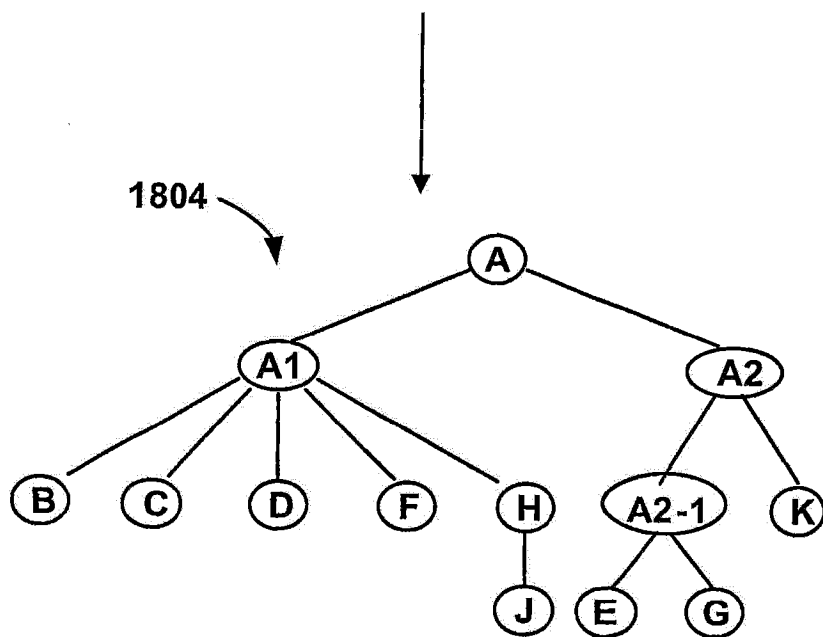


图13



归因之前的语素多分级

图14A



归因之后的语素分级

图14B

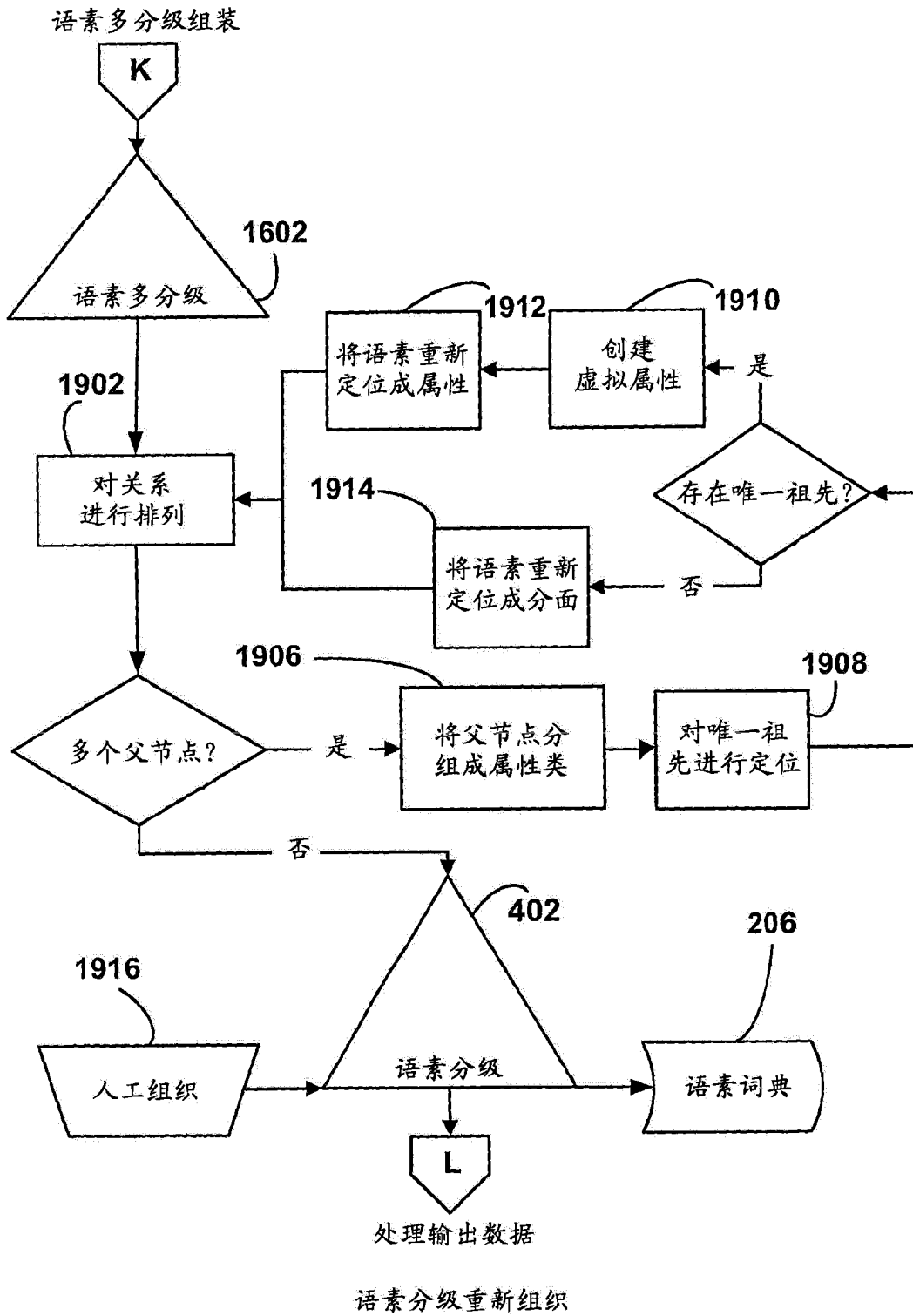
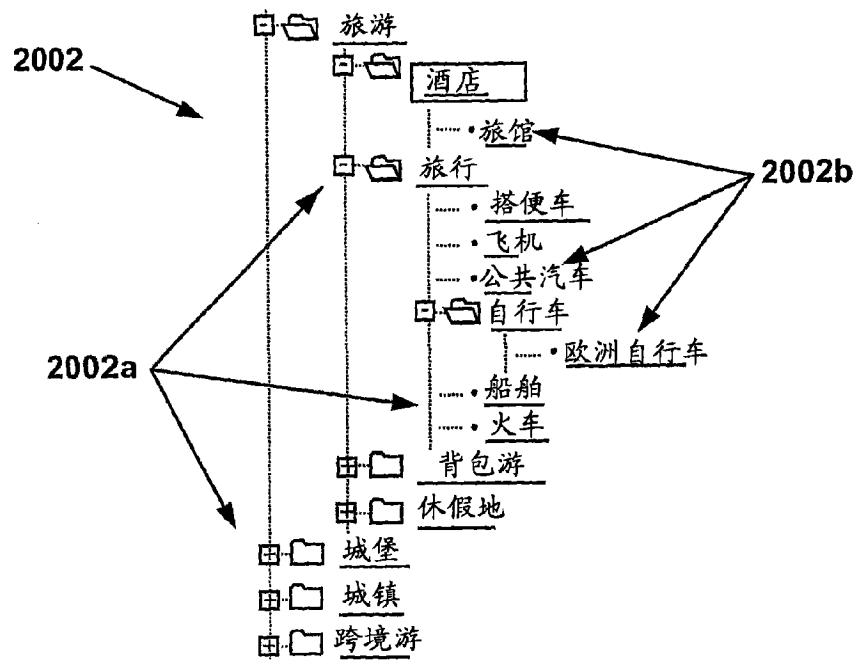
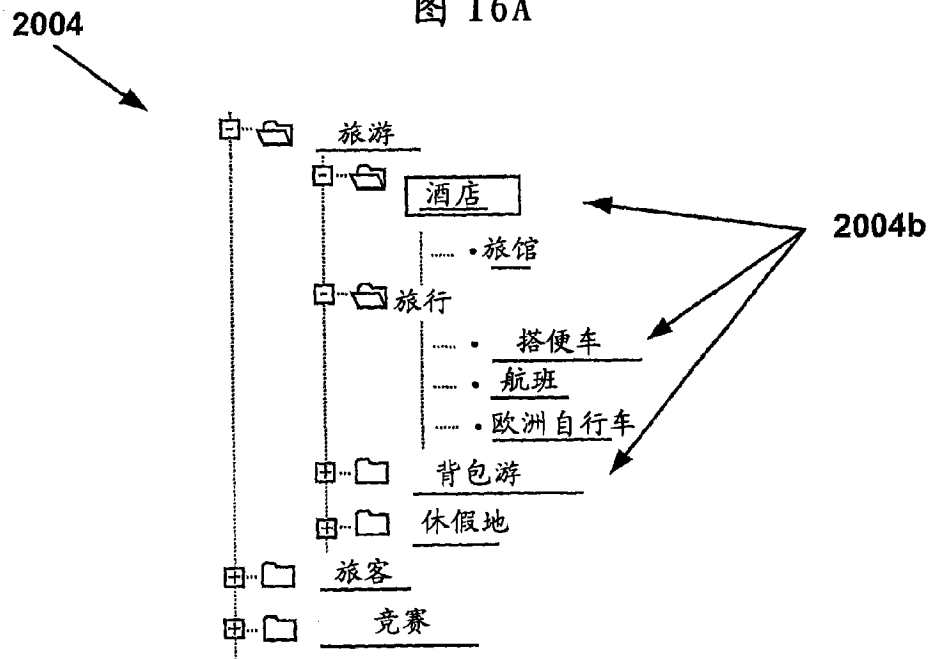


图15



语素分级

图 16A



关键词分级

图 16B

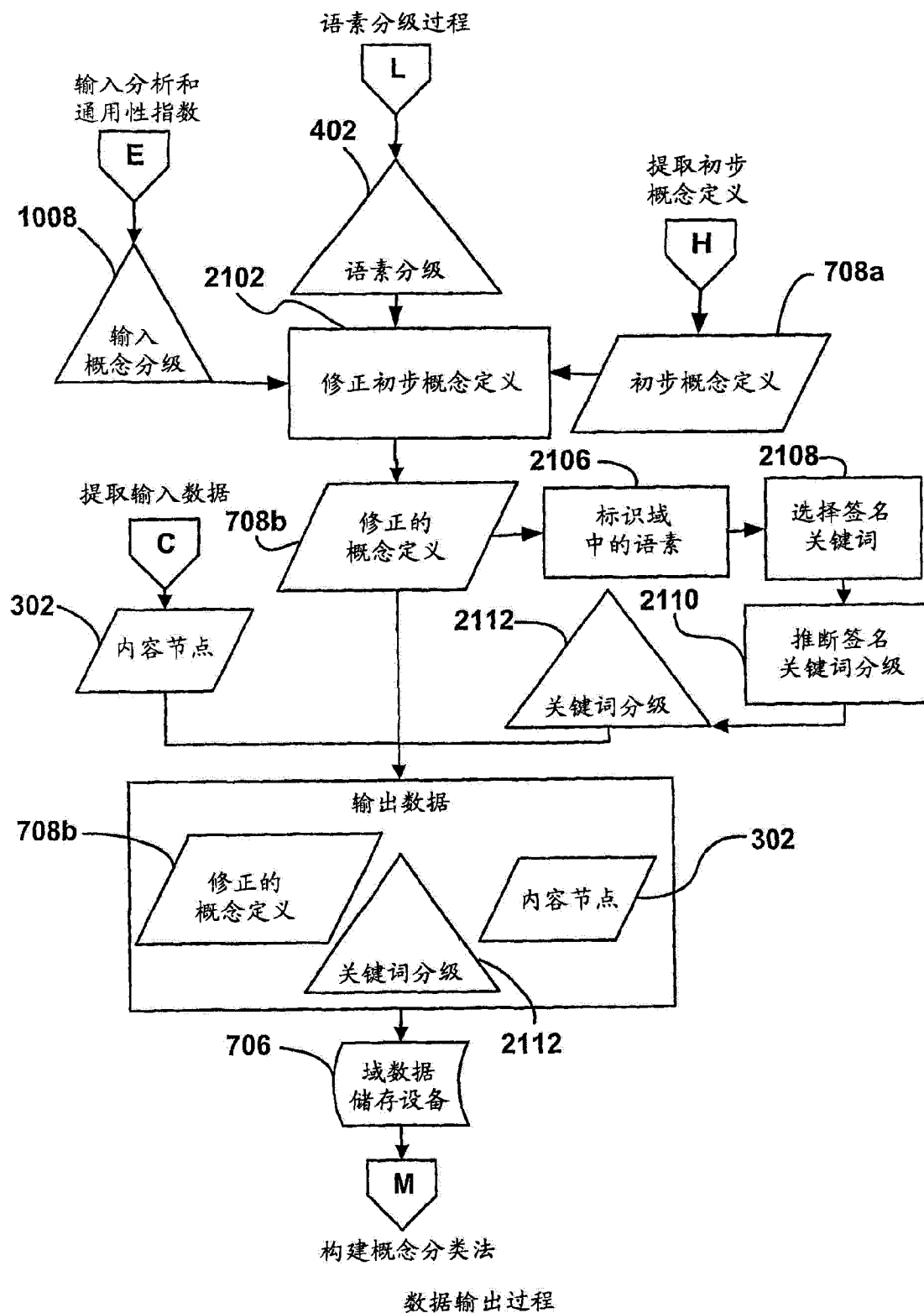


图17

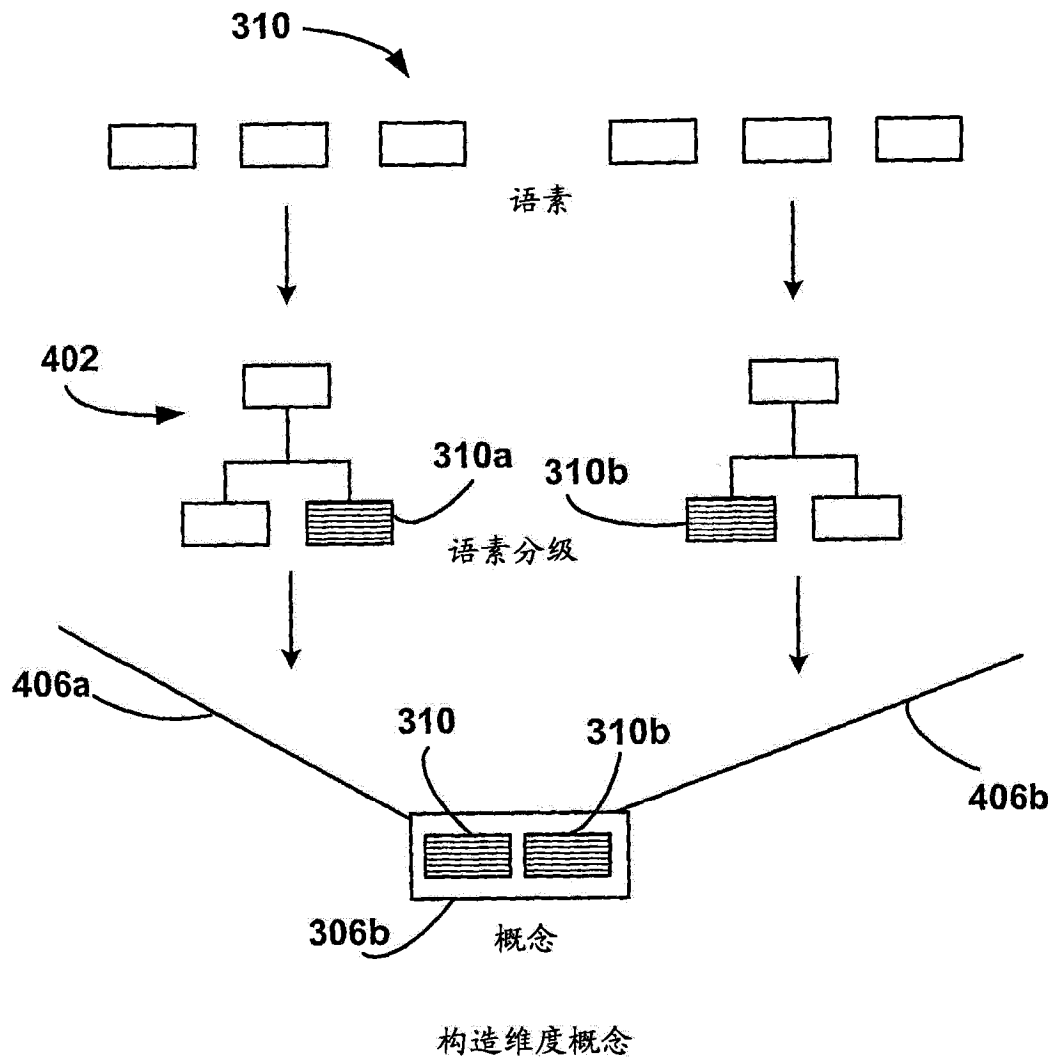
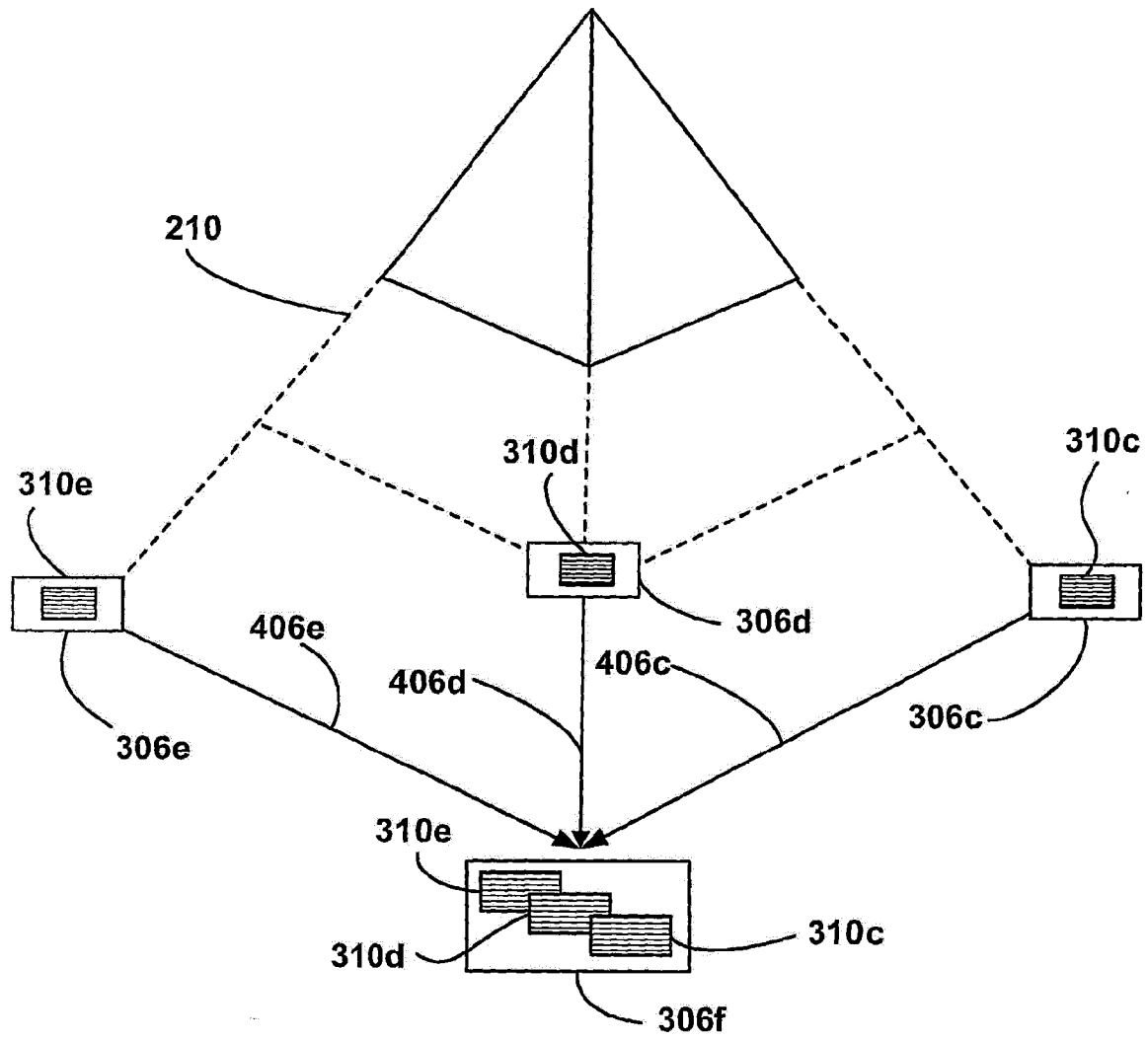


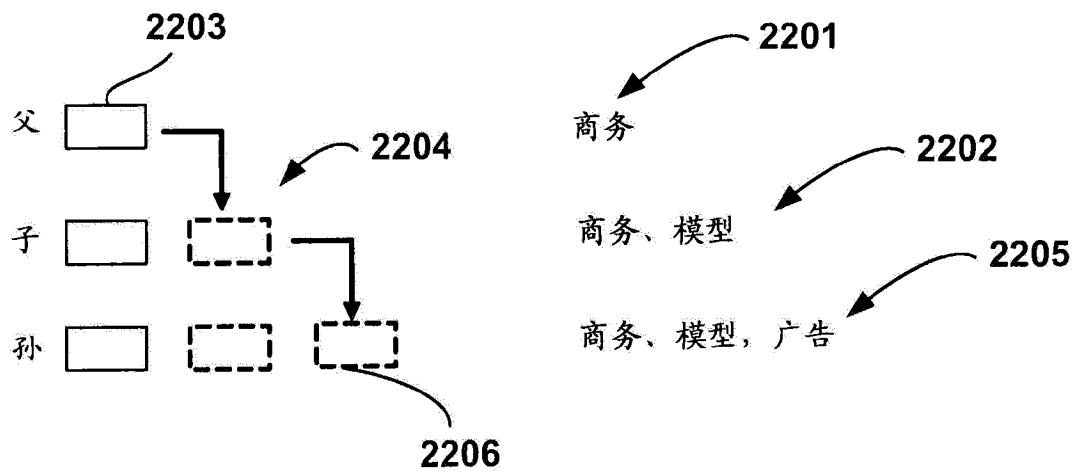
图18



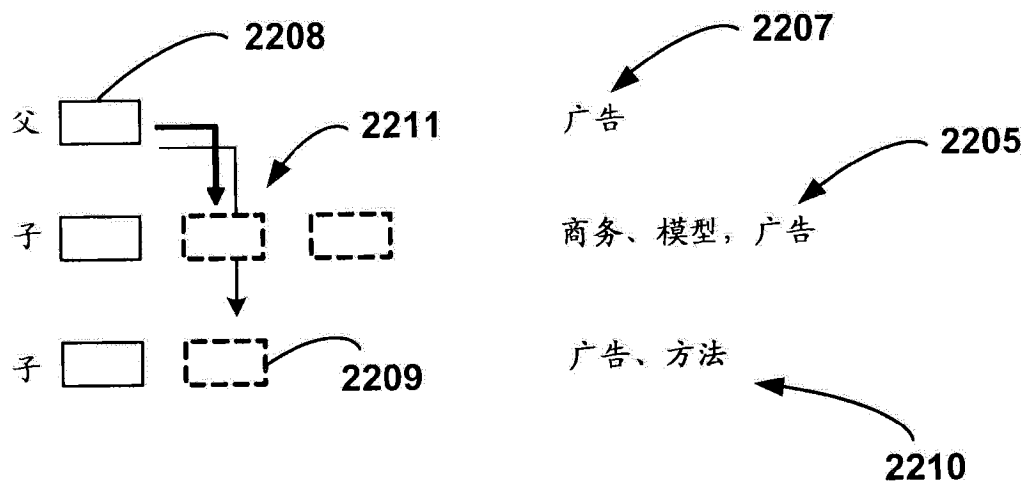
构造维度概念分类法

图19

概念分级 #1



概念分级 #2



构造隐式概念关系

图20

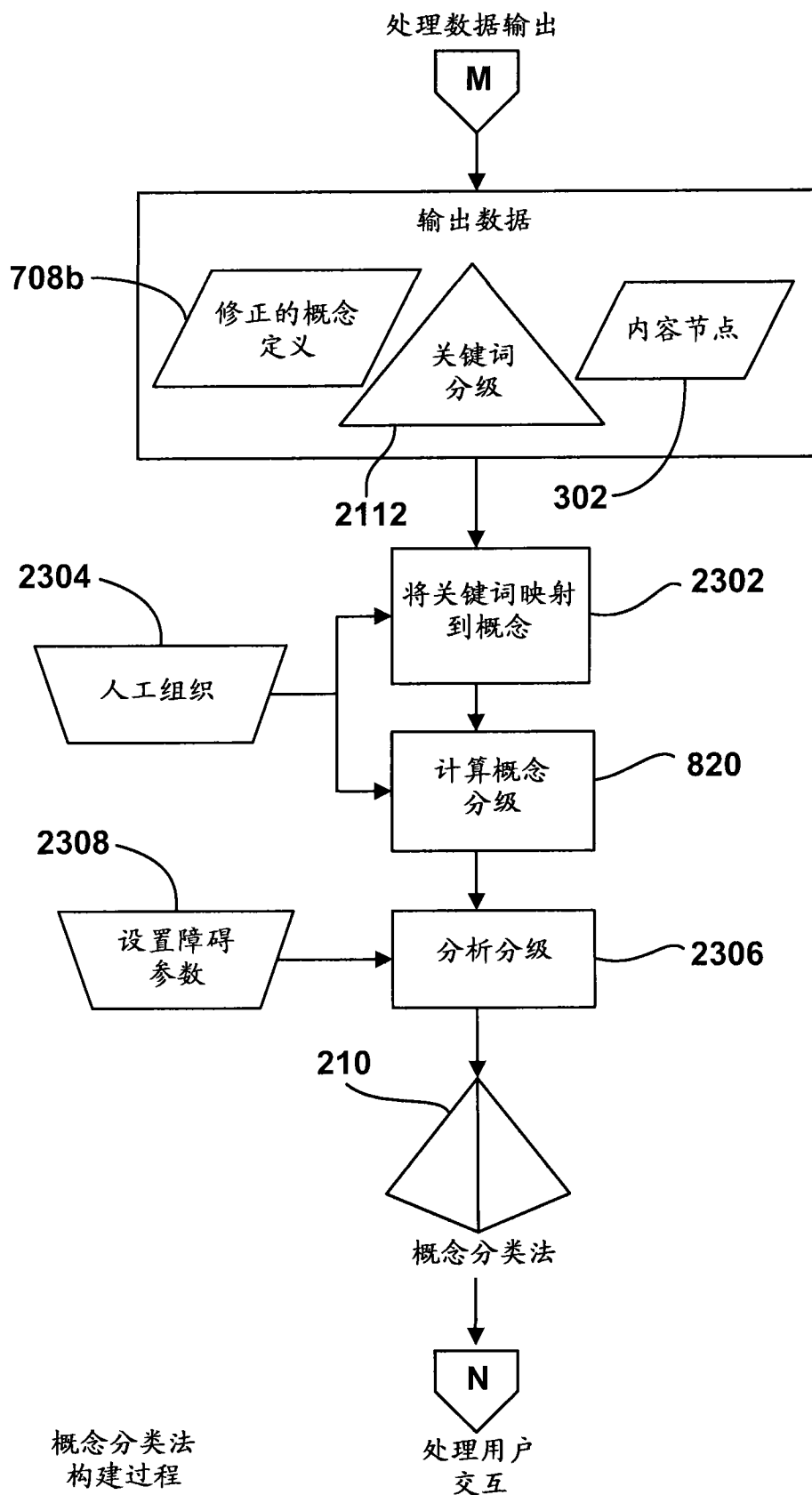


图21

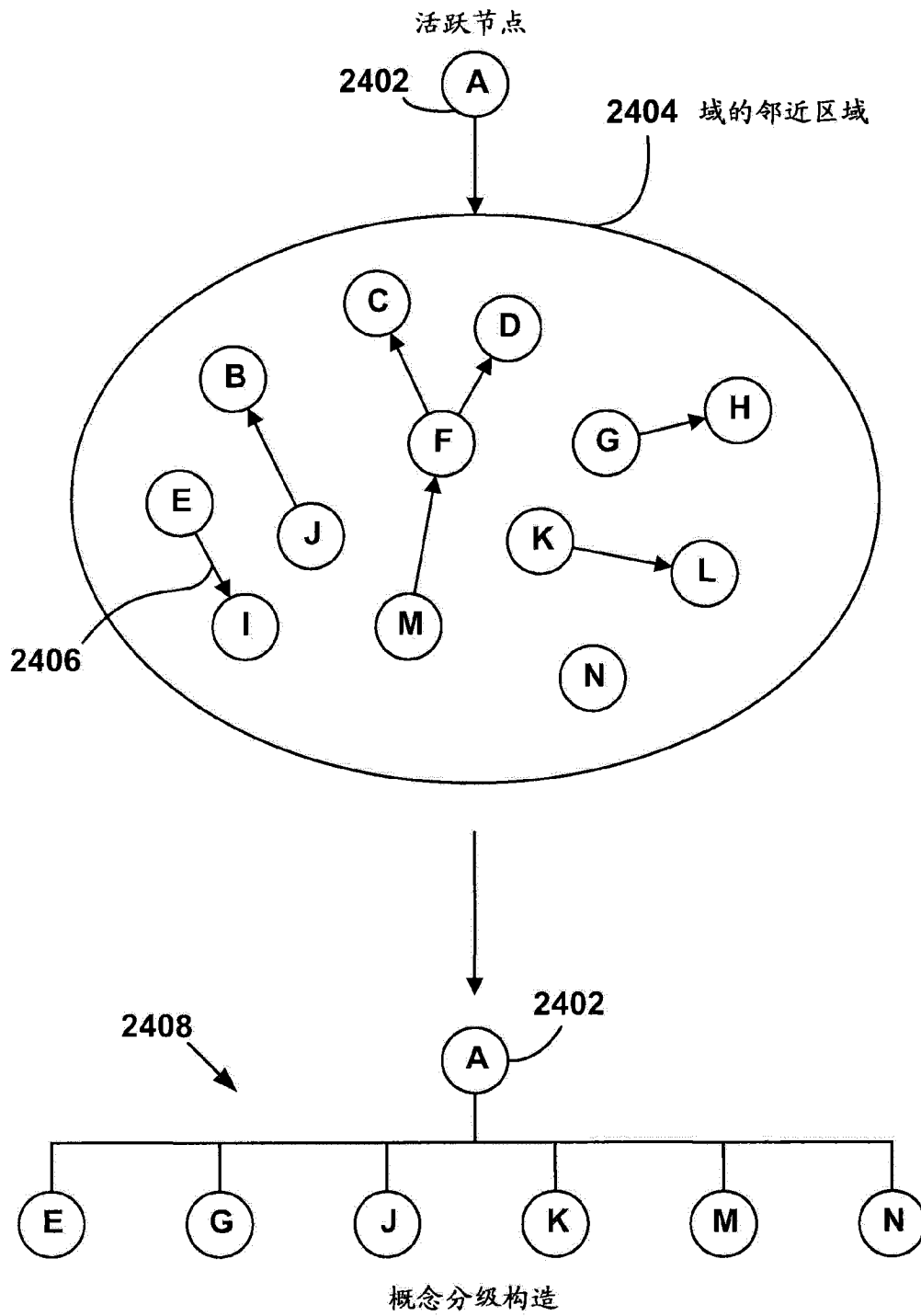


图22

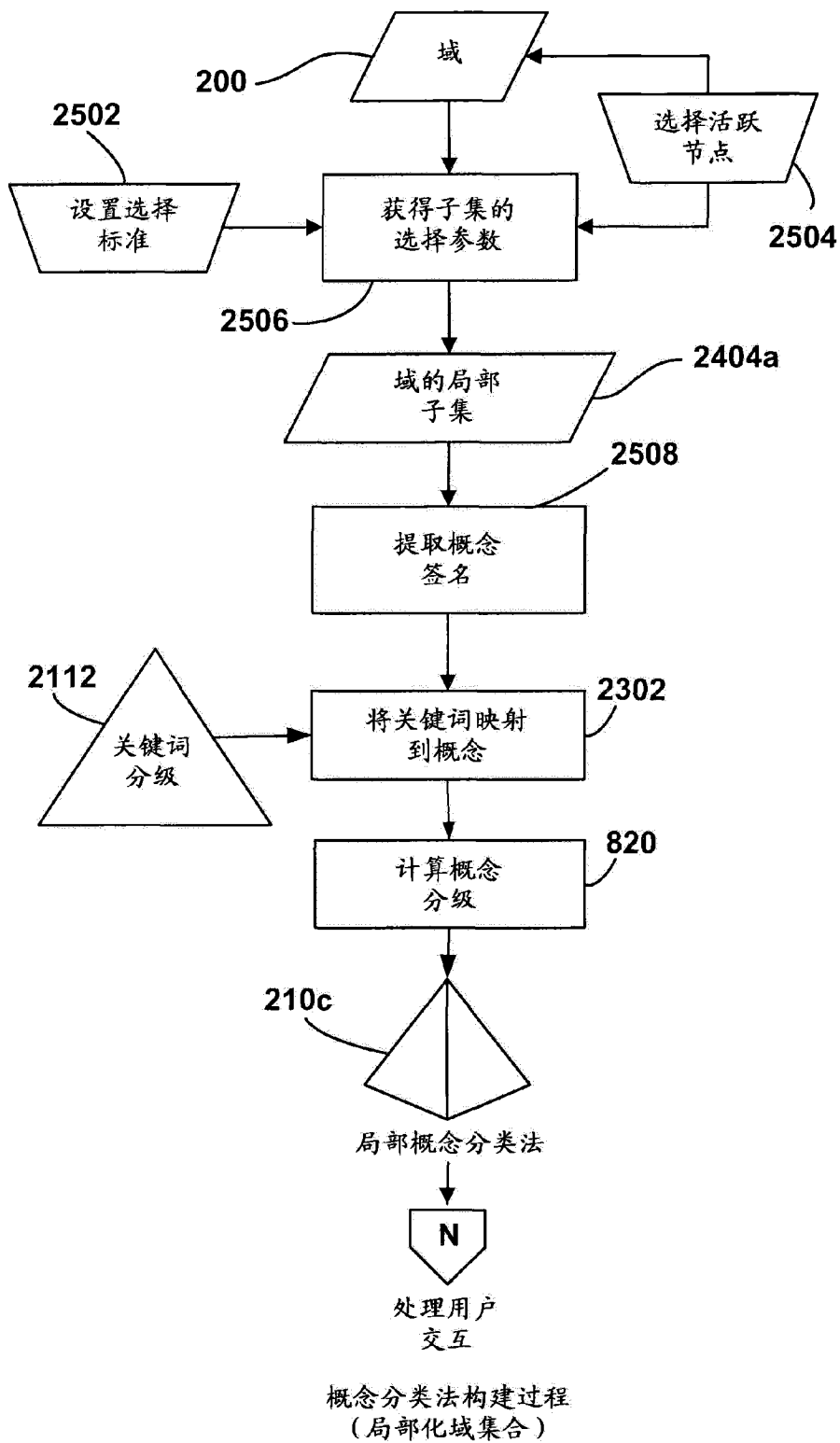
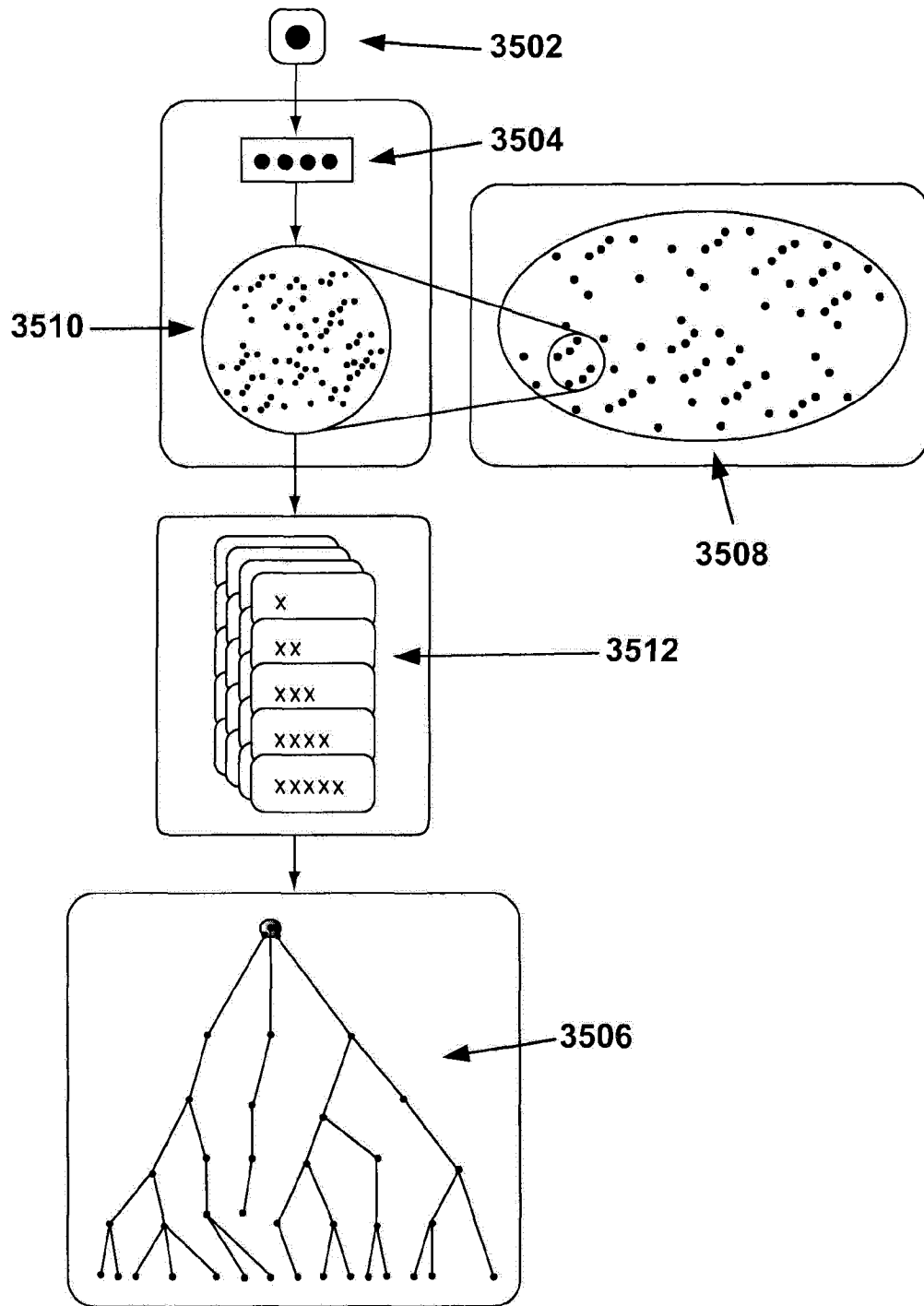
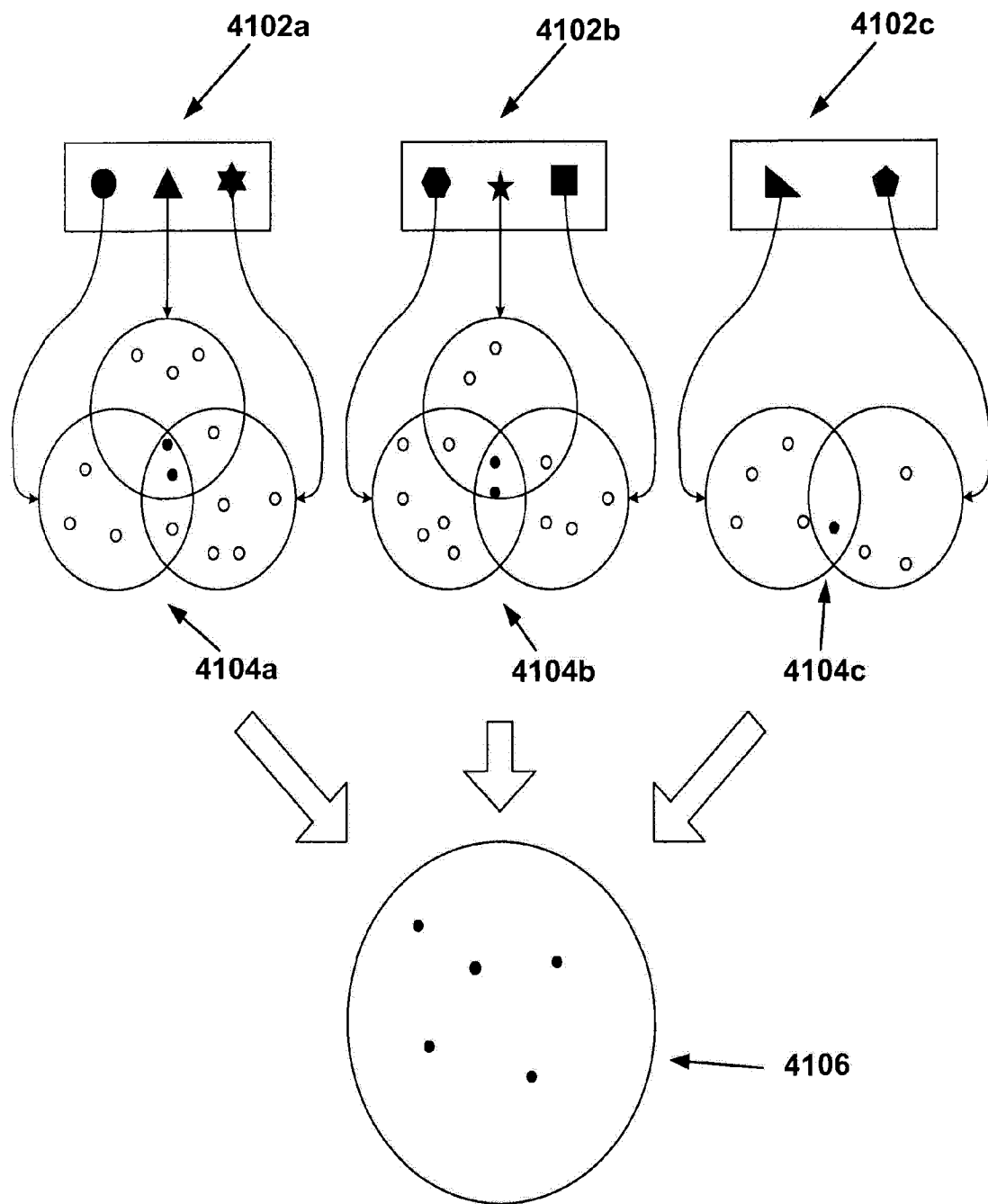


图23



动态综合模式

图24



用于动态综合的候选集组装

图25

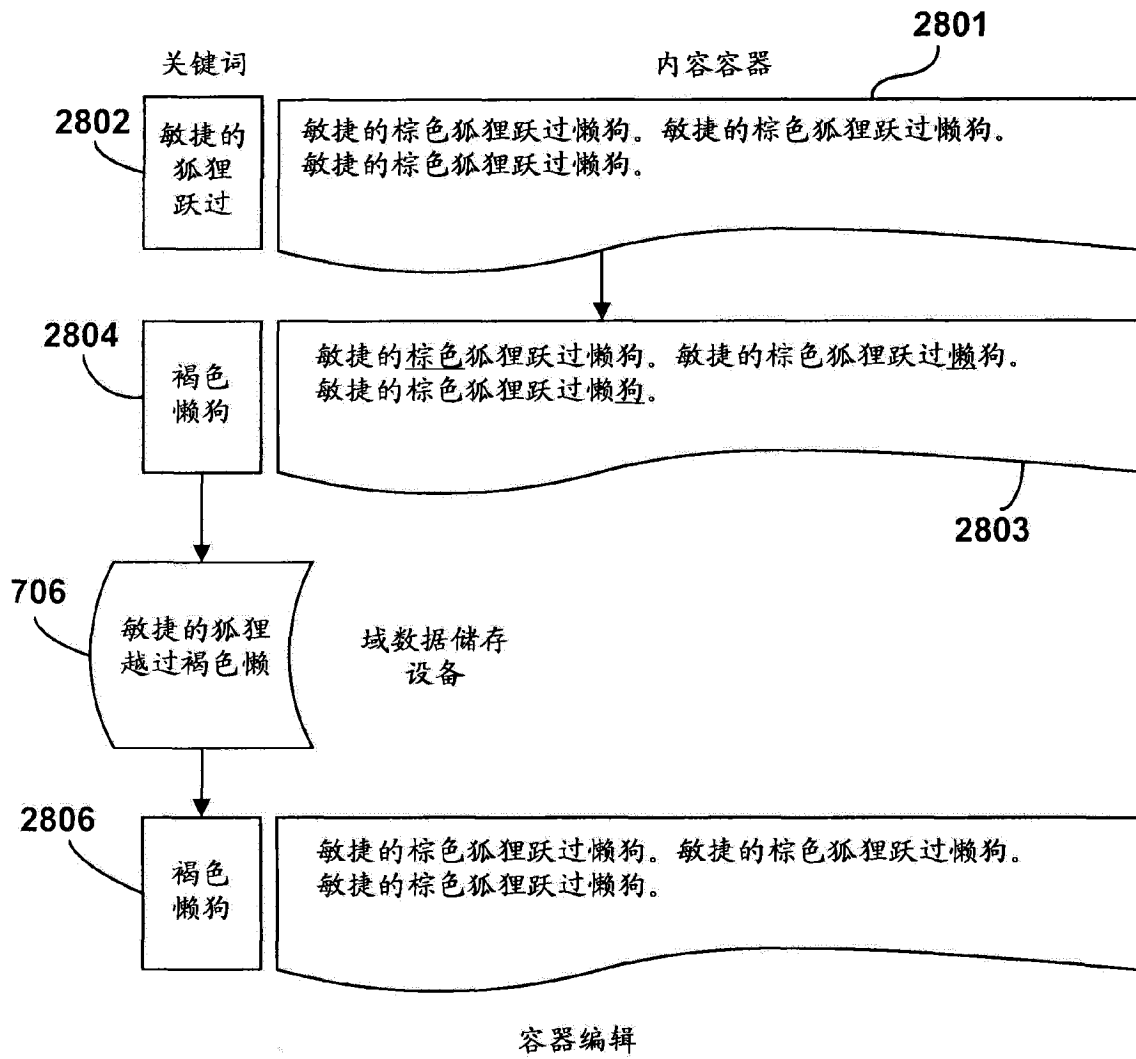


图26

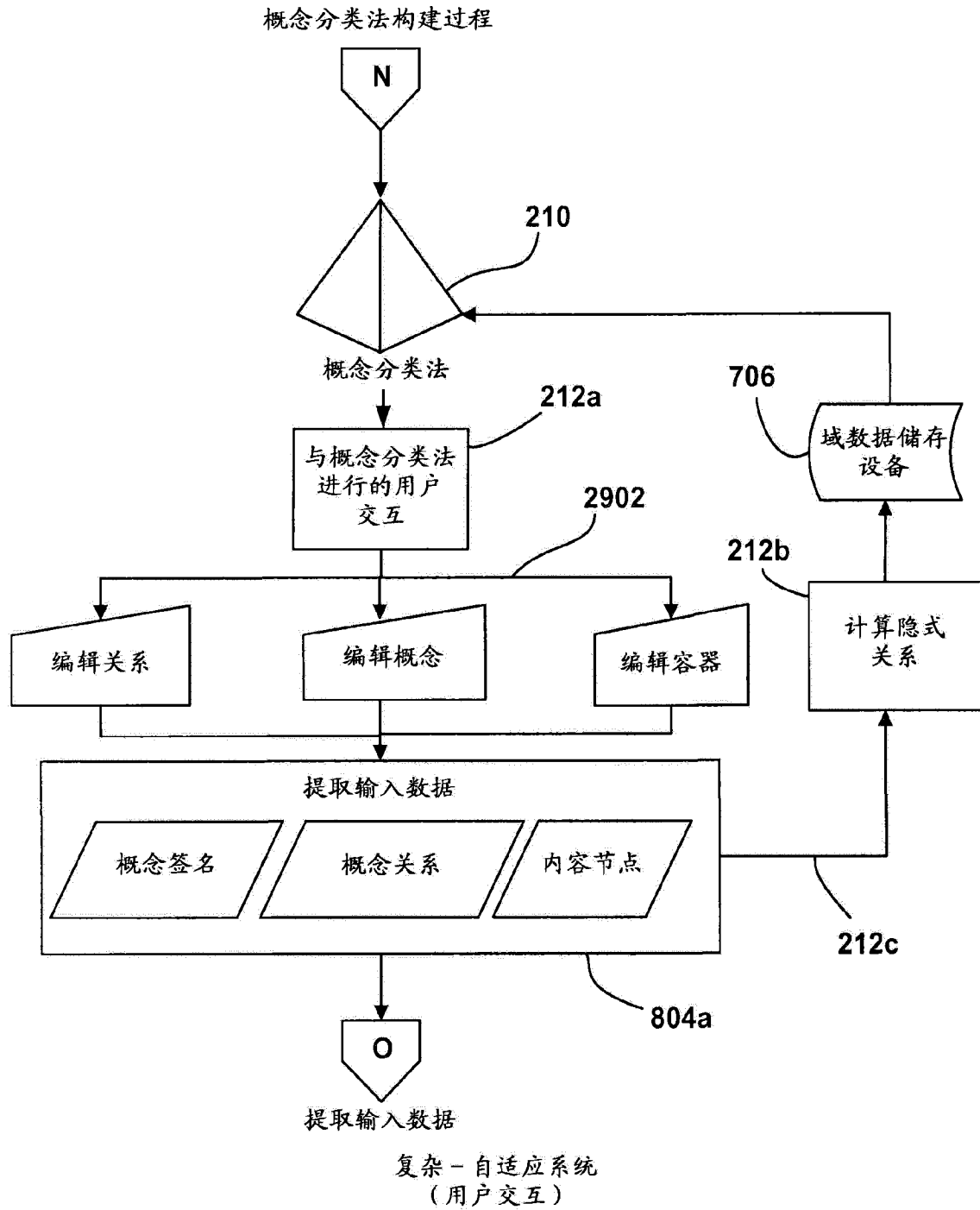


图27

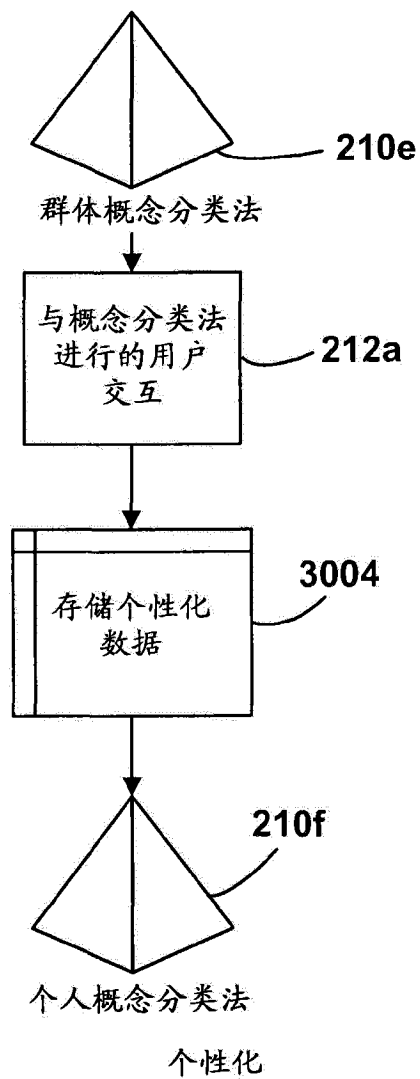
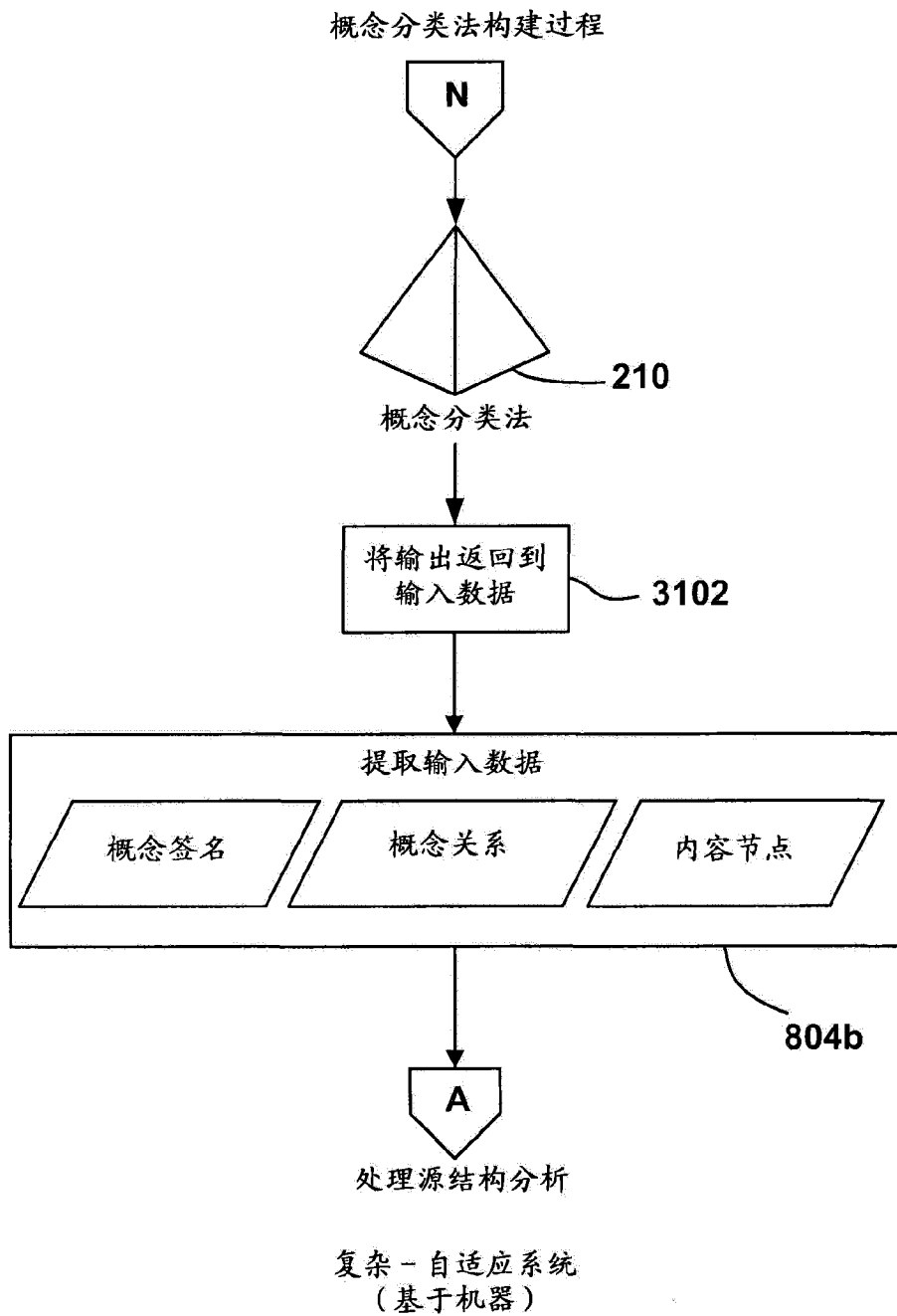


图28



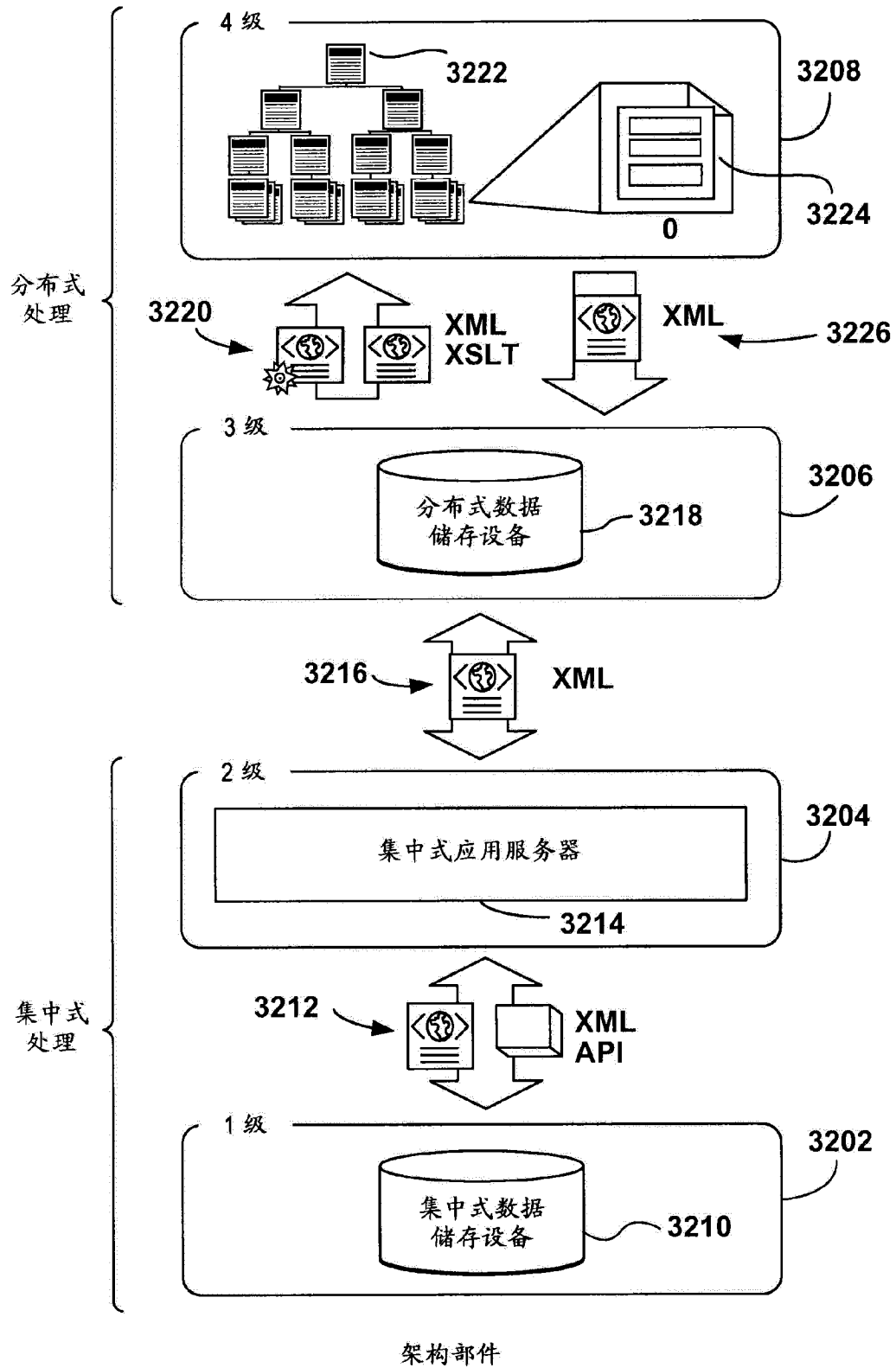
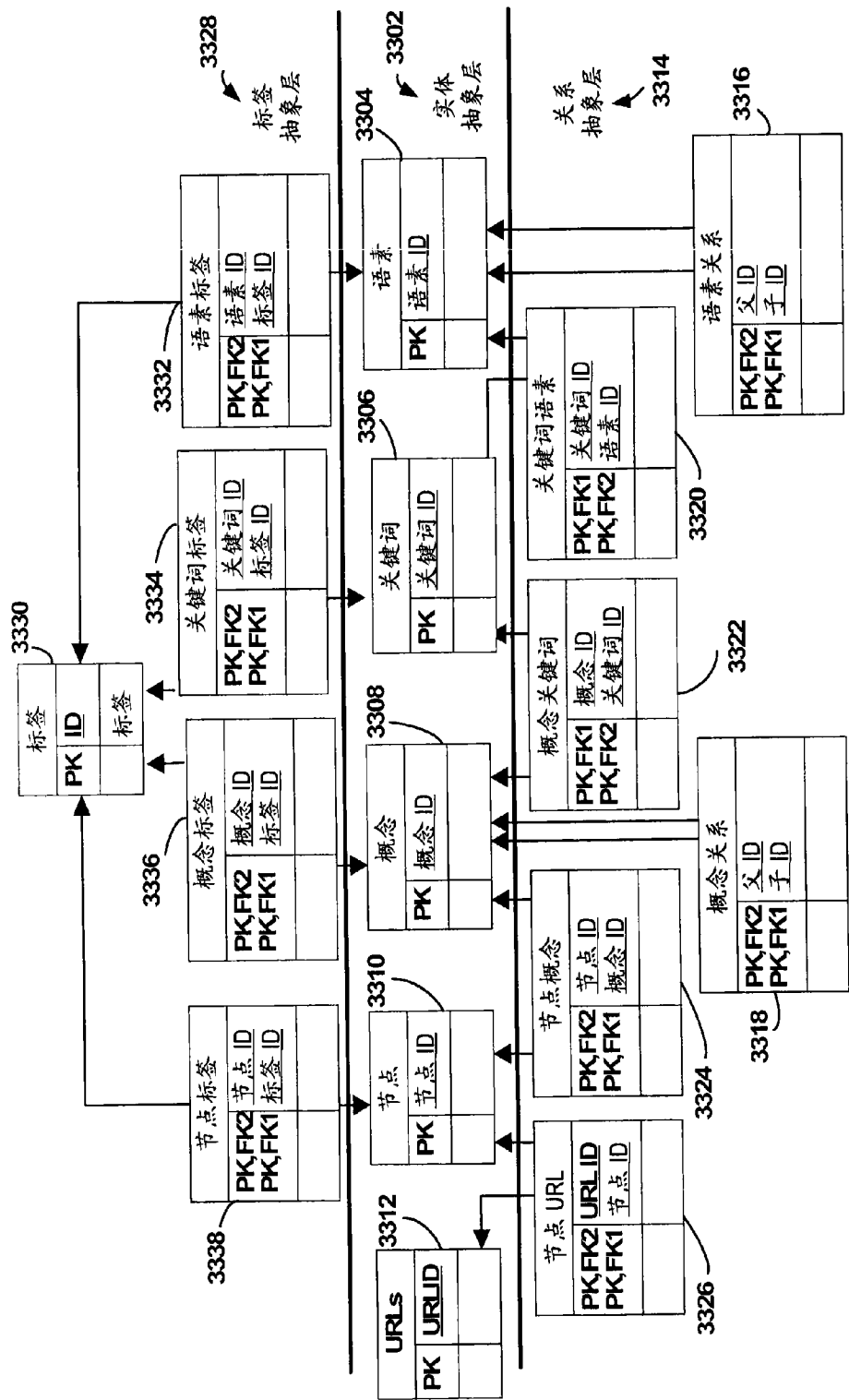


图30



简化的数据库模式

图 31

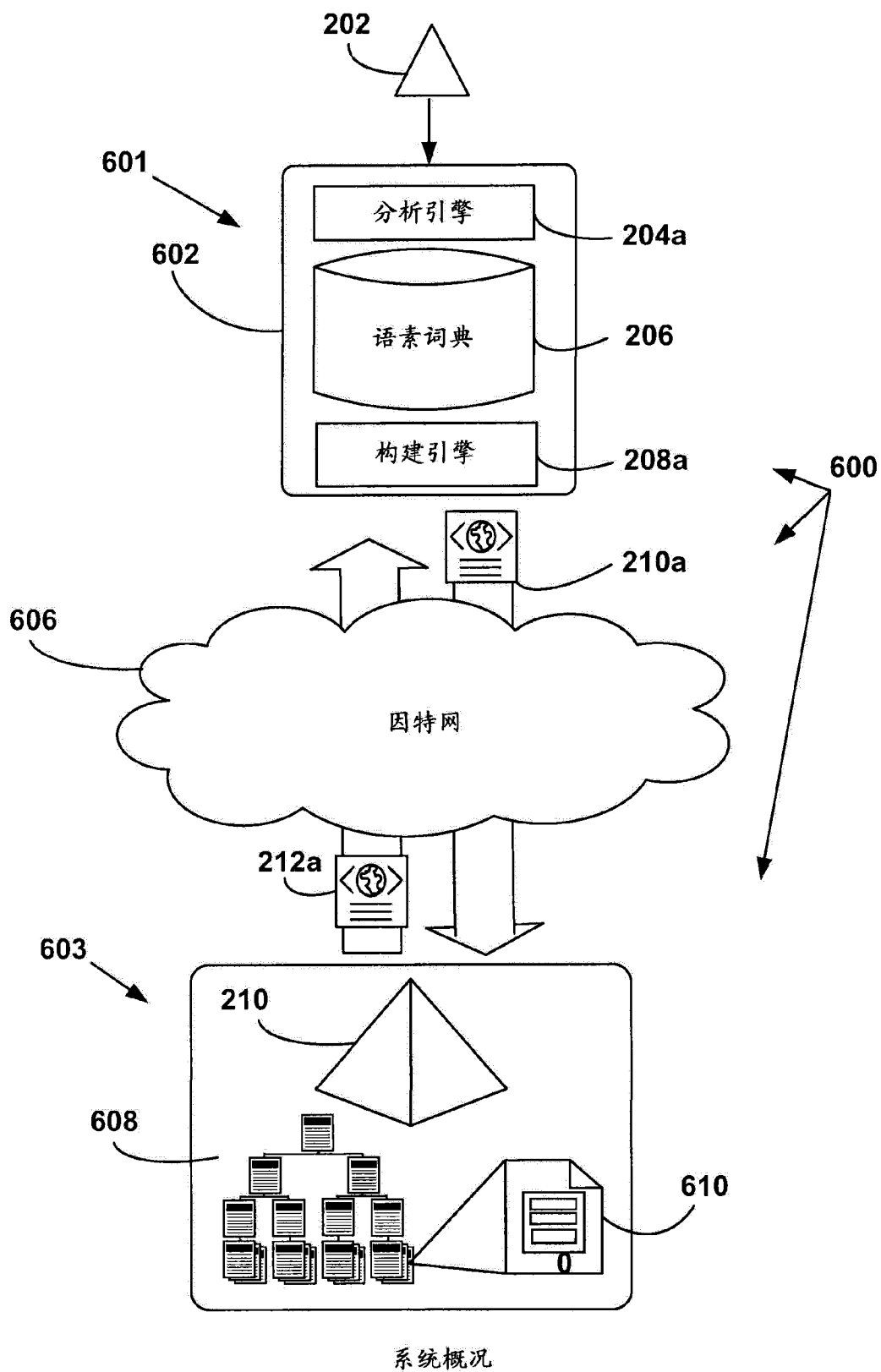


图32

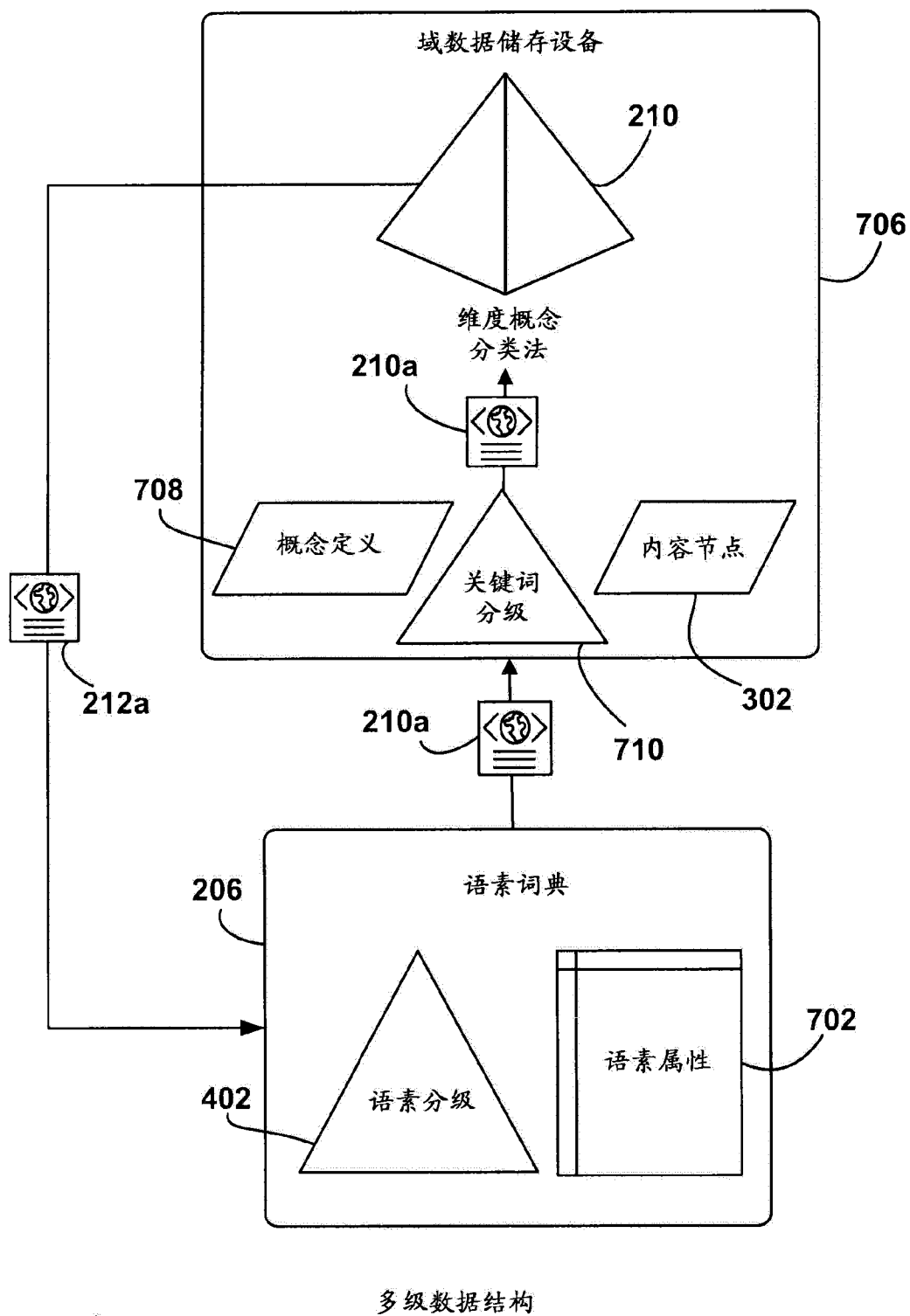
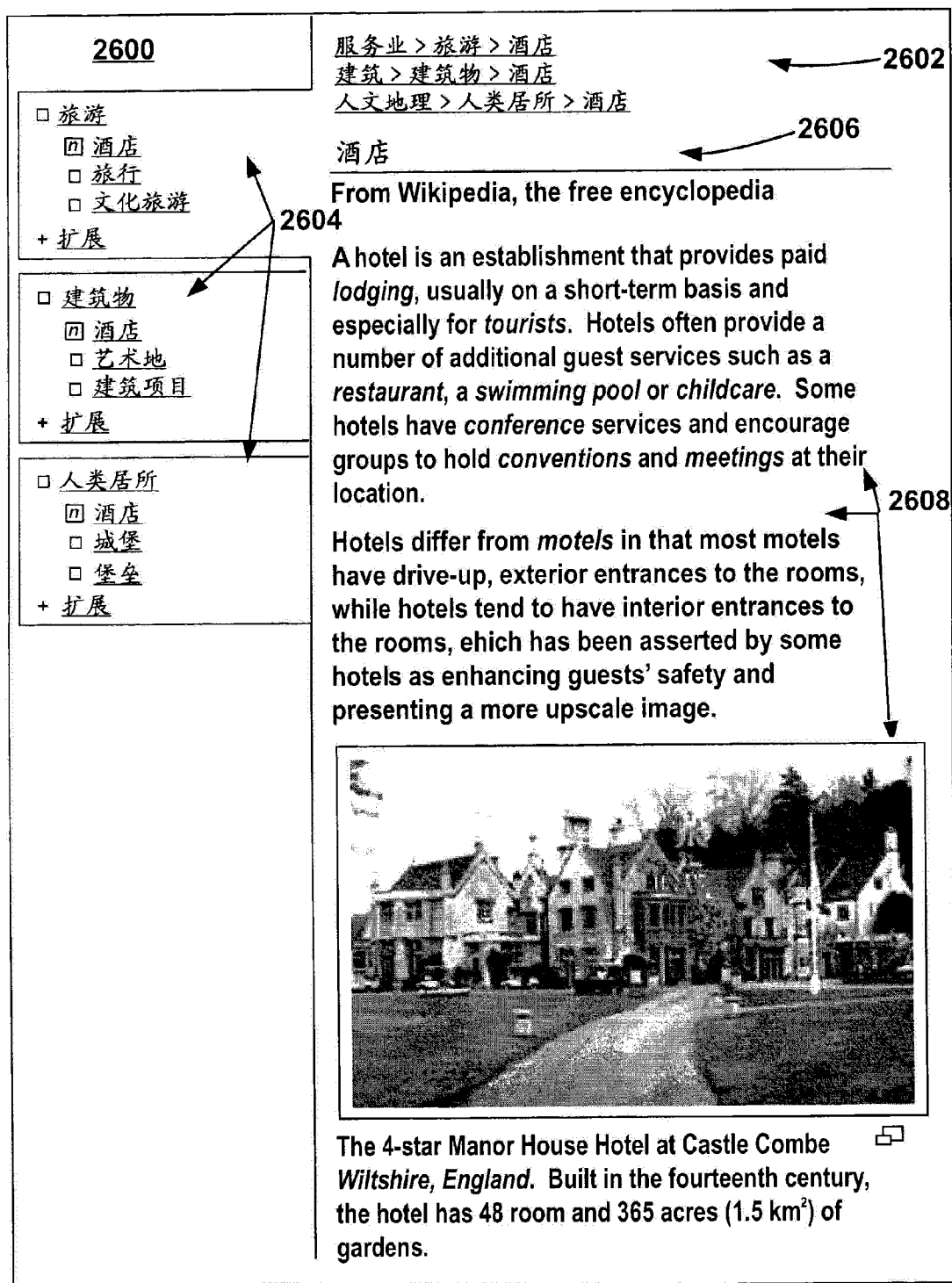
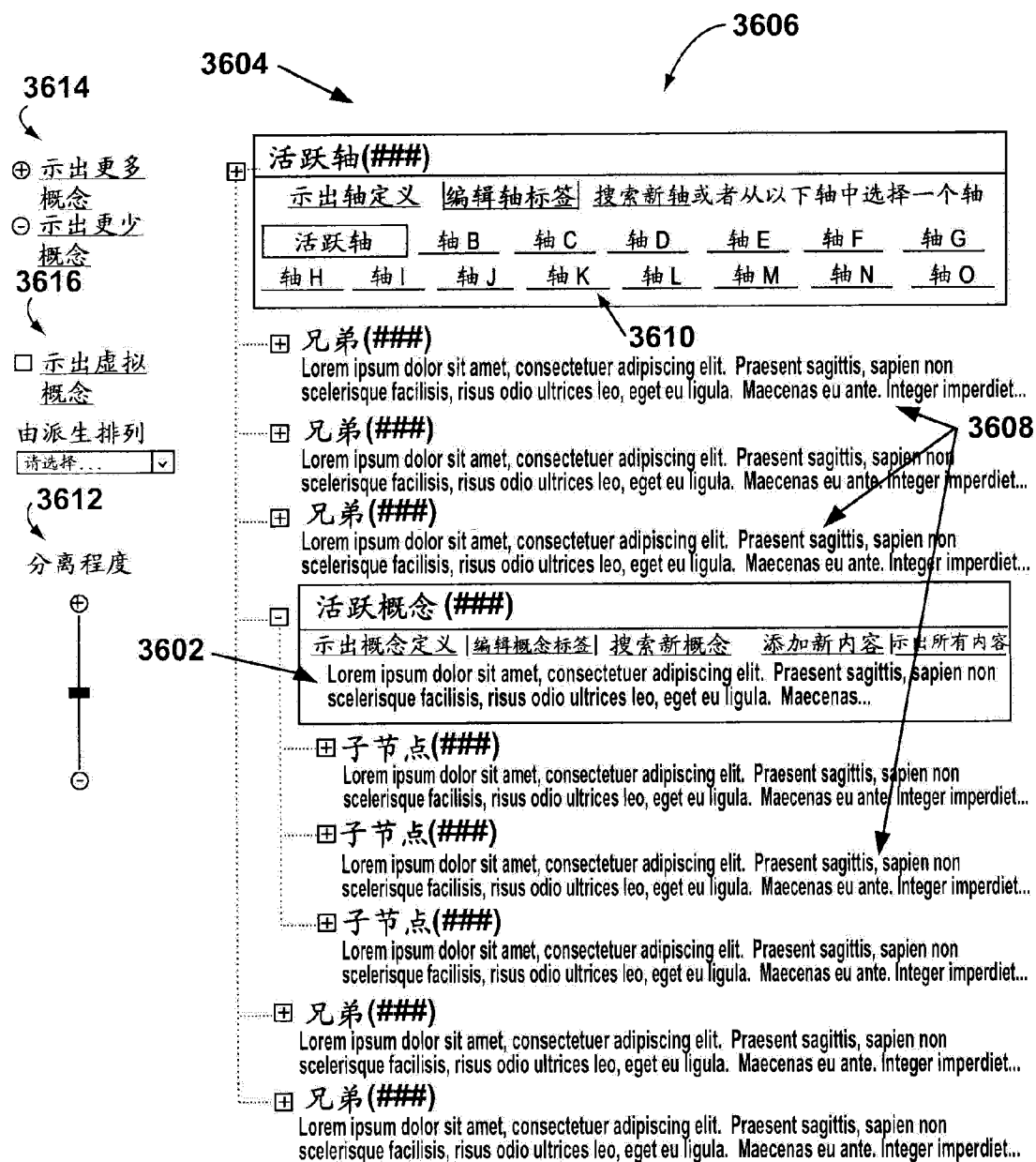


图33



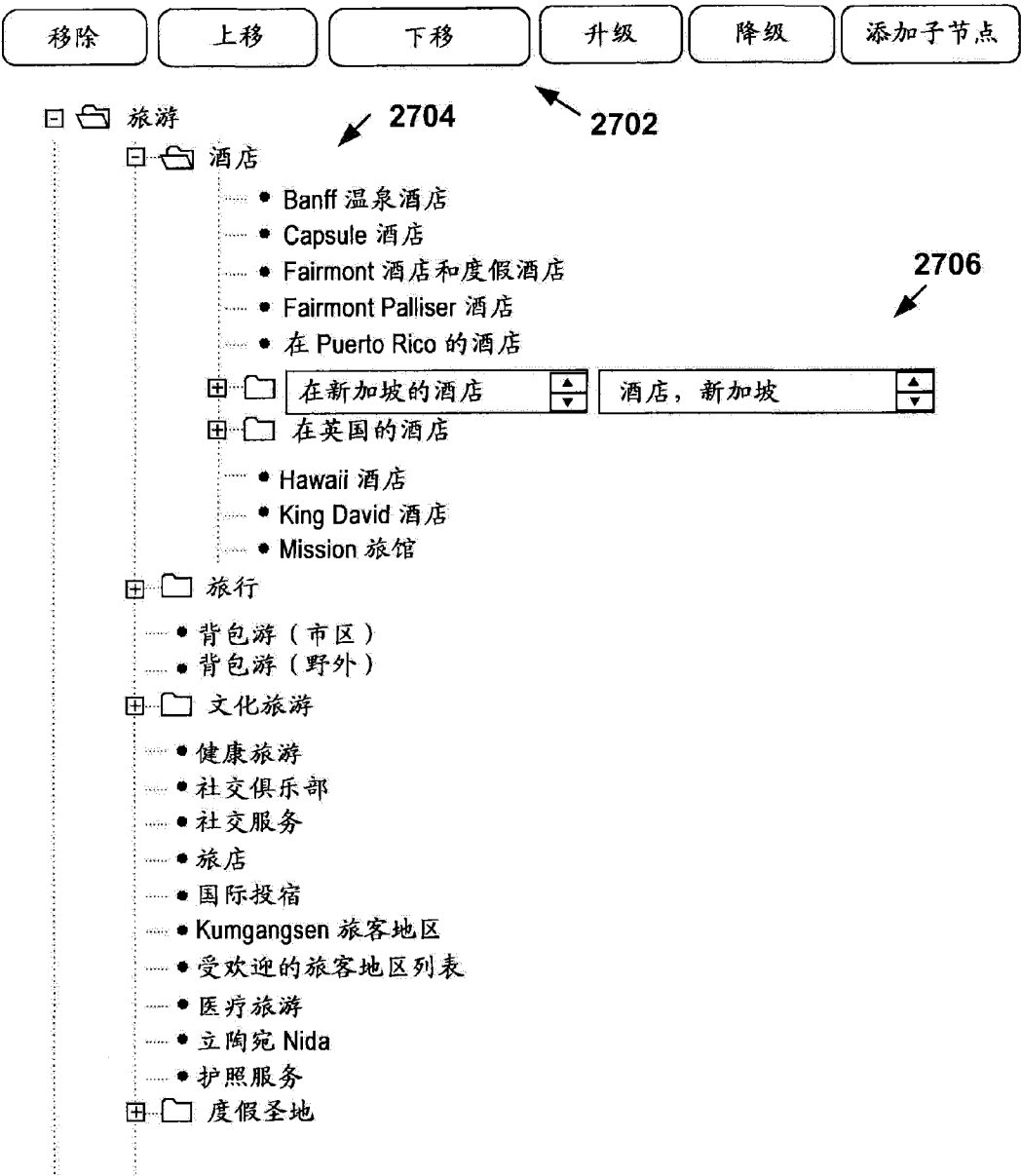
呈现概念分类法视图

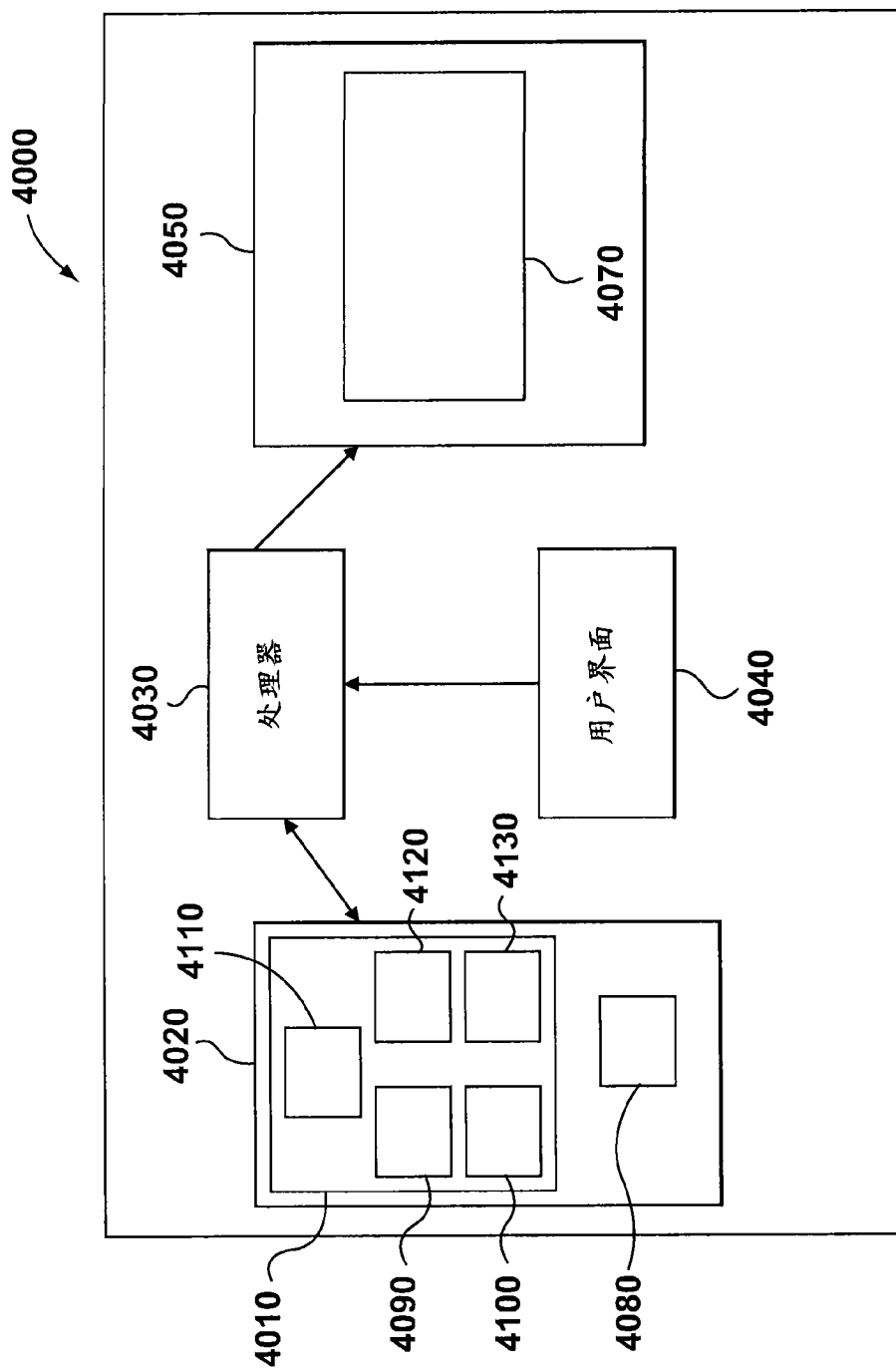
图34



动态概念综合用户界面

图35





计算机系统

图 37