

(12) 发明专利

(10) 授权公告号 CN 101339564 B

(45) 授权公告日 2012. 07. 04

(21) 申请号 200810128102. 6

US 2006/0129446 A1, 2006. 06. 15, 全文 .

(22) 申请日 2008. 07. 02

CN 1526233 A, 2004. 09. 01, 全文 .

(30) 优先权数据

US 2007/0143122 A1, 2007. 06. 21, 全文 .

2007-173637 2007. 07. 02 JP

审查员 巩瑜

(73) 专利权人 索尼株式会社

地址 日本东京

(72) 发明人 角川元辉

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 秦晨

(51) Int. Cl.

G06F 17/30 (2006. 01)

(56) 对比文件

CN 1418344 A, 2003. 05. 14, 全文 .

US 2003/0093790 A1, 2003. 05. 15, 全文 .

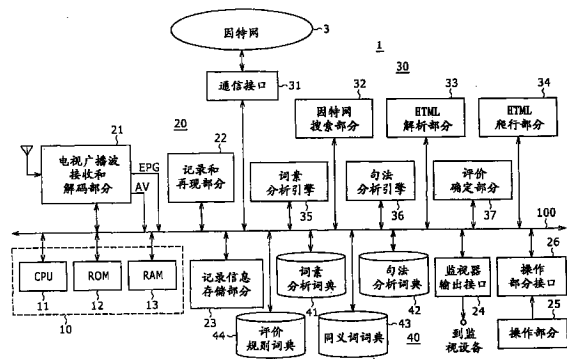
权利要求书 4 页 说明书 16 页 附图 11 页

(54) 发明名称

用于搜索内容的评价的方法和系统

(57) 摘要

本发明提供信息处理设备以及用于搜索内容评价的方法和系统。该信息处理设备包括：被配置为获取内容相关信息的获取部分；被配置为从内容相关信息中提取搜索关键字的关键字提取部分；被配置为通过因特网搜索具有网页的网站并获取关于这些网站的信息的站点搜索部分；被配置为从这些网站中选择前 N 个网站的第一站点选择部分；被配置为访问 N 个站点中的每一个以便从该 N 个网站中的每一个的网页中提取文章并选择该 N 个网站中的两个或多个作为种子站点的第二站点选择部分；以及被配置为从种子站点和链接到种子站点的下级网站中收集文章并从收集到的文章中提取内容的评价结果的评价结果获取部分。



1. 一种由信息处理设备使用的用于搜索内容的评价的系统,包括:

获取部分,该获取部分被配置为获取内容相关信息,该内容相关信息包含内容的标题和描述该内容的信息;

关键字提取部分,该关键字提取部分被配置为从由所述获取部分获取的内容相关信息中提取搜索关键字;

站点搜索部分,该站点搜索部分被配置为:通过因特网搜索具有包含由所述关键字提取部分提取的搜索关键字的网页的网站,并获取关于作为搜索结果而被找到的多个网站的信息;

第一站点选择部分,该第一站点选择部分被配置为从作为由所述站点搜索部分执行的搜索结果而被找到的所述多个网站中选择前 N 个网站,N 为预定的大于 1 的整数;

第二站点选择部分,该第二站点选择部分被配置为:访问由所述第一站点选择部分选择的 N 个网站中的每一个,以便从该 N 个网站中的每一个的网页中提取文章;并且选择该 N 个网站中的两个或更多个作为种子站点,从每一个种子站点提取的文章包含个数超过预定阈值的搜索关键字的例子;以及

评价结果获取部分,该评价结果获取部分被配置为:从由所述第二站点选择部分选择的所述两个或更多个种子站点和链接到这些种子站点的下级网站中收集文章,并从收集到的文章中获取所述内容的评价结果。

2. 根据权利要求 1 所述的系统,还包括:

存储部分,该存储部分被配置为存储由所述评价结果获取部分获取的多个内容的评价结果,从而使得每一种评价结果都与用于识别所述内容中的对应的一个内容的信息相关联;以及

评价结果发送部分,该评价结果发送部分被配置为:当通过网络接收到用于获取评价结果的获取请求时,从存储在所述存储部分中的所述多个内容的评价结果中读取由包含在获取请求中的用于识别内容的信息所识别的内容的评价结果,其中,该获取请求包含用于识别内容的信息;并且通过网络将读取的评价结果发送到该获取请求的发送者。

3. 根据权利要求 1 所述的系统,还包括:

关键字接收部分,该关键字接收部分被配置为接收附加搜索关键字的输入,其中,

所述站点搜索部分请求因特网上的搜索站点搜索具有除了包含由所述关键字提取部分提取的搜索关键字之外还包含由所述关键字接收部分接收的搜索关键字的网页的网站,并且获取关于作为该搜索的结果而被找到的多个网站的信息。

4. 根据权利要求 1 所述的系统,其中,所述第一站点选择部分包括被配置为使用收缩条件对关于从搜索站点获取的作为搜索结果而被找到的所述多个网站的信息进行搜索以从作为搜索结果而被找到的所述多个网站中选择更小数目的网站的部分。

5. 根据权利要求 4 所述的系统,其中,

由所述站点搜索部分获取的关于所述多个网站的信息是统一资源定位符,即 URL,并且由所述第一站点选择部分使用的收缩条件是包含在 URL 中的特定词。

6. 根据权利要求 1 所述的系统,其中,

在由所述第二站点选择部分执行的对种子站点的选择中,不以从每一个种子站点提取的文章包含个数超过预定阈值的搜索关键字的例子为种子站点选择的标准,而是对提取的

文章进行词素分析以便检测搜索关键字,并且,基于每一个文章的搜索关键字内容来选择种子站点,如果提取的文章的搜索关键字内容超出了预定阈值,则选择具有包含所述文章的网页的网站作为种子站点,并且

在对搜索关键字内容的计算中,至少根据搜索关键字是从内容的标题中提取的还是从描述该内容的信息中提取的,将不同的权重分配给搜索关键字,提取的文章的搜索关键字内容为所检测到的该文章的搜索关键字权重的总和与全部搜索关键字权重的总和的比值。

7. 根据权利要求 1 所述的系统,其中,所述评价结果获取部分从每一个种子站点进入到逐渐降低级别的下级网站,直到进入比该种子站点的级别低预先指定的级别数的级别的网站,并且从每一个种子站点和下级网站的网页中提取和收集文章。

8. 根据权利要求 1 所述的系统,还包括:

评价规则词典,该评价规则词典包含评价表达和评估值,每一个评估值与评价表达中的对应的一种评价表达相关联,其中,

所述评价结果获取部分从收集到的文章中检测多个评价表达,通过参考所述评价规则词典来获取检测到的评价表达的对应的评估值,并且将所获取的检测到的评价表达的评估值的总和计算为内容的评价结果。

9. 根据权利要求 8 所述的系统,其中,根据对应的评价表达表示肯定的观点还是表示否定的观点,所述评价规则词典中的评价表达的每一个评估值具有正值或负值。

10. 根据权利要求 8 所述的系统,其中,

所述评价规则词典还包含用于修饰评价表达的程度表达和用于评价表达的评估值的加权因数,每一种程度表达与加权因数中的对应的一个加权因数相关联,并且

在所述评价结果获取部分中,如果通过程度表达修饰评价表达,那么,当计算内容的评价结果时,评价表达的评估值乘以程度表达的加权因数被用作所述内容的评价结果。

11. 根据权利要求 8 所述的系统,其中,在所述评价结果获取部分中,对收集到的文章进行词素分析,从而找到评价表达,并且,对收集到的文章进行句法分析,从而检测到与找到的评价表达在语法上相关的单词,并且,如果检测到的单词是搜索关键字,那么将找到的评价表达的评估值作为内容的评价结果加上。

12. 根据权利要求 11 所述的系统,还包括:

同义词词典,用于检测与搜索关键字相关的单词,其中,

对收集到的文章进行词素分析,从而找到评价表达,并且对收集到的文章进行句法分析,从而检测到与找到的评价表达在语法上相关的单词,并且,如果检测到的单词是与由所述同义词词典检测到的搜索关键字相关的单词,那么将找到的评价表达的评估值作为内容的评价结果加上。

13. 根据权利要求 12 所述的系统,其中,根据检测到的单词为搜索关键字还是为与由所述同义词词典检测到的搜索关键字相关的单词,使找到的评价表达的评估值乘以不同的加权因数,并且,

将找到的评价表达的评估值乘以加权因数后作为内容的评价结果加上,而不是将找到的评价表达的评估值作为内容的评价结果加上。

14. 根据权利要求 11 所述的系统,还包括:

评价确定部分,该评价确定部分被配置为:通过爬行所述收集到的文章搜索包含找到的评价表达的句子,识别与找到的评价表达在语法上相关的单词,并且,如果所识别的单词是描述内容的评价的文章的搜索关键字,则存储该句子作为评价代表候补句子。

15. 根据权利要求 11 所述的系统,还包括:

评价确定部分,该评价确定部分被配置为:通过爬行所述收集到的文章搜索包含找到的评价表达的句子,识别与找到的评价表达在语法上相关的单词,并且,如果所识别的单词是与描述内容的评价的文章的搜索关键字相关的单词,则存储该句子作为评价代表候补句子。

16. 一种由信息处理设备使用的用于搜索内容的评价的系统,包括:

获取部分,该获取部分被配置为获取内容相关信息,该内容相关信息包含内容的标题和描述该内容的信息;

关键字提取部分,该关键字提取部分被配置为从由所述获取部分获取的内容相关信息中提取搜索关键字;

站点搜索请求部分,该站点搜索请求部分被配置为:请求因特网上的搜索站点搜索具有包含由所述关键字提取部分提取的搜索关键字的网页的网站,并且获取关于作为搜索结果而被找到的多个网站的信息;

第一站点选择部分,该第一站点选择部分被配置为从作为由搜索站点执行的搜索的结果而被找到的所述多个网站中选择前 N 个网站,N 为预定的大于 1 的整数;

第二站点选择部分,该第二站点选择部分被配置为:访问由所述第一站点选择部分选择的 N 个网站中的每一个,以便从该 N 个网站中的每一个的网页中提取文章;并且选择该 N 个网站中的两个或更多个作为种子站点,从每一个种子站点提取的文章包含个数超过预定阈值的搜索关键字的例子;以及

评价结果获取部分,该评价结果获取部分被配置为:从由所述第二站点选择部分选择的所述两个或更多个种子站点和链接到这些种子站点的下级网站中收集文章,并且从收集到的文章中获取所述内容的评价结果。

17. 一种由信息处理设备使用的用于搜索内容的评价的方法,该方法包括以下步骤:

获取包含内容的标题和描述该内容的信息的内容相关信息;

从在获取步骤中获取的内容相关信息中提取搜索关键字;

通过因特网搜索具有包含在提取步骤中提取的搜索关键字的网页的网站,并且获取关于作为搜索结果而被找到的所述多个网站的信息;

从在搜索步骤中作为搜索结果而被找到的多个网站中选择前 N 个网站,N 为预定的大于 1 的整数;

访问在选择步骤中选择的 N 个网站中的每一个,以便从该 N 个网站中的每一个的网页中提取文章,并选择该 N 个网站中的两个或更多个作为种子站点,从每一个种子站点中提取的文章包含个数超过预定阈值的搜索关键字的例子;以及

从在所述的选择步骤中选择的所述两个或更多个种子站点和链接到这些种子站点的下级网站中收集文章,并从收集到的文章中获取内容的评价结果。

18. 根据权利要求 17 所述的方法,还包括以下步骤:

将在所述的获取步骤中获取的多个内容的评价结果存储到存储部分中,从而使得每一

种评价结果都与用于识别所述内容中的对应的一个内容的信息相关联 ;以及

当通过网络接收到用于获取评价结果的获取请求时,从在所述的存储步骤中存储的所述多个内容的评价结果中读取由包含在获取请求中的、用于识别内容的信息所识别的内容的评价结果,并将读取的评价结果通过网络发送到该获取请求的发送者,其中,该获取请求包含用于识别内容的信息。

19. 一种用于搜索内容的评价的系统,该系统包括:

服务器设备,被配置为

获取包含内容的标题和描述该内容的信息的内容相关信息;

从获取的内容相关信息中提取搜索关键字;

通过因特网搜索具有包含提取的搜索关键字的网页的网站,并获取关于作为搜索结果而被找到的多个网站的信息;

从作为搜索结果而被找到的所述多个网站中选择前N个网站,N为预定的大于1的整数;

访问选择的N个网站中的每一个,以便从该N个网站中的每一个的网页中提取文章,并选择该N个网站中的两个或更多个作为种子站点,从每一个种子站点中提取的文章包含个数超过预定阈值的搜索关键字的例子;

从选择的所述两个或更多个种子站点和链接到这些种子站点的下级网站中收集文章,并从收集到的文章中获取内容的评价结果;

存储获取的多个内容的评价结果,从而使得每一种评价结果都与用于识别内容中的对应的一个内容的信息相关联 ;以及

当通过所述网络接收到用于获取评价结果的获取请求时,从存储的所述多个内容的评价结果中读取由包含在获取请求中的用于识别内容的信息所识别的内容的评价结果,并将读取的评价结果通过网络发送到该获取请求的发送者,其中,该获取请求包含用于识别内容的信息;

网络 ;以及

通过所述网络连接到所述服务器设备的客户端设备,被配置为

将用于获取评价结果的获取请求通过所述网络发送到所述服务器设备,所述获取请求包含用于识别内容的信息,以及

从所述服务器设备接收内容的评价结果并将评价结果通知用户。

用于搜索内容的评价的方法和系统

[0001] 相关申请的交叉引用

[0002] 本发明包括与于 2007 年 7 月 2 日提交给日本专利局的日本专利申请 JP 2007-173637 相关的主题,该日本专利申请的全部内容以引用的方式并入本文中。

技术领域

[0003] 本发明涉及用于搜索内容例如广播节目内容或存储在封装介质中的内容的评价的设备、方法和系统。

背景技术

[0004] 随着近来硬盘驱动器的容量的增加,能够记录如此大量的电视广播节目内容以至于超过一天长度的记录和再现设备已经出现了。很多这样的记录和再现设备都装备有基于用户登记的嗜好信息或关键字自动选择电视广播节目内容并记录所选择的电视广播节目内容的特征。因此,用户经常被迫去考虑要再现记录的大量电视广播节目内容中的哪些内容。

[0005] 上述类型的记录再现设备的用户能够从记录的大量电视广播节目内容中选择任意的记录的电视广播节目内容,并能够再现和观看所选择的电视广播节目内容。但是,理想的情况是,当选择要再现的记录的电视广播节目内容时,用户能够知道对记录的电视广播节目内容中的每一个的评价。对记录的电视广播节目内容中的每一个的评价可用作从大量电视广播节目内容中选择理想内容的基础。

[0006] 在因特网上,有大量网站,用户在上面写下其对各种主题感想或观点。人们已经提出了这样一种系统,该系统用于收集关于各种主题的这种感想或观点,并且使用词素分析 (morphological analysis) 技术或语法 (syntactic analysis) 分析技术来分析这些感想或观点,从而确定关于这些主题的评价 (例如,参见日本专利公布 No. 2003-248681、2003-157255、2003-157254 和 2003-157253)。

发明内容

[0007] 然而,在上面所列的文献中描述的系统没有把关于电视广播节目内容的主题信息的特性考虑在内,而是处理各种主题的通用系统。因此,为了确定评价,必须爬行通过 (crawl through) 巨大数量的类型广泛的多种网站,以便从这些网站中收集数据并存储这些数据。因此,服务器系统的规模相应地就必须很大。

[0008] 更具体来说,在因特网上有巨大数量的网站,在相关技术的上述系统中,访问这些巨量网站中的每一个以便从每一个网站的网页中收集数据并存储这些数据,并且对这样存储的包含关于这些主题的感想和观点的巨大数量的信息进行词素分析和句法分析,从而确定评价。因此,服务器系统的规模不可避免地会非常大。

[0009] 由于需要超大规模的服务器系统,因此很难使个人用户拥有的个人装置能够对内容 (例如电视广播节目内容) 的评价进行搜索。此外,在相关技术中,逐个地访问巨大数量

的网站,以便收集包含关于目标主题(即,应当搜索关于它的评价的主题)的感想和观点,因此要经过很长时间才能获得关于该评价的信息,即搜索结果。

[0010] 本发明致力于解决与现有方法和设备相关的上述问题以及其它问题,并且提供使得在相对较短的时间和相对更小规模的结构下获得内容的评价结果成为可能的设备、方法和系统。

[0011] 根据本发明的一个实施例,提供一种信息处理设备,该信息处理设备包括:获取部分;关键字提取部分;站点搜索(site search)部分;第一站点选择部分;第二站点选择部分;和评价结果获取部分。获取部分被配置为获取内容相关信息,该内容相关信息包含内容的标题和描述该内容的信息。关键字提取部分被配置为从由获取部分获取的内容相关信息中提取搜索关键字。站点搜索部分被配置为通过因特网搜索具有包含由关键字提取部分提取的搜索关键字的网页的网站,并获取关于作为搜索结果而被找到的多个网站的信息。第一站点选择部分被配置为从作为由站点搜索部分执行的搜索的结果而被找到的多个网站中选择前N个网站,N为预定的大于1的整数。第二站点选择部分被配置为访问由第一站点选择部分选择的N个网站中的每一个以便从该N个网站中的每一个的网页中提取文章,并且选择该N个网站中的两个或更多个作为种子站点(seed site),从每一个种子站点提取的文章包含个数超过预定阈值的搜索关键字的例子。评价结果获取部分被配置为从由第二站点选择部分选择的两个或更多个种子站点和链接到这些种子站点的下级网站中收集文章,并从收集到的文章中获取内容的评价结果。

[0012] 根据上述信息处理设备,关键字提取部分从要搜索其评价的内容的内容相关信息中提取搜索关键字。然后,站点搜索部分(例如搜索引擎)自动逐个地访问因特网上的网站以便从每一个网站的网页中收集信息,并检测具有包含由关键字提取部分提取的搜索关键字的网页的网站。

[0013] 然后,第一站点选择部分从具有包含搜索关键字的网页的那些网站中选择前N个网站。N为预定的大于1的整数。

[0014] 然后,第二站点选择部分访问由第一站点选择部分选择的N个网站,以便从该N个网站中的每一个的网页中提取文章,并选择该N个网站中的两个或更多个作为种子站点。从每一个种子站点中提取的文章包含个数超过预定阈值的搜索关键字的例子。这里,种子站点是指从其开始爬行通过链接到其上的较低级网站的网站。

[0015] 评价结果获取部分爬行通过由第二站点选择部分选择的种子站点和链接到种子站点的下级网站(即,逐个地访问这些网站),以便从每一个网页中收集文章。然后,评价结果获取部分对收集到的文章进行词素分析和句法分析,例如,以便获取该内容的评价结果。

[0016] 根据该信息处理设备,由第一站点选择部分和第二站点选择部分在两个步骤中从作为搜索结果而被找到的网站中选择更小数量的网站,从该更小数量的网站的网页中收集用于获取评价结果的文章,并且从这些收集到的文章中获得评价结果。因此,该信息处理设备具有比相关技术中的那些设备更小规模的结构,但是能够在相对更小的时间内获得评价结果。

[0017] 根据本发明的另一个实施例,提供一种信息处理设备,该信息处理设备包括:获取部分;关键字提取部分;站点搜索请求部分;第一站点选择部分;第二站点选择部分;和评价结果获取部分。获取部分被配置为获取内容相关信息,该内容相关信息包含内容的标题

和描述该内容的信息。关键字提取部分被配置为从由获取部分获取的内容相关信息中提取搜索关键字。站点搜索请求部分被配置为请求因特网上的搜索站点搜索具有包含由关键字提取部分提取的搜索关键字的网页的网站,并获取关于作为搜索结果而被找到的多个网站的信息。第一站点选择部分被配置为从作为由搜索站点执行的搜索的结果而被找到的多个网站中选择前N个网站,N为预定的大于1的整数。第二站点选择部分被配置为访问由第一站点选择部分选择的N个网站中的每一个以便从该N个网站中的每一个的网页中提取文章,并且选择该N个网站中的两个或更多个作为种子站点,从每一个种子站点提取的文章包含个数超过预定阈值的搜索关键字的例子。评价结果获取部分被配置为从由第二站点选择部分选择的两个或更多个种子站点和链接到这些种子站点的下级网站中收集文章,并从收集到的文章中获取内容的评价结果。

[0018] 与前面描述的信息处理设备不同,该信息处理设备没有站点搜索部分。替代的是,该信息处理设备具有站点搜索请求部分。该站点搜索请求部分请求因特网上的搜索站点对具有包含由关键字提取部分提取的搜索关键字的网页的网站进行搜索,并获取关于作为搜索结果而被找到的多个网站的信息。

[0019] 于是,与前面描述的信息处理设备的情况也一样,由第一站点选择部分和第二站点选择部分从作为搜索的结果而被找到的网站中选择更小数量的网站,并且使用所选择的更小数量的网站来获取评价结果。从自所述更小数量的网站收集的文章中获取评价信息。

[0020] 该信息处理设备没有站点搜索部分,并且,替代的是,该信息处理设备请求因特网上的搜索站点执行搜索。因此,该信息处理设备具有比前面描述的信息处理设备更小规模的结构。即,使用因特网上的搜索站点允许该信息处理设备在结构上具有减小的规模,因此该信息处理设备甚至可以被包含在个人使用的装置例如记录和再现设备中。

[0021] 根据本发明的又一个实施例,提供一种用于搜索由信息处理设备使用的内容的评价的方法,该方法包括以下步骤:获取步骤;提取步骤;搜索和获取步骤;选择步骤;访问和选择步骤;以及收集和获取步骤。获取步骤获取内容相关信息,该内容相关信息包含该内容的标题和描述该内容的信息。提取步骤从在获取步骤中获取的内容相关信息中提取搜索关键字。搜索和获取步骤通过因特网搜索具有包含在提取步骤中提取的搜索关键字的网页的网站,并获取关于作为搜索结果而被找到的多个网站的信息。选择步骤从作为搜索步骤中的搜索的结果而被找到的多个网站中选择前N个网站,N为预定的大于1的整数。访问和选择步骤访问在选择步骤中选择的N个网站中的每一个,以便从该N个网站中的每一个的网页中提取文章,并选择该N个网站中的两个或更多个作为种子站点,从每一个种子站点提取的文章包含个数超过预定阈值的搜索关键字的例子。收集和获取步骤从在所述的选择步骤中选择的两个或更多个种子站点和链接到这些种子站点的下级网站中收集文章,并从收集到的文章中获取内容的评价结果。

[0022] 根据本发明的又一个实施例,提供一种用于搜索内容的评价的系统,该系统包括服务器设备、网络、以及通过所述网络连接到所述服务器设备的客户端设备。该服务器设备被配置为:获取包含内容的标题和描述该内容的信息的内容相关信息;从获取的内容相关信息中提取搜索关键字;通过因特网搜索具有包含提取的搜索关键字的网页的网站,并获取关于作为搜索结果而被找到的多个网站的信息;从作为搜索结果而被找到的所述多个网站中选择前N个网站,N为预定的大于1的整数;访问选择的N个网站中的每一

个,以便从该N个网站中的每一个的网页中提取文章,并选择该N个网站中的两个或更多个作为种子站点,从每一个种子站点中提取的文章包含个数超过预定阈值的搜索关键字的例子;从选择的所述两个或更多个种子站点和链接到这些种子站点的下级网站中收集文章,并从收集到的文章中获取内容的评价结果;存储获取的多个内容的评价结果,从而使得每一种评价结果都与用于识别内容中的对应的一个内容的信息相关联;以及当通过所述网络接收到用于获取评价结果的获取请求时,从存储的所述多个内容的评价结果中读取由包含在获取请求中的用于识别内容的信息所识别的内容的评价结果,并将读取的评价结果通过网络发送到该获取请求的发送者,其中,该获取请求包含用于识别内容的信息。该客户端设备被配置为将用于获取评价结果的获取请求通过所述网络发送到所述服务器设备,所述获取请求包含用于识别内容的信息,以及从所述服务器设备接收内容的评价结果并将评价结果通知用户。

[0023] 根据本发明的上述实施例,在两个步骤中从作为搜索的结果而被找到的网站中选择更小数量的网站,从所述更小数量的网站的网页中收集用于获取评价结果的文章,并且从这些收集到的文章中获得评价结果。因此,实现了结构规模的减小,而且,仍然可以在相对较短的时间内获得内容的评价结果。

[0024] 附图说明

[0025] 图 1 是示出根据本发明第一实施例的信息处理设备的示例性硬件结构的框图。

[0026] 图 2 是示出包括根据第一实施例的信息处理设备的系统的结构的概要的视图。

[0027] 图 3 是示出在根据第一实施例的信息处理设备中使用的评价规则词典(reputation rule dictionary)的流程图。

[0028] 图 4 是示出用于在根据第一实施例的信息处理设备中使用的评价规则词典中的词典数据的实例的视图。

[0029] 图 5、6、7、8 和 9 均是示出由根据第一实施例的信息处理设备执行的评价搜索处理的一部分的流程图。

[0030] 图 10 是示出由根据第一实施例的信息处理设备显示的示例性评价结果的视图。

[0031] 图 11 是示出包括根据本发明第二实施例的信息处理设备的评价搜索系统的示例性结构的概要的视图。

[0032] 图 12 是示出根据第二实施例的信息处理设备的示例性硬件结构的框图。

[0033] 图 13 是示出在根据本发明的一个实施例的评价搜索系统中的客户端设备的示例性硬件结构的框图。

具体实施方式

[0034] 在下文中,将参考附图描述本发明的实施例。

[0035] 【信息处理设备的第一实施例】

[0036] 根据本发明第一实施例的信息处理设备是一种记录和再现设备,例如所谓的硬盘记录器,该记录和再现设备能够接收电视广播并将接收到的电视广播的数据存储到“记录信息存储部分”例如硬盘,并且还能够从记录信息存储部分读取数据并再现该数据。

[0037] 图 1 示出记录和再现设备 1 的硬件结构,该记录和再现设备 1 是根据第一实施例的信息处理设备。图 2 示出包括根据本实施例的记录和再现设备 1 的内容评价搜索系统的

配置的概要。

[0038] 如图 2 中所示,根据本实施例的记录和再现设备 1 连接到监视设备 2 和因特网 3。具有搜索引擎的搜索站点 4、大量的公告板站点 5、大量的博客站点 6 等连接到因特网 3。尽管没有示出,但是监视设备 2 具有用于显示图像的显示器例如液晶显示器 (LCD) 和用于输出再现声音的扩音器。

[0039] 如果搜索站点 4 通过因特网 3 从搜索请求终端例如记录和再现设备 1 接收搜索关键字,那么搜索站点 4 使用搜索引擎对具有包括该搜索关键字的网页的网站进行搜索,并将搜索结果发送到该搜索请求终端。例如,搜索结果包括作为搜索结果而被找到的统一资源定位符 (URL) 等。

[0040] 公告板站点 5 中的每一个都是提供电子公告板的网站,在电子公告板中,访问该网站的因特网用户可以随意地记入他们的观点、感想等。该公告板站点 5 由企业公司或个人管理。关于内容或其它各种主题的观点、感想等记入在公告板站点 5 中的每一个的电子公告板上。

[0041] 博客站点 6 中的每一个都是其中个人或企业公司记入日记条目 (diary entry)、对书籍或电影的评论、或者各种其它类型信息的网站。更不用说,关于这些内容的观点、感想等有时也在这些博客站点发表。博客站点的 URL 通常包含单词“blog”。

[0042] 根据第一实施例的记录和再现设备 1 没有对具有包括该搜索关键字的网页的网站执行搜索的搜索引擎,而是请求搜索站点 4 对具有包括该搜索关键字的网页的网站执行搜索。

[0043] 然后,根据本实施例的记录和再现设备 1 从搜索站点 4 获取搜索结果,即,作为搜索结果而被找到的网站,并从这些网站中选择将要用来获得正在搜索其评价的内容的评价结果 (即,关于评价的信息) 的网站。然后,根据本实施例的记录和再现设备 1 从所选择的网站的网页和链接到所选择的网站的下级网站的网页中收集书面信息,以便从该书面信息中获得上述内容的评价结果的信息。然后,记录和再现设备 1 从该书面信息中检测关于该评价的话语 (words),并基于检测到的关于该评价的话语 (words) 获得评价结果。然后,在本实施例中,记录和再现设备 1 将评价结果发送到监视设备 2,该监视设备 2 显示该评价结果或将该评价结果作为声音输出,从而将评价结果呈现给用户。

[0044] 如图 1 中所示的记录和再现设备 1 的示例性硬件结构配置用来实现上述功能。如图 1 中所示,根据本实施例的记录和再现设备 1 包括控制部分 10、“电视广播接收和记录 / 再现部分” 20、评价搜索部分 30、评价搜索用词典 40 和系统总线 100。控制部分 10、“电视广播接收和记录 / 再现部分” 20、评价搜索部分 30 和评价搜索用词典 40 与系统总线 100 相连。

[0045] 在本实施例中,控制部分 10 由微型计算机形成,并包括中央处理单元 (CPU) 11、用于程序的只读存储器 (ROM) 12 和用于工作区域的随机存取存储器 (RAM) 13。

[0046] ROM 12 存储关于电视广播节目选择的控制程序、关于电视广播节目内容的记录和再现的控制程序、用于搜索内容的评价和获取评价结果的程序,以及其它程序。CPU 11 在使用 RAM 13 作为工作区域时执行存储在 ROM 12 中的程序,从而控制上述的多种处理。

[0047] “电视广播接收和记录 / 再现部分” 20 包括“电视广播波接收和解码部分” 21、记录和再现部分 22、记录信息存储部分 23、监视器输出接口 24、操作部分 25,以及操作部分接

口 26。通过操作部分接口 26,控制部分 10 接收使用操作部分 25 输入的用户操作的输入,并确定用户操作的输入对应于什么指令。然后,基于确定结果,控制部分 10 根据该指令执行处理控制。

[0048] 在本实施例中,“电视广播波接收和解码部分”21 具有接收模拟地面电视广播、数字地面电视广播、广播卫星 (BS) 广播、通信卫星 (CS) 广播等的特征。

[0049] “电视广播波接收和解码部分”21 基于从控制部分 10 供给的电视广播节目选择控制信号接收电视广播波的电视广播节目并对其解码,并且将通过解码获得的电视广播节目内容的视频和音频信号 AV 发送到系统总线 100。

[0050] 基于用户通过使用操作部分 25 进行的广播波选择操作输入和广播节目选择操作输入,控制部分 10 生成要供给到电视广播波接收和解码部分 21 的电视广播节目选择控制信号。

[0051] 当控制部分 10 确定用户既没有通过操作操作部分 25 发出记录开始指令又没有通过操作操作部分 25 发出再现开始指令时,控制部分 10 允许已经从电视广播波接收和解码部分 21 发送到系统总线 100 的视频和音频信号 AV 通过监视器输出接口 24 发送到监视设备 2,从而在监视设备 2 的显示器上显示视频,音频由扩音器输出。

[0052] 同时,当控制部分 10 确定用户通过操作操作部分 25 发出记录开始指令时,控制部分 10 允许已经从电视广播波接收和解码部分 21 发送到系统总线 100 的视频和音频信号 AV 发送到记录和再现部分 22。

[0053] 记录和再现部分 22 包括记录编码部分和再现解码部分。当确定已经接收到记录开始指令时,记录和再现部分 22 的记录编码部分在控制部分 10 的控制下对视频和音频信号 AV 进行记录编码处理,包括数据压缩,并且,将所得到的信号写入到记录信息存储部分 23 以便存储在其中。

[0054] 在本实施例中,记录信息存储部分 23 由硬盘驱动器或大容量半导体存储器装置形成,因此能够存储相对大量的电视广播节目内容。

[0055] 在第一实施例中,“电视广播波接收和解码部分”21 具有从电视广播波中获取电子程序设计指南 (electronic programming guide) (EPG) 信息的功能。

[0056] 例如,控制部分 10 将 EPG 信息临时存储在 RAM13 中,并且,当将已经请求其记录的电视广播节目内容的记录信息存储在记录信息存储部分 23 中时,控制部分 10 从临时存储的 EPG 信息中提取内容相关信息,并将提取的内容相关信息存储在记录信息存储部分 23 中,以便与记录的电视广播节目内容相关联。内容相关信息包括记录的电视广播节目内容的标题、出现在内容中的人的姓名、关于该内容的信息的信息等。

[0057] 在第一实施例中,例如,在再现记录的内容之一前,用户使用操作部分 25 能够发出显示记录的电视广播节目内容 (在下文中称为“记录内容”) 的列表的指令。

[0058] 一旦接收到显示记录内容列表的指令,根据本实施例的控制部分 10 就将关于记录的广播节目的列表的信息发送到监视设备 2,从而使监视设备 2 的显示器显示该列表。该列表示出存储在记录信息存储部分 23 中的每一个电视广播节目内容的标题以及记录日期和时间。

[0059] 在本实施例中,记录内容的该列表不仅可以用来选择期望的记录内容来开始内容再现,而且可以用来选择期望的记录内容来发出搜索该内容的评价的指令。

[0060] 在显示记录内容的该列表后,一旦接收到搜索所选择的记录内容的评价的指令,如下所述,控制部分 10 使用评价搜索部分 30 和评价搜索用词典 40 从自因特网 3 获取的信息中获取所选择的记录内容的评价结果,并将该评价结果的信息发送到监视设备 2,从而监视设备 2 的显示器显示该评价结果的信息。参考由监视设备 2 的显示器显示的评价结果的信息,用户能够从大量记录内容中选择记录内容之一,从而开始对所选择的内容进行再现。

[0061] 当控制部分 10 确定用户通过操作操作部分 25 已经发出对从记录内容的列表中选择记录内容的再现开始指令时,控制部分 10 从记录信息存储部分 23 读取所选择的记录内容的记录信息,并将该记录信息发送到记录和再现部分 22。

[0062] 在控制部分 10 的控制下,记录和再现部分 22 允许再现解码部分操作用来对所选择的记录内容的记录信息进行再现解码处理,包括解压 / 解码,从而重建视频和音频信号 AV。在控制部分 10 的控制下,重建的视频和音频信号 AV 被发送到监视设备 2,并且,再现的视频显示在显示器上,而再现的音频由扩音器输出。

[0063] 接着,现在将在下文中对评价搜索部分 30 进行描述。在本实施例中,评价搜索部分 30 包括:用于连接到因特网 3 的通信接口 31、因特网搜索部分 32、超文本标记语言 (HTML) 解析部分 33、HTML 爬行部分 34、词素分析引擎 35、句法分析引擎 36 和评价确定部分 37。通信接口 31、因特网搜索部分 32、HTML 解析部分 33、HTML 爬行部分 34、词素分析引擎 35、句法分析引擎 36 和评价确定部分 37 中的每一个与系统总线 100 相连。

[0064] 请注意,在这里,因特网搜索部分 32、HTML 解析部分 33、HTML 爬行部分 34、词素分析引擎 35、句法分析引擎 36 和评价确定部分 37 可以在硬件中实现,或者可以通过软件应用程序实现。在它们通过软件应用程序实现的情况下,该软件应用程序预先存储在 ROM12 中。

[0065] 在本实施例中,评价搜索用词典 40 包括:词素分析词典 41、句法分析词典 42、同类词词典 (thesaurus dictionary) 43 和评价规则词典 44。这些词典 41 到 44 中的每一个都与系统总线 100 相连。

[0066] 当用户发出了搜索所选择的记录内容的评价的指令时,评价搜索部分 30 中的每一个部件都由控制部分 10 激活。

[0067] 因特网搜索部分 32 在其中存储了搜索站点 4 的 URL。如下所述,因特网搜索部分 32 生成包括从所选择的记录内容的内容相关信息中提取的搜索关键字的站点搜索请求,并且,在控制部分 10 的控制下,因特网搜索部分 32 通过通信接口 31 访问搜索站点 4,其中,该站点搜索请求的搜索评价的指令已经被发出。

[0068] HTML 解析部分 33 具有分析以 HTML 格式写入的文档 (在下文中称为“HTML 文档”;不仅该文章由文本字符构成,而且图像也可以置于 HTML 文档中) 的处理能力,并且消除标签和图像部分,从而只提取由文本字符构成的文章部分。这里使用的术语“解析”是指分析 HTML 文档,从而只提取文章部分。

[0069] HTML 爬行部分 34 执行对多个级别进行下述操作的处理:收集网站上的 HTML 文档;解析所收集到的 HTML 文档,以识别链接到第一网站的网站;收集所识别的链接到第一网站的网站上的 HTML 文档;解析所收集到的 HTML 文档,以识别包含在 HTML 文档中的链接;收集所识别的链接的目的地上的 HTML 文档;解析所收集到的 HTML 文档等等。该处理称为“爬行”。

[0070] 参考词素分析词典 41, 词素分析引擎 35 执行这样的处理, 即: 将短语或段 (passage) 划分为语素, 即在语言中有意义的最小单位; 以及识别每一个语素的语音 (speech) 部分 (例如, 名词、动词、副词、作为主语的助动词的后置词、形容词等)。用于词素分析词典 41 的词典数据可以通过因特网 3 获得, 或者可以预先为记录和再现设备 1 特定地生成。由于词素分析和词素分析词典是众所周知的, 因此这里省略对它们的详细描述。

[0071] 在其中段中的每一个单词 (word) 的语音部分和各单词已经被识别为由词素分析引擎 35 执行的词素分析的结果的情形下, 句法分析引擎 36 在参考句法分析词典 42 时执行识别根据其构造该段的规则的操作。用于句法分析词典 42 的词典数据也可以通过因特网 3 获得, 或者可以预先为记录和再现设备 1 特定地生成。由于句法分析和句法分析词典是众所周知的, 因此这里省略对它们的详细描述。

[0072] 在本实施例中, 参考评价规则词典 44, 评价确定部分 37 执行将评估值 (评价得分) 计算为评价结果的处理。在本实施例中, 用于评价规则词典 44 的词典数据预先为记录和再现设备 1 特定地生成。

[0073] 【评价规则词典的创建】

[0074] 现在参考图 3 的流程图对生成用于评价规则词典 44 的词典数据的方法的过程进行描述。图 4 示出用于评价规则词典 44 的词典数据的实例。

[0075] 首先, 在评价规则词典 44 中登记其评价表达 (expression) 和评估值 (得分) (步骤 T1)。这里使用的术语“评价表达”是指表达肯定的或否定的观点的单词或短语, 例如“有趣的”、“感动的”、“好的”、“枯燥的”、“无趣的”和“不好的”。给肯定的评价表达分配正的 (+) 评估值 (得分), 给否定的评价表达分配负的 (-) 评估值 (得分)。

[0076] 就正的评价表达而言, 如图 4 所示, 例如, 给“有趣的”和“感动的”分配“+5 点”的评估值, 给“好的”分配“+2 点”的评估值。就负的评价表达而言, 例如, 给“枯燥的”和“无趣的”分配“-5 点”的评估值, 给“不好的”分配“-2 点”的评估值。

[0077] 接着, 评价表达的程度表达和加权因数 (系数) 在评价规则词典 44 中登记 (步骤 T2)。这里使用的术语“程度表达”是指显示其所指的单词或短语的程度的单词或短语, 例如“非常”、“相当”、“完全地”和“高度地”。当程度表达出现在评价表达前面时, 程度表达的加权因数 (系数) 用作用于评价表达的评估值的系数。

[0078] 如图 4 所示, 例如, 给“非常”分配“ $\times 2$ ”的加权因数 (系数), 给“相当”分配“ $\times 3$ ”的加权因数 (系数)。

[0079] 最后, 通过词素分析引擎 35 对登记的评价表达和程度表达中的每一个进行词素分析, 并且将分析结果存储在评价规则词典 44 中, 从而与对应的评价表达或程度表达相关联, 如图 4 所示 (步骤 T3)。因此, 评价规则词典 44 的创建完成。请注意, 对评价表达和程度表达的词素分析的结果在下文中将分别称为“评价表达 (词素分析结果)”和“程度表达 (词素分析结果)”。

[0080] 同义词词典 43 是在其中登记与词条 (entry) 表达相似 (例如同义词、近义词和缩写词) 的表达 (在下文中称为“语义相似的表达”) 的词典。这种语义相似的表达最初登记在同义词词典 43 中, 但是, 如果需要的话, 允许用户另外登记同义词、近义词、缩写词等。

[0081] 当在公告板或博客上记入对内容的评论时, 用户可以使用搜索关键字的同义词、缩写词等。考虑到该事实, 准备同义词词典 43。如下所述, 在使用同义词词典 43 的情况下,

当搜索内容评价时,搜索关键字的语义相似的表达也被考虑在内。

[0082] [评价搜索处理]

[0083] 接着,现在将参考图 5、6、7、8 和 9 在下文中对搜索存储在记录信息存储部分 23 的电视广播节目内容的评价的过程进行描述,该过程是在具有上述结构的记录和再现设备 1 中执行的。

[0084] 图 5 和图 6 是示出评价搜索处理的示例性流程的流程图。在根据本实施例的记录和再现设备 1 中,在控制部分 10 的控制下,评价搜索部分 30 使用评价搜索用词典 40 执行评价搜索处理。如上所述,在本实施例中,当用户从记录内容列表中选择特定记录内容并执行开始搜索所选择的内容的评价的搜索的操作时,开始由图 5 和图 6 的流程图示出的评价搜索处理。

[0085] 图 7 到图 9 是参考特定实例补充地示出如图 5 和图 6 中所示的各个步骤的处理的流程图。虽然图 5 和图 6 的流程与图 7 到图 9 的流程不完全对应,但是这两个流程之间的对应关系是清楚的。

[0086] 一旦接收到搜索所选择的记录内容的评价的指令,控制部分 10 首先读取存储在记录信息存储部分 23 中的内容相关信息以便与所选择的记录内容相关联,并且获取已经从 EPG 获得的记录内容的标题和节目内容信息(图 5 中的步骤 S1)。

[0087] 然后,控制部分 10 将获取的内容相关信息(节目信息)供给到词素分析引擎 35,并允许词素分析引擎 35 对内容相关信息进行词素分析(图 5 中的步骤 S2)。然后,控制部分 10 选取作为词素分析结果而作为名词被检测到的词语(expression)作为搜索关键字,并将该搜索关键字供给到因特网搜索部分 32(图 5 中的步骤 S3)。

[0088] 例如,假设在步骤 S1 获取的内容相关信息是包含记录内容“Flamboyance of Royal Family”的标题和包括显示在记录内容中出现的人员的单词“Kimura Takuo”、“Minamioji Ginpachi”等的节目内容信息(参见图 7 中的步骤 S21)。在这种情况下,作为通过词素分析进行的名词提取的结果(参见图 7 中的步骤 S22)，“Flamboyance”、“Royal Family”、“Kimura Takuo”、“Minamioji Ginpachi”等作为搜索关键字提取(参见图 7 中的步骤 S23)。

[0089] 请注意,在这种情况下,关于标题的词素分析的结果在下文中将称为“标题(词素分析结果)”,而关于节目内容信息的词素分析的结果在下文中将称为“内容信息(词素分析结果)”

[0090] 在本实施例中,用于评价搜索的搜索关键字不局限于从上述内容相关信息中自动提取的那些搜索关键字,而允许用户输入用于评价搜索的附加的搜索关键字。因此,例如,控制部分 10 允许在监视设备 2 的显示器上显示提示用户输入附加的搜索关键字的消息,并且等待用户使用操作部分 25 输入搜索关键字。如果用户输入搜索关键字,那么控制部分 10 将输入的搜索关键字作为附加的搜索关键字供给到因特网搜索部分 32(图 5 中的步骤 S4)。

[0091] 请注意,用户不是必须输入附加的搜索关键字。用户可以对不需要附加的搜索关键字的效果输入指令。在这种情况下,控制部分 10 消除提示用户输入附加的搜索关键字的消息,并进入下一个步骤。在这种情况下,只有在图 5 中的步骤 S3 中提取的搜索关键字用于评价搜索。请注意,可以忽略用于接收附加的搜索关键字的步骤 S4。

[0092] 例如,假设用户已经输入了附加的搜索关键字(参见图7中的步骤S24),并且该附加的搜索关键字是“blog”。在这种情况下,将附加的搜索关键字“blog”加入到在图5中的步骤S3(=图7中的步骤S21)中所获取的搜索关键字(参见图7中的步骤S25)。

[0093] 请注意,附加的搜索关键字“blog”是用户为了从博客站点获取评价信息的目的而添加的关键字。

[0094] 应该理解,附加的搜索关键字并不局限于本实例。例如,在用户希望从公告板站点获取评价信息的情况下,用户可以输入“公告板”作为搜索关键字。应该理解,作为附加的搜索关键字,用户可以不输入关于网站的表达,而可以输入与所选择的记录内容(要搜索其评价)相关的表达。

[0095] 接着,控制部分10控制因特网搜索部分32将搜索请求发送到搜索站点。同时,因特网搜索部分32通过通信接口31和因特网3将包含所接收到的搜索关键字的搜索请求发送到预先登记的搜索站点4。

[0096] 响应该搜索请求,搜索站点4搜索具有包含接收到的搜索关键字的网页的网站,并将作为搜索结果而找到的网站的URL信息发送到记录和再现设备1。因此,控制部分10通过通信接口31接收到URL信息,并将前N个网站的URL信息临时存储到RAM13或记录信息存储部分23中的空白空间中,例如,该记录信息存储部分23由硬盘驱动器形成。

[0097] 只有前N个网站的URL被存储的事实意指从作为搜索结果而被找到的巨大数量的网站中选择有限数量的网站(评价搜索用网站的第一选择;图5中的步骤S6)。

[0098] 在图7的实例中,搜索站点使用“Flamboyance”、“RoyalFamily”、“Kimura Takuo”、“Minamioji Ginpachi”等以及外加的“blog”(参见图7中的步骤S25)作为搜索关键字执行搜索(网络搜索;参见图7中的步骤S26)。作为该搜索的结果,找到巨大数量的网站,但是,只存储其中前N个网站的URL(参见图7中的步骤S27)。

[0099] 请注意,在图5中的步骤S6的评价搜索用网站的第一选择中,可以使用包含在URL中的特定单词例如“blog”来进一步减小网站数量。包含这种特定单词的URL很可能是特定类型网站的URL,因此这种特定单词可以用来只选择特定类型的网站。例如,通过使用特定单词“blog”,即,通过只选择其URL包含特定单词“blog”的网站,可以只提取博客站点。

[0100] 在本实例中,从前N个网站中选择其URL包含单词“blog”的网站(参见图8中的步骤S31)。如果从前N个网站中选择的网站数量大于预定的有限数M($M \leq N$),那么进一步从本实例中所选择的网站中选择前M个网站(参见图8中的步骤S32)。

[0101] 请注意,可以这样安排,首先从作为搜索结果而被找到的网站中选择其URL包含单词“blog”的网站,然后,从所选择的网站中选择前N个网站。

[0102] 在如上所述的对评价搜索用网站的第一选择完成后,控制部分10依次访问所选择的网站以便从每一个网站中收集HTML文档,将收集到的HTML文档供给到HTML解析部分33,并控制HTML解析部分33以从每一个HTML文档中提取文章(图5中的步骤S7,以及图8中的步骤S33、S34和S35)。

[0103] 例如,假设收集到的HTML文档之一描述:“<HTML><BODY>Yesterday's Flamboyance of Royal Family wasquiteamusing! </BODY></HTML>”。在这种情况下,作为由HTML解析部分33所解析的结果,从该HTML文档中提取文章“Yesterday's Flamboyance of Royal Family was quite amusing!”。

[0104] 接着,控制部分 10 将所提取的文章中的每一个供给到词素分析引擎 35,并控制词素分析引擎 35 对所提取的文章进行词素分析,从而从其中只提取用于检测搜索关键字的名词(图 5 中的步骤 S8,以及图 8 中的步骤 S36 和 S37)。

[0105] 然后,控制部分 10 识别具有包含如此多以至于“搜索关键字内容”超出预定阈值 $th(\%)$ 的搜索关键字的 HTML 文档的网站,并且选择所识别的网站作为用来搜索评价的网站(第二站点选择)(图 5 中的步骤 S9,以及图 8 中的步骤 S38 和 S39)。在该第二站点选择中选择的网站将在下文中称为“种子站点”。这是因为该站点是下文描述的爬行的起始点。

[0106] 在第二站点选择中,当根据从广播节目内容的标题中是否已经提取了搜索关键字、从节目内容信息中是否已经提取了搜索关键字、或者用户是否已经另外输入了搜索关键字来给每一个搜索关键字分配不同权重时,计算搜索关键字内容。

[0107] 例如,当已经从标题中提取搜索关键字时,给检测到的该搜索关键字分配 1.5 的权重;当已经从节目内容信息中提取搜索关键字时,给检测到的该搜索关键字分配 1.0 的权重;当用户已经另外输入搜索关键字时,给检测到的该搜索关键字分配 0.8 的权重。

[0108] 例如,假设在步骤 S3 已经选择五个关键字,即,从标题中提取的“Flamboyance”和“Royal Family”、从节目内容信息中提取的“Kimura Takuo”和“Minamioji Ginpachi”、以及用户另外输入的“blog”。于是,例如,就获得的文章“A blog for everyone: Yesterday's episode of Flamboyance of Royal Family where Kimura Takuo appeared was amusing!”而言,在步骤 S8 检测到四个搜索关键字,即,“blog”、“Flamboyance”、“Royal Family”和“Kimura Takuo”。

[0109] 这里,如果检测到上述所有的五个搜索关键字,那么得分是 $1.5+1.5+1.0+1.0+0.8=5.8$ 。同时,就上述文章而言,在检测到四个搜索关键字的情况下,得分是 $1.5+1.5+1.0+0.8=4.8$ 。因此,在步骤 S9 中所计算得到的上述文章的搜索关键字内容是 $4.8/5.8 \approx 82.8(\%)$ 。

[0110] 在上述阈值 th 是 $70(\%)$ 的情况下,上述文章的搜索关键字内容超出了阈值 th 。因此,在步骤 S9,选择具有包含上述文章的网页的网站作为种子站点。

[0111] 接着,控制部分 10 指示 HTML 爬行部分 34 对预定数目的级别 执行爬行,其中,以上述方式选择的每一个种子站点(参见图 9 中的步骤 S40)作为起始站点(图 5 中的步骤 S10,以及图 9 中的步骤 S41)。

[0112] HTML 爬行部分 34 首先收集种子站点上的 HTML 文档,对收集到的 HTML 文档进行解析以识别链接,然后收集所识别的链接的目的地上的 HTML 文档,接着对收集到的 HTML 文档进行解析以识别包含在该 HTML 文档中的链接,然后收集所识别的链接的目的地上的 HTML 文档,对收集到的 HTML 文档进行解析等等。HTML 爬行部分 34 对预定数目的级别重复上述操作。

[0113] 请注意,在这种情况下,要执行爬行的级别数可以在开始评价搜索处理之前由用户指定。还应该注意,在开始评价搜索处理之前,可以允许用户指定是否应该在同一域上的链接站点的范围内或者在所有链接站点的范围内执行爬行。

[0114] 接着,控制部分 10 将在图 5 中的步骤 S10 中收集到的 HTML 文档(参见图 9 中的步骤 S42)供给到 HTML 解析部分 33,并控制 HTML 解析部分 33 以对 HTML 文档进行解析,从

而从其中提取文章（图 6 中的步骤 S11，以及图 9 中的步骤 S43）。

[0115] 接着，控制部分 10 控制词素分析引擎 35 和句法分析引擎 36 对在步骤 S11 中提取的文章分别进行词素分析和句法分析（图 6 中的步骤 S12，以及图 9 中的步骤 S44）。此时的词素分析结果将在下文中称为“爬行文章（词素分析结果）”，而此时的句法分析结果将在下文中称为“爬行文章（句法分析结果）”。

[0116] 接着，控制部分 10 将爬行文章（词素分析结果）和爬行文章（句法分析结果）供给到评价确定部分 37，并使用评价规则词典控制评价确定部分 37，以从步骤 S12 的处理结果中提取关于所选择的内容（正搜索其评价）的评估值和评价代表句子（图 6 中的步骤 S13，以及图 9 中的步骤 S46）。

[0117] 评价确定部分 37 首先通过爬行文章（词素分析结果）搜索包含任何评价表达（词素分析结果）的句子。然后，评价确定部分 37 对所找到的这些句子执行以下处理 (1)、(2) 和 (3)。

[0118] (1) 参考爬行文章（句法分析结果），评价确定部分 37 识别与评价表达（词素分析结果）在语法上相关的单词，并且，如果所识别的单词是包含在标题（词素分析结果）或内容信息（词素分析结果）中的单词，那么评价确定部分 37 确定该句子描述关于节目的评价（这类句子在下文中将称为“A 类句子”）。

[0119] (2) 参考爬行文章（句法分析结果），评价确定部分 37 识别与评价表达（词素分析结果）在语法上相关的单词，并且，如果所识别的单词是与包含在标题（词素分析结果）或内容信息（词素分析结果）中的单词相关的单词（包含在标题或内容信息中的单词的相关单词），那么评价确定部分 37 确定该句子描述关于节目的评价（这类句子在下文中将称为“B 类句子”）。

[0120] 这里，术语“相关单词”是指在语法上相关的单词，例如，同义词（例如，“Nippon”对于“Japan”）或者上位词（例如“Asia”对于“Japan”）。参考同义词词典 43 以便确定某些单词是否是相关单词。例如，考虑到与搜索关键字相关的单词可以用于节目评论中的可能性，执行该处理。

[0121] (3) 即使已经从爬行文章（词素分析结果）中作为包含评价表达（词素分析结果）的句子找到的句子不满足上述 (1) 和 (2) 中所提及的条件，那么评价确定部分 37 确定该句子描述关于节目的评价（这类句子在下文中将称为“C 类句子”）。

[0122] 在本实施例中，评价确定部分 37 分别计算肯定的评价值和否定的评价值。如果评估值是正的，评价确定部分 37 参考评价规则词典 44 并结合上述 A、B 和 C 三种类型句子将找到的评价表达的评估值（得分）加到肯定的评价值，如果评估值是负的，评价确定部分 37 将找到的评价表达的评估值（得分）加到否定的评价值。

[0123] 请注意，在本实施例中，预先将不同的加权系数分配给类型 A、B 和 C，并且，评价确定部分 37 将找到的评价表达的评估值（得分）乘以与句子类型相对应的加权系数后加到肯定的评价值或否定的评价值。在这种情况下，这些加权系数保持在评价确定部分 37 中，并且这些加权系数满足下列不等式：类型 A 的加权系数 > 类型 B 的加权系数 > 类型 C 的加权系数。

[0124] 请注意，在评价表达（词素分析结果）与上述 (1) 和 (2) 中识别的单词之间的语法关系伴随有否定表达的情况下，评价表达在意思是相反的，因此，如果评估值（得分）

是正确的,则评价确定部分 37 将评价表达的评估值(得分)加到否定的评价值,如果评估值(得分)是负的,评价确定部分 37 将评价表达的评估值(得分)加到肯定的评价值。

[0125] 此外,在本实施例中,在爬行文章(词素分析结果)中,在任何程度表达(词素分析结果)都存在于评价表达(词素分析结果)之前的情况下,评价确定部分 37 从评价规则词典 44 中认出程度表达的系数,并将评价表达的评估值(得分)乘以该系数后加到肯定的评价值或否定的评价值。

[0126] 此外,在本实施例中,评价确定部分 37 将检测到的上述三种类型的句子存储为“评价代表候补句子”。

[0127] 接着,控制部分 10 提取由评价确定部分 37 计算得到的肯定的评价值和否定的评价值以及从评价代表候补句子中选取的一个或多个句子作为评价结果,将它们转换成显示信息以便在监视设备 2 的显示器上显示,并通过监视器输出接口 24 将显示信息发送到监视设备 2(图 6 中的步骤 S14,以及图 9 中的步骤 S47)。

[0128] 图 10 示出显示的评价结果的实例。评价结果的该实例示出肯定的评价值(Positive)、否定的评价值(Negative)、几个肯定的代表句、几个否定的代表句、作为搜索结果而被找到的网站数目、以及提取的关于评价的句子的数目。

[0129] 有多种方法可用于从存储的评价代表候补句子中选取几个评价代表句子。例如,可以优选选取包含从标题中提取的搜索关键字比从节目内容信息中提取的搜索关键字多的句子。另外,可以优先选取其中在所有的评价代表候补句子中经常出现的单词具有较高频率的句子。另外,可以随机地选取几个评价代表句子。

[0130] 如上所述,根据上述第一实施例,使用搜索关键字通过因特网在两个步骤中从作为搜索结果而被找到的巨大数量的网站中选择有限数量的网站,然后,对该有限数量的网站执行爬行以便检测描述评价的句子,并且从这些检测到的句子中获得评价结果。因此,可以用相对简单的结构构造评价搜索设备。

[0131] 因此,如上述实施例一样,记录和再现设备可以装备有作为一个功能模块的评价搜索部分。因此,并不一定要使用用于确定评价的上述系统引导连接到因特网的个人电脑去搜索每一个节目的评价。与上述实施例一样,记录和再现设备自身能够通过因特网搜索每一个节目的评价和显示该评价。这对用户来说很方便。

[0132] 根据上述第一实施例的信息处理设备是记录和再现设备。但是,应该理解,根据本发明的信息处理设备不局限于记录和再现设备。还应该注意,在上述实施例中,搜索广播节目内容的评价,因此从得自 EPG 信息的信息中提取搜索关键字。但是,应该理解,这只是一个实例,对本发明来说不是必要的。例如,在内容提供服务器在其中存储了内容相关信息例如内容的标题和描述内容的信息的情况下,信息处理设备可以从内容提供服务器获取内容相关信息,以便从其中提取搜索关键字。

[0133] **【信息处理设备的第二实施例】**

[0134] 在上述第一实施例中,记录和再现设备自身形成根据本发明的信息处理设备的一个实施例。但是,可以在因特网或者局域网上提供用于评价搜索的服务器设备。例如,这样的用于评价搜索的服务器设备也可以形成根据本发明的信息处理设备的一个实施例。下文描述的第二实施例就是该情况的一个实例。

[0135] 图 11 是示出包括根据第二实施例的信息处理设备的系统的概要的视图。

[0136] 在第二实施例中,在因特网上提供“节目评价搜索存储服务器设备”7。节目评价搜索存储服务器设备 7 包括与上述记录和再现设备 1 的评价搜索部分 30 和评价搜索用词典 40 相似的部件。假设要对其进行评价搜索请求,节目评价搜索存储服务器设备 7 预先存储了内容的评价结果,并且,响应来自客户端设备例如记录和再现设备 1 的评价搜索请求,节目评价搜索存储服务器设备 7 将存储在其中的对应的评价结果发送到该客户端设备。

[0137] 因此,在第二实施例中,与第一实施例不同,记录和再现设备 1 不包括上文针对第一实施例描述的评价搜索部分 30 和评价搜索用词典 40 二者,如下所述,替换的是,记录和再现设备 1 包括作为作出评价搜索请求的客户端设备的功能模块。

[0138] 图 12 是示出节目评价搜索存储服务器设备 7 的示例性硬件结构的框图。如图 12 所示,根据本实施例的节目评价搜索存储服务器设备 7 包括控制部分 50、评价搜索部分 60、评价搜索用词典 70、评价搜索请求部分 80 和系统总线 101。控制部分 50、评价搜索部分 60、评价搜索用词典 70 和评价搜索请求部分 80 与系统总线 101 相连。

[0139] 与上述控制部分 10 一样,控制部分 50 由微型计算机形成,并包括中央处理单元(CPU)51、用于程序的 ROM52 和用于工作区域的 RAM53。ROM52 存储用于搜索内容评价和获取评价结果的程序、用于请求评价搜索的程序、以及其它程序。CPU51 在使用 RAM53 作为工作区域时执行存储在 ROM52 中的程序,从而控制上述的多种处理。

[0140] 评价搜索部分 60 具有与上述第一实施例中的评价搜索部分 30 相同的结构。在本实施例中,评价搜索部分 60 包括用于连接到因特网 3 的通信接口 61、因特网搜索部分 62、HTML 解析部分 63、HTML 爬行部分 64、词素分析引擎 65、句法分析引擎 66 和评价确定部分 67,每一个上述部分都与系统总线 101 相连。在第二实施例中,用来输入附加的搜索关键字的操作部分 68 和操作部分接口 69 形成评价搜索部分 60 的一部分。

[0141] 这里,因特网搜索部分 62、HTML 解析部分 63、HTML 爬行部分 64、词素分析引擎 65、句法分析引擎 66 和评价确定部分 67 可以在硬件中实现,或者可以通过软件应用程序实现。请注意,在它们通过软件应用程序实现的情况下,该软件应用程序预先存储在 ROM52 中。

[0142] 评价搜索用词典 70 与上述第一实施例中的评价搜索用词典 40 相同。在本实施例中,评价搜索用词典 70 包括词素分析词典 71、句法分析词典 72、同义词词典 73 和评价规则词典 74,其中的每一个都与系统总线 101 相连。

[0143] 涉及使用评价搜索部分 60 和评价搜索用词典 70 的评价搜索处理与上述第一实施例中的评价搜索处理相同。但是,服务器设备 7 对包含在 EPG 信息中的所有广播节目内容连续地执行预先在第一实施例中描述的评价搜索。服务器设备 7 还将作为该处理的结果而获得的评价结果(其包括评价评估值、评价代表句子等)存储在评价搜索请求部分 80 中的评价结果存储部分 81 中,从而与对应的广播节目内容相关联。

[0144] 评价搜索请求部分 80 包括评价结果存储部分 81、“评价结果获取请求接收部分”82 和“输出评价结果生成部分”83。“评价结果获取请求接收部分”82 通过因特网 3 从客户端设备接收评价结果获取请求,并且分析该评价结果获取请求,从而确定其所指的是哪个广播节目内容。在本实施例中,为了将评价结果获取请求所指的是哪个节目内容通知服务器设备 7,来自客户端设备的评价结果获取请求包括用于识别要搜索其评价的广播节目内容的“广播节目内容识别信息”。

[0145] 广播节目内容识别信息的实例包括:最初唯一地分配给每一个广播节目内容的

识别信息；每一个广播节目内容的标题；以及分配给每一个广播节目内容的 Gemstar 码（G 码）。

[0146] 在评价结果存储部分 81 中，每一个广播节目内容的评价结果都被存储，以便与对应的广播节目内容识别信息相关联。

[0147] 一旦通过通信接口 61 接收到评价结果获取请求，评价结果获取请求接收部分 82 从评价结果获取请求提取广播节目内容（要搜索其评价）的识别信息，并将提取到的广播节目内容的识别信息传送到输出评价结果生成部分 83。输出评价结果生成部分 83 接收广播节目内容的识别信息，并从评价结果存储部分 81 获取存储在评价结果存储部分 81 中的评价结果，以便与接收到的广播节目内容的识别信息相关联。然后，输出评价结果生成部分 83 生成要发送到作出了评价结果获取请求的客户端设备的“评价结果发送消息”。然后，输出评价结果生成部分 83 将生成的评价结果发送消息通过通信接口 61 和因特网 3 发送到客户端设备。

[0148] 在评价结果存储部分 81 没有存储与包含在评价结果获取请求中的广播节目内容（要搜索其评价）的识别信息相对应的评价结果的情况下，服务器设备 7 对该广播节目内容执行评价搜索处理，并将作为评价搜索处理结果而获得的评价结果发送到客户端设备。此时获得的评价结果也存储在评价结果存储部分 81。

[0149] 图 13 示出记录和再现设备 1 的示例性硬件结构，其是根据第二实施例的客户端设备的实例。

[0150] 如上所述，根据第二实施例的记录和再现设备 1 不包括上文针对第一实施例描述的评价搜索部分 30 和评价搜索用词典 40 二者，替换的是，记录和再现设备 1 包括评价结果获取请求部分 90，该评价结果获取请求部分 90 是作为作出评价搜索请求的客户端设备的功能模块。

[0151] 在本实施例中，评价结果获取请求部分 90 包括评价结果获取请求生成部分 91、评价结果接收部分 92 和评价结果显示信息生成部分 93。例如，当控制部分 10 接收到对从记录内容的显示列表中选择的评价进行搜索的指令时，控制部分 10 激活评价结果获取请求部分 90。

[0152] 当评价结果获取请求部分 90 被激活时，首先，在允许选择的记录内容的识别信息（例如标题信息）被包含在评价搜索请求中后，评价结果获取请求生成部分 91 将对于从记录内容的显示列表中选择的评价搜索请求通过通信接口 31 和因特网 3 发送到节目评价搜索存储服务器设备 7。请注意，因特网 3 上的节目评价搜索存储服务器设备 7 的 URL 预先已经登记和存储在记录和再现设备 1 中。

[0153] 如上所述，响应对于该记录内容的评价搜索请求，服务器设备 7 将该记录内容的评价结果发送到记录和再现设备 1。因此，评价结果接收部分 92 通过通信接口 31 接收评价结果，并将接收到的评价结果传送到评价结果显示信息生成部分 93。

[0154] 评价结果显示信息生成部分 93 接收评价结果并由其生成关于上述评价结果的显示信息，并将生成的显示信息通过监视器输出接口 24 供给到监视设备 2。

[0155] 这样，如图 10 中所示的记录内容的评价结果显示在监视设备 2 的显示器上，从而将评价结果呈现给用户。

[0156] 在第二实施例中，服务器设备 7 预先搜索广播节目内容的评价结果并将评价结果

存储在评价结果存储部分 81 中。因此,一旦从客户端设备接收到评价结果获取请求,作为对该获取请求的响应,服务器设备 7 就能够立即将存储在评价结果存储部分 81 中的评价结果发送到客户端设备。因此,用户能够快速看到记录内容的评价结果。

[0157] 请注意,根据第二实施例的服务器设备 7 不需要通过因特网连接客户端设备。或者,虽然服务器设备 7 连接到因特网,但是服务器设备 7 可以通过 LAN 连接客户端设备。

[0158] 【其它实施例或变形例】

[0159] 在上述实施例中,搜索记录内容的评价。但是,例如,也可以搜索当前正在广播的广播节目内容的评价。例如,当电视广播节目接收设备被配置为允许用户基于 EPG 从节目时间表中指定广播节目以便搜索其评价时,实现该搜索。

[0160] 在这种电视广播节目接收设备应用于第一实施例的情况下,该电视广播节目接收设备被配置为从 EPG 数据中提取要搜索其评价的广播节目的标题和相关信息,然后从其中提取搜索关键字,并且使用提取的搜索关键字执行上述评价搜索操作。

[0161] 请注意,不仅可以在选择和接收实时广播节目时应用本发明,而且也可以在接收和再现通过视频点播 (Video on Demand) 提供的内容时应用本发明。本发明也可以应用于这种情况:广播节目首先存储在服务器上,用户从存储在服务器上的广播节目中选择期望的广播节目以接收它,如服务器类型 (server-type) 的广播节目的情况一样。

[0162] 在前面的描述中,假定评价搜索的对象是广播节目内容,但是评价搜索的对象并不局限于广播节目内容。本发明也可以应用于搜索多种其它类型的内容例如音乐内容、电子书内容和视频游戏内容的评价。

[0163] 从第二实施例的描述中显然可知,根据本发明的信息处理设备不是必须对搜索其评价的内容进行再现或解码,而是仅仅需要具有能够搜索内容评价并将评价结果呈现给用户的能力即可。

[0164] 本领域技术人员应该理解,根据设计要求和其它因素,可以进行各种修改、组合、子组合和替换,只要它们在所附权利要求或其等同形式的范围内即可。

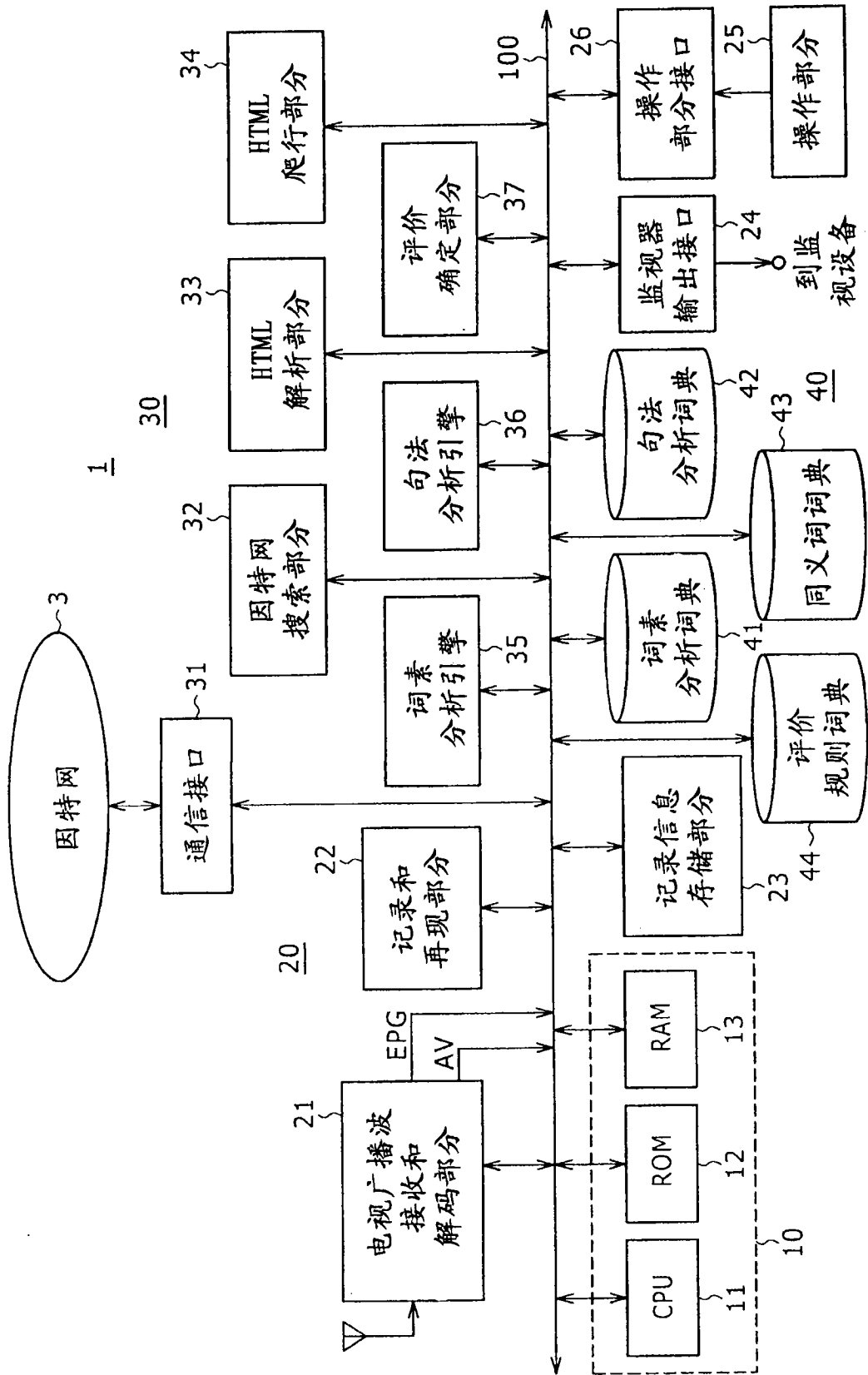


图1

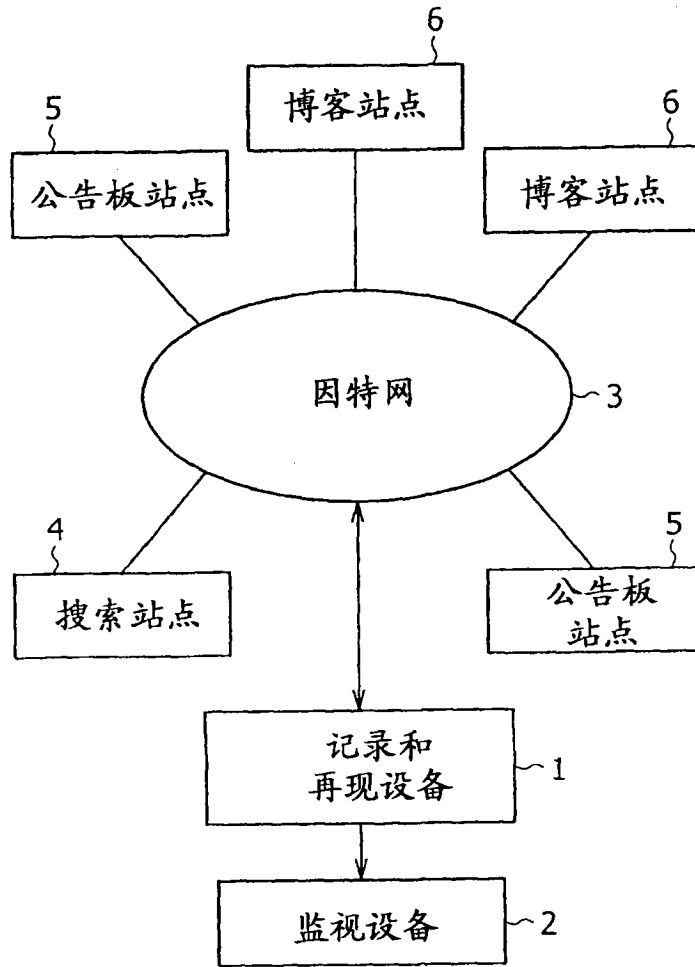


图 2

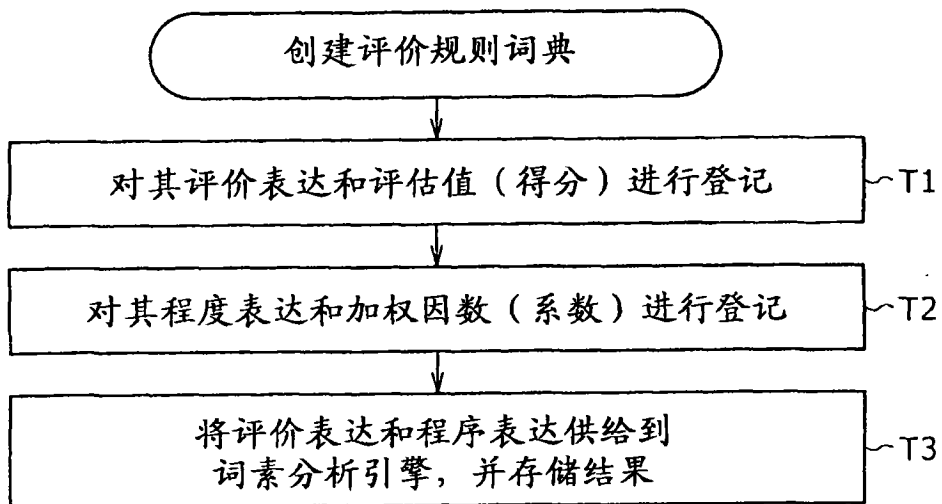


图 3

评价规则词典

类型	表达	得分/系数	词素分析结果
评价表达	有趣的	+5点	xxx@xxx@xxx
评价表达	感动的	+5点	xxx@xxx@xxx
评价表达	好的	+2点	xxx@xxx@xxx
评价表达	枯燥的	-5点	xxx@xxx@xxx
评价表达	不好的	-2点	xxx@xxx@xxx
评价表达	非常	×2	xxx@xxx@xxx
评价表达	相当	×3	xxx@xxx@xxx

图 4

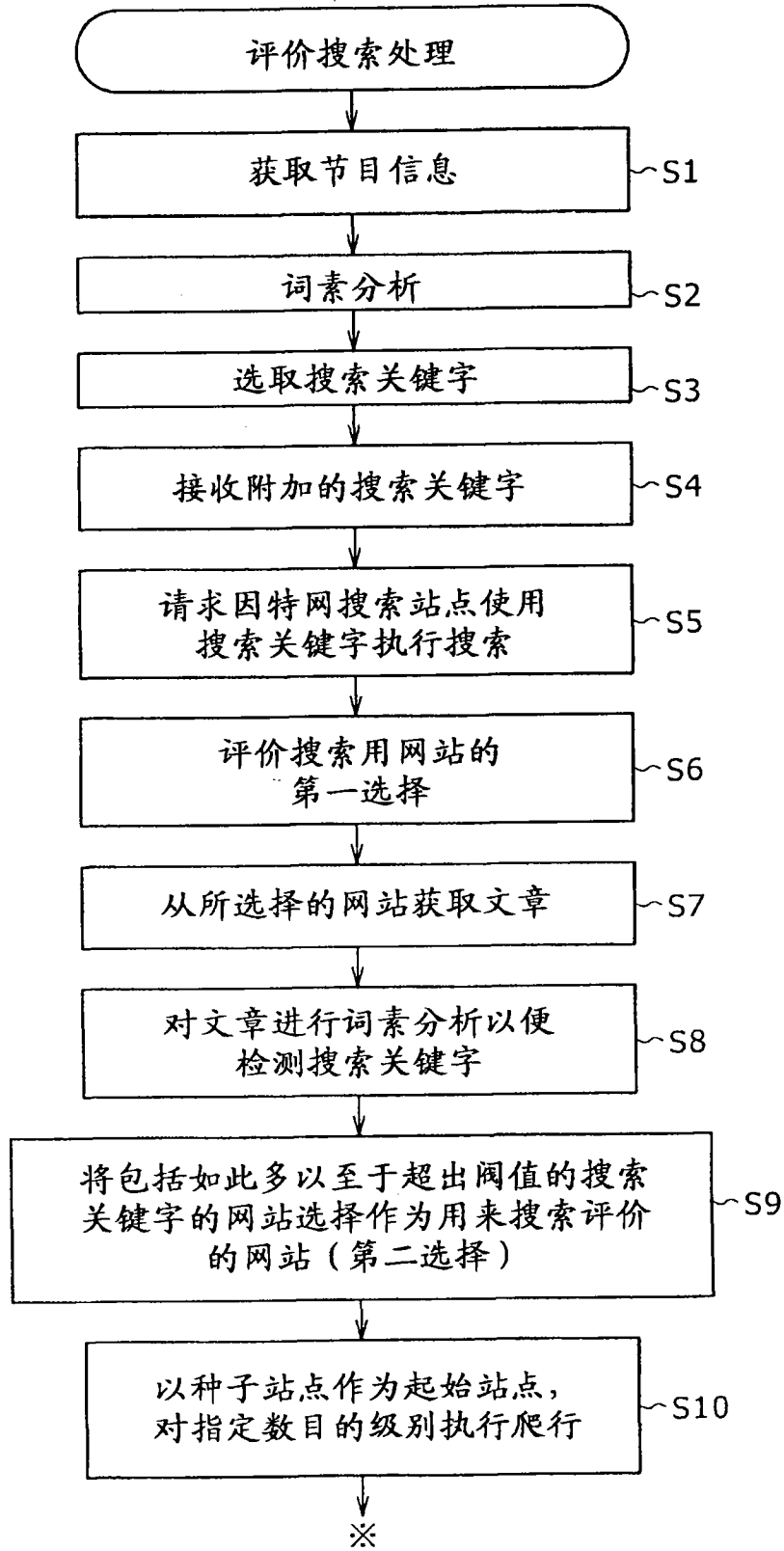


图 5

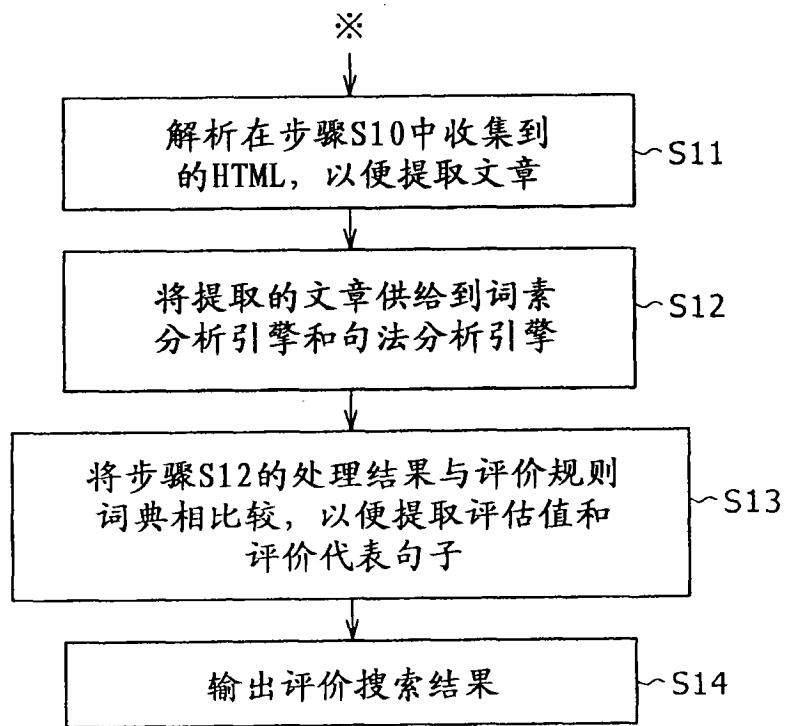


图6

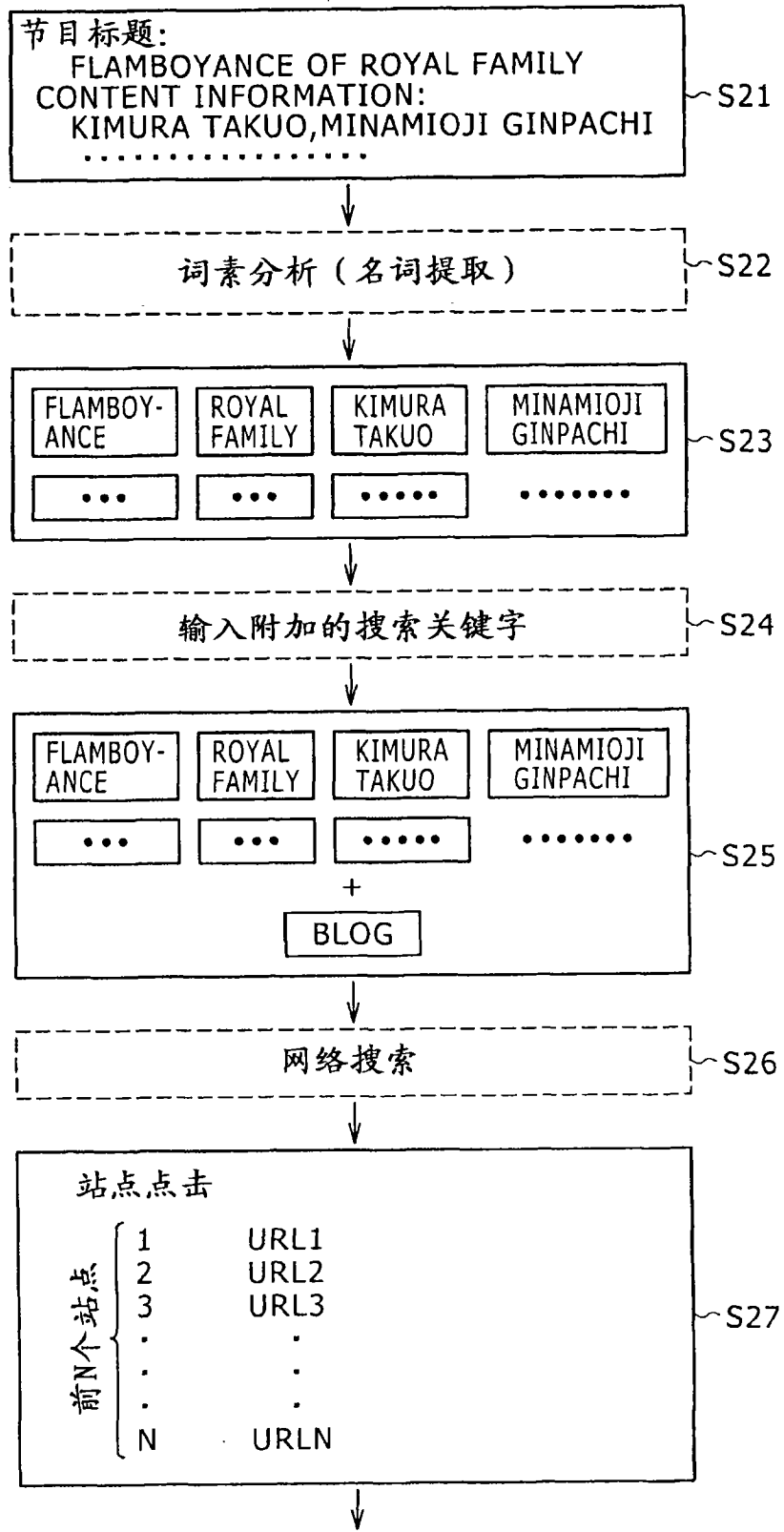


图 7

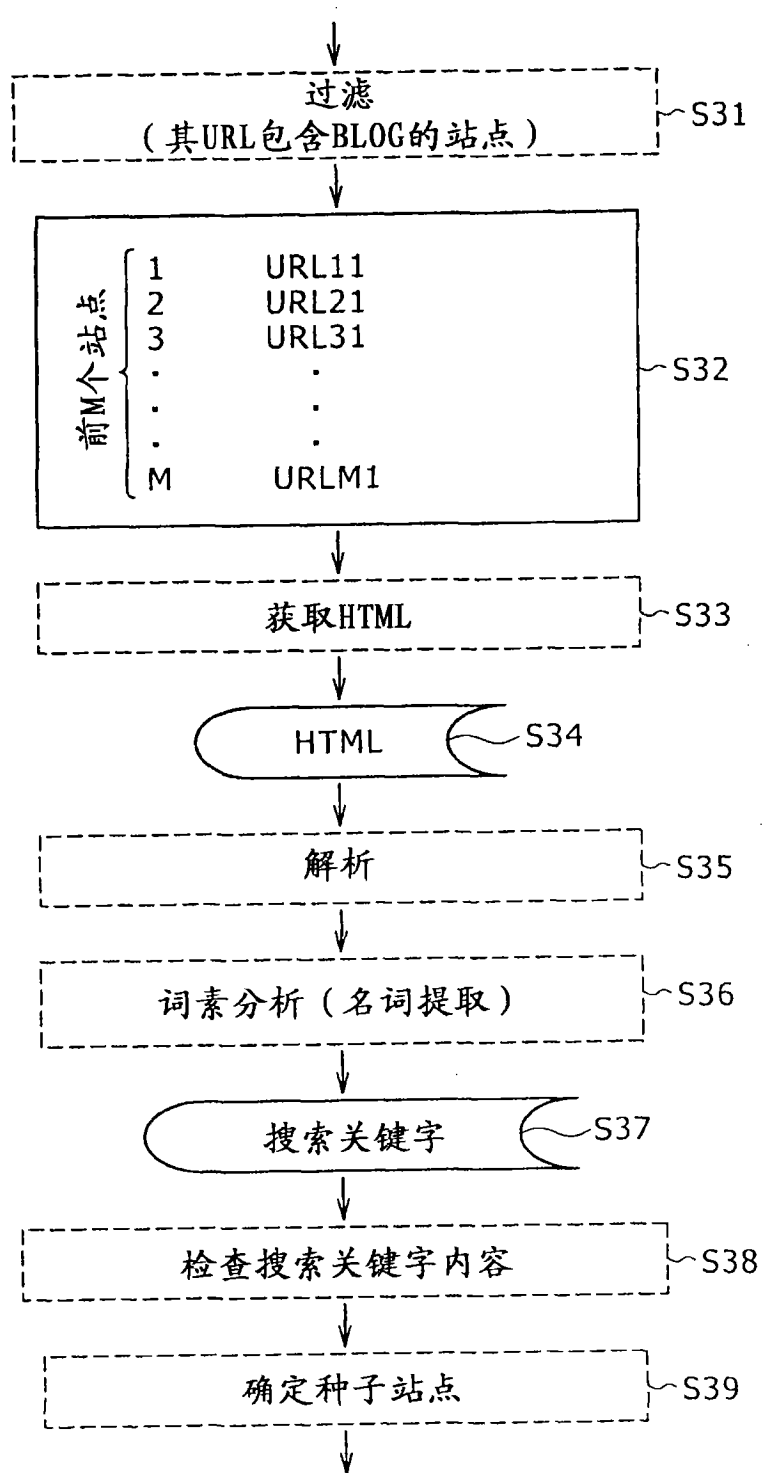


图 8

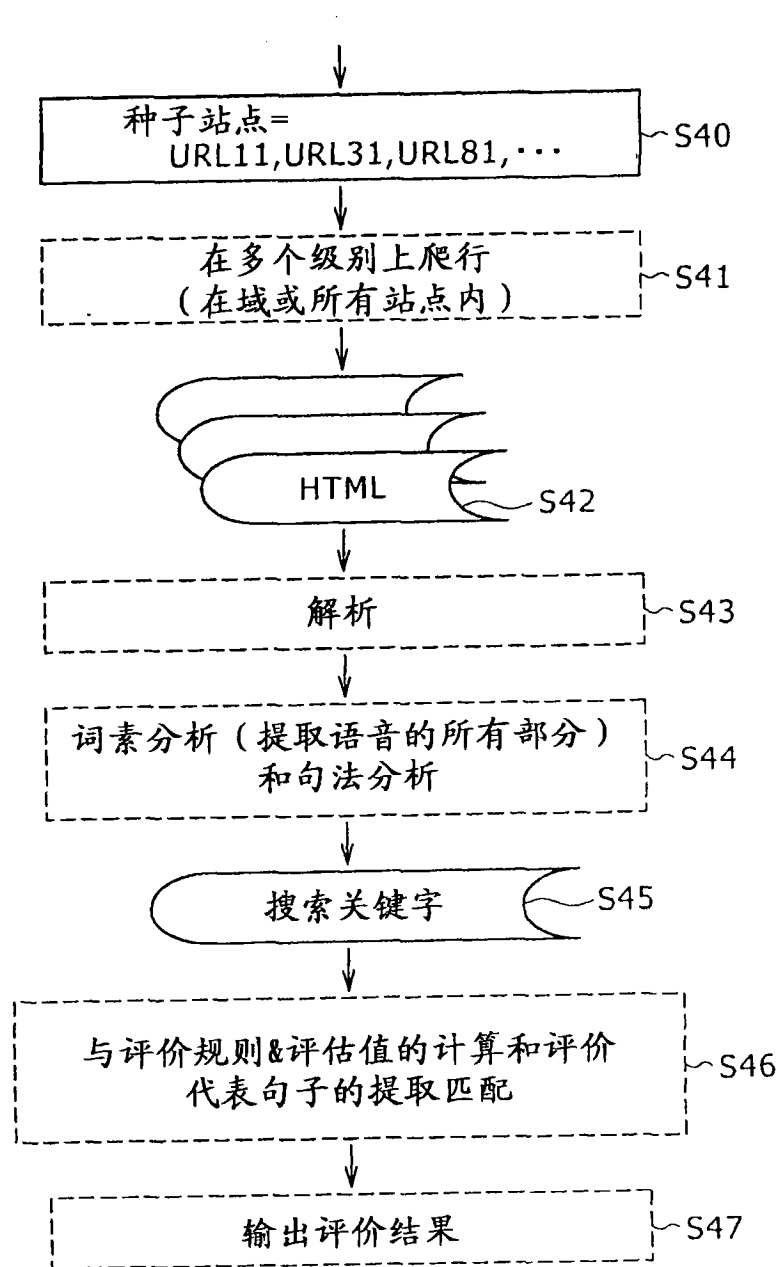


图 9

肯定的 = 80 个点
IT WAS VERY AMUSING.
I'M SURE TO WATCH NEXT EPISODE.
.....
否定的 = 10 个点
I WAS DISAPPOINTED BY SLOWNESS
OF SCENE DEVELOPMENT
.....
找到的站点数 = 2800000 个站点
评论数 = 1530 条评论

图 10

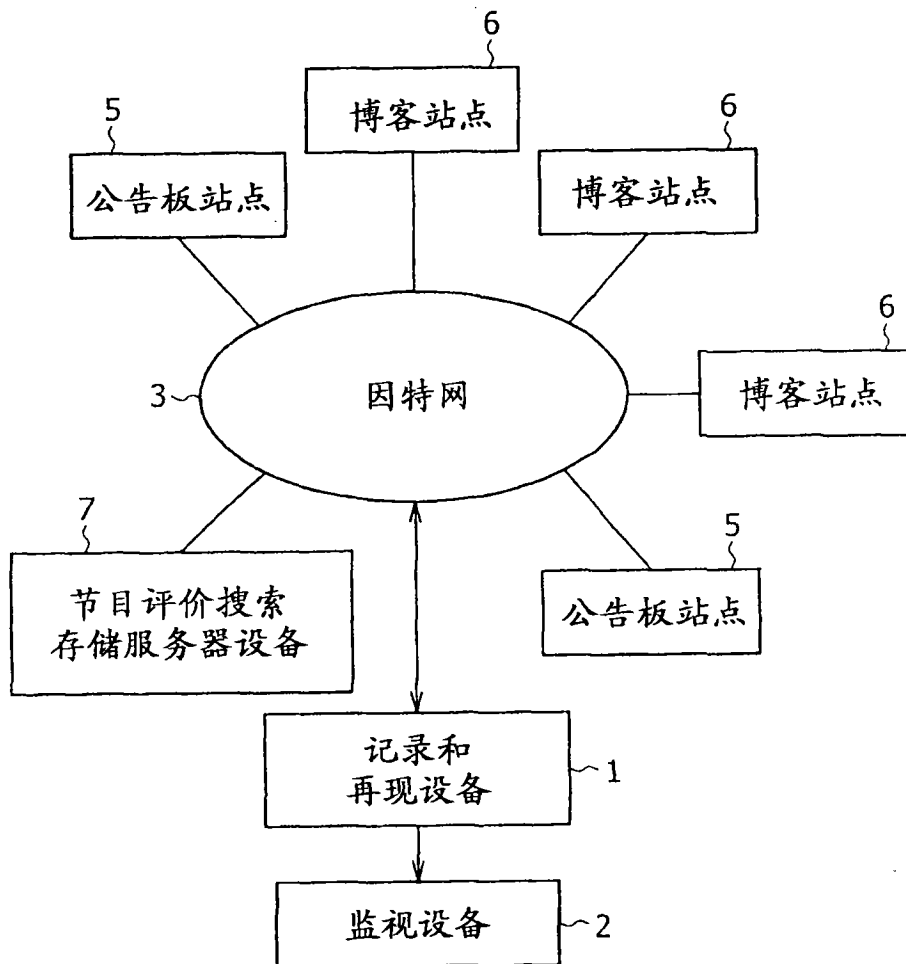


图 11

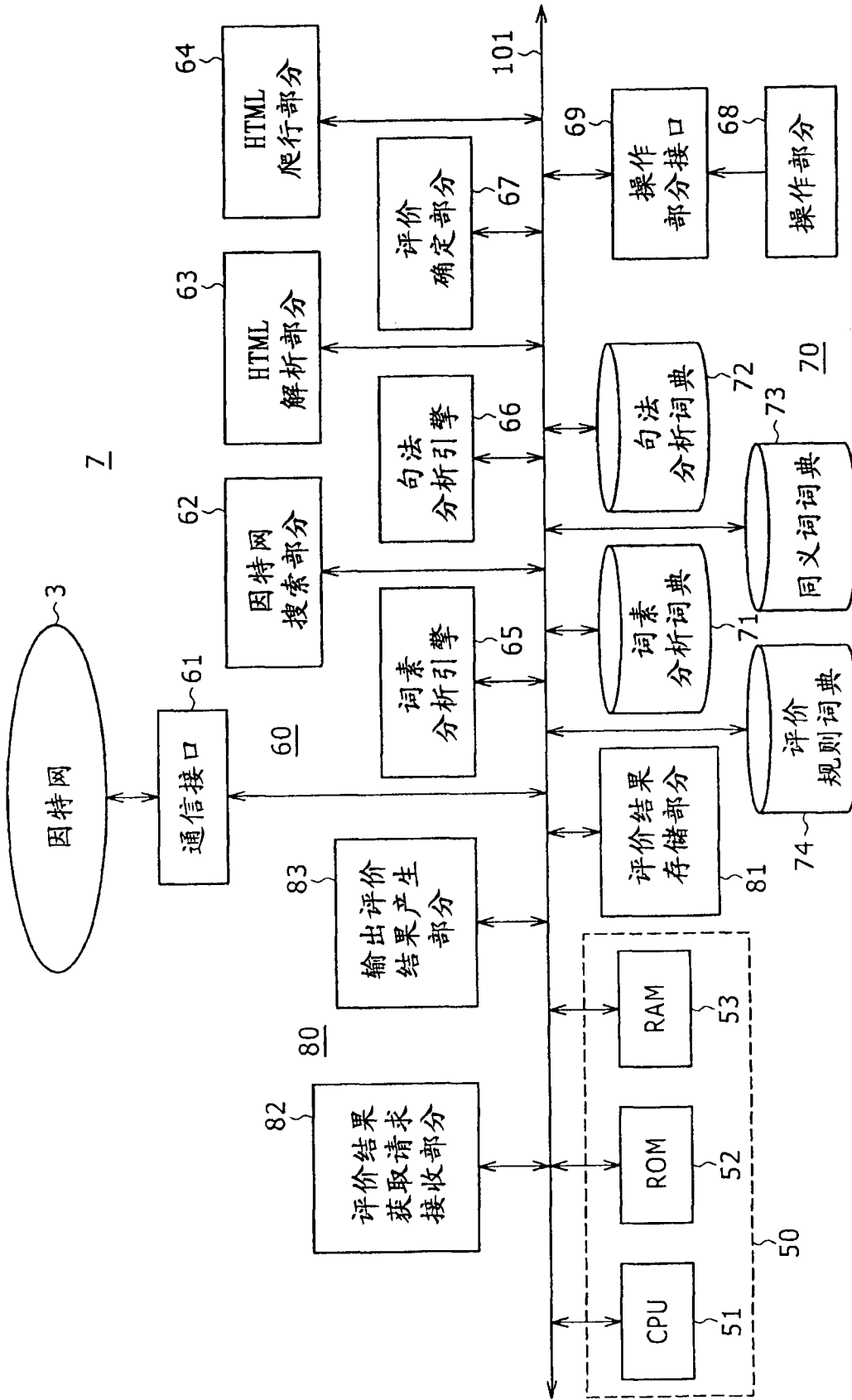


图12

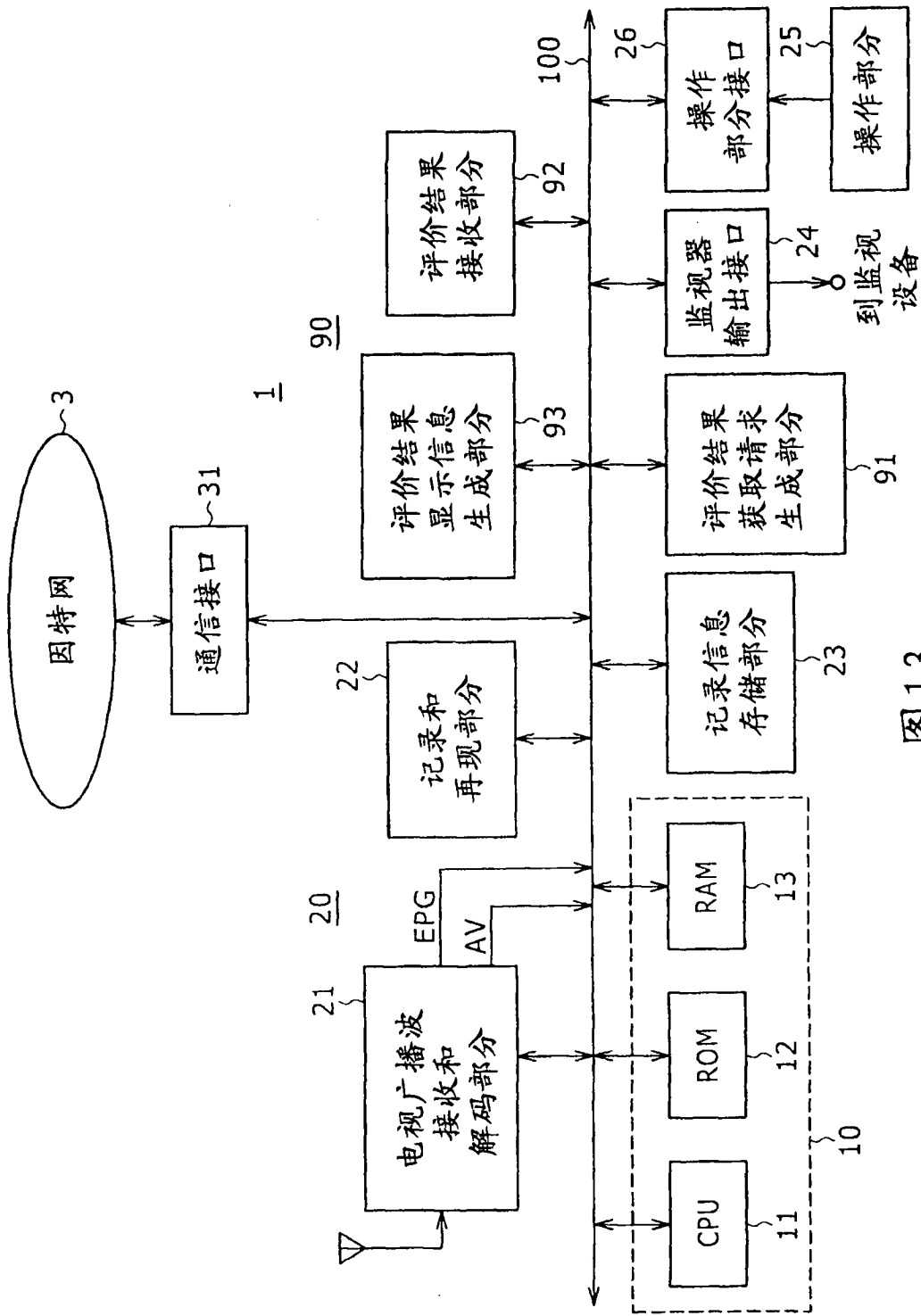


图13