



(12) 发明专利

(10) 授权公告号 CN 103201737 B

(45) 授权公告日 2016.06.29

(21) 申请号 201180053722.8

代理人 李辉 黄纶伟

(22) 申请日 2011.11.07

(51) Int. Cl.

(30) 优先权数据

G06F 17/30(2006.01)

2010-252326 2010.11.10 JP

2010-252325 2010.11.10 JP

(85) PCT国际申请进入国家阶段日

(56) 对比文件

2013.05.08

CN 101071420 A, 2007.11.14,

CN 101206673 A, 2008.06.25,

US 6169986 B1, 2001.01.02,

(86) PCT国际申请的申请数据

审查员 李腾

PCT/JP2011/075572 2011.11.07

(87) PCT国际申请的公布数据

W02012/063770 JA 2012.05.18

(73) 专利权人 乐天株式会社

地址 日本东京都

(72) 发明人 平手勇宇

(74) 专利代理机构 北京三友知识产权代理有限公司 11127

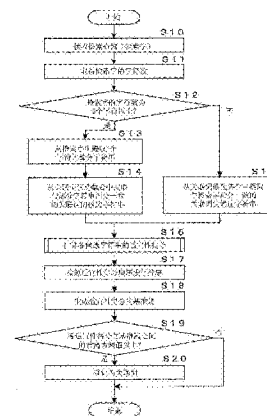
权利要求书2页 说明书20页 附图16页

(54) 发明名称

关联词登记装置、信息处理装置、关联词登记方法、关联词登记装置用程序以及记录介质

(57) 摘要

生成根据规定条件从检索查询记录(12a)中提取出关联词的候选的关联词候选集合(12b)(S1~S4),接收用户输入的检索字的检索查询(S10),根据检索字的字符串生成部分字符串(S13),根据部分字符串从关联词候选集合中提取候选字符串(S14),计算候选字符串的适宜性得分(S16),按照得分的顺序对候选字符串进行排序(S17),根据适宜性得分和排序生成针对排序的适宜性得分的基准线L1(S18),提取适宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串作为用于登记为关联词的登记字符串(S19),在关联词DB(12c)中登记所提取出的登记字符串和检索字作为关联词(S20)。



1. 一种关联词登记装置,其具有:

关联词候选集合生成单元,其根据规定条件从存储了与检索查询有关的记录的检索查询记录中提取关联词的候选,生成关联词候选集合;

接收单元,其接收用户输入的检索字的检索查询;

部分字符串生成单元,其根据所述检索字的字符串生成部分字符串;

候选字符串提取单元,其根据所述生成的部分字符串,从所述关联词候选集合中提取候选字符串;

得分计算单元,其根据所述候选字符串与所述检索字之间的相似度、所述候选字符串的使用次数和基于所述候选字符串的检索的检索结果数,计算所述候选字符串的适宜性得分;

排序单元,其按照所述得分的顺序对所述候选字符串进行排序;

判定基准生成单元,其根据所述候选字符串的适宜性得分和排序,生成针对所述排序的适宜性得分的基准线,作为是否将所述候选字符串登记为关联词的判定基准;

登记字符串提取单元,其提取所述适宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串,作为用于登记为关联词的登记字符串;以及

关联词登记单元,其在关联词数据库中登记所述提取出的登记字符串和所述检索字作为关联词,

所述部分字符串生成单元根据所述检索字的字符数设定所述部分字符串的长度和生成所述部分字符串时的检索字的字符数的移动宽度中的至少一方,以该设定的长度和该设定的移动宽度中的至少一方生成部分字符串。

2. 根据权利要求1所述的关联词登记装置,其特征在于,

所述部分字符串生成单元针对使所述检索字进行表音变换后的检索字生成所述部分字符串。

3. 根据权利要求1或2所述的关联词登记装置,其特征在于还具有参数调节单元,该参数调节单元对所述得分计算单元的所述相似度和所述使用次数的信息进行调节。

4. 根据权利要求1所述的关联词登记装置,其特征在于,

所述判定基准生成单元通过对所述排序与所述适宜性得分之间的关联进行近似的近似函数或规定范围的所述排序的数据求出所述基准线。

5. 根据权利要求1所述的关联词登记装置,其特征在于,

所述登记字符串提取单元对要提取的所述登记字符串的数量设置上限。

6. 根据权利要求1所述的关联词登记装置,其特征在于还具有检索字存储单元,该检索字存储单元存储与内部或外部的EC网站中销售的商品关联的检索字。

7. 根据权利要求1所述的关联词登记装置,其特征在于,

所述得分计算单元根据与所述检索查询有关的记录的数据计算所述适宜性得分。

8. 一种关联词登记装置中的关联词登记方法,该关联词登记装置登记关联词,该关联词登记方法包括:

关联词候选集合生成步骤,根据规定条件从存储了与检索查询有关的记录的检索查询记录中提取关联词的候选,生成关联词候选集合;

接收步骤,接收用户输入的检索字的检索查询;

部分字符串生成步骤,根据所述检索字的字符串生成部分字符串;

候选字符串提取步骤,根据所述生成的部分字符串,从所述关联词候选集合中提取候选字符串;

得分计算步骤,根据所述候选字符串与所述检索字之间的相似度、所述候选字符串的使用次数和基于所述候选字符串的检索的检索结果数,计算所述候选字符串的适宜性得分;

排序步骤,按照所述得分的顺序对所述候选字符串进行排序;

判定基准生成步骤,根据所述候选字符串的适宜性得分和排序,生成针对所述排序的适宜性得分的基准线,作为是否将所述候选字符串登记为关联词的判定基准;

登记字符串提取步骤,提取所述适宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串,作为用于登记为关联词的登记字符串;以及

关联词登记步骤,在关联词数据库中登记所述提取出的登记字符串和所述检索字作为关联词,

在所述部分字符串生成步骤中,根据所述检索字的字符数设定所述部分字符串的长度和生成所述部分字符串时的检索字的字符数的移动宽度中的至少一方,以该设定的长度和该设定的移动宽度中的至少一方生成部分字符串。

9. 根据权利要求8所述的关联词登记方法,其特征在于,

在所述部分字符串生成步骤中,针对使所述检索字进行表音变换后的检索字生成所述部分字符串。

10. 根据权利要求8或9所述的关联词登记方法,其特征在于还包括参数调节步骤,在该参数调节步骤中,对所述得分计算步骤的所述相似度和所述使用次数的信息进行调节。

11. 根据权利要求8所述的关联词登记方法,其特征在于,

在所述判定基准生成步骤中,通过对所述排序与所述适宜性得分之间的关联进行近似的近似函数或规定范围的所述排序的数据求出所述基准线。

12. 根据权利要求8所述的关联词登记方法,其特征在于,

在所述登记字符串提取步骤中,对要提取的所述登记字符串的数量设置上限。

13. 根据权利要求8所述的关联词登记方法,其特征在于还包括检索字存储步骤,在该检索字存储步骤中,存储与内部或外部的EC网站中销售的商品关联的检索字。

14. 根据权利要求8所述的关联词登记方法,其特征在于,

在所述得分计算步骤中,根据与所述检索查询有关的记录的数据计算所述适宜性得分。

15. 一种信息处理方法,参照权利要求1~7中的任意一项所述的关联词登记装置的关联词数据库进行信息处理,该信息处理方法的特征在于包括:

关联词提取步骤,参照所述关联词数据库,提取与所述接收到的检索查询的检索字对应的关联词;以及

关联词输出步骤,输出由所述关联词提取步骤提取出的关联词。

关联词登记装置、信息处理装置、关联词登记方法、关联词登记装置用程序以及记录介质

技术领域

[0001] 本发明涉及登记检索关键字作为关联词的关联词登记装置、信息处理装置、关联词登记方法、关联词登记装置用程序以及记录介质的技术领域。

背景技术

[0002] 为了补充检索关键字并提高翻译等的精度,在信息检索系统、自然语言处理系统、机械翻译系统等中利用同义词辞典这种单词数据库。研究了用于自动构建该同义词辞典的各种装置。例如,在专利文献1中公开了如下的意思辞典登记装置:利用意思辞典中的已经赋予了意思信息的辞典条目的信息,自动对输入辞典条目赋予意思信息。

[0003] 现有技术文献

[0004] 专利文献

[0005] 专利文献1:日本特开2000-268035号公报

发明内容

[0006] 发明要解决的课题

[0007] 但是,在专利文献1所记载的技术中,根据输入数据和已登记的全部数据的单纯的单词间意思的距离来判定近义,所以,近义的判定精度低,无法抑制不需要的同义词这样的关联词的登记。因此,难以构建高精度的同义词辞典等的关联词的辞典。

[0008] 本发明是鉴于这种问题而完成的,其课题的一例的目的在于,提供能够登记高精度的关联词的关联词登记装置等。

[0009] 用于解决课题的手段

[0010] 为了解决上述课题,第1方面所记载的发明是一种关联词登记装置,其具有:关联词候选集合生成单元,其根据规定条件从存储了与检索查询有关的记录的检索查询记录中提取关联词的候选,生成关联词候选集合;接收单元,其接收用户输入的检索字的检索查询;部分字符串生成单元,其根据所述检索字的字符串生成部分字符串;候选字符串提取单元,其根据所述生成的部分字符串,从所述关联词候选集合中提取候选字符串;得分计算单元,其根据所述候选字符串与所述检索字之间的相似度、所述候选字符串的使用次数和基于所述候选字符串的检索的检索结果数,计算所述候选字符串的适宜性得分;排序单元,其按照所述得分的顺序对所述候选字符串进行排序;判定基准生成单元,其根据所述候选字符串的适宜性得分和排序,生成针对所述排序的适宜性得分的基准线,作为是否将所述候选字符串登记为关联词的判定基准;登记字符串提取单元,其提取所述适宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串,作为用于登记为关联词的登记字符串;以及关联词登记单元,其在关联词数据库中登记所述提取出的登记字符串和所述检索字作为关联词。

[0011] 第2方面所记载的发明的特征在于,在第1方面所记载的关联词登记装置中,所述

部分字符串生成单元根据所述检索字的字符数设定部分字符串的长度,生成该设定的长度的部分字符串。

[0012] 第3方面所记载的发明的特征在于,在第1方面所记载的关联词登记装置中,所述部分字符串生成单元根据所述检索字的字符数设定生成部分字符串时的检索字的字符数的移动宽度,以该设定的移动宽度生成部分字符串。

[0013] 第4方面所记载的发明的特征在于,在第1~3方面中的任意一个方面所记载的关联词登记装置中,所述部分字符串生成单元针对使所述检索字进行表音变换后的检索字生成所述部分字符串。

[0014] 第5方面所记载的发明的特征在于,在第1~4方面中的任意一个方面所记载的关联词登记装置中,所述关联词登记装置还具有参数调节单元,该参数调节单元对所述得分计算单元的所述相似度和所述使用次数的信息进行调整。

[0015] 第6方面所记载的发明的特征在于,在第1~5方面中的任意一个方面所记载的关联词登记装置中,所述判定基准生成单元由对所述排序与所述适宜性得分之间的关联进行近似的近似函数或规定范围的所述排序的数据求出所述基准线。

[0016] 第7方面所记载的发明的特征在于,在第1~6方面中的任意一个方面所记载的关联词登记装置中,所述登记字符串提取单元对要提取的所述登记字符串的数量设置上限。

[0017] 第8方面所记载的发明的特征在于,在第1~7方面中的任意一个方面所记载的关联词登记装置中,所述关联词登记装置还具有检索字存储单元,该检索字存储单元存储与内部或外部的EC网站中销售的商品关联的检索字。

[0018] 第9方面所记载的发明的特征在于,在第1~8方面中的任意一个方面所记载的关联词登记装置中,所述得分计算单元根据与所述检索查询有关的记录的数据计算所述适宜性得分。

[0019] 第10方面所记载的发明的特征在于,在第1~9方面中的任意一个方面所记载的关联词登记装置中,所述关联词登记装置还具有:检索查询存储单元,其按照接收顺序存储所述接收到的检索查询;检索查询提取单元,其根据预先设定的检索查询提取条件,从所述检索查询存储单元中提取接收顺序比所述接收到的检索查询早的之前的检索查询;字符串组存储单元,其存储构成所述提取出的之前的检索查询的之前的检索字和构成所述接收到的检索查询的检索字,作为字符串组;字符串提取单元,其根据预先设定的字符串组提取开始条件,从所述字符串组存储单元中提取所述之前的检索字相同或相似的字符串组;以及关联词确定单元,其根据预先设定的登记条件,从所述提取出的字符串组确定作为关联词的字符串组,所述关联词登记单元在关联词数据库中登记所述确定的字符串组作为关联词。

[0020] 第11方面所记载的发明的特征在于,在第10方面所记载的关联词登记装置中,所述检索查询存储单元还存储用户识别信息,作为所述检索查询提取条件,所述检索查询提取单元提取输入了所述检索查询的用户的用户识别信息一致的所述之前的检索查询。

[0021] 第12方面所记载的发明的特征在于,在第10或11方面所记载的关联词登记装置中,作为所述检索查询提取条件,所述检索查询提取单元提取从所述检索查询的接收时刻起规定时间内接收到的之前的检索查询。

[0022] 第13方面所记载的发明的特征在于,在第10~12方面中的任意一个方面所记载的关联词登记装置中,所述检索查询存储单元还存储类别信息,所述检索查询提取单元作为

所述检索查询提取条件根据所述类别信息提取所述先前的检索查询。

[0023] 第14方面所记载的发明的特征在于,在第10~13方面中的任意一个方面所记载的关联词登记装置中,作为所述字符串组提取开始条件,在所述检索查询存储单元中的检索查询数或所述字符串组集合的元素数超过规定阈值的情况下,所述字符串提取单元提取所述字符串组。

[0024] 第15方面所记载的发明的特征在于,在第10~13方面中的任意一个方面所记载的关联词登记装置中,在满足所述字符串组提取开始条件后经过了规定时间的情况下,所述字符串提取单元提取所述字符串组。

[0025] 第16方面所记载的发明的特征在于,在第10~15方面中的任意一个方面所记载的关联词登记装置中,作为所述登记条件,在所述先前的检索字相同或相似的字符串组中,具有相同或相似的检索字的字符串组的数量或具有相同或相似的检索字的字符串组的比例超过规定阈值的情况下,所述关联词确定单元确定该字符串组作为关联词。

[0026] 第17方面所记载的发明是一种信息处理装置,其参照第1~16方面中的任意一个方面所记载的关联词登记装置的关联词数据库进行信息处理,其特征在于,该信息处理装置具有:关联词提取单元,其参照所述关联词数据库,提取与所述接收到的检索查询的检索字对应的关联词;以及关联词输出单元,其输出由所述关联词提取单元提取出的关联词。

[0027] 第18方面所记载的发明是一种关联词登记装置的关联词登记方法,该关联词登记装置登记关联词,该关联词登记方法包括:关联词候选集合生成步骤,根据规定条件从存储了与检索查询有关的记录的检索查询记录中提取关联词的候选,生成关联词候选集合;接收步骤,接收用户输入的检索字的检索查询;部分字符串生成步骤,根据所述检索字的字符串生成部分字符串;候选字符串提取步骤,根据所述生成的部分字符串,从所述关联词候选集合中提取候选字符串;得分计算步骤,根据所述候选字符串与所述检索字之间的相似度、所述候选字符串的使用次数和基于所述候选字符串的检索的检索结果数,计算所述候选字符串的适宜性得分;排序步骤,按照所述得分的顺序对所述候选字符串进行排序;判定基准生成步骤,根据所述候选字符串的适宜性得分和排序,生成针对所述排序的适宜性得分的基准线,作为是否将所述候选字符串登记为关联词的判定基准;登记字符串提取步骤,提取所述适宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串,作为用于登记为关联词的登记字符串;以及关联词登记步骤,在关联词数据库中登记所述提取出的登记字符串和所述检索字作为关联词。

[0028] 第19方面所记载的发明是一种关联词登记装置用程序,其使计算机作为以下单元进行工作:关联词候选集合生成单元,其根据规定条件从存储了与检索查询有关的记录的检索查询记录中提取关联词的候选,生成关联词候选集合;接收单元,其接收用户输入的检索字的检索查询;部分字符串生成单元,其根据所述检索字的字符串生成部分字符串;候选字符串提取单元,其根据所述生成的部分字符串,从所述关联词候选集合中提取候选字符串;得分计算单元,其根据所述候选字符串与所述检索字之间的相似度、所述候选字符串的使用次数和基于所述候选字符串的检索的检索结果数,计算所述候选字符串的适宜性得分;排序单元,其按照所述得分的顺序对所述候选字符串进行排序;判定基准生成单元,其根据所述候选字符串的适宜性得分和排序,生成针对所述排序的适宜性得分的基准线,作为是否将所述候选字符串登记为关联词的判定基准;登记字符串提取单元,其提取所述适

宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串,作为用于登记为关联词的登记字符串;以及关联词登记单元,其在关联词数据库中登记所述提取出的登记字符串和所述检索字作为关联词。

[0029] 第20方面所记载的发明是一种计算机可读的记录介质,其记录了使计算机作为以下单元进行工作的关联词登记装置用的程序:关联词候选集合生成单元,其根据规定条件从存储了与检索查询有关的记录的检索查询记录中提取关联词的候选,生成关联词候选集合;接收单元,其接收用户输入的检索字的检索查询;部分字符串生成单元,其根据所述检索字的字符串生成部分字符串;候选字符串提取单元,其根据所述生成的部分字符串,从所述关联词候选集合中提取候选字符串;得分计算单元,其根据所述候选字符串与所述检索字之间的相似度、所述候选字符串的使用次数和基于所述候选字符串的检索的检索结果数,计算所述候选字符串的适宜性得分;排序单元,其按照所述得分的顺序对所述候选字符串进行排序;判定基准生成单元,其根据所述候选字符串的适宜性得分和排序,生成针对所述排序的适宜性得分的基准线,作为是否将所述候选字符串登记为关联词的判定基准;登记字符串提取单元,其提取所述适宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串,作为用于登记为关联词的登记字符串;以及关联词登记单元,其在关联词数据库中登记所述提取出的登记字符串和所述检索字作为关联词。

[0030] 发明效果

[0031] 根据本发明,生成根据规定条件从存储了与检索查询有关的记录的检索查询记录中提取关联词的候选的关联词候选集合,接收用户输入的检索字的检索查询,根据检索字的字符串生成部分字符串,根据部分字符串从关联词候选集合中提取候选字符串,根据候选字符串与检索字之间的相似度、候选字符串的使用次数和基于候选字符串的检索的检索结果数,计算候选字符串的适宜性得分,按照得分的顺序对候选字符串进行排序,根据候选字符串的适宜性得分和排序,生成针对排序的适宜性得分的基准线,作为是否将候选字符串登记为关联词的判定基准,提取适宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串,作为用于登记为关联词的登记字符串,在关联词数据库中登记提取出的登记字符串和检索字作为关联词,由此,根据减少了字符数的部分字符串从关联词候选集合中提取候选字符串,所以,能够通过候选字符串的适宜性得分来抑制不需要的关联词的登记,能够提高关联词的精度。

附图说明

[0032] 图1是示出本发明的一个实施方式的关联词登记系统的概要结构例的示意图。

[0033] 图2是示出图1的关联词登记服务器的概要结构的一例的框图。

[0034] 图3是示出图1的信息提供服务器的概要结构的一例的框图。

[0035] 图4是示出图1的终端的概要结构的一例的框图。

[0036] 图5是示出图1的关联词登记系统1的第1实施方式的动作中的、关联词登记服务器的关联词候选集合的生成的动作例的流程图。

[0037] 图6是示出图1的终端中显示的网页的一例的示意图。

[0038] 图7是示出图1的关联词登记服务器中的第1实施方式的关联词的登记的动作例的流程图。

- [0039] 图8是示出部分字符串的提取的一例的示意图。
- [0040] 图9(A)和(B)是示出排序-得分/曲线的一例的示意图。
- [0041] 图10是示出图1的关联词登记服务器中的计算适宜性得分的动作例的子进程的流程图。
- [0042] 图11(A)~(C)是示出用于计算适宜性得分的可用性的图形的一例的线图。
- [0043] 图12是示出图1的关联词登记系统1的第2实施方式的动作中的、终端中显示的网页的一例的示意图。
- [0044] 图13是示出图1的关联词登记服务器的检索查询记录/数据库构建的动作例的流程图。
- [0045] 图14是示出图1的终端中显示的网页的一例的示意图。
- [0046] 图15是示出图1的关联词登记服务器的关联词的登记的动作例的流程图。
- [0047] 图16是示出字符串组的一例的示意图。

具体实施方式

[0048] 下面,参照附图对本发明的实施方式进行说明。另外,以下说明的实施方式是针对关联词登记系统应用本发明的情况下的实施方式。

[0049] [1.关联词登记系统的结构和功能概要]

[0050] 首先,使用图1对本发明的一个实施方式的关联词登记系统的结构和概要功能进行说明。

[0051] 图1是示出本实施方式的关联词登记系统1的概要结构例的示意图。

[0052] 如图1所示,关联词登记系统1具有:关联词登记服务器10(关联词登记装置的一例),其登记关联词;信息提供服务器20,其为了实现用户利用的作为信息提供网站的一例的购物网站(内部EC(electronic commerce,电子商务)网站的一例)运营而设置,用于对用户从关联词登记服务器10中登记的信息和商品的信息;以及终端30,其供用户在信息提供服务器20中进行商品等的检索。

[0053] 关联词登记服务器10和信息提供服务器20通过局域网等连接,能够相互收发数据,构成服务器系统5。而且,服务器系统5和终端30通过网络3连接,能够根据通信协议(例如TCP/IP)收发数据。另外,网络3例如通过因特网、专用通信线路(例如CATV(Community Antenna Television)线路)、移动通信网(包含基站等)以及网关等构建。

[0054] 并且,关联词登记服务器10根据终端30等中的用户的检索行动等构建关联词的数据库,支援终端30等中的用户的检索行动。

[0055] 信息提供服务器20从终端30接收包含检索字的检索查询并进行检索,对终端30发送检索结果。并且,为了构建关联词的数据库,信息提供服务器20对关联词登记服务器10发送所接收到的检索查询。并且,信息提供服务器20作为购物网站,进行商品的检索、商品信息的提供、广告信息的提供和商品的购入手续等。

[0056] 用户的终端30存在多个。

[0057] [2.各服务器的结构和功能]

[0058] (2.1关联词登记服务器10的结构和功能)

[0059] 接着,使用图2对关联词登记服务器10的结构和功能进行说明。

[0060] 图2是示出关联词登记服务器10的概要结构的一例的框图。

[0061] 如图2所示,作为计算机进行工作的关联词登记服务器10具有通信部11、存储部12、输入输出接口部13、系统控制部14。而且,系统控制部14和输入输出接口部13经由系统总线15而连接。

[0062] 通信部11与网络3连接,对与终端30等的通信状态进行控制,进而,与局域网连接,与局域网上的信息提供服务器20等的其他服务器进行数据收发。

[0063] 存储部12例如由硬盘驱动等构成,存储操作系统和服务器程序等的各种程序、数据等。另外,各种程序例如可以经由网络3从其他服务器装置等取得,也可以记录在记录介质中并经由驱动装置(未图示)读入。

[0064] 并且,在存储部12中构建存储有与从终端30接收到的检索查询有关的记录的检索查询记录/数据库12a(以下设为“检索查询记录DB12a”)、存储根据检索查询而生成的关联词的候选字符串作为关联词候选集合的关联词候选数据库12b(以下设为“关联词候选DB12b”)、存储根据检索查询而生成的关联词的关联词数据库12c(以下设为“关联词DB12c”)、字符串组数据库12d(以下设为“字符串组DB12d”)等。另外,检索查询记录DB12a、关联词候选DB12b、关联词DB12c和字符串组DB12d可以在服务器系统5内构建在关联词登记服务器10以外的其他服务器中,也可以构建在服务器系统5外部。

[0065] 在检索查询记录DB12a(检索查询存储单元的一例)中存储有与接收到检索查询的接收时刻、用户ID等的用于区分用户的用户识别信息对应的检索查询。另外,作为用户识别信息,举出购物网站等的用户ID、用户终端或访问点的识别编号、IP地址等。进而,在信息提供服务器20中,在用户进行商品检索的情况下,还与商品所属的商品类别、用户打开的Web页所示的商品类别等的类别信息对应地,将检索查询存储在检索查询记录DB12a中。

[0066] 在关联词候选DB12b中存储有根据规定条件从检索查询记录DB12a的检索查询记录中提取出的关联词的候选字符串,作为关联词候选集合。

[0067] 在关联词DB12c中存储有由关联词登记服务器10生成的关联词。

[0068] 在字符串组DB12d(字符串组存储单元的一例)中存储有接收到的检索查询的检索字和从检索查询记录DB12a中提取出的检索查询的检索字的字符串组。

[0069] 接着,输入输出接口部13进行通信部11和存储部12与系统控制部14之间的接口处理。

[0070] 系统控制部14由CPU(Central Processing Unit)14a、ROM(Read Only Memory)14b、RAM(Random Access Memory)14c等构成。通过由CPU14a读出ROM14b或存储部12中存储的各种程序并执行,系统控制部14进行关联词的登记处理等。

[0071] (2.2信息提供服务器20的结构和功能)

[0072] 接着,使用图3对信息提供服务器20的结构和功能进行说明。

[0073] 图3是示出信息提供服务器20的概要结构的一例的框图。

[0074] 如图3所示,信息提供服务器20具有通信部21、存储部22、输入输出接口部23、系统控制部24,系统控制部24和输入输出接口部23经由系统总线25而连接。另外,信息提供服务器20的结构和功能与关联词登记服务器10的结构和功能大致相同,所以,以关联词登记服务器10的各结构和各功能中的不同之处为中心进行说明。

[0075] 通信部21通过网络3或局域网等对与终端30或关联词登记服务器10等的通信状态

进行控制等。

[0076] 在存储部22中构建商品数据库(以下设为“商品DB”。)22a和会员数据库(以下设为“会员DB”。)22b等。

[0077] 在商品DB22a中,与用于识别商品的标识符即商品ID相关联地存储有商品名、种类、商品的图像、规格和商品信息、与各商品有关的广告信息等。并且,在商品DB22a中存储有通过HTML(HyperText Markup Language)、XML(Extensible Markup Language)等的标记语言等记述的商品网页的文件等。

[0078] 进而,在商品DB22a中构建用于检索商品的检索数据库。商品DB22a是存储与服务系统5内部的EC网站中销售的商品关联的检索字的检索字存储单元的一例。

[0079] 在会员DB22b中登记有进行了会员登记的用户(购物网站的利用者)的用户ID、名称、住址、电话号码、邮件地址、职业、兴趣、购买历史、用户关心的主题和种类(商品类别)等的用户信息。并且,在会员DB22b中登记有用户从终端30登录购物网站时所需要的用户ID、登录ID和密码。这里,登录ID和密码是登录处理(用户的认证处理)中使用的登录信息。

[0080] 系统控制部24由CPU24a、ROM24b、RAM24c等构成。而且,通过由CPU24a读出ROM24b和存储部22中存储的各种程序并执行,系统控制部24进行商品检索处理、用户的商品购入处理等。

[0081] (2.3终端30的结构和功能)

[0082] 接着,使用图4对终端30的结构和功能进行说明。

[0083] 图4是示出终端30的概要结构的一例的框图。

[0084] 如图4所示,作为计算机进行工作的终端30例如是个人计算机、包含智能手机的便携型无线电话机、PDA等的便携终端,具有通信部31、存储部32、显示部33、操作部34、输入输出接口部35、系统控制部36。而且,系统控制部36和输入输出接口部35经由系统总线37而连接。

[0085] 通信部31通过网络3对与信息提供服务器20等的通信进行控制。另外,在终端30为便携终端装置的情况下,通信部31具有无线通信功能,以使得与网络3的移动体通信网连接。

[0086] 存储部32例如由硬盘驱动等构成,存储操作系统、网页浏览器的程序、网页浏览器用工具条的程序等。

[0087] 显示部33例如由液晶显示元件或EL(Electro Luminescence)元件等构成。在显示部33中,通过网页浏览器显示检索画面的网页、从信息提供服务器20提供的商品检索的结果的网页。

[0088] 操作部34例如由键盘和鼠标等构成。用户通过操作部34输入应答。另外,在显示部33为触摸面板这种触摸开关方式的显示面板的情况下,操作部34取得用户接触或接近的显示部33的位置信息。

[0089] 输入输出接口部35是通信部31和存储部32与系统控制部36之间的接口。

[0090] 系统控制部36例如具有CPU36a、ROM36b、RAM36c。系统控制部36的CPU36a读出ROM36b、RAM36c、存储部32中存储的各种程序并执行。例如,系统控制部36执行网页浏览器的程序而作为网页浏览器进行工作。

[0091] [3.关联词登记系统中的第1实施方式的动作]

[0092] 接着,使用图5~图11对本发明的一个实施方式的关联词登记系统1中的第1实施方式的动作进行说明。

[0093] 图5是示出关联词登记服务器10的关联词候选集合的生成的动作例的流程图。图6是示出终端30中显示的网页的一例的示意图。图7是示出关联词登记服务器的关联词的登记的动作例的流程图。图8是示出部分字符串的提取的一例的示意图。图9是示出排序-得分/曲线的一例的示意图。图10是示出关联词登记服务器10中的计算适宜性得分的动作例的子进程的流程图。图11是示出用于计算适宜性得分的可用性的图形的一例的线图。

[0094] (3.1关联词候选集合的生成)

[0095] 使用图5对关联词候选集合的生成进行说明。

[0096] 首先,信息提供服务器20在从终端30接收到包含用户输入的检索字的检索查询时进行检索,并且对关联词登记服务器10发送该检索查询。进而,信息提供服务器20还对关联词登记服务器10发送针对检索查询的检索结果的件数。

[0097] 然后,关联词登记服务器10从信息提供服务器20接收检索查询和检索结果的件数,将检索查询与检索结果的件数对应地存储在检索查询记录DB12a中。此时,关联词登记服务器10也可以与接收到检索查询的接收时刻、用户ID、IP地址等的用于区分用户的用户识别信息对应地存储检索查询。

[0098] 接着,如图5所示,关联词登记服务器10从检索查询记录中提取检索查询(步骤S1)。具体而言,关联词登记服务器10的系统控制部14从检索查询记录DB12a中提取一个检索查询。

[0099] 接着,关联词登记服务器10判定是否是检索结果的件数为1件以上的检索查询(步骤S2)。具体而言,关联词登记服务器10的系统控制部14从检索查询记录DB12a中读出与提取出的检索查询对应的检索结果的件数,判定检索结果的件数是否为1件以上。另外,关联词登记服务器10的系统控制部14也可以构成为,为了求出符合检索查询的检索结果的件数,对信息提供服务器20发送所提取出的检索查询,从信息提供服务器20接收符合件数。

[0100] 在检索结果的件数为1件以上的情况下(步骤S2:是),关联词登记服务器10在关联词候选集合中登记所提取出的检索查询(步骤S3)。具体而言,在检索结果的件数为1件以上、即检索结果的件数不是零件的检索查询的情况下,关联词登记服务器10的系统控制部14将该检索查询(包含作为关联词的候选字符串的检索字)作为关联词候选集合,登记在关联词候选DB12b中。

[0101] 在检索结果的件数不是1件以上的情况下(步骤S2:否),关联词登记服务器10不在关联词候选集合中登记所提取出的检索查询。

[0102] 接着,判定是否存在剩余的检索查询(步骤S4)。具体而言,关联词登记服务器10的系统控制部14判定在检索查询记录DB12a中是否存在还未进行检索结果件数判定的检索查询。

[0103] 在存在剩余的检索查询的情况下(步骤S4:是),返回步骤S1,关联词登记服务器10提取下一个检索查询,在不存在剩余的检索查询的情况下(步骤S4:否),结束关联词候选集合的生成的处理。这样,关联词登记服务器10预先生成某种程度的规模的关联词候选集合。关联词登记服务器10作为关联词候选集合生成单元的一例进行工作,其根据规定条件从存储了与检索查询有关的记录的检索查询记录中提取关联词的候选,生成关联词候选集合。

[0104] 另外,信息提供服务器20在接收检索查询并进行检索时,在检索结果的件数不是零件的情况下,也可以在检索查询中附加不是零件的信息并发送到关联词登记服务器10。该情况下,关联词登记服务器10在接收到检索结果的件数不是零件的检索查询时,将其存储在关联词候选DB12b中。

[0105] (3.2关联词的登记)

[0106] 接着,使用图6~图9对关联词的登记的动作进行说明。

[0107] 首先,如图6所示,在终端30的显示部33中,用户对从信息提供服务器20提供的Web页40的检索字输入栏41输入例如“A家的餐桌”这样的检索字。点击检索按钮42进行检索,终端30对信息提供服务器20发送包含该检索字的检索查询。接着,在信息提供服务器20从终端30接收到包含用户输入的检索字的检索查询时,对关联词登记服务器10发送该检索查询。

[0108] 接着,如图7所示,关联词登记服务器10接收包含用户输入的检索字的检索查询(步骤S10)。具体而言,关联词登记服务器10的系统控制部14从信息提供服务器20通过通信部11接收检索查询。另外,如在上述关联词候选集合的生成中说明的那样,关联词登记服务器10将接收到的检索查询存储在检索查询记录DB12a中。并且,关联词登记服务器10也可以在满足步骤S2的条件的情况下,将检索查询登记在关联词候选集合中。

[0109] 接着,关联词登记服务器10取得检索字的字符数(步骤S11)。具体而言,关联词登记服务器10的系统控制部14在检索查询中包含的检索字为汉字等的情况下,变换为平假名这样的读音,或者进行罗马字变换。例如,关联词登记服务器10的系统控制部14将“A家的餐桌”表音变换为“えいけのしよくたく”,使字符数从5个字符增加到9个字符。进而,关联词登记服务器10的系统控制部14也可以将“えいけのしよくたく”罗马字变换为“eikenoshokutaku”,从而增加到15个字符。然后,关联词登记服务器10的系统控制部14取得进行平假名变换或罗马字变换等的表音变换后的检索字的字符数(例如9个字符或15个字符)。

[0110] 接着,关联词登记服务器10判定检索字的字符数是否为5个字符以上(步骤S12)。具体而言,关联词登记服务器10的系统控制部14判定表音变换后的检索字的字符数是否为5个字符以上。

[0111] 在检索字的字符数为5个字符以上的情况下(步骤S12:是),关联词登记服务器10从检索字中提取4个字符的部分字符串(步骤S13)。具体而言,如图8所示,关联词登记服务器10的系统控制部14针对检索字的字符串“eikenoshokutaku”,设部分字符串的长度(Window Size)L为4,设Window的滑动宽度(Sliding Size)S为2,生成部分字符串p1(“eike”)、部分字符串p2(“keno”)等的部分字符串并进行提取。另外,根据部分字符串的长度L,步骤S2的阈值(5个字符)被设定为比该部分字符串的长度L大的值。

[0112] 这样,关联词登记服务器10作为根据检索字的字符串生成部分字符串的部分字符串生成单元的一例进行工作。并且,关联词登记服务器10作为部分字符串生成单元的一例进行工作,其根据检索字的字符数设定部分字符串的长度,生成该设定的长度的部分字符串。并且,关联词登记服务器10作为部分字符串生成单元的一例进行工作,其根据检索字的字符数设定生成部分字符串时的检索字的字符数的移动宽度,以该设定的移动宽度生成部分字符串。并且,关联词登记服务器10作为针对使检索字进行表音变换后的检索字生成部分字符串的部分字符串生成单元的一例进行工作。

[0113] 接着,关联词登记服务器10从关联词候选集合中提取与部分字符串部分一致的关联词的候选字符串(步骤S14)。具体而言,关联词登记服务器10的系统控制部14根据图8所示的部分字符串生成检索查询Query="eike"OR"keno"OR"nosh"OR"shok"OR"okut"OR"utak",根据该检索查询,从关联词候选DB12b中提取关联词的候选字符串。即,关联词登记服务器10的系统控制部14提取与所提取出的各部分字符串部分一致的候选字符串。这里,部分一致是指,例如在部分字符串"eike"的情况下,是包含"eike"的候选字符串、包含"eike"的一部分"eik"的候选字符串。这样,关联词登记服务器10作为候选字符串提取单元的一例进行工作,其根据所生成的部分字符串,从关联词候选集合中提取候选字符串。

[0114] 并且,在检索字的字符数不是5个字符以上的情况下(步骤S12:否),关联词登记服务器10从关联词候选集合中提取与检索字部分一致的关联词的候选字符串(步骤S15)。具体而言,关联词登记服务器10的系统控制部14不提取部分字符串,而从关联词候选DB12b中提取与所接收到的检索字本身或表音变换后的检索字部分一致的候选字符串。

[0115] 另外,关联词登记服务器10的系统控制部14也可以根据相似度来提取与部分字符串或检索字相似的候选字符串。另外,作为相似度。举出Jaro-Winkler距离、莱文斯坦距离等的字符串间的距离。

[0116] 接着,关联词登记服务器10计算各关联词的候选字符串的适宜性得分(步骤S16)。具体而言,关联词登记服务器10的系统控制部14通过后述适宜性得分分子进程,计算所提取出的各候选字符串的适宜性得分。

[0117] 接着,关联词登记服务器10按照适宜性得分顺序求出排序(步骤S17)。具体而言,关联词登记服务器10的系统控制部14进行按照适宜性得分从高到低的顺序排序各候选字符串的处理。然后,关联词登记服务器10的系统控制部14取出上位一定数例如上位50位的候选字符串,进一步缩小候选字符串。这样,关联词登记服务器10作为按照得分的顺序对候选字符串进行排序的排序单元的一例进行工作。

[0118] 接着,关联词登记服务器10生成适宜性得分的基准线(步骤S18)。具体而言,关联词登记服务器10的系统控制部14根据所取出的上位的关联词的候选字符串中的适宜性得分为上位排序的字,决定并生成基准线。更具体而言,如图9(A)所示,在图示了针对排序的相对适宜性得分的曲线中,系统控制部14例如将对1位~n位的适宜性得分的点进行近似的近似函数设为基准线L1。

[0119] 这里,关于作为基准线的一例的近似函数,举出对数函数、指数函数、1次函数、2次函数、3次函数等的乘幂函数、双曲线这样的圆锥曲线的函数等。近似函数只要是对排序-得分/曲线的各点进行近似或插值的函数即可。在求解近似函数时,利用最小二乘法等的基准即可。并且,相对适宜性得分例如是用排序1位的适宜性得分的值去除各得分而得到的值。另外,在决定基准线L1时,也可以通过最小二乘法决定下位的适宜性得分、50位~10位的各得分。另外,排序-得分的曲线按照适宜性得分的排序顺序排序,所以,成为广义的递减的曲线。

[0120] 这样,关联词登记服务器10作为判定基准生成单元的一例进行工作,其根据候选字符串的适宜性得分和排序,生成针对排序的适宜性得分的基准线,作为是否将候选字符串登记为关联词的判定基准。并且,关联词登记服务器10作为通过对排序与适宜性得分之间的关联进行近似的近似函数求出基准线的判定基准生成单元的一例进行工作。并且,关

关联词登记服务器10作为通过规定范围的排序的数据求出基准线的判定基准生成单元的一例进行工作。

[0121] 接着,关联词登记服务器10判定各适宜性得分与基准线之间的背离是否为阈值以上(步骤S19)。具体而言,关联词登记服务器10的系统控制部14判定基准线L1与排序1位的得分之间的差分是否为阈值以上。更具体而言,关联词登记服务器10的系统控制部14从排序1位的适宜性得分中减去在基准线L1的式子中代入排序的值而得到的值,计算差分。然后,如果差分为阈值 θ 以上,则提取候选字符串作为关联词的登记字符串。这里,改变对适宜性得分进行调节的参数等,通过仿真,例如如基准线的0.1倍那样,预先求出阈值。该情况下,关联词登记服务器10的系统控制部14提取适宜性得分为基准线的1.1倍以上的候选字符串作为关联词。

[0122] 这样,关联词登记服务器10作为登记字符串提取单元的一例进行工作,其提取适宜性得分与基准线L1之间的背离为预先设定的阈值 θ 以上的候选字符串,作为用于登记为关联词的登记字符串。

[0123] 在各适宜性得分与基准线之间的背离为阈值以上的情况下(步骤S19:是),关联词登记服务器10登记与基准线之间的背离为阈值以上的关联词的候选字符串作为关联词(步骤S20)。具体而言,关联词登记服务器10的系统控制部14在各适宜性得分与基准线之间的背离为阈值以上的情况下,在关联词DB12c中登记排序1位的候选字符串,作为接收到的检索字的关联词的登记字符串。另外,在存在多个(例如3个)各适宜性得分与基准线之间的背离为阈值以上的候选字符串的情况下,关联词登记服务器10的系统控制部14不限于排序1位的候选字符串,也可以如排序1位~3位的候选字符串、1位和3位的候选字符串那样存储规定数的候选字符串,还可以如排序2位那样存储规定顺位的候选字符串。这样,关联词登记服务器10作为在关联词数据库中登记所提取出的登记字符串和检索字作为关联词的关联词登记单元的一例进行工作。

[0124] 在差分不是阈值以上的情况下(步骤S19:否),关联词登记服务器10结束处理。

[0125] 接着,对所构建的关联词数据库的利用进行说明。

[0126] 信息提供服务器20在根据从用户的终端30接收到的检索查询进行检索时,作为关联词提取单元的一例,参照关联词登记服务器10的关联词DB12c,提取与检索查询的检索字对应的关联词。例如,信息提供服务器20对关联词登记服务器10发送所接收到的检索查询。关联词登记服务器10从信息提供服务器20接收检索查询,从关联词DB12c中提取与检索查询的检索字对应的登记字符串。然后,关联词登记服务器10作为关联词输出单元的一例,对信息提供服务器20发送所提取出的登记字符串作为关联词。信息提供服务器20对终端30发送所接收到的关联词。然后,如图6所示,终端30如“也许是‘○○▲’?”那样在关联词显示栏43中显示所接收到的关联词。另外,也可以如“排序1的关联词”、“排序2的关联词”那样显示多个关联词。并且,不限于检索结果的件数为零件或较少的情况,也可以显示关联词。

[0127] (3.3适宜性得分的计算的子进程)

[0128] 接着,使用图10和图11对适宜性得分的计算的子进程进行说明。

[0129] 在步骤S14或步骤S15中提取出关联词的候选字符串后,如图10所示,关联词登记服务器10计算所提取出的关联词的候选字符串与检索字的距离(步骤S30)。具体而言,关联词登记服务器10的系统控制部14例如根据式(1),计算检索字u与候选字符串w之间的距离

Distance(w,u)。

[0130] [数1]

[0131] $Distance(w,u) = \alpha \cdot Dist(w,u) + \beta \cdot DistYomi(w,u) \cdots (1)$

[0132] 这里,Distance(w,u)是检索字u与候选字符串w之间的距离,是表示它们以何种程度相似的指标(相似度的一例)。另外,Dist(w,u)是候选字符串w与检索字u之间的Jaro-Winkler距离,Dist.Yomi(w,u)是候选字符串w的读音与检索字u的读音之间的Jaro-Winkler距离。系数 α 和系数 β 存在 $\alpha + \beta = 1$ 的关系,在希望看重读音时,增大系数 β 的比例(例如 $\alpha < \beta, \alpha = 0.3, \beta = 0.7$)。另外,Distance(w,u)作为距离,优选为能够由0~1表现的Jaro-Winkler距离,但是,不限于Jaro-Winkler距离,也可以是莱文斯坦距离、表示检索字u与候选字符串w之间的相似度的指标。

[0133] 接着,关联词登记服务器10取得各关联词的候选字符串的出现频度(步骤S31)。具体而言,关联词登记服务器10的系统控制部14计算检索查询记录中的各关联词的候选字符串的出现频度。更具体而言,关联词登记服务器10的系统控制部14根据式(2),根据检索查询记录DB12a的数据,针对各候选字符串w计算P(w)。

[0134] [数2]

[0135] $P(w) = \log_{10}(Occ.Count(w)) \cdots (2)$

[0136] 这里,P(w)涉及候选字符串w在检索查询记录DB12a的检索记录数据中被检索到几次的出现频度。另外,Occ.Count(w)是检索查询记录中的候选字符串w的出现次数(候选字符串的使用次数)。在使用检索查询记录DB12a的检索查询记录数据的情况下,P(w)较高的字是各种用户在检索中经常输入的一般单词。这样,关联词登记服务器10的系统控制部14作为适宜性得分计算单元的一例,根据检索查询存储单元(检索查询记录DB12a)的数据计算适宜性得分。

[0137] 接着,关联词登记服务器10根据各关联词的候选字符串的商品DB22a中的检索件数计算检索可能性(可用性)(步骤S32)。具体而言,首先,关联词登记服务器10的系统控制部14对信息提供服务器20发送各关联词的候选字符串w,信息提供服务器20的系统控制部24求出商品DB22a中的检索件数(检索结果数),对关联词登记服务器10发送各候选字符串w的检索件数。

[0138] 然后,关联词登记服务器10的系统控制部14根据式(3),根据检索件数计算各候选字符串w的检索可能性的值。

[0139] [数3]

[0140] $Availability(w) = \begin{cases} 1 & (\#of\ Search\ Result(w) > 0) \\ 0 & (\#of\ Search\ Result(w) = 0) \end{cases} \cdots (3)$

[0141] 并且,Availability(w)是基于候选字符串w的检索的检索结果数(检索符合数)的权重系数,例如由下式表现。这里,#of Search Result(w)是候选字符串w的检索结果数。另外,可用性(Availability)如图11(A)所示,除了式(3)以外,如图11(B)(C)所示存在各种变化。

[0142] 接着,关联词登记服务器10根据检索字与关联词的候选字符串之间的距离、各关联词的候选字符串的出现频度、各关联词的候选字符串的检索可能性,计算各关联词的候选字符串的适宜性得分(步骤S33)。具体而言,关联词登记服务器10的系统控制部14根据式

(4)计算各候选字符串w的适宜性得分。

[0143] [数4]

$$[0144] \quad \text{Score}(w|u) = \frac{P(w) + \gamma}{\text{Distance}(w,u) + \delta} \cdot \text{Availability}(w) \cdots \cdots (4)$$

[0145] 这里， γ 和 δ 是用于调整适宜性得分的参数。参数 γ 是针对 $P(w)$ 的调整值，表示允许对不经常输入的字进行修正的程度。不经常输入的字 $P(w)$ 的值较小，所以，在检索中难以搜索到新词汇、新商品。在希望在检索中尽可能搜索到不经常使用的字的情况下，通过服务器侧的设定来增大参数 γ 的值。

[0146] 并且，参数 δ 是针对 $\text{Distance}(w,u)$ 的调整值，主要防止 $\text{Distance}(w,u)$ 成为零而使得分的值发散。另外，关联词登记服务器10预先设定式(4)中的参数 γ 和 δ 、式(1)中的参数 α 和 β (例如 $\gamma > \delta$ 、 $\gamma > 1$ 、 $\delta < 1$ 、 $\gamma = 2$ 、 $\delta = 0.01$)。

[0147] 如上所述，关联词登记服务器10作为得分计算单元的一例进行工作，其根据候选字符串与所述检索字之间的相似度、候选字符串的使用次数和基于候选字符串的检索的检索结果数，计算候选字符串的适宜性得分。并且，关联词登记服务器10作为对相似度和使用次数的信息进行调节的参数调节单元的一例进行工作。

[0148] 以上，根据本实施方式，生成根据检索结果为1以上(规定条件的一例)而从存储了与检索查询有关的记录的检索查询记录DB12a中提取出关联词的候选字符串的关联词候选集合，将其存储在关联词候选DB12b中，接收用户输入的检索字的检索查询，根据检索字的字符串生成部分字符串，根据部分字符串从关联词候选DB12b的关联词候选集合中提取候选字符串，根据候选字符串与检索字之间的相似度、候选字符串的使用次数和基于候选字符串的检索的检索结果数，计算候选字符串的适宜性得分，按照得分的顺序对候选字符串进行排序，根据候选字符串的适宜性得分和排序，生成针对排序的适宜性得分的基准线，作为是否将候选字符串登记为关联词的判定基准，提取适宜性得分与基准线之间的背离为预先设定的阈值以上的候选字符串，作为用于登记为关联词的登记字符串，在关联词DB12c中登记所提取出的登记字符串作为检索字的关联词，由此，根据减少了字符数的部分字符串从关联词候选集合中提取候选字符串，所以，能够通过候选字符串的适宜性得分来抑制不需要的关联词的登记，能够提高关联词的精度。进而，根据本实施方式，能够使处理高速化。

[0149] 并且，当根据关联词对用户提示检索字时，能够使用户快速找到目标检索对象。

[0150] 并且，在关联词登记服务器10根据检索字的字符数设定部分字符串的长度、并生成该设定的长度的部分字符串的情况下，通过决定部分字符串的长度，能够实现处理的高效化，能够实现处理的高速化和关联词精度的提高。

[0151] 并且，在关联词登记服务器10根据检索字的字符数设定生成部分字符串时的检索字的字符数的移动宽度、并以该设定的移动宽度生成部分字符串的情况下，能够进一步使处理高速化。

[0152] 并且，在关联词登记服务器10针对使检索字进行表音变换后的检索字生成部分字符串的情况下，即使是较短的字，也能够高效提取部分字符串，能够实现处理的高速化和关联词精度的提高。

[0153] 并且，关联词登记服务器10作为参数调节单元，在还具有对得分计算单元的相似度和使用次数的信息进行调节的参数 δ 、 γ 的情况下，通过对参数 δ 、 γ 进行调整，能够根据

用户的特性和状况改变适宜性得分,能够进一步提高关联词的精度。

[0154] 并且,关联词登记服务器10作为判定基准生成单元,通过对排序与适宜性得分之间的关联进行近似的近似函数求出基准线,由此,作为脱离近似函数的异常值,容易检测关联词的登记字符串,能够进一步提高关联词的精度。

[0155] 并且,关联词登记服务器10作为判定基准生成单元,在通过规定范围(例如上位)的排序的数据求出基准线L1的情况下,能够从上位的排序中检测脱离基准线L1的候选字符串,所以,能够更加适当地提取处于上位排序的候选字符串。

[0156] 在还具有商品DB22a作为存储与在内部或外部的EC网站中销售的商品相关联的检索字的检索字存储单元的一例的情况下,特别地,使用商品DB22a的数据,能够容易地计算可用性,能够反映在适宜性得分中。

[0157] 并且,在关联词登记服务器10还具有检索查询记录DB12a作为存储与检索查询有关的记录的检索查询存储单元的一例、根据检索查询记录DB12a的数据计算适宜性得分的情况下,当从检索查询记录DB12a中蓄积的字中提取候选字符串时,能够一次性地缩小到用户兴趣集中的候选字符串。特别地,在基于最近的记录的情况下,能够提取特别反映了用户喜好的候选字符串。

[0158] 并且,在步骤S19,关联词登记服务器10作为登记字符串提取单元,也可以对要提取的登记字符串的数量设置上限。该情况下,能够抑制不需要的关联词的登记,能够提高关联词的精度。

[0159] 并且,在信息提供服务器20参照关联词数据库而提取与所接收到的检索查询的检索字对应的关联词并进行输出的情况下,能够对用户提示能够使用户快速找到目标检索对象的高精度的关联词。

[0160] 另外,也可以根据所取出的上位的关联词的候选字符串中的适宜性得分为下位排序的候选字符串来决定基准线。更具体而言,如图9(B)所示,在图示了针对排序的相对适宜性得分的曲线中,关联词登记服务器10的系统控制部14例如决定连结50位和10位的适宜性得分而得到的直线(基准线L2)。

[0161] 基准线L2在x轴为排序、y轴为相对适宜性得分的曲线中由 $y=ax+b$ 的直线的式子表现。这里,相对适宜性得分例如是用排序1位的适宜性得分的值去除各得分而得到的值。另外,在决定基准线L2时,可以通过最小二乘法决定下位的适宜性得分、50位~10位的各得分。这里,如图9(B)所示,与上位的适宜性得分相比,下位的适宜性得分成为按照排序顺序而相互近似的适宜性得分。

[0162] 这里,基准线L2是对排序与适宜性得分之间的关联进行近似的近似函数的一例,是连结50位和10位的适宜性得分而得到的直线。另外,作为近似函数,除了直线(1次函数)以外,也可以是对数函数、指数函数、2次函数或3次函数等的乘幂函数、双曲线这样的圆锥曲线的函数等。

[0163] 在决定基准线L2后,关联词登记服务器10求出基准线与y轴(相对适宜性得分的轴)之间的截距b的值,将对截距b的值乘以常数而得到的值设为阈值 θ 。

[0164] 另外,在步骤S13中,关联词登记服务器10也可以在部分字符串的提取中,根据检索字的字符数而动态变更部分字符串的长度L和Window的滑动宽度S。例如,在检索字的字符数较多的情况下,延长部分字符串的长度L,扩大滑动宽度S。作为基准值,在针对字符数

15设为L=4、S=3的情况下,在检索字的字符数为40时,如L=8、S=3那样延长字符串的长度,如L=4、S=5那样扩大滑动宽度,如L=8、S=5那样扩大字符串的长度和滑动宽度。这样,在关联词登记服务器10根据检索字的字符数而动态变更部分字符串的长度L和Window的滑动宽度S的情况下,能够实现处理的高速化和关联词精度的提高。

[0165] 并且,购物网站也可以是与网络3连接的外部EC网站,信息提供服务器20不限于商品检索,也可以是一般的检索服务器。该情况下,根据外部EC网站等中的检索服务器的检索件数计算检索可能性(可用性)。

[0166] [4.关联词登记系统中的第2实施方式的动作]

[0167] 接着,使用图6和图12~图16对本发明的一个实施方式的关联词登记系统1中的第2实施方式的动作进行说明。另外,针对与所述第1实施方式的动作相同或对应的部分使用相同标号,主要说明不同之处。其他实施方式和变形例也同样。

[0168] (4.1检索查询记录和字符串组集合的生成)

[0169] 首先,使用图6和图12~图14对检索查询记录和字符串组集合的生成进行说明。

[0170] 另外,从以某种程度构建了检索查询记录DB12a的状态起进行说明。在该检索查询记录DB12a中,与用户识别信息一起存储有与检索查询的接收时刻对应的检索查询。进而,还与商品的类别信息对应地存储有检索查询(类别信息在后面叙述)。

[0171] 作为该用户识别信息的一例,例如在来自登录到购物网站的用户A的终端30的检索查询的情况下,举出用户A的用户ID、存在于从用户A的终端30发送的检索查询的头中的IP地址。

[0172] 关于该接收时刻,举出信息提供服务器20从终端30接收到检索查询的接收时刻、关联词登记服务器10从信息提供服务器20接收到该检索查询的接收时刻。如果是信息提供服务器20从终端接收到检索查询的时刻,则信息提供服务器20还对关联词登记服务器10发送检索查询的接收时刻。并且,接收时刻可以是在检索查询记录DB12a中存储检索查询时标注的时间戳。

[0173] 首先,当用户(例如用户A)利用用户A的终端30登录购物网站时,如图6所示,从信息提供服务器20对该终端30发送购物网站的主页即Web页40的信息。然后,在终端30的显示部33中显示Web页40。该Web页40是购物网站的主页,显示有主类别显示45,扩展有针对各类别的链接。进而,显示有该购物网站的关联群组企业的商品和服务的类别显示46,扩展有针对各类别的链接。并且,在Web页40中显示有表示用户A的信息“A,你好”。

[0174] 接着,如图6所示,当在Web页40中点击“家电”的商品类别的链接时,如图12所示,在显示部33中显示“家电”的Web页50。在该Web页50中显示主类别“家电”的类别显示55的子类别显示56。

[0175] 当在该Web页50的检索字记入栏51中输入检索字“▼▼○”并点击检索按钮52时,终端30的系统控制部36通过通信部31对信息提供服务器20发送检索查询。该检索查询具有所输入的检索字、Web页50的类别信息(表示主类别“家电”的类别显示65的类别识别编号等)、作为用户识别信息的一例的(用户A的)用户ID等。

[0176] 接着,信息提供服务器20从终端30接收该检索查询,将其发送到关联词登记服务器10,并且参照商品DB22a进行商品的检索。并且,确定该检索查询的发送方的IP地址,也将该IP地址发送到关联词登记服务器10。该IP地址是用户识别信息的一例,在用户未登录的

情况下,代替用户ID而使用。

[0177] 接着,如图13所示,关联词登记服务器10接收包含检索字的检索查询(步骤S41)。具体而言,关联词登记服务器10的系统控制部14通过通信部11从信息提供服务器20接收检索查询,将接收到的检索查询和检索查询的接收时刻对应地存储在检索查询记录DB12a中。进而,关联词登记服务器10的系统控制部14与IP地址对应地将检索查询存储在检索查询记录DB12a中。另外,关联词登记服务器10的系统控制部14也可以使检索查询包含IP地址。并且,由于得知检索查询的受理的前后关系即可,所以,关联词登记服务器10的系统控制部14也可以不附加接收时刻,而附加接收到的顺序,将检索查询存储在检索查询记录DB12a中。这样,关联词登记服务器10的系统控制部14作为按照接收顺序存储所接收到的检索查询的检索查询存储单元的一例进行工作。

[0178] 接着,关联词登记服务器10判定在检索查询记录中是否存在用户ID一致的先前的检索查询(步骤S42)。具体而言,关联词登记服务器10的系统控制部14从检索查询记录DB12a中提取与用户A的用户ID一致、且接收时刻比接收到的检索查询早的先前(接收顺序较早的一例)的检索查询。然后,关联词登记服务器10的系统控制部14在能够提取先前的检索查询的情况下,判定为在检索查询记录中存在先前的检索查询,在无法提取的情况下,判定为不存在先前的检索查询。另外,关联词登记服务器10的系统控制部14也可以从检索查询记录DB12a中提取与接收到的检索查询的IP地址一致、且接收时刻比接收到的检索查询早的先前的检索查询。

[0179] 这里,与用户ID一致的情况或与IP地址一致的情况是预先设定的检索查询提取条件的一例。这样,关联词登记服务器10作为检索查询提取单元的一例进行工作,作为检索查询提取条件,其提取输入了检索查询的用户的用户识别信息一致的先前的检索查询。

[0180] 在检索查询记录中存在用户ID一致的先前的检索查询的情况下(步骤S42:是),关联词登记服务器10判定接收时刻和先前的检索查询的接收时刻是否为规定时间内(步骤S43)。具体而言,关联词登记服务器10的系统控制部14提取所接收到的检索查询的接收时刻与先前的检索查询的接收时刻之差为规定时间内(例如3秒~60秒之间)的先前的检索查询。然后,关联词登记服务器10的系统控制部14在能够提取先前的检索查询的情况下,判定为先前的检索查询为规定时间内,在无法提取的情况下,判定为不是规定时间内。

[0181] 这里,规定时间内是预先设定的检索查询提取条件的一例。这样,关联词登记服务器10作为检索查询提取单元的一例进行工作,其根据预先设定的检索查询提取条件,从检索查询存储单元中提取接收时刻比接收到的检索查询早的先前的检索查询。并且,关联词登记服务器10作为检索查询提取单元的一例进行工作,作为检索查询提取条件,其提取接收时刻为从检索查询的接收时刻起的规定时间内的先前的检索查询。

[0182] 在接收时刻和先前的检索查询的接收时刻为规定时间内的情况下(步骤S43:是),关联词登记服务器10存储所接收到的检索字和接收时刻最近的先前的检索字的字符串组(步骤S44)。具体而言,关联词登记服务器10的系统控制部14选择提取出的先前的检索查询中的、构成所接收到的检索查询的检索字和构成接收时刻最近的先前的检索查询的检索字,将这些先前的检索字和接收到的检索字的字符串组存储在字符串组DB12d中。

[0183] 另外,作为预先设定的检索查询提取条件的一例,关联词登记服务器10的系统控制部14也可以进一步缩小为类别信息(例如主类别)一致的先前的检索查询。

[0184] 接着,关联词登记服务器10与用户ID和接收时刻对应地将检索查询存储在检索查询记录中(步骤S45)。具体而言,关联词登记服务器10的系统控制部14与类别信息、用户ID(也可以是IP地址)和接收时刻对应地将检索查询存储在检索查询记录DB12a中。并且,在检索查询记录中不存在用户ID一致的先前的检索查询的情况下(步骤S42:否)、接收时刻和先前的检索查询的接收时刻不是规定时间内的情况下(步骤S43:否),关联词登记服务器10的系统控制部14也与用户ID和接收时刻对应地将检索查询存储在检索查询记录DB12a中。

[0185] 这样,关联词登记服务器10作为字符串组存储单元的一例进行工作,其存储构成所提取出的先前的检索查询的先前的检索字和构成所接收到的检索查询的检索字,作为字符串组。

[0186] 接着,在信息提供服务器20进行了商品检索后,作为检索结果,如图14所示,将Web页60的信息发送到终端30。

[0187] 接着,终端30从信息提供服务器20接收Web页60的信息,将其显示在显示部33中。

[0188] 用户A查看Web页60的检索结果,在存在目标商品的情况下,点击商品显示栏63的商品,但是,在不存在目标商品的情况下,进一步在检索字输入栏61中输入检索字,当点击检索按钮62时,如上所述,终端30对信息提供服务器20发送包含检索字“▲▲0■”和类别信息(“家电”)的检索查询。然后,信息提供服务器20接收该检索查询,发送到关联词登记服务器10,进行检索。如上所述,关联词登记服务器10接收检索查询(步骤S41),进行步骤S42~步骤S45的处理。

[0189] 针对来自其他用户B等的检索查询也进行上述处理,在检索查询记录DB12a中生成类别信息、按照每个用户ID或每个IP地址而同时与接收时刻对应的检索查询的检索查询记录,在字符串组DB12d中生成汇集了字符串的字符串组集合。

[0190] (4.2关联词的登记)

[0191] 接着,使用图15和图16对关联词的登记进行说明。

[0192] 首先,如图15所示,关联词登记服务器10判定是否满足提取开始条件(步骤S50)。具体而言,作为提取开始条件的一例,关联词登记服务器10的系统控制部14判定检索查询记录DB12a中的检索查询记录的检索查询数(总检索查询数)、字符串组DB12d中存储的字符串组集合的元素数(字符串组集合的总数)是否超过规定阈值。并且,关联词登记服务器10的系统控制部14也可以判定从上次满足该字符串组提取开始条件起是否经过了规定时间。

[0193] 在满足提取开始条件的情况下(步骤S50:是),关联词登记服务器10从字符串组集合中提取先前的检索字相同的字符串组(步骤S51)。具体而言,如图16所示,关联词登记服务器10的系统控制部14从字符串组DB12d中提取先前的检索字71相同(检索字相同)的字符串组70。在不满足提取开始条件的情况下(步骤S50:否),结束关联词登记的处理。

[0194] 另外,关联词登记服务器10的系统控制部14也可以根据预先设定的字符串组提取开始条件,从字符串组DB12d中提取先前的检索字相似的字符串组。检索字彼此相似例如是指,基于Jaro-Winkler距离、莱文斯坦距离等的字符串间的距离的相似度的值为规定阈值以上的情况。例如,关联词登记服务器10的系统控制部14还提取具有与先前的检索字71相似的先前的检索字“▼▼●”的字符串组。

[0195] 这样,关联词登记服务器10作为字符串提取单元的一例进行工作,其根据预先设定的字符串组提取开始条件,从字符串组存储单元中提取所述先前的检索字相同或相似的

字符串组。

[0196] 接着,关联词登记服务器10判定提取出的字符串组的总数是否为规定以上(步骤S52)。具体而言,关联词登记服务器10的系统控制部14判定与某个先前的检索字71有关的字符串组的总数是否为规定以上(例如3个以上)。如图16所示,与先前的检索字71有关的字符串组有<先前的检索字71-检索字71>的字符串组70、<先前的检索字71-检索字72>的字符串组70、<先前的检索字71-检索字73>的字符串组70合计3个。

[0197] 在提取出的字符串组的总数为规定以上的情况下(步骤S52:是),关联词登记服务器10判定所提取出的字符串组中是否存在满足登记条件的字符串组(步骤S53)。具体而言,作为登记条件的一例,关联词登记服务器10的系统控制部14判定各字符串组70的数量是否为规定数以上(例如10个以上)。这样,关联词登记服务器10的系统控制部14判定在先前的检索字71相同或相似的字符串组70中具有相同或相似的检索字72、73、74的字符串组的数量是否超过规定阈值。关联词登记服务器10的系统控制部14确定超过规定阈值的字符串组作为关联词。另外,如果存在与检索字73“▲▲▲”相似的检索字“▲▲△”,则关联词登记服务器10的系统控制部14也可以计数为相同字符串组。

[0198] 这样,关联词登记服务器10作为关联词确定单元的一例进行工作,其根据预先设定的登记条件,从提取出的字符串组中确定作为关联词的字符串组。并且,关联词登记服务器10作为关联词确定单元的一例进行工作,作为登记条件,在先前的检索字相同或相似的字符串组中具有相同或相似的检索字的字符串组的数量超过规定阈值的情况下,确定该字符串组作为关联词。

[0199] 并且,作为登记条件的一例,关联词登记服务器10的系统控制部14也可以判定在先前的检索字71相同或相似的字符串组中具有相同或相似的检索字的字符串组的比例是否超过规定阈值(例如80%)。

[0200] 这样,关联词登记服务器10作为关联词确定单元的一例进行工作,作为登记条件,在先前的检索字相同或相似的字符串组中具有相同或相似的检索字的字符串组的比例超过规定阈值的情况下,确定该字符串组作为关联词。

[0201] 接着,在存在满足登记条件的字符串组的情况下(步骤S53:是),关联词登记服务器10登记字符串组作为关联词(步骤S54)。具体而言,如图16所示,由于<先前的检索字71-检索字71>的字符串组70为3个(15%)、<先前的检索字71-检索字72>的字符串组70为16个(80%)、<先前的检索字71-检索字73>的字符串组70为1个(5%),所以,关联词登记服务器10的系统控制部14将<先前的检索字71-检索字72>的字符串组70作为关联词而登记在关联词DB12c中。

[0202] 这样,关联词登记服务器10作为登记所确定的字符串组作为关联词的关联词登记单元的一例进行工作。

[0203] 另外,在不存在满足登记条件的字符串组的情况下(步骤S53:否),关联词登记服务器10判定不满足登记条件的字符串组是否登记在关联词DB中(步骤S55)。具体而言,关联词登记服务器10的系统控制部14判定已经登记的<先前的检索字-检索字>的字符串组70的比例是否在字符串组DB12d中减少而穿过规定阈值(例如80%)。在具有相同或相似的先前的检索字的其他字符串组增加的情况下产生这种情况。

[0204] 在不满足登记条件的情况下(步骤S55:是),关联词登记服务器10从关联词DB中删

除不满足登记条件的字符串组(步骤S56)。具体而言,关联词登记服务器10的系统控制部14在关联词DB12c中,从已经登记的字符串组中删除不满足登记条件的字符串组。

[0205] 在步骤S54之后,关联词登记服务器10判定是否针对全部字符串组进行了提取(步骤S57)。具体而言,关联词登记服务器10的系统控制部14判定是否针对字符串组DB12d的字符串组集合中的全部先前的检索字进行了步骤S51的字符串组的提取处理,在未对全部字符串组进行提取的情况下(步骤S57:否),返回步骤S11,在针对全部字符串组进行了提取的情况下(步骤S57:是),结束关联词登记的处理。

[0206] 以上,根据本实施方式,接收用户输入的检索字的检索查询,根据预先设定的检索查询提取条件,从检索查询记录中提取在时间上在接收到的检索查询之前取得的先前的检索查询,将提取出的先前的检索查询的先前的检索字和接收到的检索查询的检索字作为字符串组而存储在字符串组DB12d中,在满足预先设定的字符串组提取开始条件时,提取先前的检索字相同或相似的字符串组,将满足预先设定的登记条件的提取出的字符串组作为关联词而登记在关联词DB12c中,由此,在字符串组提取开始条件下蓄积一定量的接收到的检索查询和基于检索查询提取条件的先前的检索查询的字符串组,从该蓄积的字符串组中,将满足登记条件的字符串组作为关联词而登记在关联词DB12c中,所以,能够提高关联词的精度。

[0207] 并且,关联词登记服务器10还存储用户识别信息,作为检索查询提取条件,在提取输入了检索查询的用户的用户识别信息一致的先前的检索查询的情况下,能够生成同一用户的字符串组,所以,能够进一步提高关联词的精度。

[0208] 并且,作为检索查询提取条件,在提取接收时刻为从检索查询的接收时刻起的规定时间内的先前的检索查询的情况下,关联词登记服务器10能够提取规定时间内的关联性高的先前的检索查询,所以,能够进一步提高关联词的精度。

[0209] 并且,关联词登记服务器10还存储类别信息,作为检索查询提取条件,在根据类别信息提取先前的检索查询的情况下,能够将类别相同或关联的检索字彼此作为字符串组,所以,能够进一步提高关联词的精度。

[0210] 并且,作为字符串组提取开始条件,在字符串组集合的元素数超过规定阈值的情况下,关联词登记服务器10提取字符串组时,能够蓄积某种程度的规模的字符串组,能够从中提取适当的字符串组,所以,能够进一步提高关联词的精度。

[0211] 并且,作为登记条件,在先前的检索字相同或相似的字符串组中具有相同或相似的检索字的字符串组的数量超过规定阈值的情况下,关联词登记服务器10登记字符串组作为关联词时,能够选择高精度的字符串组,能够进一步提高关联词的精度。

[0212] 并且,作为登记条件,在先前的检索字相同或相似的字符串组中具有相同或相似的检索字的字符串组的比例超过规定阈值的情况下,关联词登记服务器10登记该字符串组作为关联词时,能够选择高精度的字符串组,能够进一步提高关联词的精度。

[0213] 另外,在步骤S51中,作为字符串组提取开始条件,也可以在检索查询记录DB12a中的检索查询数超过规定阈值的情况下,关联词登记服务器10提取字符串组。该情况下,能够蓄积某种程度的规模的字符串组,能够从中提取适当的字符串组,所以,能够进一步提高关联词的精度。

[0214] 并且,在步骤S50中,也可以在从满足字符串组提取开始条件起经过了规定时间的

情况下,关联词登记服务器10提取字符串组。该情况下,能够蓄积某种程度的规模的字符串组,能够从中提取适当的字符串组,所以,能够进一步提高关联词的精度。

[0215] 进而,也可以组合关联词登记系统中的第1和第2实施方式的动作来登记关联词。该情况下,也可以以能够区分如第1实施方式的动作那样考虑了字符串彼此的距离的关联词和如第2实施方式的动作那样考虑了使用检索查询的顺序的关联词的方式,在关联词DB12c中构建关联词数据库。

[0216] 作为关联词数据库的活用方法,例如,首先,在终端30中输入“○▽”进行检索的情况下,信息提供服务器20参照在关联词DB12c中通过第1实施方式的动作而构建的关联词数据库,如图6所示,作为关联词,在Web页40中显示“也许‘○○▲’?”。接着,在终端30中选择“○○▲”进行检索的情况下,信息提供服务器20也可以参照在关联词DB12c中通过第2实施方式的动作而构建的关联词数据库,作为关联词,在Web页40中显示“是‘…▲·’么?”。

[0217] 另外,代替这样阶段地显示关联词,也可以同时显示这些关联词。

[0218] 进而,本发明不限于上述各实施方式。上述各实施方式是例示,具有实质上与本发明的权利要求范围所记载的技术思想相同的结构、发挥相同作用效果的方式均包含在本发明的技术范围内。

[0219] 标号说明

[0220] 1:关联词登记系统;10:关联词登记服务器(关联词登记装置);12a:检索查询记录DB;12b:关联词候选DB;12c:关联词DB;12d:字符串组DB;20:信息提供服务器(信息处理装置);22a:商品DB;30:终端。

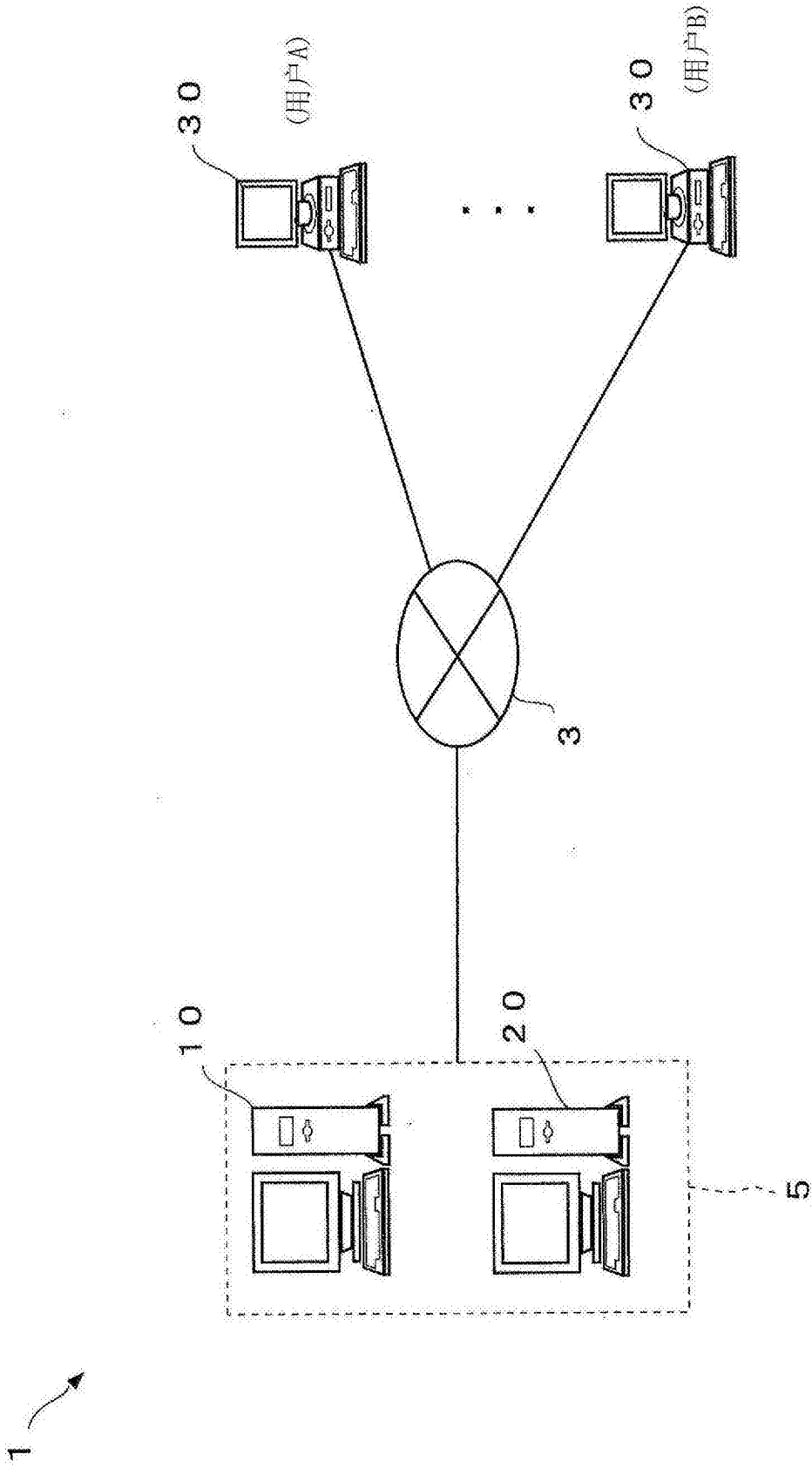


图1

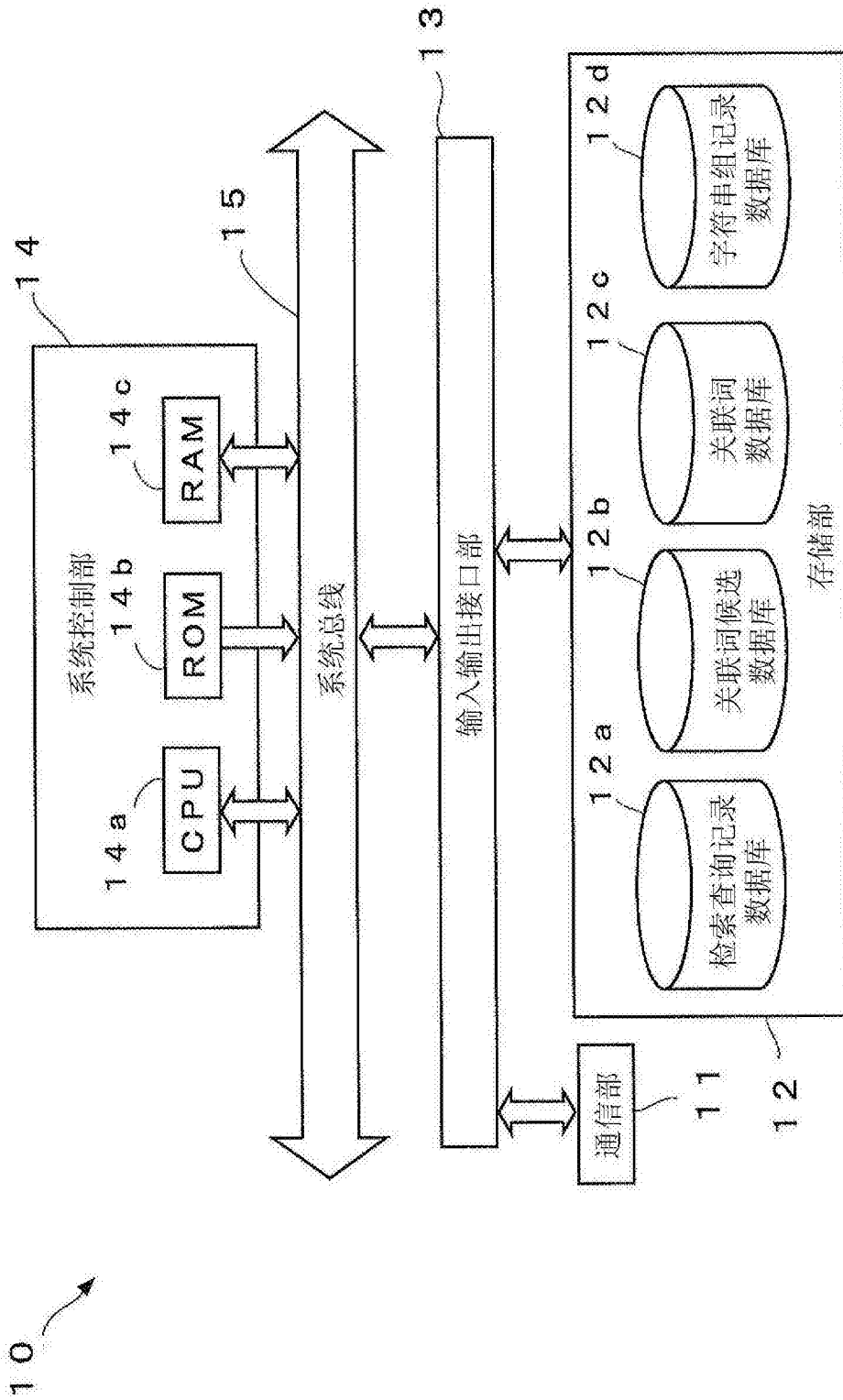


图2

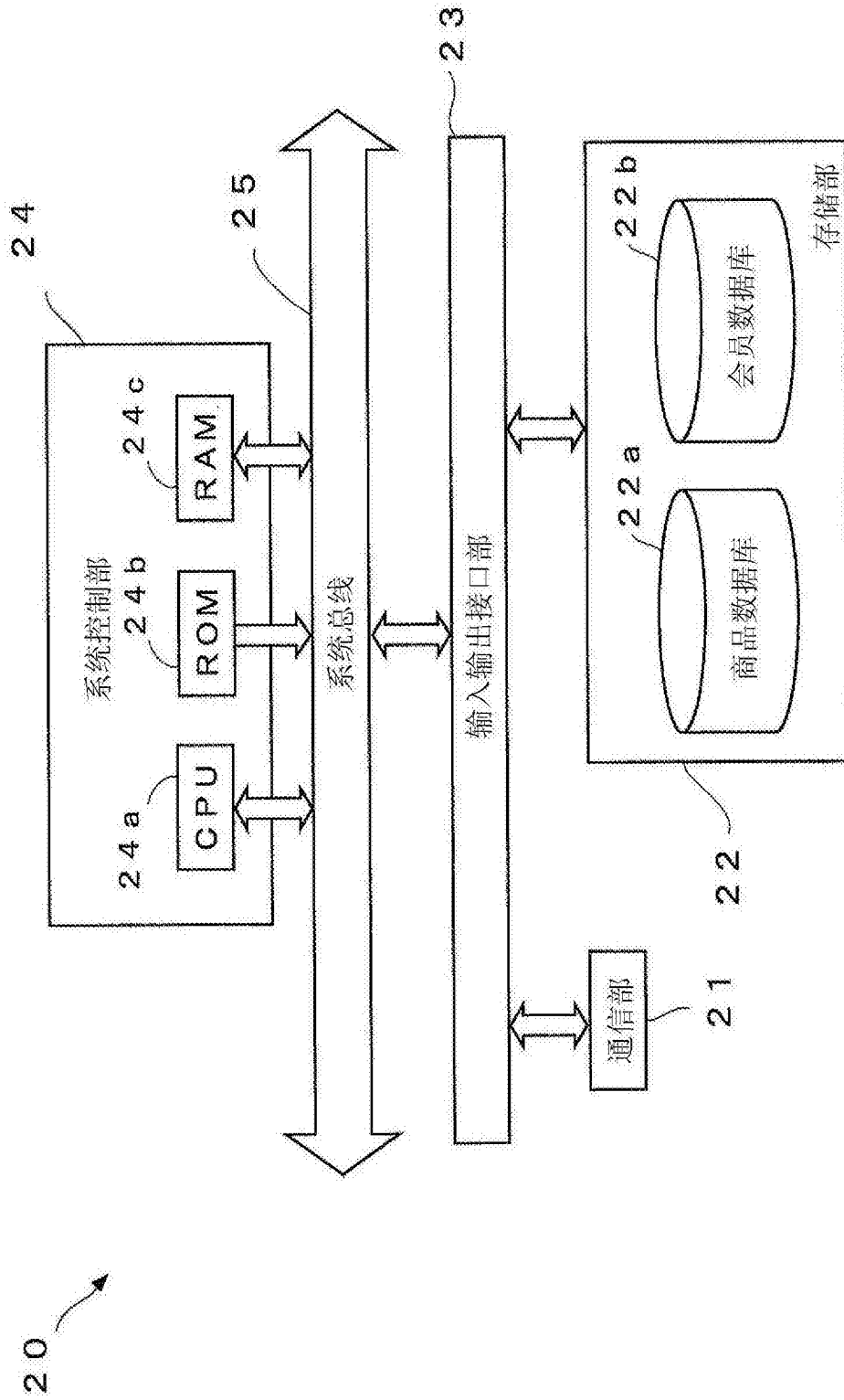


图3

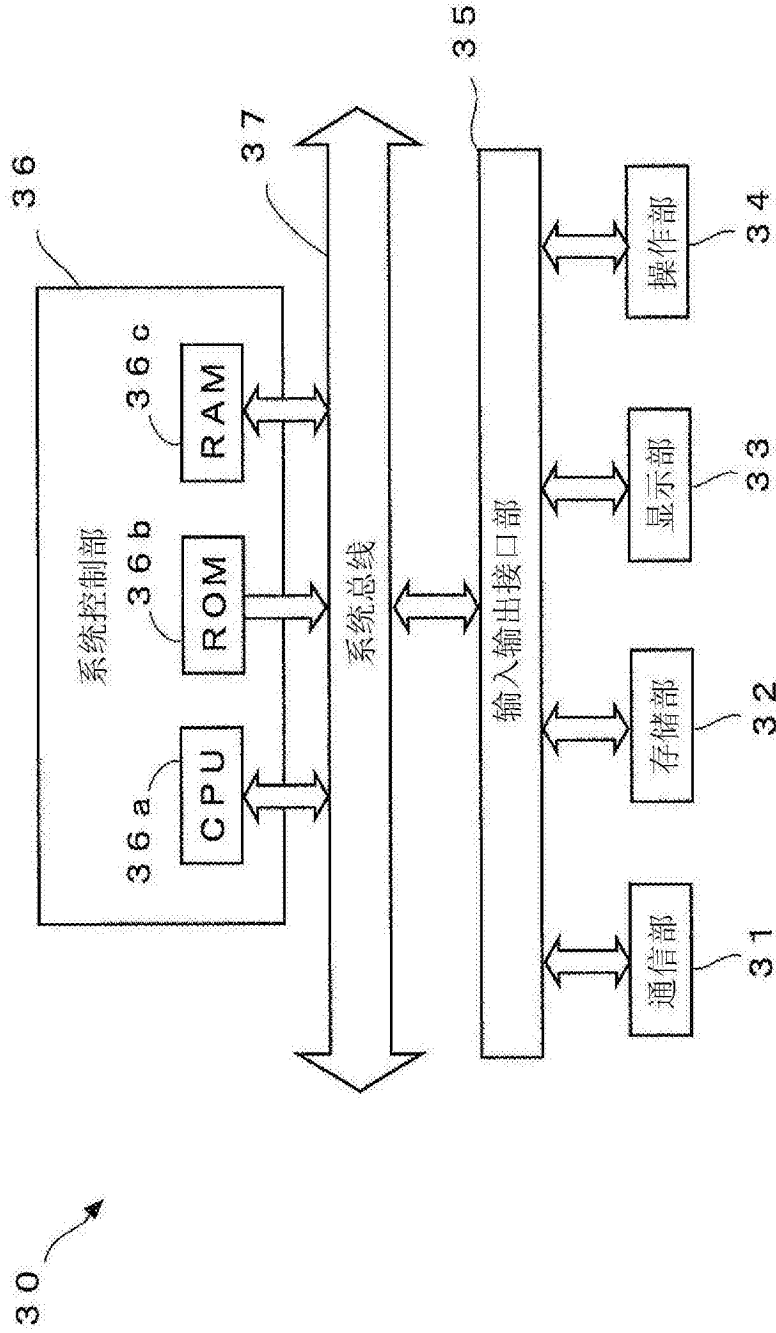


图4

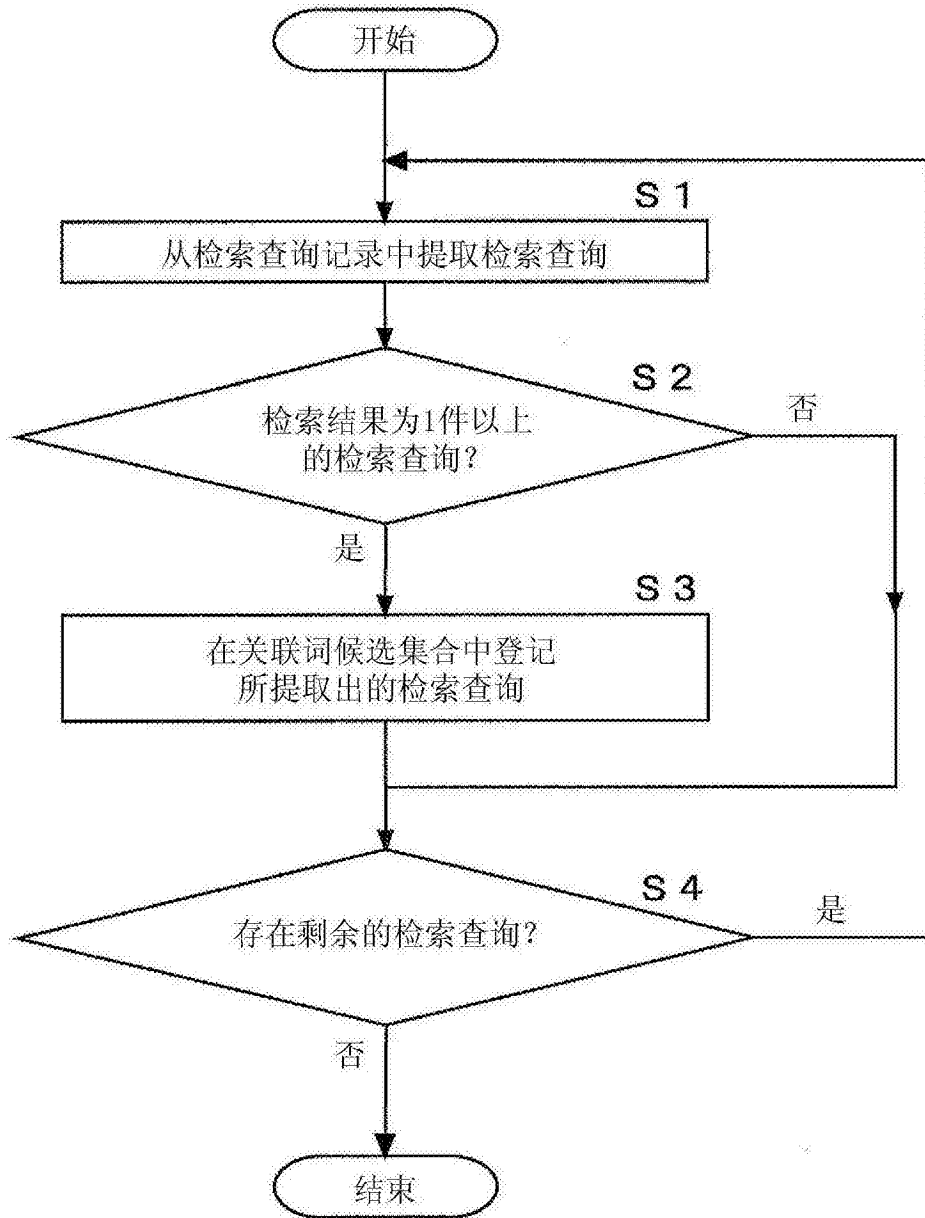


图5

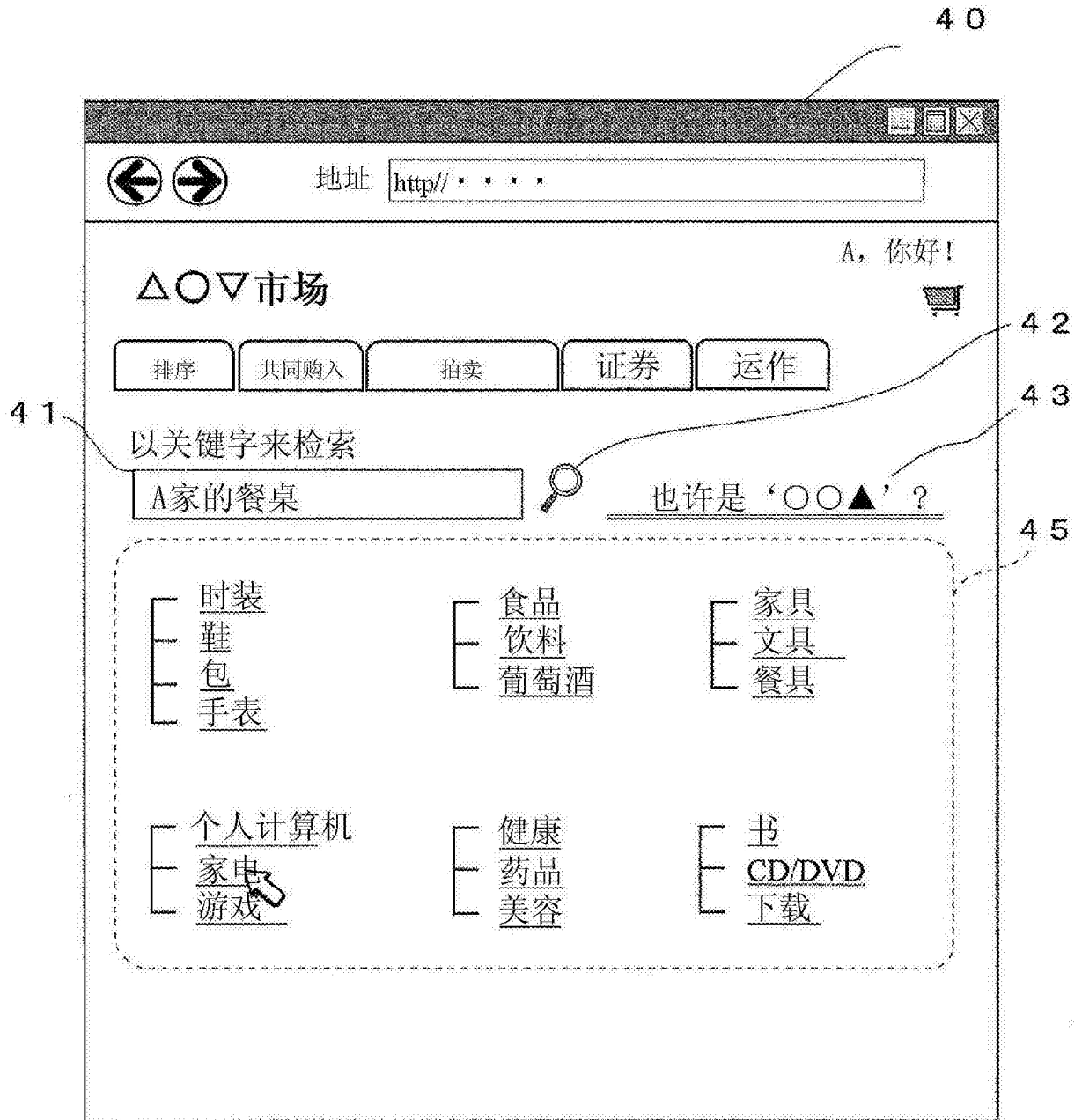


图6

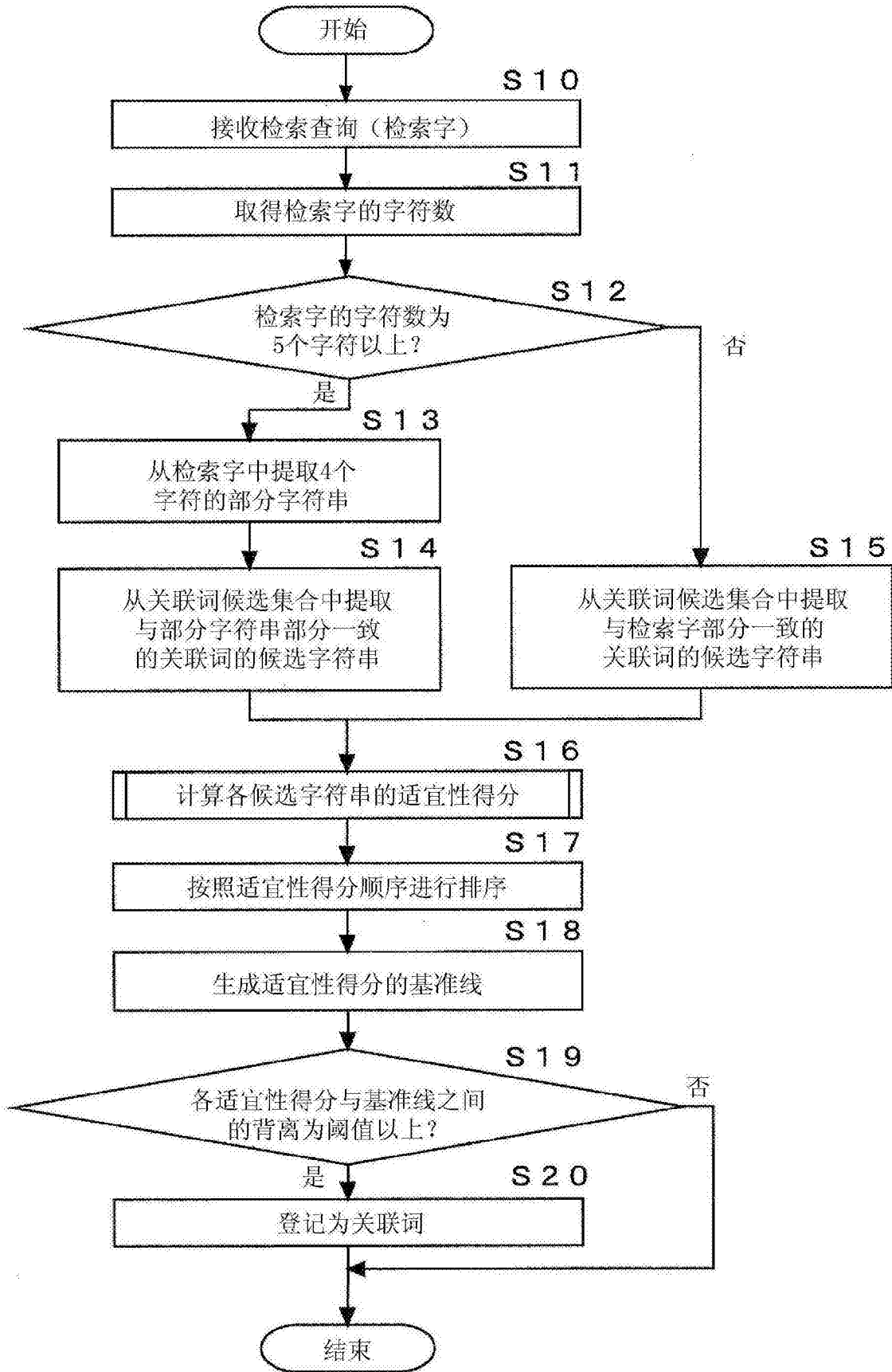


图7

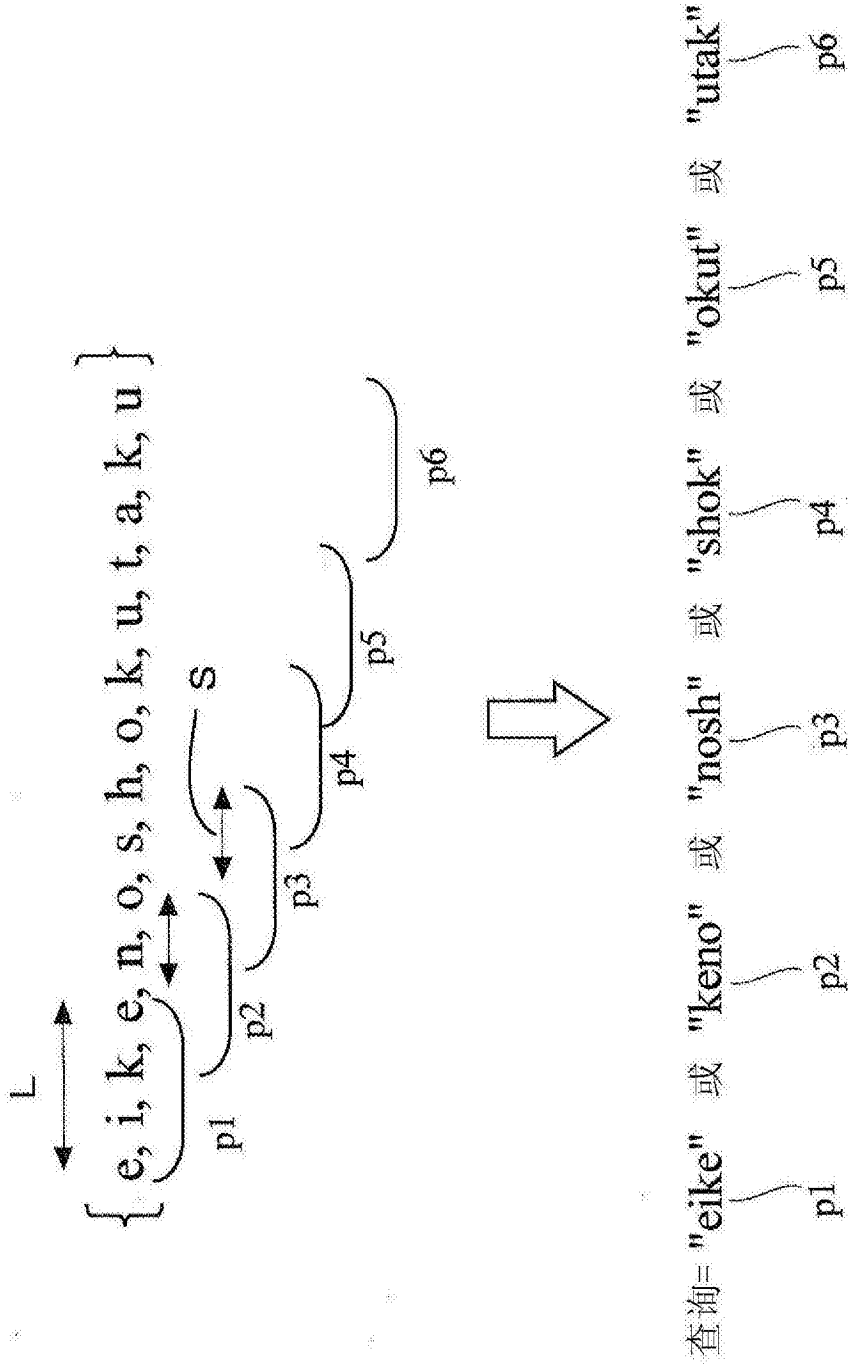
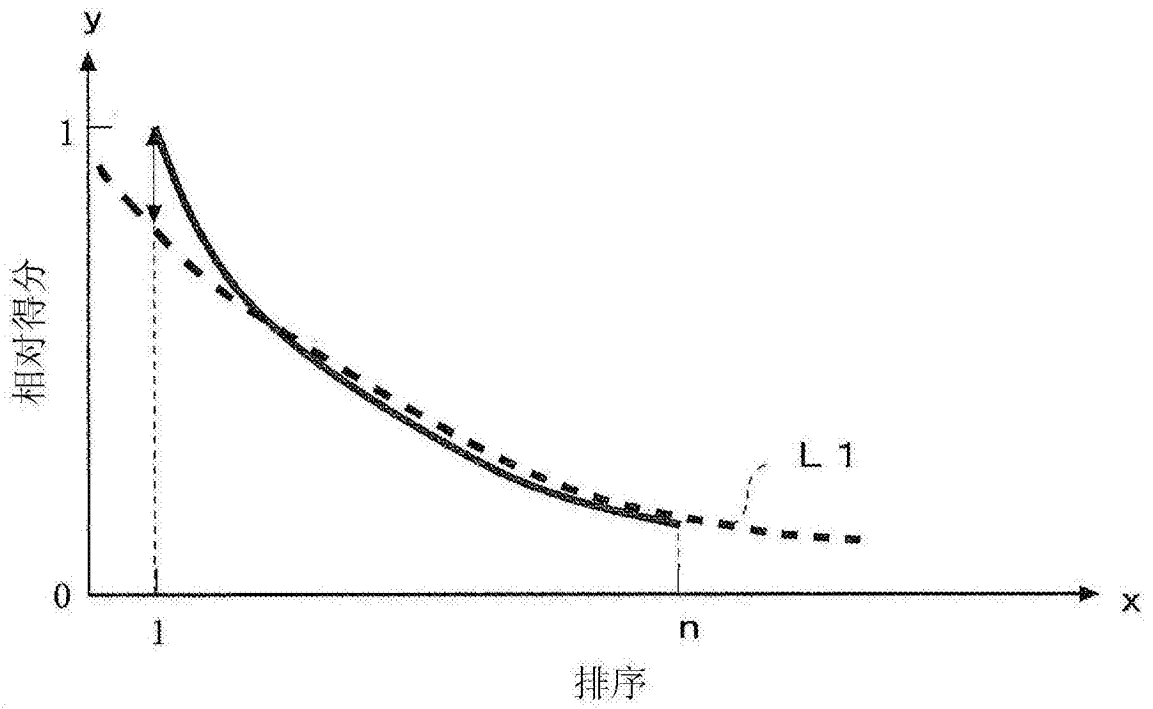


图8

(A)



(B)

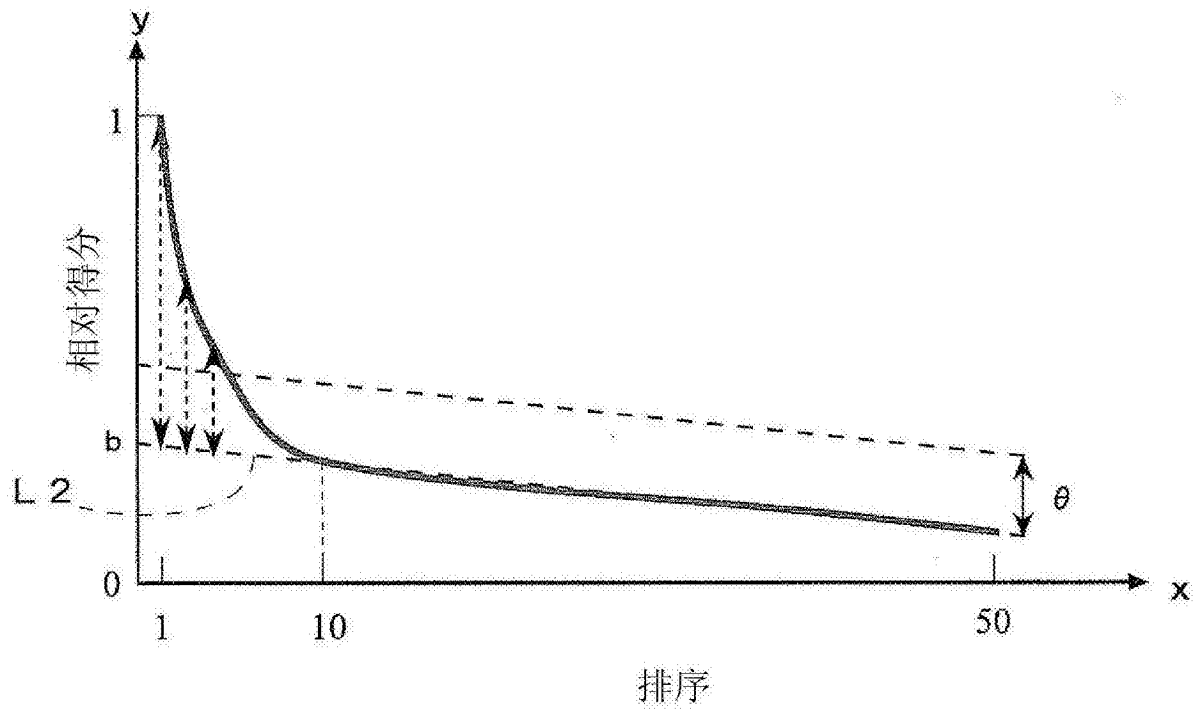


图9

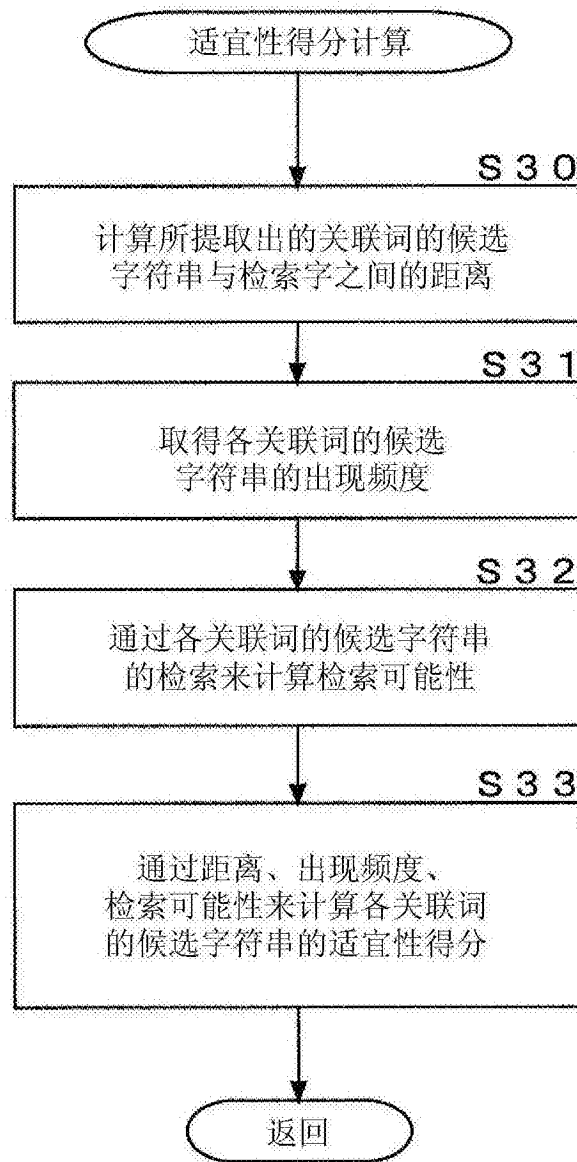


图10

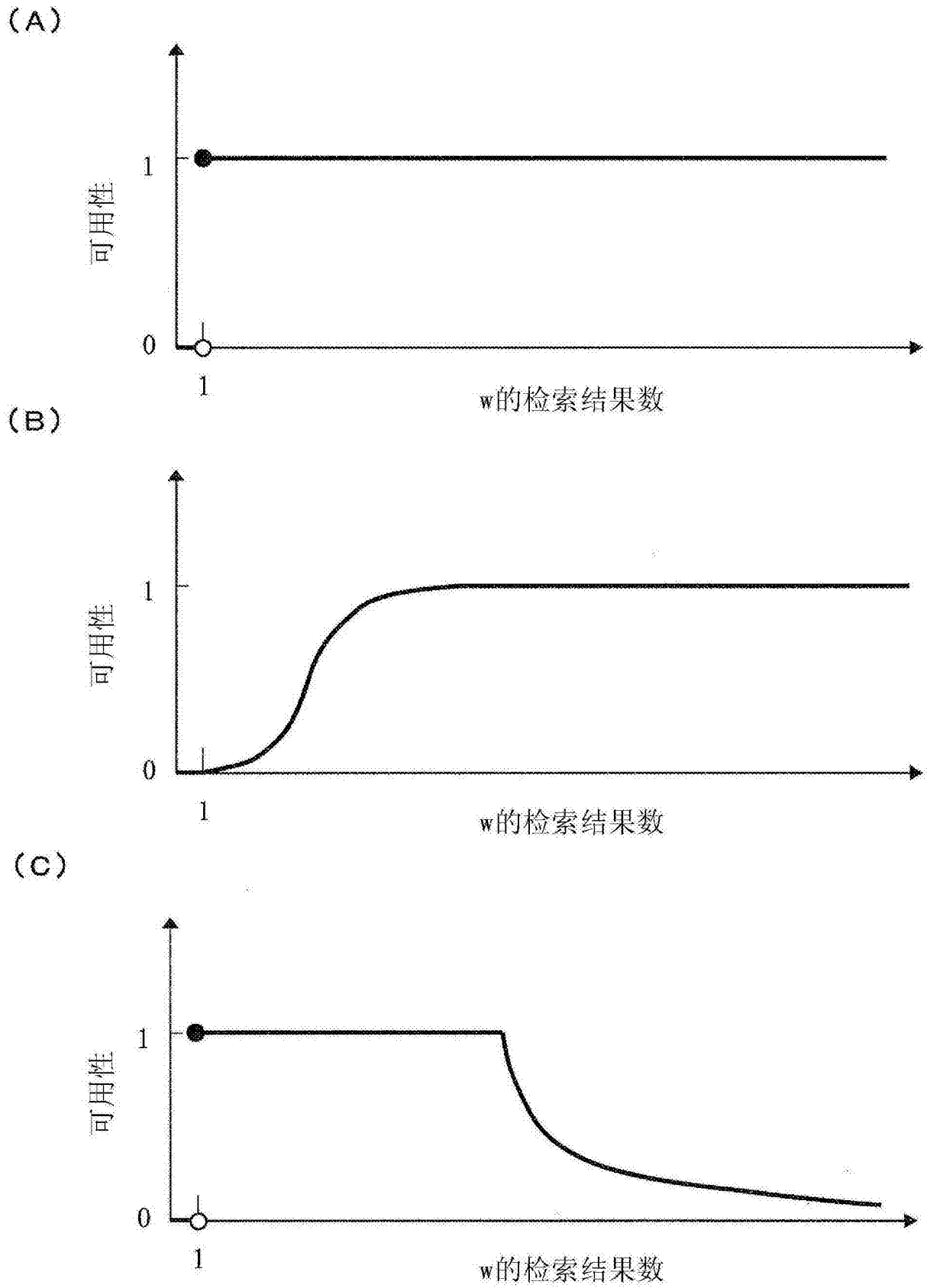


图11

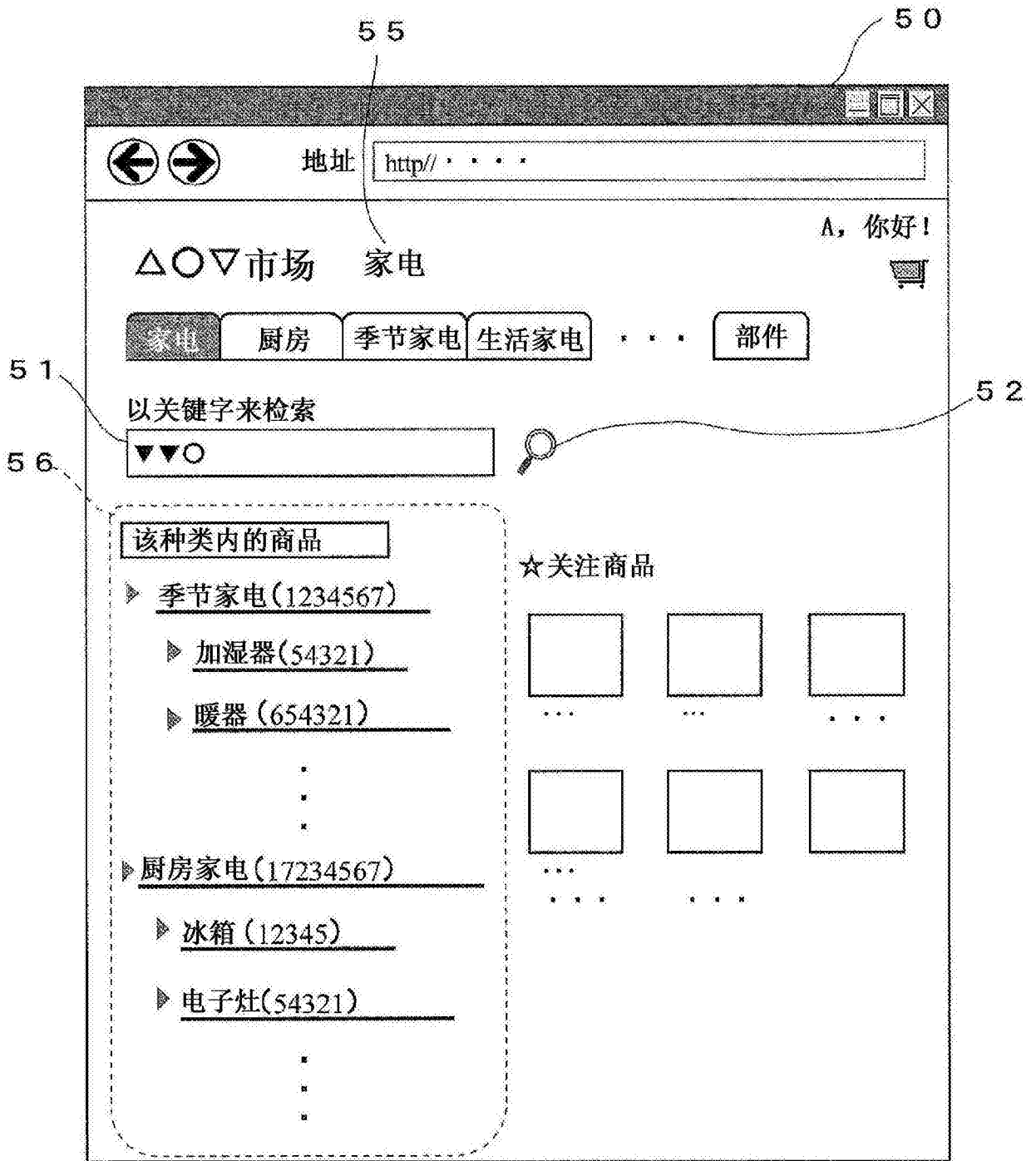


图12

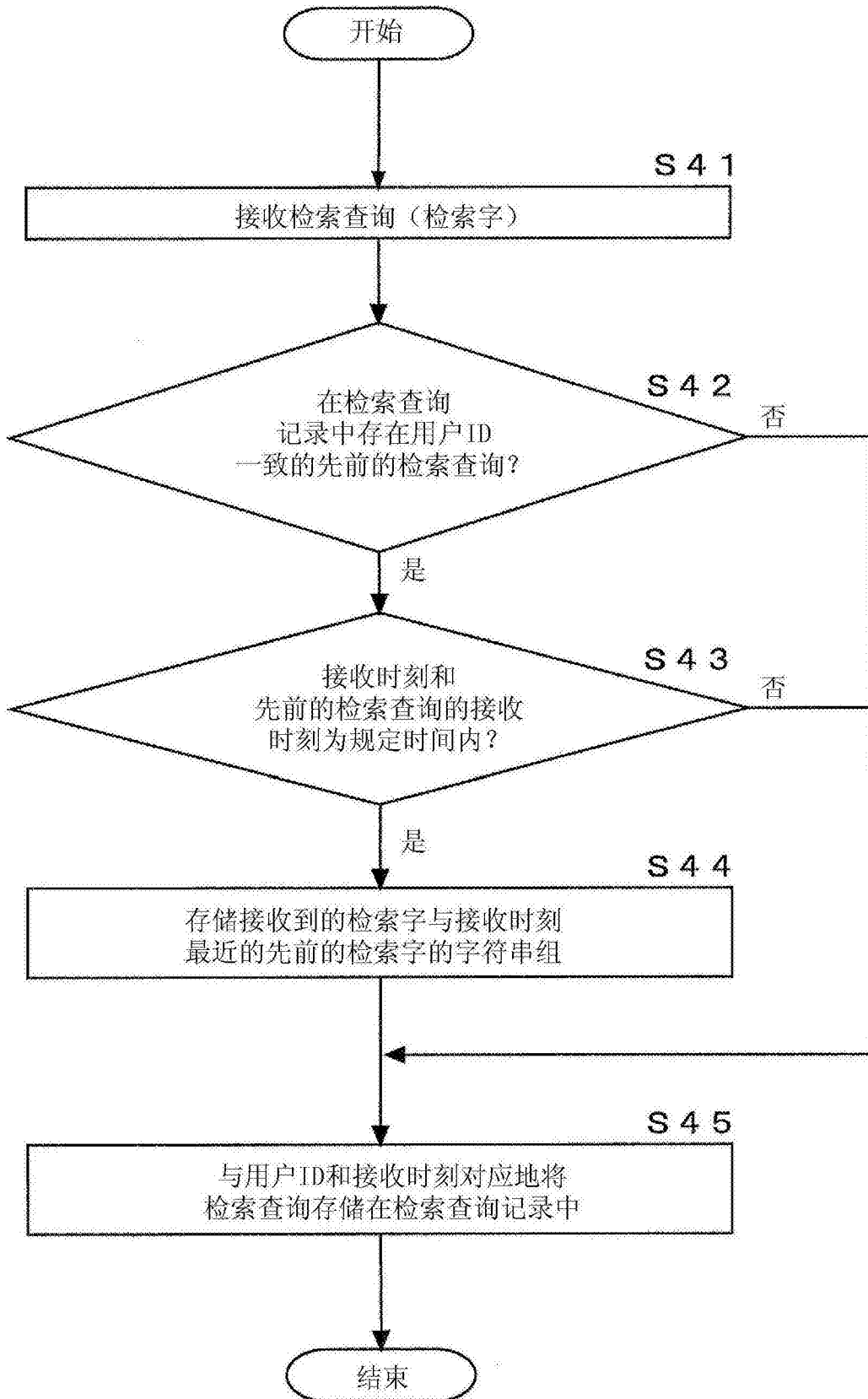


图13

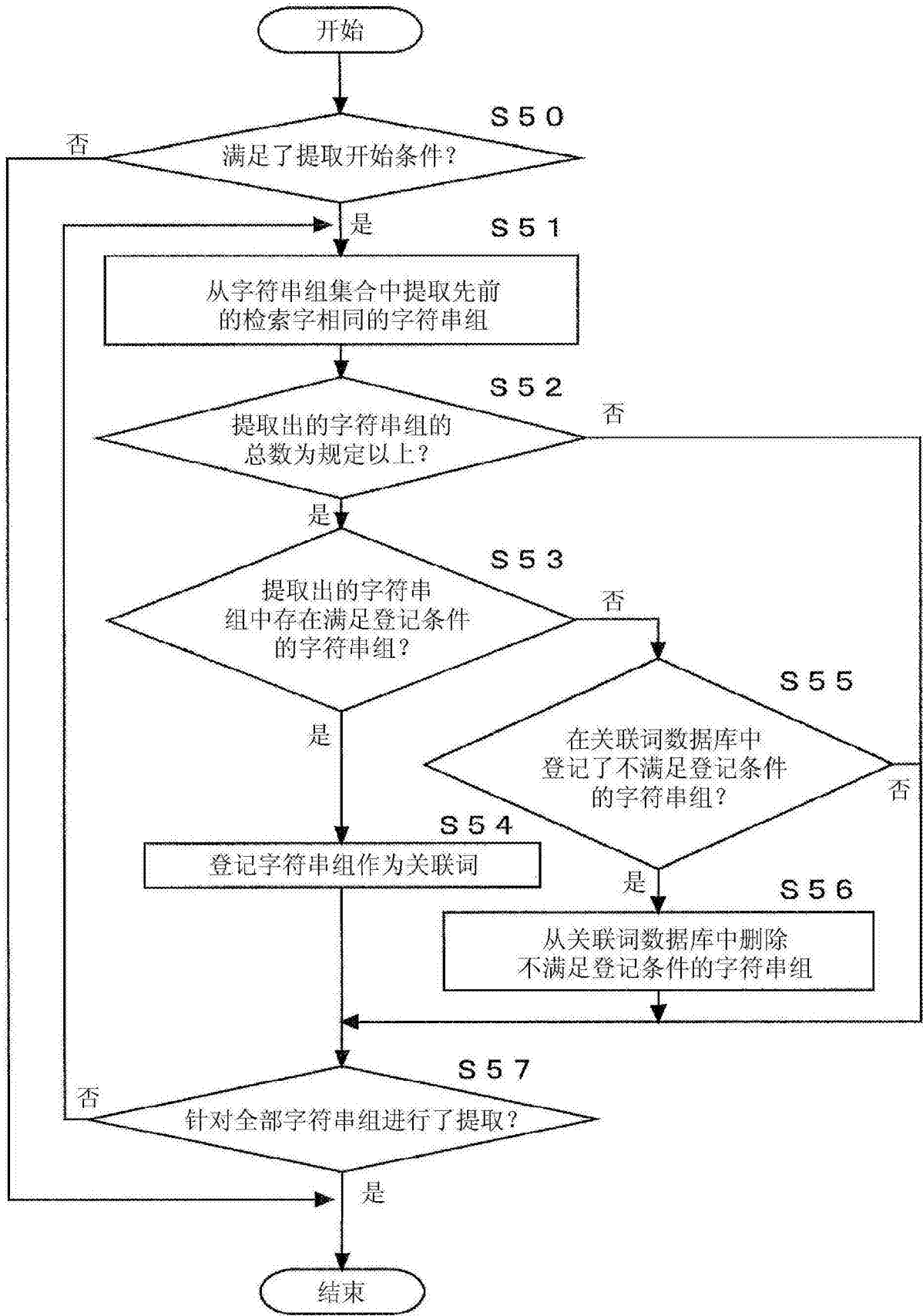


图15

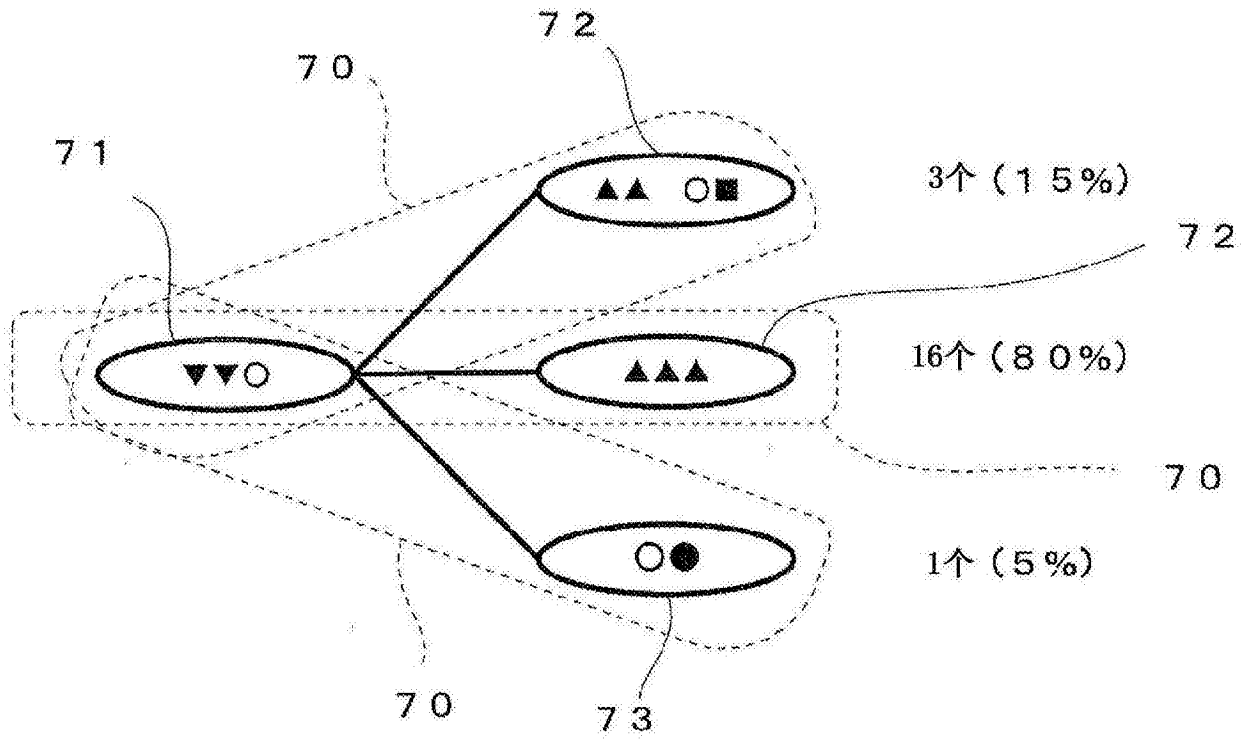


图16