[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
G06F 7/00

[12] 发明专利申请公开说明书

[21] 申请号 02828375.9

[43] 公开日 2005 年 8 月 17 日

[11] 公开号 CN 1656442A

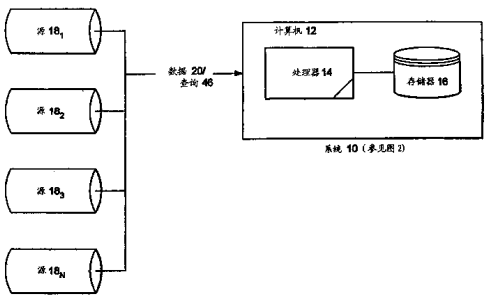
[22] 申请日 2002.12.27 [21] 申请号 02828375.9
[30] 优先权
[32] 2001.12.28 [33] US [31] 60/344,067
[86] 国际申请 PCT/US2002/041630 2002.12.27
[87] 国际公布 WO2003/058427 英 2003.7.17
[85] 进入国家阶段日期 2004.8.27
[71] 申请人 杰佛里·詹姆斯·乔纳斯
地址 美国内华达州
[72] 发明人 杰佛里·詹姆斯·乔纳斯
[74] 专利代理机构 北京三友知识产权代理有限公司
代理人 李 辉

权利要求书 30 页 说明书 10 页 附图 7 页

[54] 发明名称 实时数据入库

[57] 摘要

公开了一种用于对进入数据库的数据和数据库中的数据并进行处理，并检索所处理的数据的方法和系统。该数据包括多个实体的多个标识。该方法和系统包括：(a)对进入数据库的数据以及数据库中的数据进行处理，(b)在数据库中进行存储之前对接收数据进行增强，(c)根据接收数据和没有任何数据丢失的现有数据中的记录之间的关联确定并匹配记录，(d)根据用户定义的警告规则和关联启动警告，(e)当用于匹配记录的标识随后被确定为公共交叉实体而不是一般特有实体时自动停止另外的匹配并分离先前匹配的记录，(f)接收用于检索存储在数据库中的处理后数据的数据查询，(g)利用相同的算法处理这些查询，以及(h)将处理后的数据传送给使用同一算法的另一数据库。



ISSN1008-4274

1．一种处理数据的方法，该方法包括以下步骤：

接收包括具有至少一个标识的至少一条记录的数据，各条记录表示多个实体中的至少一个实体；

利用一算法来处理所接收的数据；

在数据库中存储处理后的数据；

接收用于检索所述数据库中存储的数据的至少一部分的数据查询；以及

利用所述算法处理所述查询。

2．根据权利要求 1 所述的方法，其中所述实体是个人。

3．根据权利要求 1 所述的方法，其中所述实体是个人财产。

4．根据权利要求 3 所述的方法，其中所述个人财产是车辆。

5．根据权利要求 1 所述的方法，其中所述实体是不动产。

6．根据权利要求 1 所述的方法，其中所述实体是机构。

7．根据权利要求 1 所述的方法，其中所述实体是化学化合物。

8．根据权利要求 1 所述的方法，其中所述实体是有机化合物。

9．根据权利要求 1 所述的方法，其中所述实体是蛋白质。

10．根据权利要求 1 所述的方法，其中所述实体是生物结构。

11．根据权利要求 1 所述的方法，其中所述实体是生物统计值。

12．根据权利要求 1 所述的方法，其中所述实体是原子结构。

13．根据权利要求 1 所述的方法，进一步包括在利用算法处理所述接收数据之前将接收数据转换成标准化消息格式的步骤。

14．根据权利要求 1 所述的方法，其中利用算法处理接收数据的步骤包括保留各条记录的属性。

15．根据权利要求 14 所述的方法，其中保留各条记录的属性的步骤包括保留以下识别信息：

提供各条记录的源系统，以及

表示所述源系统中的各条记录的唯一标识。

16．根据权利要求14所述的方法，其中保留各条记录的属性的步骤包括保留查询系统和特定用户的识别信息。

17．根据权利要求1所述的方法，其中利用算法处理接收数据的步骤包括在所述数据库中进行存储和在所述数据库中进行查询中的一个之前对接收数据进行分析。

18．根据权利要求17所述的方法，其中在所述数据库中进行存储和在所述数据库中进行查询中的一个之前对接收数据进行分析的步骤包括将至少一个所述标识与以下之一进行比较：

用户定义的标准，以及

次级数据库和列表之一中的至少一个数据集。

19．根据权利要求18所述的方法，其中所比较的标识是所述多个实体中的至少一个实体的名称，并且所述数据集位于名称根列表中。

20．根据权利要求18所述的方法，其中所比较的标识是所述多个实体中的至少一个实体的地址，并且所述数据集位于地址列表中。

21．根据权利要求18所述的方法，其中将所述多个标识中的至少一个标识与用户定义的标准进行比较的步骤包括根据所述用户定义的标准对至少一个标识进行格式化。

22．根据权利要求18所述的方法，其中在所述数据库中进行存储和在所述数据库中进行查询中的一个之前对接收数据进行分析的步骤包括对所述接收数据进行增强。

23．根据权利要求22所述的方法，其中对接收数据进行增强的步骤包括：

对所述次级数据库和列表之一中的至少一个数据集进行查询以获得所述接收数据的附加标识，以及

使用所述附加标识补充所述接收数据。

24．根据权利要求23所述的方法，其中查询至少一个数据集的步骤包括：

所述次级数据库中的至少一个数据集利用所述算法来查询附加数据库，以找出与所接收的标识中的至少一个相关的附加标识；以及

使用所述次级数据库中的所述附加标识来补充所述接收数据。

25. 根据权利要求 17 所述的方法，其中在所述数据库中进行存储和在所述数据库中进行查询中的一个之前对接收数据进行分析的步骤包括生成所述标识的散列键。

26. 根据权利要求 1 所述的方法，其中利用算法处理接收数据的步骤包括根据用户定义的标准将处理后的查询存储在所述数据库中。

27. 根据权利要求 26 所述的方法，其中所述用户定义的标准包括有效日期。

28. 根据权利要求 1 所述的方法，其中接收包括具有至少一个标识的至少一条记录的数据的步骤、利用算法处理接收数据的步骤、以及将处理后的数据存储在数据库中的步骤是实时执行的，其中各条记录表示多个实体中的至少一个。

29. 根据权利要求 1 所述的方法，其中接收包括具有至少一个标识的至少一条记录的数据的步骤、利用算法处理接收数据的步骤、以及将处理后的数据存储在数据库中的步骤是分批执行的，其中各条记录表示多个实体中的至少一个。

30. 根据权利要求 1 或 17 所述的方法，其中利用算法处理接收数据的步骤包括：

从所述数据库种检索一组附加记录，该组附加记录具有与所述接收数据中的标识相似的标识；

分析所检索到的记录组的各个标识，以与所述接收数据的至少一部分匹配；

将所述接收数据的至少一部分与所检索到的记录组的至少一个已分析记录进行匹配，该已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录；

分析在接收数据的至少一部分中是否包含至少一个先前没有存储在所检索到的记录组的所述至少一个已分析记录中的标识，其中该已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录；以及

重新分析所检索到的记录组的各个标识，以与以下各项进行匹配：

所述接收数据的至少一部分，和

所检索到的记录组的所述已分析记录，其中所述已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录；以及

在所述数据库中存储所述匹配记录。

31．根据权利要求30所述的方法，其中将接收数据的至少一部分与至少一个已分析记录进行匹配的步骤包括分配持续键。

32．根据权利要求30所述的方法，其中利用算法处理接收数据的步骤还包括在重新分析所检索到的记录组的各个标识以进行匹配之前从所述数据库中检索一组附加记录，该组附加记录具有与以下各项中的标识相似的标识：

所述接收数据的至少一部分，以及

所检索到的记录组的所述已分析记录，所述已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录。

33．根据权利要求32所述的方法，其中利用算法处理接收数据的步骤包括重复以下步骤：

从所述数据库中检索一组记录；

分析所检索到的记录组的各个标识；

匹配所述接收数据的至少一部分；

分析在所述接收数据的至少一部分中是否存在至少一个先前没有存储的标识；

从所述数据库中检索一组附加记录；以及

重新分析所检索到的记录组的各个标识以进行匹配，直到找不到另外的匹配。

34．根据权利要求30所述的方法，其中利用算法处理接收数据的步骤包括：

确定特定的标识是否为以下各项之一：

表示至少两个不同实体的公共交叉记录，和

表示特定实体的一般特有记录；

以及

如果确定特定的标识是表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则分离先前基于该特定标识而匹配的记录。

35．根据权利要求34所述的方法，其中利用算法处理接收数据的步骤包括：如果确定特定的标识是表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则禁止基于该特定标识对记录进行任何另外的匹配。

36．根据权利要求34所述的方法，其中利用算法处理接收数据的步骤包括将分离后的记录作为接收数据进行重新处理。

37．根据权利要求34所述的方法，其中确定特定的标识是表示至少两个不同实体的公共交叉记录还是表示特定实体的一般特有记录的步骤、以及将先前匹配的记录分离的步骤是实时执行的。

38．根据权利要求34所述的方法，其中确定特定的标识是表示至少两个不同实体的公共交叉记录还是表示特定实体的一般特有记录的步骤、以及将先前匹配的记录分离的步骤是分批执行的。

39．根据权利要求30所述的方法，其中利用算法处理接收数据的步骤包括：

将接收数据与至少一个所存储的记录进行比较以确定是否存在关联；以及

为每一个被确定为反映了与接收数据的至少一部分的关联的存储记录生成关联记录。

40．根据权利要求39所述的方法，其中利用算法处理接收数据的步骤包括为每一个关联记录生成至少一个置信度指示符。

41．根据权利要求40所述的方法，其中比较接收数据的步骤、生成关联记录的步骤和生成至少一个置信度指示符的步骤是实时执行的。

42．根据权利要求40所述的方法，其中比较接收数据的步骤、生成关联记录的步骤和生成至少一个置信度指示符的步骤是分批执行的。

43．根据权利要求40所述的方法，其中至少一个所述的置信度指示

符表示以下两者之间关联的可能性：

由与所述接收数据的所述部分有关联的特定记录表示的实体，以及由所述接收数据的所述部分表示的实体。

44．根据权利要求 40 所述的方法，其中至少一个所述的置信度指示符表示以下的可能性：

由与所述接收数据的所述部分有关联的特定记录表示的实体，和由所述接收数据的所述部分表示的实体是相同的。

45．根据权利要求 40 所述的方法，其中利用算法处理接收数据的步骤包括：分析所述关联记录以确定所述关联记录是否反映了至少一个先前未确定的关联。

46．根据权利要求 45 所述的方法，其中分析所述关联记录的步骤包括对反映至少一个级别的分离程度的关联记录进行分析。

47．根据权利要求 46 所述的方法，其中对反映至少一个级别的分离程度的关联记录进行分析的步骤包括对满足至少一个用户定义的标准的关联记录进行分析。

48．根据权利要求 47 所述的方法，其中对满足至少一个用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为最大级别的分离程度。

49．根据权利要求 47 所述的方法，其中对满足至少一个用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为具有大于一个最小值的置信度指示符的关联记录。

50．根据权利要求 45 所述的方法，其中利用算法处理接收数据的步骤还包括根据至少一个用户定义的警告规则发出警告。

51．根据权利要求 50 所述的方法，其中根据至少一个用户定义的警告规则发出警告的步骤包括通过电子通信装置传送所述警告。

52．根据权利要求 51 所述的方法，其中所述电子通信装置包括电子邮件系统。

53．根据权利要求 51 所述的方法，其中所述电子通信装置包括电话。

54．根据权利要求 51 所述的方法，其中所述电子通信装置包括传呼

机。

55. 根据权利要求 51 所述的方法，其中所述电子通信装置包括个人数字助理。

56. 根据权利要求 50 所述的方法，其中分析所述关联记录的步骤包括：

在至少一个次级数据库上复制所述关联记录；

根据工作负荷标准将接收数据分配给至少一个次级数据库以进行分析；以及

从至少一个次级数据库发出满足所述用户定义警告规则的标准的警告。

57. 根据权利要求 1 或 28 所述的方法，其中利用算法处理接收数据的步骤还包括利用该算法将所存储的处理后数据传送给至少一个次级数据库。

58. 根据权利要求 57 所述的方法，其中将所存储的处理后数据传送给至少一个次级数据库的步骤是实时执行的。

59. 根据权利要求 57 所述的方法，其中将所存储的处理后数据传送给至少一个次级数据库的步骤是分批执行的。

60. 一种处理数据的方法，包括以下步骤：

接收包括具有至少一个标识的至少一条记录的数据，各条记录表示多个实体中的至少一个实体；

利用一算法进行以下处理：

从数据库中检索一组附加记录，该组附加记录具有与接收数据中的标识相似的标识，

分析所检索到的记录组的各个标识以与接收数据的至少一部分进行匹配，

将接收数据的所述至少一部分与所检索到的记录组的至少一条已分析记录进行匹配，该已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录，

分析在所述接收数据的至少一部分中是否包含至少一个先前没有存

储在所检索到的记录组的所述至少一个已分析记录中的标识，其中该已分析记录被确定为反映了具有表示多个实体中的同一个实体的标识的记录；以及

重新分析所检索到的记录组中的各个标识以匹配以下各项：

所述接收数据的至少一部分，以及

所检索到的记录组的所述已分析记录，其中该已分析记录被确定为反映了具有表示多个实体中的同一个实体的标识的记录；以及

在所述数据库中存储匹配记录。

61．根据权利要求 60 所述的方法，其中利用算法的步骤包括分配持续键。

62．根据权利要求 60 所述的方法，其中利用算法的步骤还包括在重新分析所检索到的记录组中的各个标识以进行匹配之前从所述数据库中检索一组附加记录，该组附加记录具有与以下各项中的标识相似的标识：

所述接收数据的至少一部分，以及

所检索到的记录组的所述已分析记录，所述已分析记录被确定为反映了具有表示多个实体中的同一个实体的标识的记录。

63．根据权利要求 60 或 62 所述的方法，其中利用算法的步骤包括重复以下步骤：

从数据库中检索一组附加记录；

分析所检索到的记录组中的各个标识；

匹配接收数据的至少一部分；

分析接收数据的至少一部分中是否包括至少一个先前没有存储的标识；

从所述数据库中检索一组附加记录；以及

重新分析所检索到的记录组中的各个标识以进行匹配，直到找不到另外的匹配。

64．根据权利要求 63 所述的方法，其中接收数据的步骤、利用算法的步骤和存储匹配记录的步骤是实时执行的。

65．根据权利要求 63 所述的方法，其中接收数据的步骤、利用算法

的步骤和存储匹配记录的步骤是分批执行的。

66．根据权利要求 60 所述的方法，其中利用算法的步骤还包括：

确定特定标识是否为以下各项之一：

表示至少两个不同实体的公共交叉记录，以及

表示特定实体的一般特有记录；以及

如果特定标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则分离先前基于该特定标识而匹配的记录。

67．根据权利要求 66 所述的方法，其中利用算法的步骤包括：如果特定标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则禁止基于该特定标识对记录进行任何另外的匹配。

68．根据权利要求 66 所述的方法，其中利用算法的步骤包括将所分离的记录作为接收数据重新进行处理。

69．根据权利要求 66 所述的方法，其中确定特定标识是表示至少两个不同实体的公共交叉记录还是表示特定实体的一般特有记录的步骤以及分离先前匹配的记录的步骤是实时执行的。

70．根据权利要求 66 所述的方法，其中确定特定标识是表示至少两个不同实体的公共交叉记录还是表示特定实体的一般特有记录的步骤以及分离先前匹配的记录的步骤是分批执行的。

71．根据权利要求 60 所述的方法，其中利用算法的步骤包括：

将接收数据与至少一个所存储的记录进行比较以确定是否存在关联；以及

为被确定为与接收数据的至少一部分存在关联的各条存储记录创建关联记录。

72．根据权利要求 71 所述的方法，其中利用算法的步骤包括为各条关联记录创建至少一个置信度指示符。

73．根据权利要求 72 所述的方法，其中比较接收数据的步骤、创建关联记录的步骤和创建至少一个置信度指示符的步骤是实时执行的。

74．根据权利要求 72 所述的方法，其中比较接收数据的步骤、创建关联记录的步骤和创建至少一个置信度指示符的步骤是分批执行的。

75．根据权利要求 72 所述的方法，其中至少一个所述的置信度指示符表示以下两者之间关联的可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，以及

由接收数据的所述部分表示的实体。

76．根据权利要求 72 所述的方法，其中至少一个所述的置信度指示符表示以下可能性：

由与接收数据的所述部分存在关联的特定记录所表示的实体，和

由接收数据的所述部分表示的实体是相同的。

77．根据权利要求 72 所述的方法，其中利用算法的步骤包括分析所述关联记录以确定所述关联记录是否反映了先前未确定的至少一个关联。

78．根据权利要求 77 所述的方法，其中分析所述关联记录的步骤包括对反映至少一个级别的分离程度的关联记录进行分析。

79．根据权利要求 78 所述的方法，其中对反映至少一个级别的分离程度的关联记录进行分析的步骤包括对满足至少一个用户定义的标准的关联记录进行分析。

80．根据权利要求 79 所述的方法，其中对满足至少一个用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为最大级别的分离程度。

81．根据权利要求 79 所述的方法，其中对满足至少一个用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为具有大于一个最小值的置信度指示符的关联记录。

82．根据权利要求 77 所述的方法，其中利用算法的步骤还包括根据至少一个用户定义的警告规则发出警告。

83．根据权利要求 82 所述的方法，其中根据至少一个用户定义的警告规则发出警告的步骤包括通过电子通信装置传送所述警告。

84．根据权利要求 83 所述的方法，其中所述电子通信装置包括电子

邮件系统。

85. 根据权利要求 83 所述的方法，其中所述电子通信装置包括电话。

86. 根据权利要求 83 所述的方法，其中所述电子通信装置包括传呼机。

87. 根据权利要求 83 所述的方法，其中所述电子通信装置包括个人数字助理。

88. 根据权利要求 82 所述的方法，其中分析所述关联记录的步骤包括：

在至少一个次级数据库上复制所述关联记录；

根据工作负荷标准将接收数据分配至所述至少一个次级数据库以进行分析；以及

从所述至少一个次级数据库发出基于所述用户定义警告规则的警告。

89. 根据权利要求 60 所述的方法，还包括在利用算法的步骤之前将接收数据转换成标准化消息格式的步骤。

90. 根据权利要求 60 所述的方法，其中利用算法的步骤包括保留各个标识的属性。

91. 根据权利要求 90 所述的方法，其中保留各个记录的属性的步骤包括保留以下各项的识别信息：

提供各条记录的源系统，以及

表示所述源系统中的各条记录的唯一标识。

92. 根据权利要求 90 所述的方法，其中保留各个标识的属性的步骤包括保留查询系统和特定用户的识别信息。

93. 根据权利要求 60 所述的方法，其中利用算法的步骤包括在数据库中进行存储和在数据库中进行查询中的一个之前分析接收数据。

94. 根据权利要求 93 所述的方法，其中在所述数据库中进行存储和在所述数据库中进行查询中的一个之前分析接收数据的步骤包括将所述标识中的至少一个与以下各项之一进行比较：

用户定义的标准，和

次级数据库和列表之一中的至少一个数据集。

95．根据权利要求 94 所述的方法，其中所比较的标识是所述多个实体中的至少一个实体的名称，并且所述数据集位于名称根列表中。

96．根据权利要求 94 所述的方法，其中所比较的标识是所述多个实体中的至少一个实体的地址，并且所述数据集位于地址列表中。

97．根据权利要求 94 所述的方法，其中将至少一个标识与用户定义的标准进行比较的步骤包括根据用户定义的标准对至少一个标识进行格式化。

98．根据权利要求 93 所述的方法，其中在所述数据库中进行存储和在所述数据库中进行查询中的一个之前分析接收数据的步骤包括对接收数据进行增强。

99．根据权利要求 98 所述的方法，其中对接收数据进行增强的步骤包括：

对所述次级数据库和列表之一中的至少一个数据集进行查询以获得所述接收数据的附加标识；以及

使用所述附加标识补充所述接收数据。

100．根据权利要求 99 所述的方法，其中查询至少一个数据集的步骤包括：

所述次级数据库中的至少一个数据集利用所述算法查询多个附加数据库以找到与所接收的标识中的至少一个相关的附加标识；以及

使用所述次级数据库中的附加标识来补充所述接收数据。

101．根据权利要求 93 所述的方法，其中利用算法的步骤包括创建所述标识的散列键。

102．根据权利要求 60 所述的方法，其中利用算法的步骤包括根据用户定义的标准在所述数据库中存储处理后的查询。

103．根据权利要求 102 所述的方法，其中所述用户定义的标准包括有效日期。

104．根据权利要求 60 所述的方法，其中利用算法的步骤还包括利用所述算法将所存储的处理后的数据传送给至少一个次级数据库。

105．根据权利要求 104 所述的方法，其中将所存储的处理后的数据传送给至少一个次级数据库的步骤是实时执行的。

106．根据权利要求 104 所述的方法，其中将所存储的处理后的数据传送给至少一个次级数据库的步骤是分批执行的。

107．一种分离先前匹配的记录的方法，该方法包括以下步骤：

确定表示至少一个实体的至少一条记录中的特定标识是否是以下各项之一：

表示至少两个不同实体的公共交叉记录，和

表示特定实体的一般特有记录；以及

如果特定的标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则分离先前基于该特定标识而匹配的记录。

108．根据权利要求 107 所述的方法，还包括如果特定标识被确定为表示多个实体的公共交叉记录而不是表示一个实体的一般特有记录，则禁止基于该特定标识对记录进行任何另外的匹配。

109．根据权利要求 107 所述的方法，还包括重新处理所分离的记录的步骤。

110．根据权利要求 107 所述的方法，其中确定特定标识是表示至少两个不同实体的公共交叉记录还是表示特定实体的一般特有记录的步骤是实时执行的。

111．根据权利要求 107 所述的方法，其中确定特定标识是表示至少两个不同实体的公共交叉记录还是表示特定实体的一般特有记录的步骤是分批执行的。

112．一种处理数据库中的数据的方法，该方法包括以下步骤：

接收数据，该数据包括具有至少一个标识的至少一条记录，各条记录表示多个实体中的至少一个实体；

实时地比较接收数据和数据库中所存储的至少一条记录以确定是否存在关联；

实时地为数据库中所存储的被确定为与接收数据的至少一部分存在

关联的各条记录创建关联记录；以及

在所述数据库中存储各条关联记录。

113．根据权利要求 112 所述的方法，还包括实时地为各条关联记录创建至少一个置信度指示符的步骤。

114．根据权利要求 113 所述的方法，其中至少一个置信度指示符表示以下两者之间的关联的可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，以及

由接收数据的所述部分表示的实体。

115．根据权利要求 113 所述的方法，其中至少一个置信度指示符表示以下的可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，和

由接收数据的所述部分表示的实体是相同的。

116．根据权利要求 112 或 113 所述的方法，还包括分析所述关联记录以确定这些关联记录是否反映至少一个先前未确定的关联的步骤。

117．根据权利要求 116 所述的方法，其中分析所述关联记录的步骤包括对反映至少一个级别的分离程度的关联记录进行分析。

118．根据权利要求 117 所述的方法，其中对反映至少一个级别的分离程度的关联记录进行分析的步骤包括对满足至少一个用户定义标准的关联记录进行分析。

119．根据权利要求 118 所述的方法，其中对满足至少一个用户定义标准的关联记录进行分析的步骤包括将所分析的关联记录限定为最大级别的分离程度。

120．根据权利要求 118 所述的方法，其中对满足至少一个用户定义标准的关联记录进行分析的步骤包括将所分析的关联记录限定为具有大于一个最小至的置信度指示符的关联记录。

121．根据权利要求 116 所述的方法，还包括根据至少一个用户定义的警告规则发出警告的步骤。

122．根据权利要求 121 所述的方法，其中根据至少一个用户定义的警告规则发出警告的步骤包括通过电子通信装置传送所述警告。

123. 根据权利要求 122 所述的方法，其中所述电子通信装置包括电子邮件系统。

124. 根据权利要求 122 所述的方法，其中所述电子通信装置包括电话。

125. 根据权利要求 122 所述的方法，其中所述电子通信装置包括传呼机。

126. 根据权利要求 122 所述的方法，其中所述电子通信装置包括个人数字助理。

127. 根据权利要求 121 所述的方法，还包括以下步骤：

在至少一个次级数据库上复制所述关联记录；

根据工作负荷标准将接收数据分配给所述至少一个次级数据库以进行分析；以及

从所述至少一个次级数据库中发出满足用户定义警告规则的标准的警告。

128: 对于用于处理数据的系统，一种包含程序指令的计算机可读介质，计算机执行该程序指令以执行包括以下步骤的方法：

接收包括具有至少一个标识的至少一条记录的数据，各条记录表示多个实体中的至少一个实体；

利用一个算法处理接收数据；

在数据库中存储处理后的数据；

接收用于检索所述数据库中存储的至少一部分所述数据的数据查询；以及

利用所述算法处理所述查询。

129. 根据权利要求 128 所述的计算机可读介质，其中所述实体是人。

130. 根据权利要求 128 所述的计算机可读介质，其中所述实体是个人财产。

131. 根据权利要求 130 所述的计算机可读介质，其中所述所述个人财产是车辆。

132. 根据权利要求 128 所述的计算机可读介质，其中所述实体是不

动产。

133. 根据权利要求 128 所述的计算机可读介质，其中所述实体是机构。

134. 根据权利要求 128 所述的计算机可读介质，其中所述实体是化学化合物。

135. 根据权利要求 128 所述的计算机可读介质，其中所述实体是有机化合物。

136. 根据权利要求 128 所述的计算机可读介质，其中所述实体是蛋白质。

137. 根据权利要求 128 所述的计算机可读介质，其中所述实体是生物结构。

138. 根据权利要求 128 所述的计算机可读介质，其中所述实体是生物统计值。

139. 根据权利要求 128 所述的计算机可读介质，其中所述实体是原子结构。

140. 根据权利要求 128 所述的计算机可读介质，其中所述方法进一步包括在利用算法处理接收数据之前将接收数据转换成标准化消息格式的步骤。

141. 根据权利要求 128 所述的计算机可读介质，其中所述利用算法处理接收数据的步骤包括保留各条记录的属性。

142. 根据权利要求 141 所述的计算机可读介质，其中保留各条记录的属性的步骤包括保留以下各项的识别信息：

提供各条记录的源系统，以及

表示所述源系统中的记录的唯一标识。

143. 根据权利要求 141 所述的计算机可读介质，其中保留各条记录的属性的步骤包括保留查询系统和特定用户的识别信息。

144. 根据权利要求 128 所述的计算机可读介质，其中利用算法处理接收数据的步骤包括在所述数据库中进行存储和在所述数据库中进行查询中的一个之前分析接收数据。

145. 根据权利要求 144 所述的计算机可读介质，其中在所述数据库中进行存储和在所述数据库中进行查询中的一个之前分析接收数据的步骤包括将所述多个标识中的至少一个标识与以下各项之一进行比较：

用户定义的标准，和

所述数据库与列表之一中的至少一个数据集。

146. 根据权利要求 145 所述的计算机可读介质，其中所比较的标识是所述多个实体中的至少一个实体的名称，并且所述数据集位于名称根列表中。

147. 根据权利要求 145 所述的计算机可读介质，其中所比较的标识是所述多个实体中的至少一个实体的地址，并且所述数据集位于地址列表中。

148. 根据权利要求 145 所述的计算机可读介质，其中将至少一个标识与用户定义的标准进行比较的步骤包括根据用户定义的标准对至少一个标识进行格式化。

149. 根据权利要求 144 所述的计算机可读介质，其中在所述数据库中进行存储和在所述数据库中进行查询中的一个之前分析接收数据的步骤包括对接收数据进行增强。

150. 根据权利要求 149 所述的计算机可读介质，其中对接收数据进行增强的步骤包括：

对所述次级数据库和列表之一中的至少一个数据集进行查询以获得所述接收数据的附加标识，以及

使用所述附加标识补充所述接收数据。

151. 根据权利要求 150 所述的计算机可读介质，其中查询至少一个数据集的步骤包括：

至少一个数据库中的至少一个数据集利用所述算法查询附加数据库，以找到与所接收的标识中的至少一个相关的附加标识；以及

使用至少一个附加数据库中的附加标识补充所述接收数据。

152. 根据权利要求 144 所述的计算机可读介质，其中在所述数据库中进行存储和在所述数据库中进行查询中的一个之前分析接收数据的步

骤包括创建所述标识的散列键。

153. 根据权利要求 128 所述的计算机可读介质，其中利用算法处理接收数据的步骤包括根据用户定义的标准在所述数据库中存储处理后的查询。

154. 根据权利要求 153 所述的计算机可读介质，其中所述用户定义的标准包括有效日期。

155. 根据权利要求 128 所述的计算机可读介质，其中接收包括具有至少一个标识的至少一个记录的数据的步骤、利用算法处理接收数据的步骤以及将处理后的数据存储在数据库中的步骤是实时执行的，其中各条记录表示多个实体中的至少一个实体。

156. 根据权利要求 128 所述的计算机可读介质，其中接收包括具有至少一个标识的至少一个记录的数据的步骤、利用算法处理接收数据的步骤以及将处理后的数据存储在数据库中的步骤是分批执行的，其中各条记录表示多个实体中的至少一个实体。

157. 根据权利要求 128 或 144 所述的计算机可读介质，其中利用算法处理接收数据的步骤包括：

从所述数据库中检索一组具有与接收数据中的标识相似的标识的附加记录；

分析所检索到的记录组中的各个标识以与接收数据的至少一部分进行匹配；

对接收数据的至少一部分和所检索到的记录组中的至少一条已分析记录进行匹配，所述已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录；

分析在接收数据的所述至少一部分中是否包含至少一个先前没有存储在所检索到的记录组的所述至少一条已分析记录中的标识，其中该已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录；以及

重新分析所检索到的记录组中的各个标识以匹配以下各项：

接收数据的至少一部分，和

所检索到的记录组中被确定为反映了具有表示所述多个实体中同一个实体的标识的记录的已分析记录；以及

将匹配的记录存储在所述数据库中。

158．根据权利要求157所述的计算机可读介质，其中将接收数据的至少一部分与至少一个已分析记录进行匹配的步骤包括分配持续键。

159．根据权利要求157所述的计算机可读介质，其中利用算法处理接收数据的步骤还包括在重新分析所检索到的记录组中的各个标识以进行匹配之前从所述数据库中检索一组附加记录，该组附加记录具有与以下各项中的标识相似的标识：

接收数据的至少一部分，以及

所检索到的记录组中被确定为反映了具有表示所述多个实体中同一个实体的标识的记录的已分析记录。

160．根据权利要求159所述的计算机可读介质，其中利用算法处理接收数据包括重复以下步骤：

从所述数据库中检索一组记录；

分析所检索到的记录组中的各个标识；

匹配接收数据的至少一部分；

分析接收数据的所述至少一部分中是否包含至少一个之前没有存储的标识；

从所述数据库中检索一组附加记录；以及

重新分析所检索到的记录组中的各个标识以进行匹配，直到找不到另外的匹配。

161．根据权利要求157所述的计算机可读介质，其中利用算法处理接收数据的步骤还包括：

确定特定标识是否为以下各项之一：

表示至少两个不同实体的公共交叉记录，以及

表示特定实体的一般特有记录；以及

如果特定标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则分离先前基于该特定标识而匹配的

记录。

162. 根据权利要求 161 所述的计算机可读介质，其中利用算法处理接收数据的步骤包括如果特定标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则禁止基于该特定标识对记录进行任何另外的匹配。

163. 根据权利要求 161 所述的计算机可读介质，其中利用算法处理接收数据的步骤包括将所分离的记录作为接收数据重新进行处理。

164. 根据权利要求 161 所述的计算机可读介质，其中确定特定标识是表示至少两个不同实体的公共交叉记录还是表示特定实体的一般特有记录的步骤以及分离先前匹配的记录的步骤是实时执行的。

165. 根据权利要求 161 所述的计算机可读介质，其中确定特定标识是表示至少两个实体的公共交叉记录还是表示特定实体的一般特有记录的步骤以及分离先前匹配的记录的步骤是分批执行的。

166. 根据权利要求 157 所述的计算机可读介质，其中利用算法处理接收数据的步骤包括：

将接收数据与至少一个所存储的记录进行比较以确定是否存在关联；以及

为被确定为与接收数据的至少一部分存在关联的各条所存储的记录创建关联记录。

167. 根据权利要求 166 所述的计算机可读介质，其中利用算法处理接收数据的步骤包括为每一个关联记录创建至少一个置信度指示符。

168. 根据权利要求 167 所述的计算机可读介质，其中比较接收数据的步骤、创建关联记录的步骤和创建至少一个置信度指示符的步骤是实时执行的。

169. 根据权利要求 167 所述的计算机可读介质，其中比较接收数据的步骤、创建关联记录的步骤和创建至少一个置信度指示符的步骤是分批执行的。

170. 根据权利要求 167 所述的计算机可读介质，其中至少一个所述的置信度指示符表示以下两者之间的关联的可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，以及

由接收数据的所述部分表示的实体。

171．根据权利要求 167 所述的计算机可读介质，其中至少一个所述的置信度指示符表示以下可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，和

由接收数据的所述部分表示的实体是相同的。

172．根据权利要求 167 所述的计算机可读介质，其中利用算法处理接收数据的步骤包括分析所述关联记录以确定所述关联记录是否反映至少一个先前未确定的关联。

173．根据权利要求 172 所述的计算机可读介质，其中分析所述关联记录的步骤包括对反映至少一个级别的分离程度的关联记录进行分析。

174．根据权利要求 173 所述的计算机可读介质，其中对反映至少一个级别的分离程度的关联记录进行分析的步骤包括对满足用户定义的标准的关联记录进行分析。

175．根据权利要求 174 所述的计算机可读介质，其中对满足用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为最大级别的分离程度。

176．根据权利要求 174 所述的计算机可读介质，其中对满足用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为具有大于一个最小值的置信度指示符的关联记录。

177．根据权利要求 172 所述的计算机可读介质，其中利用算法处理接收数据的步骤还包括根据至少一个用户定义的警告规则发出警告。

178．根据权利要求 177 所述的计算机可读介质，其中根据至少一个用户定义的警告规则发出警告的步骤包括通过电子通信装置传送该警告。

179．根据权利要求 178 所述的计算机可读介质，其中所述电子通信装置包括电子邮件系统。

180．根据权利要求 178 所述的计算机可读介质，其中所述电子通信装置包括电话。

181. 根据权利要求 178 所述的计算机可读介质，其中所述电子通信装置包括传呼机。

182. 根据权利要求 178 所述的计算机可读介质，其中所述电子通信装置包括个人数字助理。

183. 根据权利要求 177 所述的计算机可读介质，其中分析关联记录的步骤包括：

在至少一个次级数据库上复制所述关联记录；

根据工作负荷标准将接收数据分配给所述至少一个次级数据库以进行分析；以及

从所述至少一个次级数据库中发出满足用户定义警告规则的标准的警告。

184. 根据权利要求 128 或 155 所述的计算机可读介质，其中利用算法处理接收数据的步骤还包括利用该算法将所存储的处理后的数据传送给至少一个次级数据库。

185. 根据权利要求 184 所述的计算机可读介质，其中将所存储的处理后的数据传送给至少一个次级数据库的步骤是实时执行的。

186. 根据权利要求 184 所述的计算机可读介质，其中将所存储的处理后的数据传送给至少一个次级数据库的步骤是分批执行的。

187. 对于用于对进入数据库的数据和数据库中的数据进行处理的系统，一种包含程序指令的计算机可读介质，计算机执行该程序指令以执行包括以下步骤的方法：

接收包括具有至少一个标识的至少一个记录的数据，各条记录表示多个实体中的至少一个实体；

利用一算法执行以下步骤：

从数据库中检索一组附加记录，该组附加记录具有与接收数据中的标识相似的标识，

分析所检索到的记录组中的各个标识以与接收数据的至少一部分进行匹配，

对接收数据的至少一部分和所检索到的记录组中的至少一个已分析

记录进行匹配，该已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录；

分析接收数据的所述至少一部分中是否包含至少一个先前没有存储在所检索到的记录组中至少一个已分析记录中的标识，其中该已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录；以及

重新分析所检索到的记录组中的各个标识以匹配以下各项：

接收数据的至少一部分，以及

所检索到的记录组中的所述已分析记录，所述已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录；以及

将匹配的记录存储在所述数据库中。

188. 根据权利要求 187 所述的计算机可读介质，其中利用算法对接收数据和至少一个已分析记录进行匹配的步骤包括分配持续键。

189. 根据权利要求 187 所述的计算机可读介质，其中利用算法的步骤还包括在重新分析所检索到的记录组中的各个标识以进行匹配之前从数据库中检索一组附加记录，该组附加记录具有与以下各项中的标识相似的标识：

接收数据的至少一部分，以及

所检索到的记录组中的所述已分析记录，所述已分析记录被确定为反映了具有表示所述多个实体中的同一个实体的标识的记录。

190. 根据权利要求 187 或 189 所述的计算机可读介质，其中利用算法的步骤包括重复以下步骤：

从数据库中检索一组附加记录；

分析所检索到的记录组中的各个标识；

匹配接收数据的至少一部分；

分析接收数据的至少一部分中是否包含至少一个先前没有存储的标识；

从所述数据库中检索一组附加记录；以及

重新分析所检索到的记录组中的各个标识以进行匹配，直到找不到

另外的匹配。

191. 根据权利要求 190 所述的计算机可读介质，其中接收数据的步骤、利用算法的步骤和存储匹配记录的步骤是实时执行的。

192. 根据权利要求 190 所述的计算机可读介质，其中接收数据的步骤、利用算法的步骤和存储匹配记录的步骤是分批执行的。

193. 根据权利要求 187 所述的计算机可读介质，其中利用算法的步骤包括：

确定特定的标识是否为以下各项之一：

表示至少两个不同实体的公共交叉记录，以及

表示特定实体的一般特有记录；以及

如果特定标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则分离先前基于该特定标识而匹配的记录。

194. 根据权利要求 193 所述的计算机可读介质，其中利用算法的步骤包括如果特定标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则禁止基于该特定标识对记录进行任何另外的匹配。

195. 根据权利要求 193 所述的计算机可读介质，其中利用算法的步骤包括将所分离的记录作为接收数据重新进行处理。

196. 根据权利要求 193 所述的计算机可读介质，其中确定特定标识是表示至少两个实体的公共交叉记录还是表示特定实体的一般特有记录的步骤以及分离先前匹配的记录的步骤是实时执行的。

197. 根据权利要求 193 所述的计算机可读介质，其中确定特定标识是表示至少两个实体的公共交叉记录还是表示特定实体的一般特有记录的步骤以及分离先前匹配的记录的步骤是分批执行的。

198. 根据权利要求 187 所述的计算机可读介质，其中利用算法的步骤包括：

将接收数据与至少一个所存储的记录进行比较以确定是否存在关联；以及

为被确定为与接收数据的至少一部分存在关联的每一个所存储的记录创建关联记录。

199. 根据权利要求 198 所述的计算机可读介质，其中利用算法的步骤包括为每一个关联记录创建至少一个置信度指示符。

200. 根据权利要求 199 所述的计算机可读介质，其中比较接收数据的步骤、创建关联记录的步骤和创建至少一个置信度指示符的步骤是实时执行的。

201. 根据权利要求 199 所述的计算机可读介质，其中比较接收数据的步骤、创建关联记录的步骤和创建至少一个置信度指示符的步骤是分批执行的。

202. 根据权利要求 199 所述的计算机可读介质，其中至少一个置信度指示符表示以下两者之间的关联的可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，以及

由接收数据的所述部分表示的实体。

203. 根据权利要求 199 所述的计算机可读介质，其中至少一个置信度指示符表示以下的可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，和

由接收数据的所述部分表示的实体是相同的。

204. 根据权利要求 199 所述的计算机可读介质，其中利用算法的步骤包括分析所述关联记录以确定所述关联记录是否反映先前未确定的至少一个关联。

205. 根据权利要求 204 所述的计算机可读介质，其中分析所述关联记录的步骤包括对反映至少一个级别的分离程度的关联记录进行分析。

206. 根据权利要求 205 所述的计算机可读介质，其中对反映至少一个级别的分离程度的关联记录进行分析的步骤包括对满足至少一个用户定义的标准的关联记录进行分析。

207. 根据权利要求 206 所述的计算机可读介质，其中对满足用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为最大级别的分离程度。

208. 根据权利要求 206 所述的计算机可读介质，其中对满足用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为具有大于一个最小值的置信度指示符的关联记录。

209. 根据权利要求 204 所述的计算机可读介质，其中利用算法的步骤还包括根据至少一个用户定义的警告规则发出警告。

210. 根据权利要求 209 所述的计算机可读介质，其中根据至少一个用户定义的警告规则发出警告的步骤包括通过电子通信装置传送警告。

211. 根据权利要求 210 所述的计算机可读介质，其中所述电子通信装置包括电子邮件系统。

212. 根据权利要求 210 所述的计算机可读介质，其中所述电子通信装置包括电话。

213. 根据权利要求 210 所述的计算机可读介质，其中所述电子通信装置包括传呼机。

214. 根据权利要求 210 所述的计算机可读介质，其中所述电子通信装置包括个人数字助理。

215. 根据权利要求 209 所述的计算机可读介质，其中分析所述关联记录的步骤包括：

在至少一个次级数据库上复制所述关联记录；

根据工作负荷标准将接收数据分配给至少一个次级数据库以进行分析；以及

从至少一个次级数据库发出基于所述用户定义警告规则的警告。

216. 根据权利要求 187 所述的计算机可读介质，还包括在利用算法之前将接收数据转换成标准化消息格式的步骤。

217. 根据权利要求 187 所述的计算机可读介质，其中利用算法的步骤包括保留各个标识的属性。

218. 根据权利要求 217 所述的计算机可读介质，其中保留各条记录的属性的步骤包括保留以下各项的识别信息：

提供各条记录的源系统，以及

表示所述源系统中的各条记录的唯一标识。

219. 根据权利要求 217 所述的计算机可读介质，其中保留各条记录的属性的步骤包括保留查询系统和特定用户的识别信息。

220. 根据权利要求 187 所述的计算机可读介质，其中利用算法的步骤包括在数据库中进行存储和在数据库中进行查询中的一个之前分析接收数据。

221. 根据权利要求 220 所述的计算机可读介质，其中在数据库中进行存储和在数据库中进行查询中的一个之前分析接收数据的步骤包括将所述多个标识中的至少一个与以下各项之一进行比较：

用户定义的标准，以及

数据库和列表之一中的至少一个数据集。

222. 根据权利要求 221 所述的计算机可读介质，其中所比较的标识是所述多个实体中的至少一个实体的名称，并且所述数据集位于名称根列表中。

223. 根据权利要求 221 所述的计算机可读介质，其中所比较的标识是所述多个实体中的至少一个实体的地址，并且所述数据集位于地址列表中。

224. 根据权利要求 221 所述的计算机可读介质，其中将所述标识中的至少一个与用户定义的标准进行比较的步骤包括根据该用户定义的标准对至少一个标识进行格式化。

225. 根据权利要求 220 所述的计算机可读介质，其中在数据库中进行存储和在数据库中进行查询中的一个之前分析接收数据的步骤包括对接收数据进行增强。

226. 根据权利要求 225 所述的计算机可读介质，其中对接收数据进行增强的步骤包括：

对所述次级数据库和列表之一中的至少一个数据集进行查询以获得所述接收数据的附加标识；以及

使用所述附加标识补充接收数据。

227. 根据权利要求 226 所述的计算机可读介质，其中查询至少一个数据集的步骤包括：

至少一个数据库中的至少一个数据集利用所述算法查询附加数据库以找到与所接收的标识中的至少一个相关的附加标识；以及

使用至少一个附加数据库中的所述附加标识来补充接收数据。

228. 根据权利要求 220 所述的计算机可读介质，其中利用算法的步骤包括创建所述标识的散列键。

229. 根据权利要求 187 所述的计算机可读介质，其中利用算法的步骤包括根据用户定义的标准在所述数据库中存储处理后的查询。

230. 根据权利要求 229 所述的计算机可读介质，其中所述用户定义的标准包括有效日期。

231. 根据权利要求 187 所述的计算机可读介质，其中利用算法的步骤还包括利用所述算法将所存储的处理后的数据传送给至少一个次级数据库。

232. 根据权利要求 231 所述的计算机可读介质，其中将所存储的处理后的数据传送给至少一个次级数据库的步骤是实时执行的。

233. 根据权利要求 231 所述的计算机可读介质，其中将所存储的处理后的数据传送给至少一个次级数据库的步骤是分批执行的。

234. 对于用于分离先前匹配的记录的系统，一种包括程序指令的计算机可读介质，计算机执行该程序指令以执行包括以下步骤的方法：

确定表示至少一个实体的至少一条记录中的特定标识是否为以下各项之一：

表示至少两个不同实体的公共交叉记录，以及

表示特定实体的一般特有记录；以及

如果特定标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则分离先前基于该特定标识而匹配的记录。

235. 根据权利要求 234 所述的计算机可读介质，还包括如果特定标识被确定为表示至少两个不同实体的公共交叉记录而不是表示特定实体的一般特有记录，则禁止基于该特定标识对记录进行任何另外的匹配。

236. 根据权利要求 234 所述的计算机可读介质，还包括重新处理所

分离的记录的步骤。

237. 根据权利要求 234 所述的计算机可读介质，其中确定特定标识是表示至少两个不同实体的公共交叉记录还是表示特定实体的一般特有记录的步骤以及分离先前匹配的记录的步骤是实时执行的。

238. 根据权利要求 234 所述的计算机可读介质，其中确定特定标识是表示至少两个实体的公共交叉记录还是表示特定实体的一般特有记录的步骤以及分离先前匹配的记录的步骤是分批执行的。

239. 对于用于处理数据库中的数据的系统，一种包括程序指令的计算机可读介质，计算机执行所述程序指令以执行包括以下步骤的方法：

接收包括具有至少一个标识的至少一条记录的数据，各条记录表示多个实体中的至少一个实体；

将接收数据与存储在数据库中的至少一条记录进行比较，以实时地确定是否存在关联；

为存储在数据库中的被确定为与接收数据的至少一部分存在关联的各条记录实时地创建关联记录；以及

将各条关联记录存储在数据库中。

240. 根据权利要求 239 所述的计算机可读介质，还包括为各条关联记录实时地创建至少一个置信度指示符的步骤。

241. 根据权利要求 240 所述的计算机可读介质，其中至少一个置信度指示符表示以下两者之间的关联的可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，以及

由接收数据的所述部分表示的实体。

242. 根据权利要求 240 所述的计算机可读介质，其中至少一个置信度指示符表示以下的可能性：

由与接收数据的所述部分存在关联的特定记录表示的实体，和

由接收数据的所述部分表示的实体是相同的。

243. 根据权利要求 239 或 240 所述的计算机可读介质，还包括分析所述关联记录以确定所述关联记录是否反映先前未确定的至少一个关联的步骤。

244. 根据权利要求 243 所述的计算机可读介质，其中分析所述关联记录的步骤包括对反映至少一个级别的分离程度的关联记录进行分析。

245. 根据权利要求 244 所述的计算机可读介质，其中对反映至少一个级别的分离程度的关联记录进行分析的步骤包括对满足至少一个用户定义的标准的关联记录进行分析。

246. 根据权利要求 245 所述的计算机可读介质，其中对满足至少一个用户定义的标准的关联记录进行分析的步骤包括将所分析的关联记录限定为最大级别的分离程度。

247. 根据权利要求 245 所述的计算机可读介质，其中对满足至少一个用户定义标准的关联记录进行分析的步骤包括将所分析的关联记录限定为具有大于一个最小值的置信度指示符的关联记录。

248. 根据权利要求 243 所述的计算机可读介质，还包括根据至少一个用户定义的警告规则发出警告的步骤。

249. 根据权利要求 248 所述的计算机可读介质，其中根据至少一个用户定义的警告规则发出警告的步骤包括通过电子通信装置传送警告。

250. 根据权利要求 249 所述的计算机可读介质，其中所述电子通信装置包括电子邮件系统。

251. 根据权利要求 249 所述的计算机可读介质，其中所述电子通信装置包括电话。

252. 根据权利要求 249 所述的计算机可读介质，其中所述电子通信装置包括传呼机。

253. 根据权利要求 249 所述的计算机可读介质，其中所述电子通信装置包括个人数字助理。

254. 根据权利要求 248 所述的计算机可读介质，还包括以下步骤：

在至少一个次级数据库上复制所述关联记录；

根据工作负荷标准将接收数据分配给至少一个次级数据库以进行分析；以及

从所述至少一个次级数据库发出满足用户定义警告规则的标准的警告。

实时数据入库

本申请要求 2001 年 12 月 28 日在美国专利局提交的临时申请 No.60/344067 的优先权。

技术领域

本发明总体上涉及用于处理和检索数据仓库中的数据的方法、程序和系统，更具体地，涉及用于处理进入数据仓库的数据和数据仓库内的数据、查询数据仓库中的数据以及分析数据仓库中的数据的方法、程序和系统。

背景技术

数据仓库是基于计算机的数据库，设计为存储记录并对通常来自多个源的查询进行响应。这些记录与诸如个人、机构和财产等的实体相对应。每条记录包括该实体的标识，例如个人的姓名、地址或帐户信息。

遗憾的是，由于创建、维持和/或提高一定的数据质量、完整性和性能等方面的某些局限，当前的数据仓库系统的有效性不高。这些局限还增大了实施、校正和维护这些系统所需的风险、成本和时间。

这些问题和局限性包括但不限于以下各项：（a）与源自各种数据源的不同格式或冲突格式相关的难题，（b）由于在接收时丢失的信息而导致的不完整数据，（c）由于（通常较少）误差或拼写错误而导致的反映同一实体的多条输入记录，（d）识别多条记录是否反映同一实体和/或在多条记录之间是否存在某种关联的能力不足，（e）当合并被确定为反映同一实体的两条记录或者丢弃一条记录时发生的数据丢失，（f）当合并后的记录后来被确定为反映两个独立实体时随后分离记录的能力不足，（g）根据用户定义的警告规则实时地发出警告的能力不足，（h）从利用与用于处理所接收数据的算法或转换处理不同的算法或转换处理进行的

查询中获得的不适当的结果，以及（i）不能根据预定的标准（例如，在特定时间段内）来保持持续的查询。

例如，当接收个人的标识并将其存储在数据库中时：（a）可以以逗号分隔的格式获得来自一个源的记录，而以另一种数据格式接收另一个源的记录；（b）可能会丢失来自各种记录的数据，例如电话号码、地址或一些其它的识别信息；或者（c）由于一条记录对应于当前姓名而另一条记录对应于婚前姓名，所以可能会无察觉地接收反映同一个人的两条记录。在后一情况下，系统可能确定应该合并这两条记录或者丢弃一条记录（该记录可能来自于可靠性较低的源）。但是，在合并处理中，当前系统通常会丢弃数据，这就使系统不具备随后在确定这两条记录反映两个独立实体时将两条记录分开的能力。

另外，当接收标识并将其存储在数据库中时，计算机可以在将数据加载到数据库中之前执行转换和增强（enhancing）处理。但是，当前系统的查询工具使用很少的（如果有的话）用于接收并处理所接收数据的转换和增强处理，造成这些查询的结果不一致，因此不适当、不充分并潜在地存在错误。

类似地，当前的数据仓库系统不具有充分识别实体之间的关联或实时地确定这些实体是否反映同一实体的必要工具。例如，一个人可以具有与第二个人相同的地址，而第二个人可以具有与第三个人相同的电话号码。在这种情况下，确定第一个人与第三个人存在某种关联的可能性是很有利的，尤其是实时确定。

此外，当前数据仓库系统在识别实体间的不适当或冲突关系以及根据用户定义的警告规则实时地提供警告方面的能力有限。这种有限的能力出于几个因素，包括但不限于不能有效地识别如上所述的关联性。

另外，当前的数据仓库系统不能首先转换和增强记录，并且随后在预定的时间中保持持续的查询。在各种情况（包括但不限于罪犯调查中识别人名的情况）下，持续查询是很有利的。识别任何与该人相符的匹配的查询可能最初找不到结果并且实际上在当前系统中会抛弃已查询到的数据。但是，以与所接收数据（其中已查询到的数据可以用于匹配其

它接收数据或查询，并可提供更好的结果）相同的方式加载查询是很有利的。

因此，当前数据仓库系统的任何或所有问题和局限性（不管在此是否区分）都降低了数据仓库的准确性、可靠性和及时性，并极大地降低了性能。实际上，这些问题可能会导致不适当的结果和基于这些结果的不正确判断。

提供本发明以解决这些和其它问题。

发明内容

本发明的目的是提供一种处理进入数据库的数据和数据库中的数据的方法、程序和系统。该方法优选地包括以下步骤：（a）接收多个实体的数据，（b）利用一算法处理所接收的数据，（c）在数据库中存储经处理的数据，（d）接收用于检索存储在数据库中的数据的数据查询，以及（e）利用相同的算法处理这些查询。

所述数据包括具有表示一个或更多个实体的一个或更多个标识的一个或更多条记录。这些实体可以是个人、财产、机构、蛋白质或者能够由标识数据表示的其它事物。

所述算法包括接收已转换成标准消息格式的数据，并保留这些标识的属性，例如源系统、源系统的唯一标识值、查询系统和/或用户。

算法处理包括在存储到数据库中或在数据库中进行查询之前分析数据，其中该分析步骤可以包括：（a）将一个或更多个标识与用户定义的标准或数据库、列表或其它电子格式中的一个或更多个数据集进行比较，（b）根据用户定义的标准对标识进行格式化，（c）在进行存储或查询之前，通过在其它数据库（这些数据库可以具有与第一数据库相同的算法并以级联的方式继续查找）或列表中的一个或多个数据集中查询附加标识，以使用任何附加标识来补充所接收的数据，由此来增强数据，（d）生成标识的散列键（hash key），以及（d）根据用户定义的标准（例如规定的时间段）存储处理后的查询。

进一步的，该方法、程序和系统包括：（a）利用一算法来处理数据

并匹配记录，其中该算法处理包括：（i）从数据库中检索一组记录，该组记录包括与所接收数据中的标识相似的标识，（ii）对所检索到的记录组进行分析以匹配所接收的数据，（iii）将所接收的数据与被确定为反映同一实体的检索记录进行匹配，（iv）分析是否有任何新的标识被添加到任何匹配的记录中，（v）重新查找所检索到的记录组的其它记录以与任何匹配记录进行匹配，以及（b）在数据库中存储匹配记录。另外，该算法可以包括：（a）从数据库中检索一组附加记录，该组附加记录包括与匹配记录中的标识相似的标识，（b）重复这些步骤：检索记录、匹配分析、匹配相同的实体记录、分析新的标识、重新查找所检索到的记录直到找不到其它匹配为止，以及（c）将持续键（persistent key）分配给这些记录。可以分批地或实时地执行这些处理。

进一步地，该方法、程序和系统包括：确定特定的标识是多个实体之间公共的还是某一实体特有的，并且如果用于匹配这些记录的特定标识后来被确定为是多个实体之间公共的而不是某一实体特有的，则分离先前匹配的记录。这种确定和分离步骤可以实时地或分批地执行。该确定和分离步骤可以包括：根据被确定为在多个实体之间是公共的而不是某一实体特有的标识而停止任何另外的匹配，以及重新处理任何所分离的记录。

进一步地，将所接收的数据与至少一个其它先前存储的记录进行比较，以确定多个实体之间的关联性，并且为存在关联的每两个实体创建一条关联记录。该关联记录可以包括一个或多个置信度指示符，用于表示两个实体之间关联的可能性或者两个实体相同的可能性。关联记录也可以对于接收数据中包含的各个实体具有参考作用，或者是指定的。对这些关联记录进行分析以根据用户定义的标准来确定是否存在任何先前未知的关联记录。这些关联记录反映了第一分离程度，其可分析为仅包括那些符合预定标准的记录，例如，分离程度测试的最大次数或者最小关联程度和/或相似性置信度指示符。可以根据用户定义的警告规则发出识别关联记录组的警告。可以通过各种电子通信装置（例如电子邮件消息、电话、个人数字助理或传呼机消息）来传送该警告。

进一步地，该方法包括：（a）在一个或更多个数据库中复制这些关联记录，（b）根据工作负荷标准将所接收的数据分配给一个或更多个附加数据库以进行分析，以及（c）从这些附加数据库中发出任何警告。

进一步地，该方法和系统包括：将所存储的数据传送到另一个数据库，该数据库使用与第一数据库相同的算法。可以实时地或分批地执行处理和传送步骤。

下面将参照附图和说明书对本发明的这些和其它方面与特征进行说明。

附图说明

图 1 是根据本发明的系统的方框图；

图 2 是在图 1 所示系统模块中处理数据的流程图；

图 3 是图 2 所示的处理算法模块的流程图；以及

图 4 是图 3 所示的用于分析所存储的已分析记录的模块的流程图。

具体实施方式

尽管本发明易于以许多不同的形式实施，但在附图中所示并将在此详细说明的是本发明的具体实施例，应该理解的是，本公开被认为是本发明原理的范例，而并不旨在将本发明限定为所示的具体实施例。

在图 1-4 中示出了用于处理进入数据库的数据和数据库中的数据以及用于检索处理后的数据的数据处理系统 10。系统 10 包括至少一个具有处理器 14 和存储器 16 的传统计算机 12。存储器 16 用于存储可执行软件以操作系统 10，以及用于存储数据库和随机存取存储器中的数据。但是，可以将软件存储或设置在任何其它计算机可读介质（例如 CD、DVD 或软盘）上。计算机 12 可以从多个源 18_1-18_n 接收输入。

数据包括具有一个或更多个标识的一个或更多条记录，该一个或更多个标识表示一个或更多个实体。这些实体可以是个人、机构、财产、蛋白质、化学物质或有机化合物、生物统计或原子结构或者可以由标识数据表示的其它事物。个人型实体的标识可以包括个人姓名、（一个或多

个）地址、（一个或多个）电话号码、（一个或多个）信用卡号、社会保险号、职业信息、飞行常客（frequent flyer）或其它忠诚计划（loyalty program）或者帐户信息。一般特有标识是特定实体所特有的那些标识，例如个人实体的社会保险号。

系统 10 从多个源 18_1-18_n 接收数据，并利用算法 22 处理所接收的数据 20。该算法存储在存储器 16 中，并由处理器 14 进行处理或执行。

接收数据 20，包括但不限于接收数据的属性（例如，源系统标识），可以类似地以多种数据格式进行接收。在由算法 22 进行处理之前，将接收数据 20 转换成标准化消息格式 24，例如通用消息格式（Universal Message Format）。

此后，如图 3 所示，算法 22 接收标准化数据 26 并在通过以下步骤在数据库中进行存储或查询之前分析 28 所接收的数据 26：（a）将所接收的数据 26 与用户定义的标准或规则进行比较以执行多个功能，这些功能包括但不限于：（i）名称标准化 30（例如，对照根名（root name）列表），（ii）地址规范化（hygiene）32（例如，对照邮政编码），（iii）字段测试或转换 34(例如，比较性别字段以确定 M/F 或者将 Male 转换成 M 等），（iv）用户定义的格式化 36（例如，以 999-99-9999 格式对所有的社会保险号进行格式化），(b)通过使系统 10 访问一个或更多个数据库 40（数据库 40 可以包括与第一数据库相同的算法，由此使该系统以级联的方式访问附加数据库）来增强数据 38，以查找可以对所接收的数据 26 进行补充 42 的附加信息（该信息可以作为接收数据 20 提交），以及（c）创建分析后的数据 44 的散列键。可将任何新的、修改后的或增强后的数据存储在新创建的字段中以保持原始数据的完整性。例如，如果以标准化格式 26 接收到姓名“Bobby Smith”，则可以对照根名列表 30 将姓名“Bobby”标准化成姓名“Robert”，并存储在新创建的标准姓名字段中。另外，如果接收到 Bobby Smith 的姓名和地址 26，则系统 10 可访问传统的基于因特网的人员查找数据库 40，以获取 Bobby Smith 的电话号码，随后可以根据用户定义的标准 36 以标准方式对该电话号码进行格式化。此外，可以将地址字段对照地址列表 32，以将文字“Street”添加到标准化地址

的结尾处。然后根据增强后的数据创建 44 散列键并将其存储在新创建的字段中。

系统 10 还从多个源 18_1-18_n 接收查询 46，并利用相同的算法 22 分析和处理所接收的查询 46。例如，如果接收到对“Bobby Smith”的查询 46，则将所接收的姓名“Bobby”标准化为姓名“Robert”的相同算法 22 也将查询姓名“Bobby” 标准化为查询姓名“Robert”。实际上，系统 10 加载并存储与接收数据 20 相同的接收查询 46，保持查询系统和用户的全部属性。因此，当系统 10 处理接收查询 46 时，算法 22 可以搜索其它数据库 40（例如，公共记录数据库），以查找丢失的信息。查询结果 94 可以比精确匹配更宽泛，并且可以包括多个关联匹配。例如，如果查询的是“Bobby Smith”，则查询结果 94 可以包括那些曾使用过 Bobby Smith 的信用卡或者曾居住在 Bobby Smith 的地址的人的记录。

算法 22 还在接收到任何接收数据 26 时执行以下的功能：（a）确定在数据库中是否存在与对应于该接收数据的实体相匹配的现有记录，以及（b）如果存在，则将所接收的数据与该现有记录相匹配。例如，算法从数据库中检索一组记录 48（该组记录包括与所接收的数据中的标识相似的标识）来找到可能的候选项，并根据一般特有标识 52 对所检索到的记录组进行分析，以找出标识了与接收数据相对应的已存储记录的匹配 50。如果识别到匹配 54，则该算法分析该匹配记录是否包含任何新的或先前未知的标识 56。如果存在新的或先前未知的标识 56，则算法 22 将分析该新的或先前未知的标识 58，并根据该匹配记录中的新的或先前未知的标识来添加或更新候选列表/关联记录 70，并确定是否存在任何另外的匹配 50。重复该处理直到找不到其它匹配为止。然后该匹配处理为所有的匹配记录 60 分配相同的持续键。另外，如果对于任何记录都没有找到匹配，则为不匹配记录分配其自己的持续键 62。这些记录保留了数据的全部属性并且匹配处理不会由于合并、清除或删除功能而丢失任何数据。

例如，如果记录#1 具有个人姓名、电话号码和地址，而记录#2 具有相同的姓名和信用卡号。不知道他们是否是同一个人，所以必须将这两

条记录保持分离。然后接收到记录#3 的数据，包括个人姓名（与记录#1 相同）、地址（与记录#1 相同）、电话号码（与记录#1 相同）和信用卡号。因为#1 和#3 的姓名、电话号码和地址匹配，所以系统 10 可以确定#1 和#3 描述的是同一个人，所以该算法将#1 的数据与#3 的数据进行匹配。随后系统 10 重新运行该算法，将匹配记录#1 与候选列表的其它记录或包括与该匹配记录相似的标识的附加记录进行比较。因为匹配记录#1 的姓名和信用卡号码与记录#2 的姓名和信用卡号码匹配，所以这两条记录也匹配。然后再次将匹配记录与候选列表或检索到的附加记录进行比较以查找匹配 54，直到不能获得匹配为止。

有时，系统 10 可能确定两条记录是错误地匹配。例如，社会保险号被认为是个人的一般特有标识，所以往往根据相同的社会保险号来匹配记录。但是，在某些情况下，该编号可能随后被确定为是多个实体之间公共的而不是某一实体特有的。例如，考虑一数据输入操作，其具有作为必需字段的社会保险号的记录字段，但不知道这些个人的社会保险号的数据输入操作员仅仅为每个人输入号码“123-45-6789”。

在这种情况下，社会保险号在这些个人型实体之间将是公共的，并且不再是这些个人的一般特有标识。因此：（a）将当前已知的公共标识添加到公共标识的列表中，并且所有进一步的处理将不尝试检索候选列表的记录或者不根据该已知的公共标识来生成关联记录 70，由此停止所有进一步的匹配 64，并且（b）必须分离所有基于该错误的社会保险号匹配的记录以反映匹配前的数据，由此要求先前的数据没有丢失。为了实现后一目的，系统 10 根据数据的全部属性，将根据错误假设 66 产生的任何匹配分离为该错误假设之前的点，而不丢失任何数据。因此，如果“Bobby Smith”（其已经标准化成“Robert Smith”）的记录#1 已经与“Robert Smith”的记录#2 匹配，并且随后确定这是两个不同的个人，并且需要将它们分离为原始的记录#1 和 2，则该算法将认识到记录#1 的标准化“Robert Smith”是“Bobby”。此外，可以实时地或分批地执行确定和分离步骤。而且，可以重新提交分离后的记录作为要在系统中进行处理的新的接收数据。

存在需要对关联（甚至是不太明显的关联）进行评估 68 的情况。例如，个人#1 和#2 可以分别与机构#3 存在关联。因此可能（很有可能）在个人#1 和#2 之间存在关联。该关联可扩展到多个分离程度。因此，系统 10 将所有的接收数据与存储数据中的所有记录进行比较，并为各个实体之间存在某种关联的每一对记录创建关联记录 70。该关联记录 70 可以包括关联类型（例如，父亲、共事者）、置信度指示符（该置信度指示符是表示两个实体的关联强度的分值）72 和所分配的持续键 60 或 62。例如，置信度指示符 72 可以包括关联分值和相似性分值。关联分值是例如在 1 到 10 之间的指示符，表示个人#1 和个人#2 之间存在关联的可能性。相似性分值也是例如在 1 到 10 之间的指示符，表示个人#1 与个人#2 是同一个人。在以上所述的匹配处理的过程中可以识别置信度指示符 72。

系统 10 还对接收数据 20 和查询 46 进行分析，以根据具有大于预定值的置信度指示符的关联记录和/或小于预定数值的分离程度的关联记录来确定是否存在满足用户定义警告规则 74 的标准的情况（例如，两个实体间的不适当关联或者特定形式的行为）。例如，系统 10 可以包括欺骗性信用卡的列表，该列表可以用于确定任何接收数据或查询是否包含欺骗性信用卡号列表中的信用卡号。另外，用户定义警告规则 74 可以对接收数据和查询进行报告。例如，在输入新经销商的数据时，如果确定该新经销商与当前雇员具有相同的地址，则可能存在一个警告规则，表示雇主可能想要调查的经销商和雇员之间存在关联。一旦确定要触发用户定义警告规则的情况，系统 10 就发出警告 74，该警告可通过各种媒介（例如，通过电子邮件的消息）传送，或传送到便携通信装置，例如字符数字传呼机、个人数字助理或者电话。

例如，根据用户定义的警告规则，对于具有大于 7 的关联可能性置信度指示符 76 的所有记录，对于最大 6 级的分离程度 78，系统 10 将：（a）从个人#1 开始，（b）查找与#1 相关的具有大于 7 的置信度指示符 76 的所有其它个人 80，（c）分析所有的第一级分离个人 80，并确定置信度指示符 84 大于 7 的与第一级分离个人 80 相关的所有个人 82，并且（d）重复该处理直到满足这 6 级分离参数为止 78。该系统将以电子方式向相

关个人或分离系统发送警告 74（该警告可以包括根据用户定义标准获得的所有记录），以使得能够进行进一步的操作。

另外，关联记录 70 可以在几个数据库上进行复制。当接收到接收数据 20 时，系统将对各个其它数据库的工作负荷特性进行系统评估，并将匹配的/相关的/已分析记录分配给最可能有效分析所存储的已分析记录 68 的数据库。然后根据源自其它数据库的任何结果发出任何警告 74。

最后，可以以实时或分批处理的方式，根据可利用相同算法 92 的级联数据库公布列表 86 将处理后的数据传送 88 给附加数据库。通过这种方式，随后可将所传送的数据 88 用于与附加数据库和任何后续数据库中的数据（可包括不同的数据）进行匹配，以识别这些数据的关联、匹配或处理。例如，可以将根据本地数据库中的置信度指示符的匹配记录传送 88 到区域数据库，以与利用相同算法 92 的数据进行比较和匹配。此后，可以将从该区域数据库获得的处理数据传送 88 到国家局（national office）。通过在各个步骤中合并所处理的数据，尤其是实时地合并，机构或系统用户将能够确定不适当的或冲突的数据，以提示进一步的操作。

可使用传统的软件代码来实现上述方法、程序和系统的多个功能方面。该代码可以设置在任何计算机可读介质上，以由单个计算机或诸如互联网的分布式计算机网络使用。

通过以上说明，可以知道在不背离本发明的精神和范围的情况下可以进行多种变化和修改。应该理解，对于在此所述的具体设备并不旨在限定的目的，也不应推断为限定的目的。显然，所附权利要求涵盖了落入其范围内的所有这样的修改。

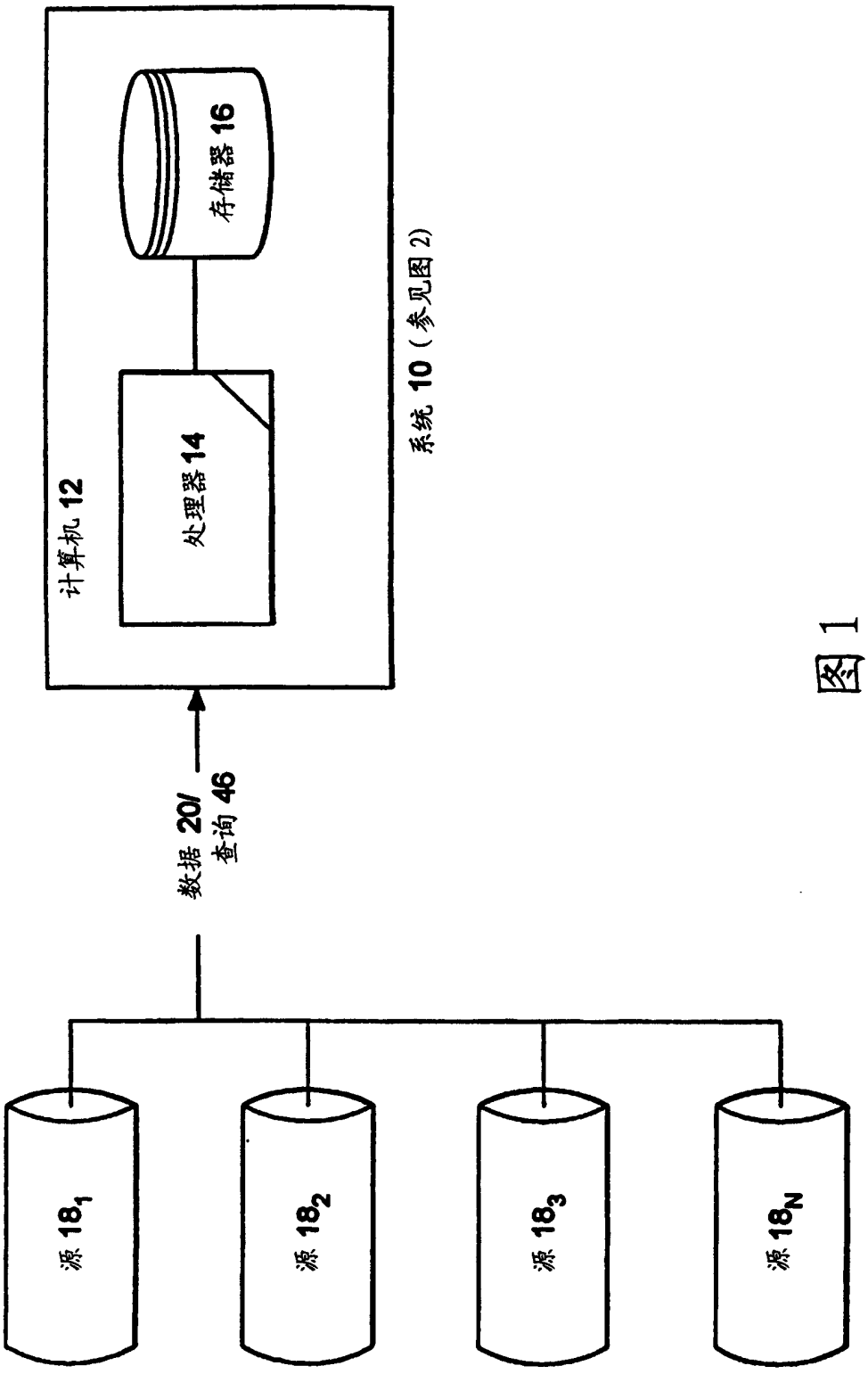


图 1

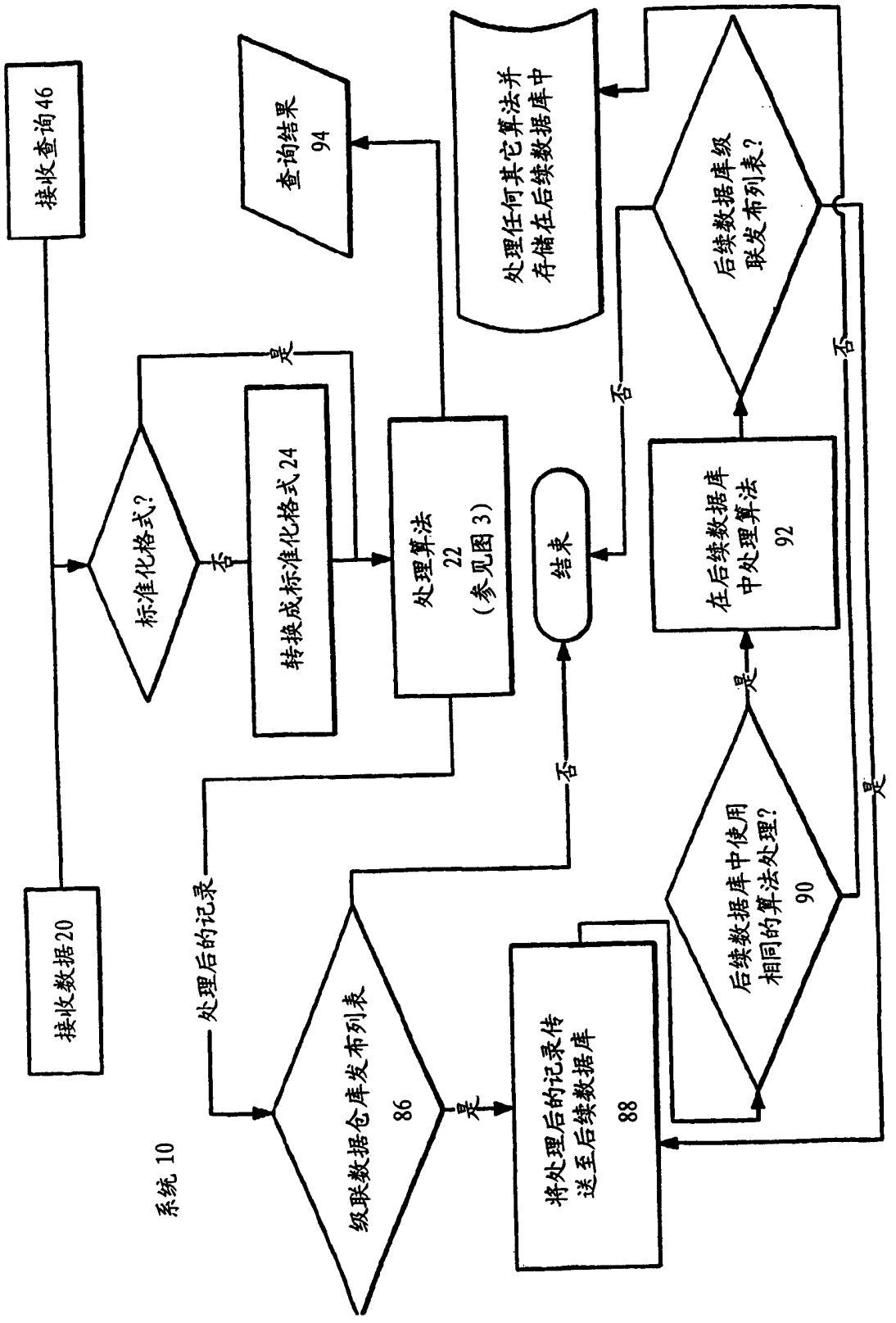

接收查询46
查询结果
94
处理任何其它算法并
存储在后续数据库中
后续数据库级
联发布列表?
标准化格式?
否
是
转换成标准化格式24
处理算法
22
(参见图3)
结束
否
在后续数据库
中处理算法
92
是
否
处理后的记录
级联数据仓库发布列表
86
是
将处理后的记录传
送至后续数据库
88
后续数据库中使用
相同的算法处理?
90
是
否
接收数据20
系统 10

图 2

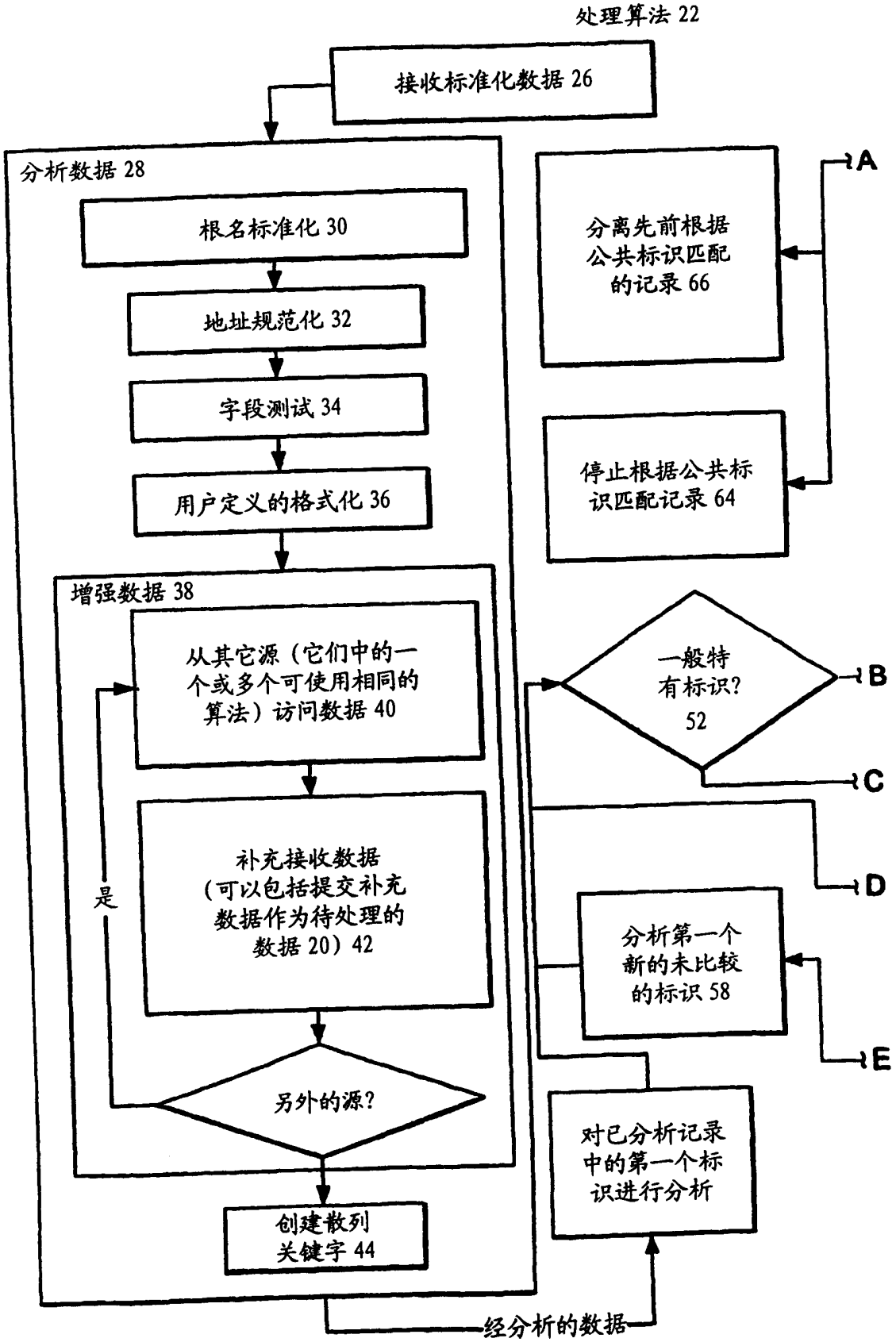

处理算法 22
接收标准化数据 26
分析数据 28
根名标准化 30
地址规范化 32
字段测试 34
用户定义的格式化 36
增强数据 38
从其它源（它们中的一个或多个可使用相同的算法）访问数据 40
补充接收数据（可以包括提交补充数据作为待处理的数据 20）42
是
另外的源?
创建散列关键字 44
经分析的数据
分离先前根据公共标识匹配的记录 66
A
停止根据公共标识匹配记录 64
一般特有标识? 52
B
C
D
分析第一个新的未比较的标识 58
E
对已分析记录中的第一个标识进行分析

图 3A

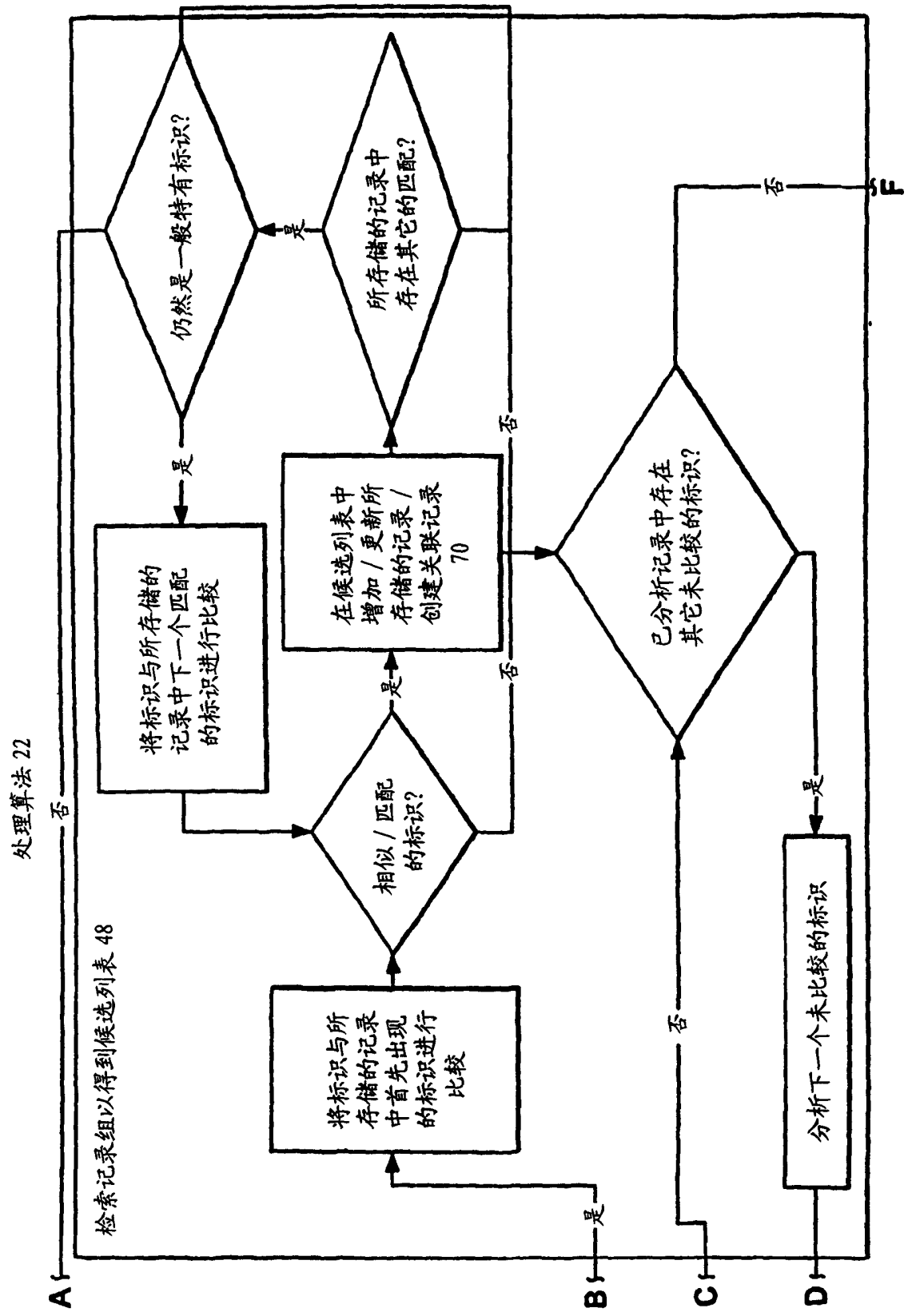

处理算法 22
检索记录组以得到候选列表 48
将标识与所存储的记录中首先出现的标识进行比较
相似 / 匹配的标识?
将标识与所存储的记录中下一个匹配的标识进行比较
仍然是一般特有标识?
在候选列表中增加 / 更新所存储的记录 / 创建关联记录 70
所存储的记录中存在其它的匹配?
已分析记录中存在其它未比较的标识?
分析下一个未比较的标识
是
否
A
B
C
D
F

图 3B

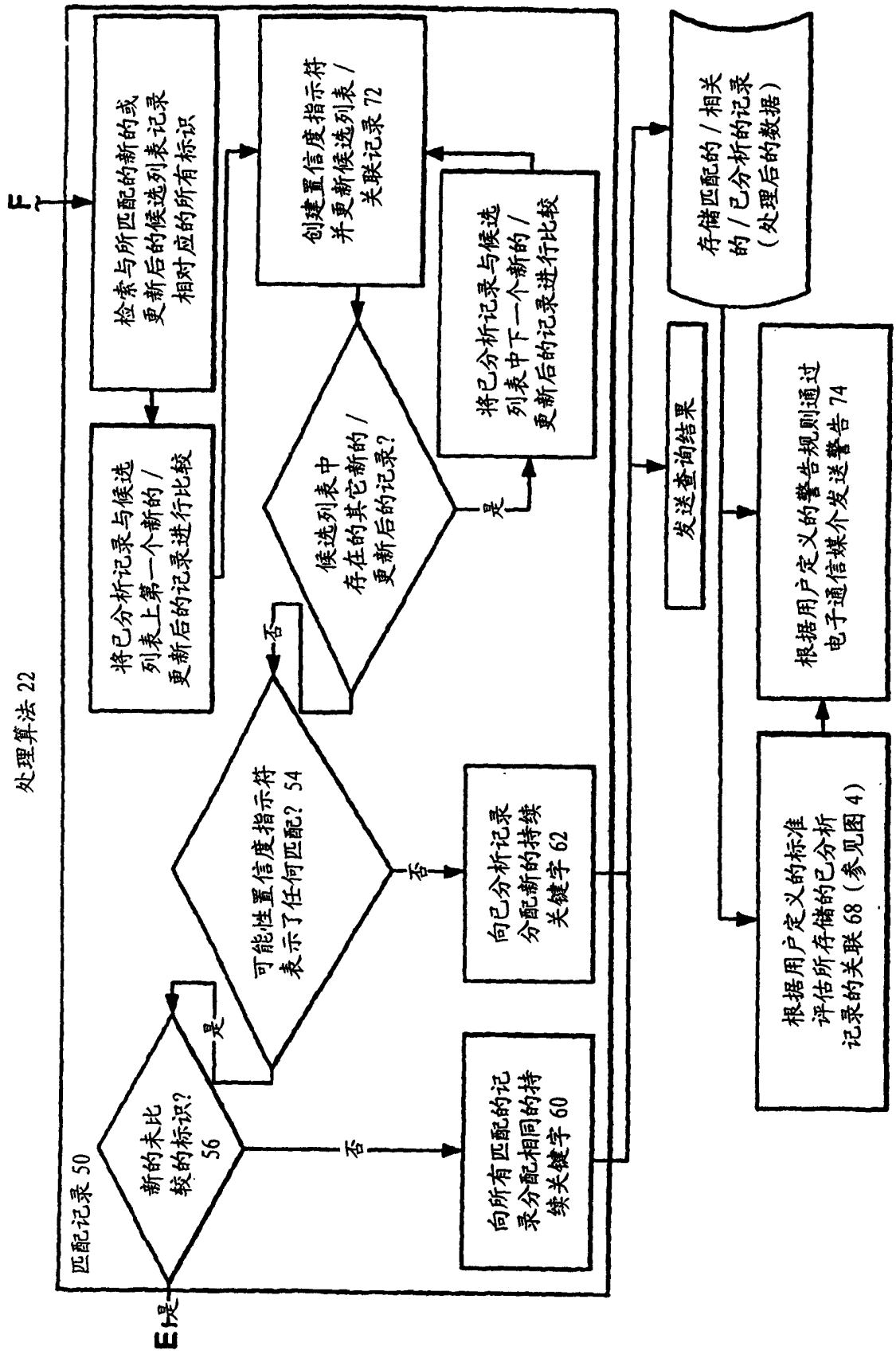

处理算法 22
匹配记录 50
F
检索与所匹配的新的或更新后的候选列表记录相对应的所有标识
将已分析记录与候选列表上第一个新的/更新后的记录进行比较
创建置信度指示符并更新候选列表/关联记录 72
候选列表中存在的其它新的/更新后的记录?
是
将已分析记录与候选列表中下一个新的/更新后的记录进行比较
否
可能性置信度指示符表示了任何匹配? 54
否
向已分析记录分配新的持续关键字 62
新的未比较的标识? 56
是
否
向所有匹配的记录分配相同的持续关键字 60
E
是
发送查询结果
存储匹配的/相关的/已分析的记录(处理后的数据)
根据用户定义的警告规则通过电子通信媒介发送警告 74
根据用户定义的标准评估所存储的已分析记录的关联 68(参见图 4)

图 3C

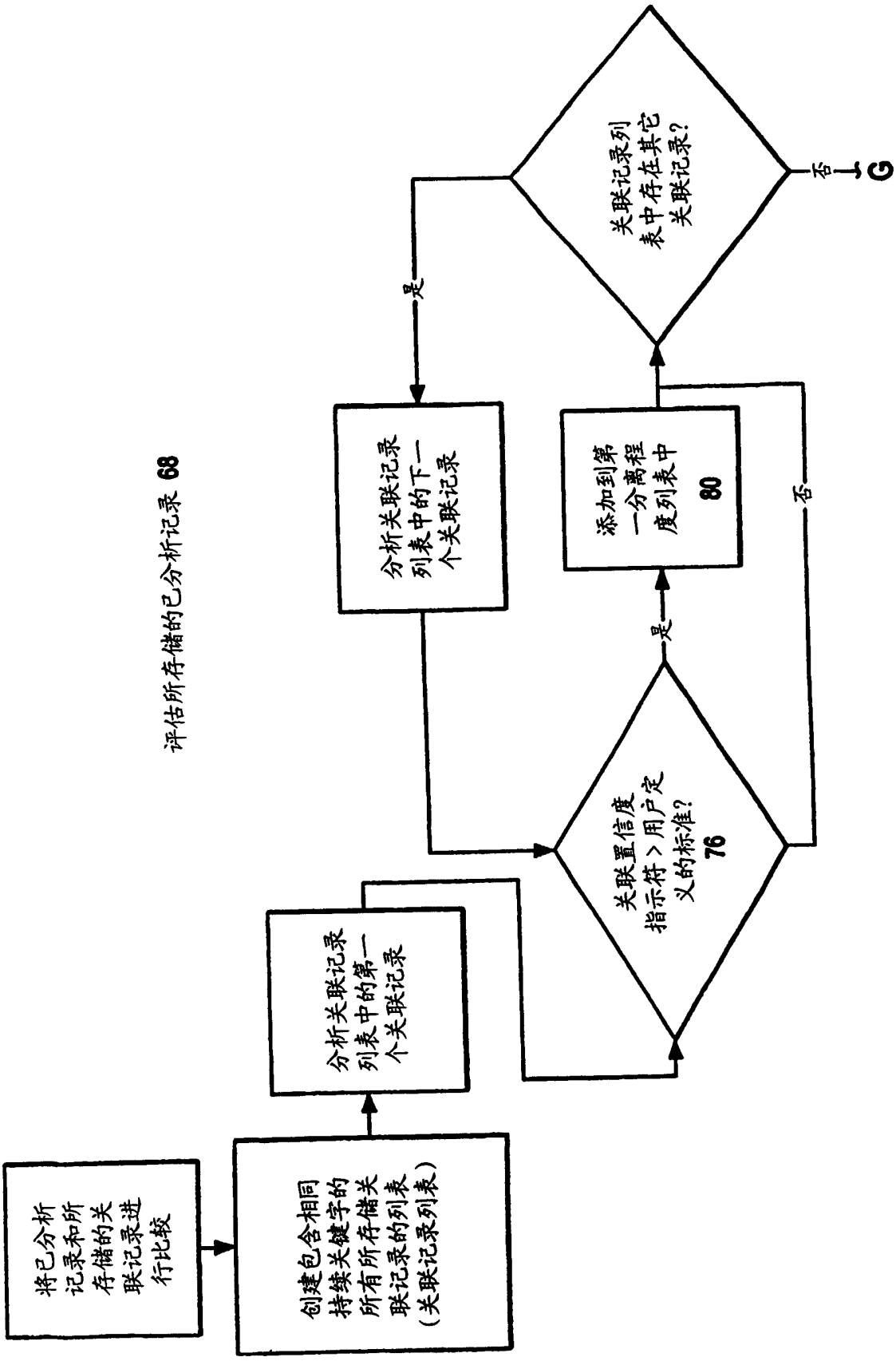

评估所存储的已分析记录 68
将已分析记录和所存储的关联记录进行比较
创建包含相同持续关键字的所有所存储关联记录的列表（关联记录列表）
分析关联记录列表中的第一个关联记录
关联置信度指示符 > 用户定义的标准? 76
是
否
添加到第一分离程度列表中 80
分析关联记录列表中的下一个关联记录
关联记录列表中存在其它关联记录?
是
否
G

图 4A

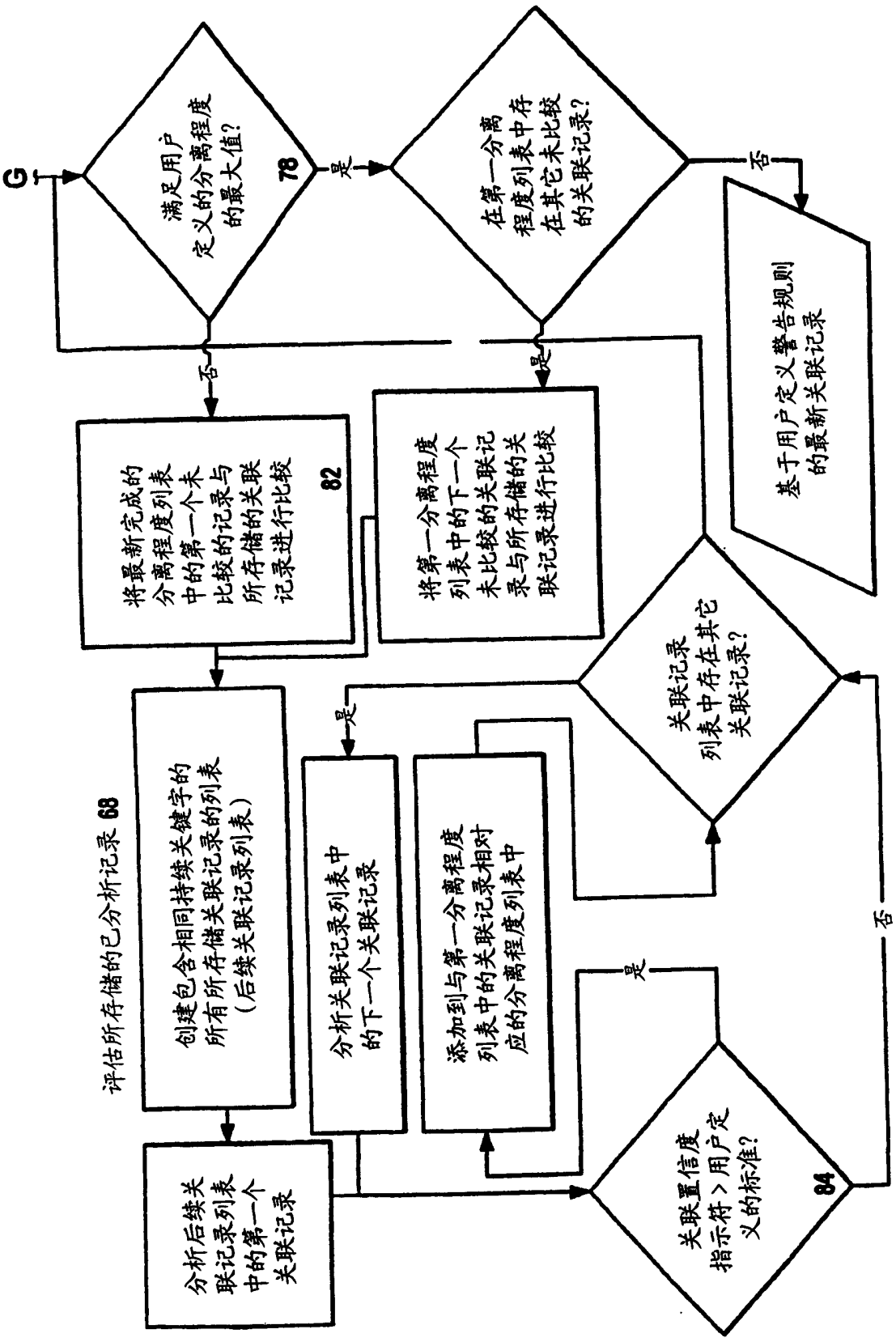

G
满足用户定义的分离程度的最大值? 78
是
在第一分离程度列表中存在其它未比较的关联记录?
否
基于用户定义警告规则的最新关联记录
否
将最新完成的分离程度列表中的第一个未比较的记录与所存储的关联记录进行比较 82
是
将第一分离程度列表中的下一个未比较的关联记录与所存储的关联记录进行比较
评估所存储的已分析记录 68
创建包含相同持续关键字的所有所存储关联记录的列表(后续关联记录列表)
关联记录列表中存在其它关联记录?
是
分析关联记录列表中的下一个关联记录
添加到与第一分离程度列表中的关联记录相对应的分离程度列表中
分析后续关联记录列表中的第一个关联记录
关联置信度指示符 > 用户定义的标准? 84
是
否

图 4B