

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 17/30 (2006.01)



## [12] 发明专利申请公布说明书

[21] 申请号 200580011783.2

[43] 公开日 2007 年 4 月 4 日

[11] 公开号 CN 1942877A

[22] 申请日 2005.3.31

[21] 申请号 200580011783.2

[30] 优先权

[32] 2004.3.31 [33] JP [31] 107139/2004

[86] 国际申请 PCT/JP2005/006317 2005.3.31

[87] 国际公布 WO2005/096182 日 2005.10.13

[85] 进入国家阶段日期 2006.10.19

[71] 申请人 松下电器产业株式会社

地址 日本大阪府

[72] 发明人 野本昌子

[74] 专利代理机构 北京市柳沈律师事务所

代理人 邵亚丽 李晓舒

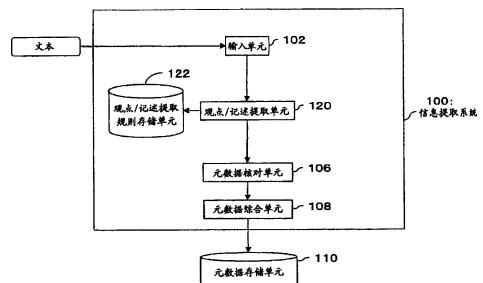
权利要求书 3 页 说明书 26 页 附图 29 页

[54] 发明名称

信息提取系统

[57] 摘要

将与文本中表现的事物有关的事实和意见等记述内容，使事实和意见的观点和记述附带对应地进行提取。为此，设为具有以下结构：输入单元，输入文本；观点/记述提取规则存储单元，存储用于特定被记述在文本中的表现的观点和与该观点有关的记述的组的观点/记述提取规则；观点/记述提取单元，从文本中的字符串的句法的属性或者含义的属性的至少一个属性中，利用所述观点/记述提取规则，将观点及其记述的组附带对应，提取作为赋予了用于识别它们的识别信息的要素元数据；以及元数据存储单元，存储所述观点/记述提取单元提取的要素元数据，将与文本中表现的事物有关的事实和意见等记述内容构成作为观点和记述的组，并且整理为事实和意见附带对应的形式来提取。



1、一种信息提取系统，包括：

输入单元，输入文本；

观点/记述提取规则存储单元，存储用于特定被记述在文本中的表现的观点和与该观点有关的记述的组的观点/记述提取规则；

观点/记述提取单元，从对被输入所述输入单元中的文本中的字符串赋予的句法的属性或者含义的属性的至少一个属性中，利用所述观点/记述提取规则，将观点及其记述的组提取作为附带对应的要素元数据；以及

元数据存储单元，存储所述观点/记述提取单元提取的要素元数据。

2、如权利要求 1 所述的信息提取系统，

所述句法的属性至少包括字符串列表或者词类分类的其中一个。

3、如权利要求 1 所述的信息提取系统，

所述含义的属性至少包含含义分类。

4、如权利要求 1 所述的信息提取系统，

所述观点/记述提取单元在将附带对应的观点和记述的组作为要素元数据提取时，赋予用于识别附带对应的观点和记述的组的识别信息（要素元数据 ID）来提取。

5、如权利要求 1 所述的信息提取系统，还包括：

属性赋予单元，从文本提取任意的字符串，并至少利用用于特定字符串的含义分类的含义属性赋予规则，输出对每个字符串赋予了含义属性的附带含义的属性文本。

6、如权利要求 1 所述的信息提取系统，

所述观点/记述提取单元在文本中不表现观点，仅表现记述的情况下，将记述的含义的属性作为观点，提取观点和记述的组。

7、如权利要求 1 所述的信息提取系统，还包括：

元数据核对单元，分别核对所述观点/记述提取单元提取的要素元数据的观点之间和记述之间，并且推定关联性；以及

元数据综合单元，根据所述被推定的关联性，综合具有关联性的要素元数据，并且输出综合元数据。

8、如权利要求 6 所述的信息提取系统，

所述元数据核对单元在核对所述观点/记述提取单元提取的要素元数据的观点和记述时，至少利用构成观点、记述的字符串的含义的属性来进行核对。

9、如权利要求7所述的信息提取系统，还包括：

话题事物推定单元，对于由所述观点/记述提取单元提取的要素元数据，利用用于推定话题的事物的话题事物推定规则，推定被作为话题的事物，

所述元数据存储单元将要素元数据与由所述话题事物推定单元推定的话题的事物一起对应存储。

10、如权利要求9所述的信息提取系统，

所述话题事物推定单元在推定要素元数据的话题的事物时，

从被存储在所述元数据存储单元中的要素元数据的观点/记述和含义的属性推定话题的事物。

11、如权利要求9或10所述的信息提取系统，

所述元数据核对单元在进行观点和记述的核对时，以通过所述话题事物推定单元推定的话题事物单位进行所述观点和记述的核对。

12、如权利要求1所述的信息提取系统，

所述观点/记述提取规则包含用于特定作为与文本的作者有关的信息的用户信息的规则的用户观点/记述提取规则，

所述观点/记述提取单元利用所述用户观点/记述提取规则，提取作为与用户信息有关的要素元数据的用户元数据。

13、如权利要求1所述的信息提取系统，

所述观点/记述提取规则包含用于特定作为与文本的书目事项有关的信息的来源信息的规则的来源观点/记述提取规则，

所述观点/记述提取单元利用所述来源观点/记述提取规则，提取作为与来源信息有关的要素元数据的来源元数据。

14、如权利要求9所述的信息提取系统，

所述元数据核对单元还包括：

客观性/可靠性判定单元，至少利用要素元数据、或者用户元数据、或者来源元数据中的一种以上，判定观点、记述的客观性/可靠性；以及

客观性/可靠性判定规则存储单元，存储用于判定所述观点、记述的客观性/可靠性的客观性/可靠性判定规则。

15、如权利要求9所述的信息提取系统，

所述话题事物推定单元在推定要素元数据的话题的事物时，除了被存储在所述元数据存储单元中的要素元数据，还利用来源元数据或者用户元数据的至少其中一个来推定话题的事物。

16、如权利要求14所述的信息提取系统，还包括：

元数据输出形式生成单元，将所述元数据整理为表形式从而生成元数据表；以及

元数据输出单元，对用户提示生成的元数据表。

17、如权利要求16所述的信息提取系统，

还包括：处理来自用户的请求的用户请求处理单元，

所述元数据输出形式生成单元利用与通过所述用户请求处理单元输入的用户请求一致的元数据来生成元数据表。

18、一种信息提取方法，包括：

输入文本的步骤；

参照用于特定被记述在文本中的表现的观点和与该观点有关的记述的组的观点/记述提取规则的步骤；以及

从对被输入所述输入单元中的文本中的字符串赋予的句法的属性或者含义的属性的至少一个属性中，利用所述观点/记述提取规则，将观点及其记述的组提取作为附带对应的要素元数据的步骤。

## 信息提取系统

### 技术领域

本发明涉及从文本提取与事物有关而被书写的事实和意见等记述内容的信息提取系统。

### 背景技术

作为以往的信息提取系统，已知从文本提取关键词（keyword）的系统、提取固有名和数值表现等的系统、提取与 5W1H 等事实有关的信息的系统、以及提取意见和评判等的系统。狭义的信息提取如非专利文献 1 中被介绍的那样，在提取文本的中心的信息的系统中，典型的是以特定的领域的文本为对象来准备应提取的信息的模板（template）（或者构架（frame）），提取该信息。另一方面，近年来进行着要提取文本中的意见和评判的研究。例如，专利文献 1 是从文件集合中提取与利用者指定的事物有关的意见的文献。

专利文献 1：特开 2003-203136 号公报

非专利文献 1：长尾他著“自然语言处理”岩波书店 (pp. 438-441, 1996)

但是，虽然在专利文献 1 那样以往的意见信息提取系统中可以提取与事物有关的意见，但是存在不能将与事物有关而被书写的事实或意见的观点与记述对应而提取的课题。

### 发明内容

本发明是鉴于上述问题而完成的，其第 1 目的是提供一种信息提取系统，将事实和意见的观点与记述附带对应地提取与在文本中表现的事物有关的事实和意见等记述内容。

本发明的第 2 目的是提供一种信息提取系统，在提取所述事实和意见等记述内容时，能将其整理为容易进行事实和意见的对应和关联性的比较的形式来进行提取。

为了解决上述课题，本发明的信息提取系统设为以下结构：即包括：输入单元，输入文本；观点/记述提取规则存储单元，存储用于特定被记述在文

本中的表现的观点和与该观点有关的记述的组的观点/记述提取规则；观点/记述提取单元，从对被输入所述输入单元中的文本中的字符串赋予的句法的属性或者含义的属性的至少一个属性中，利用所述观点/记述提取规则，将观点及其记述的组提取作为附带对应的要素元数据；以及元数据存储单元，存储所述观点/记述提取单元提取的要素元数据。

按照该结构，将与文本中表现的事物有关的事实和意见等记述内容作为观点和记述的组来构成，并且可以将事实和意见附带对应来提取。进而，可以整理为对于在之后的处理中提取的事实和意见，容易进行关联性的比较的形式。

如以上说明的那样，本发明的信息提取系统具有以下效果：通过利用被记述在文本中的表现的观点和与该观点有关的记述的组的观点/记述提取规则来将观点及其记述的组附带对应地提取，可以将与在文本中表现的事物有关的事实和意见的记述内容作为观点和记述的组来对应提取。

本发明的上述目的和优点可以通过参照附图进行说明的以下的实施方式来进一步明白。

#### 附图说明

图1是表示本发明的实施方式1的信息提取系统的结构的方框图。

图2(a)～图2(c)是表示实施方式1的信息提取系统中的到从文本提取要素元数据为止的一连串的处理的流程的说明图。

图3(a)～图3(b)是表示实施方式1的信息提取系统中的观点/记述提取规则和规则的结构要素定义的例子的图。

图4是表示实施方式1的信息提取系统中的综合元数据的例子的图。

图5是表示本发明的实施方式2的信息提取系统的结构的方框图。

图6(a)～图6(b)是表示实施方式2的信息提取系统中的、被输入的文本和被赋予了含义属性的文本的例子的图。

图7(a)～图7(b)是表示实施方式2的信息提取系统中的、含义属性赋予规则的例子和含义属性赋予规则构成要素定义的例子的图。

图8(a)～图8(b)是表示实施方式2的信息提取系统中的、附带含义属性文本的例子和观点/记述认定例的图。

图9(a)～图9(b)是表示实施方式2的信息提取系统中的、观点/记

述提取规则和规则的构成要素定义的例子的图。

图 10 是表示实施方式 2 的信息提取系统中的、要素元数据提取结果的例子的图。

图 11 是表示实施方式 2 的信息提取系统中的、综合元数据的例子的图。

图 12 是本发明的实施方式 3 的信息提取系统的结构的方框图。

图 13 (a) ~ 图 13 (b) 是表示实施方式 3 的信息提取系统中的观点/记述的认定结果和要素元数据的提取结果的图。

图 14 (a) ~ 图 14 (b) 是表示实施方式 3 的信息提取系统中的话题事物推定规则和话题事物推定规则结构要素定义的例子的图。

图 15 是表示实施方式 3 的信息提取系统中的推定的话题事物的例子的图。

图 16 是表示实施方式 3 的信息提取系统中的综合元数据的例子的图。

图 17 是表示实施方式 3 的信息提取系统中的元数据输出形式的例子的图。

图 18 是表示本发明的实施方式 4 的信息提取系统的结构的方框图。

图 19 (a) ~ 图 19 (d) 是表示实施方式 4 的信息提取系统的、表示文本的来源信息、用户信息的例子和附带含义属性的来源消息、附带含义属性的用户信息的例子的图。

图 20 (a) ~ 图 20 (b) 是表示实施方式 4 的信息提取系统的来源信息含义属性赋予规则、用户含义属性赋予规则的例子的图。

图 21 (a) ~ 图 21 (b) 是表示实施方式 4 的信息提取系统的来源观点/记述提取规则、用户观点/记述提取规则的例子的图。

图 22 (a) ~ 图 22 (b) 是表示实施方式 4 的信息提取系统的来源元数据提取结果、用户元数据提取结果的例子的图。

图 23 是表示实施方式 4 的信息提取系统的客观性/可靠性判定规则和可靠性/可靠性判定规则结构要素定义的例子的图。

图 24 (a) ~ 图 24 (b) 是表示实施方式 4 的信息提取系统的文本的例子和附带含义属性文本的例子的图。

图 25 (a) ~ 图 25 (b) 是表示实施方式 4 的信息提取系统的观点/记述提取规则例和观点/记述结构要素定义例的图。

图 26 是表示实施方式 4 的信息提取系统的要素元数据提取结果的例子的

图。

图 27 是表示实施方式 4 的信息提取系统的客观性/可靠性判定结果的例子的图。

图 28 是表示实施方式 4 的信息提取系统的元数据综合结果的例子的图。

图 29 是表示实施方式 4 的信息提取系统的元数据输出形式的例子的图。

#### 标号说明

100、200、300、400 信息提取系统

102 输入单元

106 元数据核对单元

108 元数据综合单元

110 元数据存储单元

120 观点/记述提取单元

122 观点/记述提取规则存储单元

202 属性赋予单元

204 含义属性赋予规则存储单元

206 附带含义属性文本存储单元

302 用户请求处理单元

304 元数据输出形式生成单元

306 元数据输出单元

310 话题事物推定单元

312 话题事物推定规则存储单元

412 客观性/可靠性判定单元

414 客观性/可靠性判定规则存储单元

#### 具体实施方式

以下，参照附图详细地说明本发明的实施方式。

##### (实施方式 1)

图 1 是表示本发明的实施方式 1 的信息提取系统的结构的方框图。本实施方式的信息提取系统 100 是用于将与被输入的文本中表现的事物有关的事实和意见等记述内容构成为观点和记述的组，并将其整理为容易进行事实和意见的附带对应和关联性的比较的形式来提取的系统。信息提取系统 100 包

括：输入文本的输入单元 102；对用于特定文本中记述的表现的观点和与该观点有关的记述的组的观点/记述规则进行存储的观点/记述提取规则存储单元 122；根据前述文本中的字符串的句法的属性，利用前述观点/记述提取规则将观点及其记述附带对应，作为赋予了用于识别它们的识别信息的要素元数据来提取的观点/记述提取单元 120；分别核对观点/记述提取单元 120 提取的要素元数据的观点之间、记述之间，并且推定要素元数据的关联性的元数据核对单元 106；根据前述被推定出的关联性来综合具有关联性的要素元数据的元数据综合单元 108；存储作为由元数据综合单元 108 综合的要素元数据的综合元数据的元数据存储单元 110。

而且，信息提取系统 100 的硬件结构是任意的，不被特别限制。例如，信息提取系统 100 可以通过具有 CPU 和存储装置（ROM、RAM、硬盘及其它各种存储介质）的计算机来实现。这样，在信息提取系统 100 通过计算机实现的情况下，通过计算机执行记述了该信息提取系统 100 的动作的程序来进行规定的动作。

在该信息提取系统 100 中，首先，接受由输入单元 102 输入的文本。在观点/记述提取规则存储单元 122 中，存储有用于特定被写在该文本中的表现的规定和与该观点有关的记述的组的观点/记述提取规则。观点/记述提取单元 120 参照被存储在观点/记述提取规则存储单元 122 中的观点/记述提取规则，根据前述文本中的字符串的句法的属性，使与事物有关而被记述的内容作为观点及其记述的组来附带对应。接着，作为对被附带对应的观点及其记述的组赋予了用于识别它们的识别信息的要素元数据 ID 的要素元数据来进行提取。然后，元数据核对单元 106 分别比较、核对被提取出的要素元数据的观点间、记述间，推定关联性。进而，元数据综合单元 108 根据元数据核对单元 106 推定的关联性，综合具有关联性的要素元数据，并将其作为综合元数据存储在元数据存储单元 110 中。

这里，所谓元数据，一般是表示与目录（contents）的内容和书目事项等目录有关的信息的数据。在本发明中，将与在文本中表现的事物有关的事实和意见等目录的内容有关的记述内容作为观点和记述的组来构成的结构事物元数据的基本单位，特别称为要素元数据。上述所谓事实和意见的语言中的“事实”，意味着不管谁见到都可客观地认定为相同的情况，例如指事物的名称（包括固有名称）和日期时间、或者数量这样的情况。所谓“意见”，意味

着对于各个事物，各人怎样考虑，或者感想，或者评价这样的见解，例如指重、轻、热、不充分这样的情况。所谓“观点”，意味着与事物有关的事实和意见，着眼于事物的哪样的点，或者从哪样的观点叙述这样的情况。而且，所谓“记述”，意味着从上述观点具体用哪样的表现方式在文本中被表达。但是，有时构成要素元数据的观点和记述在文本中仅表现其中一个。而且，在对于一个观点存在多种记述的情况下，对一个观点提取多个记述。而且，在要素元数据中，不仅观点和记述的组，也可以还包含它们的属性和话题等关联信息。而且，在多个要素元数据的观点和记述以及它们的关联信息中，将综合了关联的内容的要素元数据称为综合元数据。

要素元数据假设被赋予作为识别信息的要素元数据 ID。要素元数据 ID 是为了识别出现了要素元数据的文本，以及各个要素元数据而被赋予各个要素元数据的要素元数据的识别信息。而且，所谓字符串的句法的属性，是与字符串的句子结构的功能有关的属性，至少由词类分类信息、或者与字符串列表有关的信息的其中一个指定。与字符串列表有关的信息，被用于一部分词的边界的认定中，例如通过将字种用作与字符串列表有关的信息，即使是未实施句子结构解析的文本，也可以进行名词连续和助词的区分的认定等简易的解析。

接着，对于具有上述结构的信息提取系统 100，利用具体例更详细地进行说明。图 2 是表示直至从被输入的文本中提取与事物有关而被表现的事实和意见等内容作为要素元数据为止的一连串的处理的概要的说明图。在图 2 中，在图 2 (a) 中表示输入文本例，在图 2 (b) 中表示观点/记述认定例，在图 2 (c) 中表示要素元数据提取结果例。

首先，观点/记述提取单元 120 参照存储在观点/记述提取规则存储单元 122 中的观点/记述提取规则，调查从输入单元 102 输入的文本内的字符串是否具有观点/记述提取规则的规范 (pattern) 指定的句法的属性。在图 3 中表示观点/记述提取规则和规则的构成要素定义的例子。这里，所谓规则的结构要素定义，是在规则中，将规范等在记述中使用的字符串作为结构要素预先定义，如果在规则中记述结构要素名，则将其视为相当于用该结构要素名定义的字符串。结构要素名的定义方法，只要能够进行结构要素名和字符串或者字符串规范的列表的附带对应则不被特别限定。例如，可以将与结构要素名对应的字符串或者字符串规范的列表记述在一个文件中，也可以对应的字

字符串或者字符串规范的列表记述在另外的多个文件中。而且，在此后的规则例中利用同样的结构要素的情况下省略定义。在各规则中表示用于提取观点/记述的规范，以及在规范中相当于观点、记述的位置。

图3(a)所示的观点/记述提取规则是利用字符串的句法的属性来提取观点/记述的规则。观点/记述提取规则的规范中，相当于观点/记述的字符串或者其周边的字符串的句法的属性由字符串列表或者词类分类指定。在用字符串列表指定句法的属性的情况下，规则的规范中，作为包含“は”那样的字符串，以及“がも”（“が”“も”中的哪一个的意思）那样的字符串的正规表现的规范进行记述，或者如“汉字/片假名连续1”那样被预先定义的结构要素名指定。在用词类分类指定综合的属性的情况下，例如“形容动词词尾1”“形容词词尾1”那样预先定义对应于词类分类名的结构要素名，指定被定义的结构要素名。

而且，作为字符串的句法的属性的指定方法，在上述说明中使用了字符串列表和词类分类，但是本发明不限于此，另外也可以例如使用句子结构的关系。而且，在使用字符串列表和词类分类的情况下，它们的指定方法也不限于上述的方法，也可以是其它方法。而且，也可以取代句法的属性而使用含义的属性来指定，也可以指定句法的属性和含义的属性两者，进而也可以除了它们还指定统计的属性等其它属性。而且，在上述说明中仅用规则规范指定了适用规则的条件，但是也可以另外指定与规范的一部分有关的制约，也可以在规范以外进行指定。

而且，在图3(a)中，在规范中相当于观点和记述的位置用“( )”标记，被标记的部分从开头依次被参照为\$1、\$2、…。例如，在规则1的情况下，<“は”><汉字/平假名连续1><“が”或者“も”>、<英数字连续1>、<“と”>、<汉字/平假名连续1>、<形容词词尾1>以该顺序在文本中出现的情况下，与规则的规范一致。在相当于文本中的该规范的字符串中，规范中的相当于用最初“( )”括住的<汉字/平假名连续1>的部分被作为\$1参照。而且，相当于用第二个“( )”括住的<英数字连续1>的部分被作为\$2参照，相当于用第三个“( )”括住的<汉字/平假名连续1><形容词词尾1>的部分被作为\$3参照。按照规则，由\$1参照的部分被作为观点提取，由\$2、\$3参照的部分被作为记述提取。而且，规则的记法不限定于上述那样，也可以用其它的记法。

在将图3(a)的规则1应用在图2(a)的文本1的情况下，第一句的“开口部”相当于观点、“30cm”和“相当大”相当于记述。图2(b)的观点/记述认

定例是对文本内的观点/记述赋予识别用的观点/记述对 ID 号，将观点的表现的开始和结束用<VIEW (观点/记述对的号码)>…<VIEW (观点/记述对的号码)>标记，将记述的表现的开始和结束用<DESC (观点/记述对的号码)>…<DESC (观点/记述对的号码)>标记的例子。而且，观点/记述对 ID 号码的提供方法只要是可以唯一地确定观点/记述对则不进行特别限定。例如，也可以将文本的识别信息和文本内的观点/记述对的号码组合。

而且，例如如“容量大至 20 升”那样，对于一个观点（在本例中为“容量”），在“20 升”“大的”那样存在多个记述的情况下，将它们认定为对于相同观点的两个不同的记述。在本发明的观点和记述的提取规则例中，对于相同观点认定多个不同的记述的情况下，利用记号‘||’将这些记述表示为例如‘\$1||\$2’（这里，\$1、\$2 是记述）那样。

另一方面，例如如“容量对于旅行用来说小”那样，对于一个观点（在本例中为“容量”），在用途被限定为“旅行用”时如“小”那样，在记述间存在限定的关系的情况下，也可以集中多个记述（在本例中为“旅行用”和“小”）作为一个记述处理。在本发明的观点和记述的提取规则例中，对于相同观点集中关联的多个记述而认定为一个记述的情况下，将这些记述利用记号‘&&’例如表示为‘\$1&&\$2’（这里，\$1、\$2 是记述）那样。

接着，观点/记述提取单元 120 对被认定为相当于上述观点/记述提取规则的观点/记述的组，赋予用于识别出现了观点/记述对的文本和各个观点/记述对的要素元数据 ID，按照规则提取。在图 2 (c) 的要素元数据提取结果的表中表示观点/记述的提取例。在该提取结果表中，要素元数据 ID 的最上段中记载的“1 - 1a”中，左侧的“1”表示从文本 1 提取了该观点“开口部”/记述“30cm”的情况。右侧的“1a”中的“1”表示观点“开口部”/记述“30cm”是在检索了文本 1 时第一个（即最初）找到的观点/记述，“a”表示是第一个记述。

而且，在本例中，虽然用所谓<文本 ID> - <观点/记述对在文本内的号码>的形式来赋予要素元数据 ID，但是要素元数据 ID 的形式只要是可以进行文本的识别和观点/记述对的识别，则不限定于此。而且，句法的属性的赋予方法也不限于上述的方法，也可以进行句子结构分析和词态素分析。而且，上述的说明是观点/记述提取单元 120 利用观点/记述提取规则之间判定字符串的句法的属性的例子，但是本发明不限于该方法，也可以对被输入的文本预先赋予句法的属性，也可以由属性赋予单元（后述）赋予句法的属性。

接着，元数据核对单元 106 分别比较/核对被提取出的要素元数据的观点间/记述间，推定要素元数据的关联性。观点/记述的核对方法只要是至少使用构成观点、记述的字符串的句法的属性来进行核对，则没有特别限定。例如，利用词典、同义词辞典等来比较观点或者记述的构成词的概念的类似性的方法，并且，进而除了这些方法，还可以使用由观点或者记述的构成词的句子结构的关系推定类似度的方法等。这里，假设从观点和记述取出除去了助词和词尾的构成词，并且利用使用了元数据核对单元 106 内具有的词典而调查的结果，核对构成词间的句子结构的关系、构成词是否同义。首先，从图 2 (a) 的文本 1、文本 2 的观点取出的构成词间的句子结构的关系如下所示。

开口部→(构成词): 开口、部(句子结构的关系)连体修饰

拉链的开闭→(构成词): 拉链、开闭(句子结构的关系)连体修饰

皮的触感→(构成词): 皮、触感(句子结构的关系)连体修饰

皮的手感→(构成词): 皮、手感(句子结构的关系)连体修饰

色调→(构成词): 色调

接着，观点“皮的触感”、“皮的手感”的构成词中，“触感”“手感”通过词典认定为同义词，其它的构成词“皮”和句子结构的关系也一致，所以判定为两个观点“皮的触感”、“皮的手感”为同义，具有关联性。而且，对于记述也一样，在求出同义的记述时，判定要素元数据 ID1-3 的“温和湿润”和要素元数据 ID2-2 的“温和湿润”的记述为同义，具有关联性。而且，要素元数据的关联性的判定方法只要根据观点和记述的核对结果来进行判定，则不限于上述的方法，也可以是其它方法。例如，在观点和记述的概念的类似性被数值化的情况下，也可以将观点或者记述的数值位于一定范围内的要素元数据判定为“具有关联性”。

接着，根据要素元数据间的关联性，元数据综合单元 108 综合要素元数据，作为综合元数据存储在元数据存储单元 110 中。元数据的综合方法不被特别限定，这里，设为：

(1) 综合具有同义的观点的元数据

(2) 如果在具有同义的观点的元数据中存在同义的记述则进行综合

在图 2 的例子中，由于观点中“皮的触感”和“皮的手感”被判定为同义，所以综合这些观点，例如作为“皮的触感”。而且，与这些观点成对的记述“温和湿润”和“光滑”不被视为同义，所以不综合。图 4 表示这样进行了综合处理

后的综合元数据的例子。而且，在上述的说明中，说明了输入多个文本的情况，但是也可以输入一个文本。

这样，按照本实施方式，将与在文本中表现的事物有关的事实和意见的记述内容构成为观点和记述的组，并且将其整理为容易进行事实和意见的对应和关联性的比较的形式而提取，利用该提取结果，进而将事实和意见附带对应，可以综合关联的事实和意见。

### (实施方式2)

图5是表示本发明的实施方式2的信息提取系统的结构的方框图。该信息提取系统200具有与对应于图1所示的实施方式1的信息提取系统100相同的基本结构，对于相同的结构要素赋予相同的标号，省略其说明。

本实施方式的特征是具有：属性赋予单元202，对从输入单元102输入的文本的字符串赋予含义的属性；含义属性赋予规则存储单元204，存储了用于对前述字符串赋予含义属性的含义属性赋予规则；附带含义属性文本存储单元206，存储由属性赋予单元202赋予的附带含义属性文本。属性赋予单元202的处理结果，即被赋予了含义属性的文本（附带含义属性文本）被存储在附带含义属性文本存储单元206中。这时，观点/记述提取单元120对被存储在附带含义属性文本存储单元206中的附带含义属性文本进行观点/记述提取。

属性赋予单元202认定文本中的事物名、数值关联表现（时间、数量、金额等）等字符串，对它们赋予含义的属性。作为对事物名和数量表现赋予含义的属性的方法，虽然没有特别限定，但是例如可以使用：对每个关键词利用记载了其含义属性的辞典，以及利用了在文献“福本他：固有名詞抽出における日本語と英語の比較”，情報処理学会研究会報告98-NL-126, pp.107-114, 1998”中所示的固有名词提取技术的方法等。

这里，所谓含义的属性，例如是将事物名和数量表现通过表现的含义进行分类的含义分类。在含义的属性具有详细度的水平时，以及相应的表现为一般的表现的其它表现，需要表示被正规化的形式时，也可以将详细度水平和被正规化的表现作为含义的属性的详细信息一起记录。

在以下，说明属性赋予单元202利用含义属性赋予规则对事物名和数量表现赋予含义的属性的例子。

首先，属性赋予单元202参照被存储在含义属性赋予规则存储单元204

中的含义属性赋予规则，对从输入单元 102 输入的文本内的字符串，调查是否具有对应于规则的含义的属性的表现。其结果，对文本中的字符串中对应的表现和含义属性进行标记，并作为附带含义属性文本存储在附带含义属性文本存储单元 206 中。在图 6 (a) 中，表示被输入的文本例，在图 6 (b) 中表示被赋予含义的属性的文本的例子。而且，在图 7 中表示含义属性赋予规则的例子和含义属性赋予规则的构成要素定义的例子。而且，构成要素的定义方法，只要是可以进行构成要素名和字符串或者字符串规范的列表的附带对应就没有特别限定。例如，可以将构成要素名和对应的字符串或者字符串规范的列表记述作为一个文件，以及将对应的字符串或者字符串规范的列表记述在其它的多个文件中。而且，在此后的规则例中使用同样的构成要素的情况下省略定义。

在图 7 的含义属性赋予规则例中，表示在文本中的字符串中用于检测具有对应含义属性的表现的规范，对与各规范一致的表现的对象部分赋予的含义属性的含义分类和详细信息。在规则规范中，赋予含义属性的字符串和字符串列表被指定“数字连续”等字符串规范、或者“产品分类名”等对应于词列表的预先定义的构成要素名。而且，规则规范以及对象部分的 \$1、\$2 等记法与图 3 的规则一样。在该例子中，详细信息中“val”表示数值表现被正规化的值，“unit”是数量单位的表现的正规化形式，“type”表示含义的属性的下位分类。

在将图 7 的规则应用在图 6 (a) 的文本 1 的情况下，通过规则 1，“20 升”的含义属性中含义分类被识别为 QUANT(数量)，详细信息被识别为 [unit = 1 (单位为‘1’的意思)， val = 20 (数值为‘20’的意思)]。而且，通过规则 2，在“容量”的含义属性中，含义分类被识别作为 QUANT\_TYPE (数量分类)。而且，通过规则 3，“A 公司”的含义属性中，含义分类被识别为 ORGANIZATION (组织名)，详细信息被识别为 [type = company (种类是‘公司名’的意思)]。被识别出的结果，被赋予各个对应的含义属性的含义分类和详细信息，作为图 6 (b) 所示的附带含义属性文本被存储在附带含义属性文本存储单元 206 中。

而且，含义属性赋予规则的记法不限于上述的记法，也可以是其它的记法。而且，作为含义属性赋予规则的规范的记述方法，在上述说明中使用了与字符串规范或词列表对应的构成要素名，但是也可以使用其它的记述方法。

而且，作为应用含义属性赋予规则的条件的指定方法，虽然在上述说明中仅使用了规范，但是本发明不限于此，也可以是其它方法。例如，除了规范，也可以作为另外指定与规范的一部分有关的制约，也可以利用规范以外的指定方法。而且，也可以将被预先赋予了含义属性的文本直接输入观点/记述提取单元 120 中。

接着，观点/记述提取单元 120 从被存储在附带含义属性文本存储单元 206 中的附带含义属性文本中将观点/记述的组与含义的属性一起作为要素元数据提取。在图 8 (a) 中表示附带含义属性文本的例子，在图 8 (b) 中表示观点/记述认定例。而且，在图 9 中表示用于提取观点/记述的观点/记述提取规则的例子和观点/记述提取规则的构成要素的定义例。关于规则的记法、构成要素的定义方法与图 3 一样，省略说明。

图 9 所示的观点/记述提取规则和实施方式 1 的图 3 所示的观点/记述提取规则的不同之处在于，在图 9 中，被赋予文本的含义属性作为规范的一部分被记述。例如，在图 9 的规则 1 中，由<QUANT\_TYPE>，</QUANT\_TYPE>包围的标记 (tag) 开始记号以外的任意字符串，即，被赋予了所谓 QUANT\_TYPE (数量分类) 的含义属性的字符串被指定作为观点。而且，由<QUANT>，</QUANT>包围的标记开始记号以外的任意字符串，即，被赋予所谓 QUANT (数量) 的含义属性的字符串被指定作为对应于前述观点的第一个记述。在将图 9 的规则 1 应用于文本 1 的情况下，被赋予了 QUANT\_TYPE 的含义属性的“容量”相当于观点，被赋予了 QUANT 的含义属性的“20 升”相当于与该观点对应的一个记述，“大的”相当于第二个记述。接着，在将图 9 的规则 3 应用于图 8 (a) 的文本 1 的情况下，被赋予了 ORGANIZATION 的含义的属性的字符串“A 公司”相当于记述。虽然与该记述对应的观点未被表现在文本中，但是按照图 9 的规则 3，如果将含义的属性的别名识别为观点，则“公司名”被认定为观点。同样，在图 10 中表示以下例子，即对于图 8 (a) 的附带含义属性的文本 1、2，观点/记述提取单元 120 应用图 9 的规则，将观点和记述与它们的含义属性的含义分类和详细信息一起，赋予作为识别信息的要素元数据 ID，从而作为要素元数据提取的结果的例子。

而且，在上述的说明中，设为属性赋予单元 202 作为赋予字符串的含义的属性的情况，但是本发明不限于此。属性赋予单元 202 也可以对文本赋予

句法的属性和含义的属性中的至少一个，也可以观点/记述提取单元 120 利用观点/记述提取规则或者其它的规则赋予句法的属性和含义的属性中的至少一个，也可以对被输入的文本预先赋予句法的属性和含义的属性中的至少一个。

而且，在上述的说明中，设为作为含义的属性赋予含义分类和详细信息的情况，但是，只要被赋予的含义的属性包含含义分类，则本发明不限于此，例如也可以赋予详细信息以外的其它的含义的信息。

接着，元数据核对单元 106 分别比较、核对被提取的要素元数据的观点间、记述间，推定关联性。本实施方式中的元数据核对单元 106 的核对方法和实施方式 1 的不同之处在于，在核对时使用要素元数据的观点和记述的含义属性这一点。这里，在核对图 10 的要素元数据的观点间、记述间从而求同义的观点和记述时，除了实施方式 1 的方法，还在满足以下的条件的情况下认定为同义的观点或者记述。

在含义分类为“产品名”的表现中，仅在表现中英数字的边界是否插入有“-”这一点有所不同。

通过以上的方法，在图 10 的要素元数据的观点或者记述中，1-2 和 2-1 的观点“产品分类”和记述“包”、1-3 的观点“产品名”和记述“A200”和 2-2 的观点“产品名”和记述“A-200”由于各自同义而被判定为具有关联性的观点和记述，1-4a 和 1-4b 和 2-3 的观点“容量”由于是同义的观点而被判定为具有关联性。

而且，元数据的观点和记述的核对方法，以及要素元数据关联性的判定方法不限于上述的方法。观点和记述的核对方法例如也可以使用词典、同义词辞典等比较核对观点或者记述的构成词的概念性的类似性的方法，以及由观点或者记述的构成词的句子结构的关系推定类似度的方法等。而且，要素元数据的关联性的判定方法不限于上述的方法，例如，在观点和记述的概念性的类似性被数值化的情况下，也可以将观点或者记述的数值在一定范围内的要素元数据判定为“具有关联性”。

接着，元数据综合单元 108 根据前述要素元数据的关联性，与实施方式 1 一样，综合要素元数据，作为综合元数据存储在元数据存储单元 110 中。这里，假设综合满足与实施方式 1 一样的条件的观点和记述，详细的说明省略。在图 11 中表示，在图 10 的要素元数据中，综合关联的要素元数据而将

其存储在元数据存储单元 110 中的综合元数据的例子。在图 11 中，作为同义的观点和记述的 1-2 和 2-1 的观点“产品分类”和记述“包”、1-3 的观点“产品名”和记述“A200”、以及 2-2 的观点“产品名”和记述“A-200”被各自综合。而且，可知作为三个不同的记述的 1-4a 的“20 升”、1-4b 的“大的”、2-3 的“不足的”的观点“容量”被综合，作为数量的“20 升”作为该产品的容量被表现为“大的”、“不足的”，在图 8 的文本 1 和文本 2 中接受有不同的评价。

这样，按照本实施方式，可以将与在附带含义属性的文本的字符串中表现的事物有关的事实和意见的记述内容与观点和记述的含义属性一起容易地进行提取。而且，在利用该提取结果，更详细地判定了关联性以后，通过综合关联的事实和意见，可以容易的进行事实和意见的附带关联和关联性的比较。

### ( 实施方式 3 )

图 12 是表示本发明的实施方式 3 的信息提取系统的结构的方框图。该信息提取系统 300 具有与图 5 所示的对应于实施方式 2 的信息提取系统 200 同样的基本结构，对同样的构成要素赋予相同的标号，并省略其说明。

本实施方式的特征是具有以下单元：用户请求处理单元 302，处理来自用户的请求；元数据输出形式生成单元 304，整理元数据而生成元数据的输出形式；元数据输出单元 306，对用户提示元数据输出形式生成单元 304 生成的元数据的输出形式；话题事物推定单元 310，推定观点/记述提取单元 120 提取的要素元数据的话题的事物；以及话题事物推定规则存储单元 312，存储用于推定话题的事物的规则的话题事物推定规则。

这里，所谓“话题事物”是指各要素元数据对哪个事物进行记述的要素元数据的话题的事物名。该话题事物从表示事物名的要素元数据的其中一个记述中被选择。成为话题事物的候选的事物名不被特别限定，但是有人名、地名、组织名、事件名、生物或者人工物的名称以及它们的分类（例如：产品名、产品分类）等。

对于具有上述结构的信息提取系统 300，利用具体例更详细地进行说明。现在，假设存在以下的文本 1、2。

文本 1：“包 A200 的容量不足，包 A300 的容量太大。”

文本 2：“包 A200 的容量为 20 升，包 A300 的容量为 30 升。”

前述文本从输入单元 102 被输入，由属性赋予单元 202 赋予含义属性，

由观点/记述提取单元 120 认定观点/记述，到提取要素元数据为止的处理的流程与实施方式 2 一样，省略说明。在图 13 (a) 中表示对上述文本赋予含义分类，并且认定了观点/记述的结果的例子，在图 13 (b) 中表示要素元数据的提取结果的例子。

接着，话题事物推定单元 310 按照话题事物推定规则存储单元 312 中存储的通话事物推定规则，推定文本内的话题事物。话题事物的推定方法只要是使用话题事物推定规则，则没有特别限定。话题事物推定单元 310 可以使用话题事物推定规则来直接推定话题事物，也可以首先决定作为话题事物候选的要素元数据的种类，并在之后利用话题事物推定规则进行推定。这时，在被输入的文本例如存在具有公司名和人名等多种话题的可能性的情况下，希望设想多个话题事物推定候选，以便前述话题事物推定单元 310 可以选择适当的话题事物。例如，话题事物候选被规定为作为“规定是产品名或者人名”的要素元数据的记述的情况下，假设观点为产品名或者人名的要素元数据的记述被规定为话题事物的候选。这时，文本 1、2 都是在观点中具有产品名的要素元数据的记述，“A200”、“A300”成为话题事物候选。

以下，对话题事物推定单元 310 按照话题事物推定规则存储单元 312 中存储的话题事物推定规则，推定文本内的话题事物的情况进行说明。这里，假设通过与在条件单元中记述的规范进行匹配来进行话题事物推定，在图 14 (a) 中表示话题事物推定规则的例子，在图 14 (b) 中表示话题事物推定规则构成要素定义的例子。而且，规则的条件单元的规范的记法和构成要素的定义方法与图 3 基本相同，但是，在图 14 的规则 2、规则 3 中作为条件不仅是规范，还将规范的一部分字符串相同的情况加到条件中。

利用图 14 (a) 的规则，从图 13 (a) 的文本推定图 13 (b) 的要素元数据。例如，如对文本 1 应用图 14 的规则 1，则首先作为第二个记述的 <DESC2><PROD\_NAME>A200</PROD\_NAME></DESC2> 与规则 1 的条件单元中记述的规范一致，按照该规则，这里，相当于\$1 的“A200”的话题事物被推定为“A200”自身。同样，图 15 表示利用图 14 的规则，从图 13 (a) 的文本 1、2 推定出图 13 (b) 的要素元数据的话题的例子。对图 15 的要素元数据 ID1-1、1-4、2-1、2-4 的要素元数据应用图 14 的规则 3，对图 15 的 ID1-2、1-5、2-2、2-5 的要素元数据应用图 14 的规则 1，对图 15 的 ID1-3、1-6、2-3、2-6 的元数据应用图 14 (a) 的规则 2。

而且，话题事物的推定方法只要利用话题事物推定规则就不限于上述的方法，例如，也可以在与上述不同的记法的规则中指定要素元数据的观点/记述、或者句法的属性、或者含义的属性、或者其它属性。而且，也可以按照话题事物候选的种类而应用不同的规则。

接着，元数据核对单元 106 分别比较、核对被提取出的要素元数据的观点间、记述间，推定关联性。要素元数据的观点/记述的核对方法与实施方式 1 或 2 基本相同，在本实施方式中，进一步利用话题事物的推定结果进行核对。

在图 15 的例子中，要素元数据 ID1-1、1-2、1-3、2-1、2-2、2-3 具有相同的话题事物“A200”，1-4、1-5、1-6、2-4、2-5、2-6 具有相同的话题事物“A300”。对于每个具有相同的话题事物的要素元数据，与实施方式 1 一样，在求同义的观点或记述时，首先，对于话题事物为“A200”的要素元数据，具有同义的观点或记述的要素元数据是 1-1 和 2-1、1-2 和 2-2。而且，具有同义的观点的要素元数据得到 1-3 和 2-3。前者的观点和记述、后者的观点被推定出各自具有关联性。

同样，对于话题事物为“A300”的要素元数据，具有同义的观点和记述的要素元数据是 1-4 和 2-4、1-5 和 2-5。而且，具有同义的观点的要素元数据得到 1-6 和 2-6。前者的观点和记述、后者的观点被推定出各自具有关联性。

而且，元数据核对单元 106 的核对方法以及关联性的推定方法不限于上述方法。在上述说明中，对于每个具有相同话题事物的要素元数据，求同义的观点和记述，但是，例如也可以在求出具有同义的观点和记述的要素元数据后，求具有相同话题事物者，也可以进一步利用要素元数据的含义属性等。

接着，与实施方式 1 同样，元数据综合单元 108 综合要素元数据，作为综合元数据存储在元数据存储单元 110 中。要素元数据的综合的方法未被限定，这里作为例子有以下三种：

(1) 综合具有相同话题的事物，

(2) 在相同的话题中统一具有同义的观点的要素元数据，

(3) 在相同的话题中在具有同义的观点的要素元数据中如果存在同义的记述则进行统一。

对利用了该例子的情况进行说明。在图 15 的要素元数据中，具有相同话题事物的 1-1、1-2、1-3、2-1、2-2、2-3 通过上述 (1) 综合话题事物。同样，

1-4、1-5、1-6、2-4、2-5、2-6 话题事物也被综合。接着，具有相同话题事物和同义的观点的要素元数据 1-1 和 2-1、1-2 和 2-2、1-3 和 2-3、1-4 和 2-4、1-5 和 2-5、1-6 和 2-6 按照上述（2），各自的话题事物和观点被综合。进而，在具有相同话题事物、具有同义的观点和记述的要素元数据 1-1 和 2-1、1-2 和 2-2、1-4 和 2-4、1-5 和 2-5 按照上述（3），各个话题事物和观点和记述被综合。

如上述那样，在图 16 中表示元数据综合单元 108 综合了从文本 1、2 提取的图 15 的要素元数据的结果、被存储在元数据存储单元 110 指定综合元数据的例子。由该综合结果可知，根据“A200”的“容量”为“20 升”而被评价为“不足够”，另一方面，根据“A300”的“容量”为“30 升”而被评价为“太大”。而且，元数据的综合方法不限定于上述方法，只要是根据元数据核对单元 106 推定的要素元数据的观点和记述的关联性进行综合的方法，则其它方法也可以。例如，可以首先综合具有同义的观点和记述的要素元数据，之后综合具有相同话题事物的要素元数据。

接着，用户请求处理单元 302 输入用户请求，将用户请求的输出形式输出到元数据输出形式生成单元 304。元数据输出形式生成单元 304 参照存储在元数据存储单元 110 中的综合元数据，以用户请求的输出形式生成元数据，并通过元数据输出单元 306 对用户提示。

这里，说明按照用户请求的指定，作为元数据的输出形式的一例，生成元数据表的情况。首先，通过用户请求处理单元 302 输入用户请求。被输入用户请求处理单元 302 的用户请求设为指定包含了话题事物的要素元数据的一部分、或者它们的组合的其中一个。当前，作为用户请求的例子，例如指定“（话题事物：A200）并且（观点：容量）”的条件作为要素元数据应满足的条件。用户请求处理单元 302 检查被指定的用户请求的指定形式，并且在没有问题时将用户请求发送到元数据输出形式生成单元 304。

而且，在本例中，虽然假设用户请求以上述的形式被输入，但是，用户请求也可以以自由的文本（例如：“想知道 A200 的容量”）的方式被输入。在后者的情况下，用户请求处理单元 302 也可以直接解析文本从而取出上述的条件。而且，用户请求处理单元 302 也可以将用户输入输入单元 102 中的用户请求的文本暂时发送，由通过观点/记述提取单元 120 提取的要素元数据、以及它们的句子结构的关系来解析被指定的条件的内容。

元数据输出形式生成单元 304 按照从用户请求处理单元 302 获得的用户请求的指定内容，从被存储在元数据存储单元 110 中的综合元数据中挑选相应的要素元数据，使被挑选出的元数据对应生成为输出形式。例如，如果在用户请求的内容中存在话题事物的指定，则从综合元数据中挑选在话题中具有该话题事物的要素元数据，进一步挑选满足被指定的观点或记述的条件的要素元数据，生成将它们作为对象的元数据表。元数据输出单元 306 输出被生成的元数据表。

在图 17 中表示从图 16 的综合元数据中仅取出满足用户请求(话题事物：A200)并且(观点：容量)的要素元数据而作成的元数据表的例子。这时，仅话题事物为“A200”并且观点为“容量”的要素元数据作为表被输出。而且，在上述的说明中，元数据的输出形式作为元数据的表进行了说明，但是，输出形式也可以是表以外的其它形式。

这样，按照本实施方式，可以将与文本中表现的事物有关的事实和意见的记述内容与被推定的话题的事物一起，容易地进行事实和意见的附带对应。而且，可以利用该提取结果，进一步对每个话题事物将事实和意见精密地附带对应，在更详细地判定了关联性以后，综合关联的事实和意见，可以整理为容易进行关联性的比较的形式来提取。

而且，通过对用户提示将包含话题事物的要素元数据根据用户的指定而整理的元数据输出形式，可以整理并提示用户请求的信息。

#### (实施方式 4)

图 18 是表示本发明的实施方式 4 的信息提取系统的结构的方框图。该信息提取系统 400 具有与图 12 所示的实施方式 3 对应的信息提取系统 300 相同的基本结构，对同样的构成要素赋予同样的标号，省略其说明。

本实施方式的特征是：前述输入单元 102 还接受来源信息和用户信息，元数据核对单元 106 包括：客观性/可靠性判定单元 412，利用要素元数据、来源信息意见用户信息来判定观点/记述的客观性和可靠性；客观性/可靠性判定规则存储单元 414，存储了用于评价客观性和可靠性的客观性/可靠性判定规则。

这里，所谓来源信息，是指与被输入的文本有关的书目事项的信息，并且将文本中的来源信息的记述称为来源信息记述。作为来源信息的例子，有文本的种类、取得元、作者分类、作者、组织名、作成日期时间等。来源信

息记述，只要是可以进行与文本的附带对应的形式，则可以作为输入文本的一部分被输入，也可以与输入文本分别输入。来源信息记述的公文格式虽然没有特别限定，但是设为与文本的识别信息一起被输入。

而且，所谓用户信息，是指与被输入的文本的作者有关的信息，并且将文本中表现的用户信息的记述称为用户信息记述。作为用户信息的例子，有用户的性别、年龄、职业、上班地、兴趣等。用户信息记述只要是可以进行与文本的附带对应的形式，则可以作为文本的一部分被输入，也可以与输入文本分别输入。用户信息记述的公文格式虽然没有特别限定，但是设为与文本的识别信息一起被输入。

而且，将构成来源信息记述、用户信息记述作为观点和记述的组者称为各个来源元数据、用户元数据。对来源元数据和用户元数据赋予用于识别对应的文本和各个来源元数据或者用户元数据的来源元数据 ID、或者用户元数据 ID。来源元数据 ID 和用户元数据 ID 的公文格式虽然没有特别限定，但是由于需要取得与文本的对应关系，所以希望设为对应的文本 ID 能够推定的公文格式。

客观性/可靠性判定单元 412 利用要素元数据、来源元数据或者用户元数据的其中一个判定要素元数据的观点/记述的客观性和可靠性，并且将判定结果设为要素元数据的评价数据。元数据综合单元 108 除了要素元数据，还可以在综合元数据的结果中包含来源元数据、用户元数据以及要素元数据的评价数据。而且，用户可以从用户请求处理单元 302 不仅利用要素元数据，还利用源元数据、用户元数据以及要素元数据的评价数据来指定必要的信息，得到对应的元数据输出形式。

接着，利用具体例更详细地对具有上述结构的信息提取系统 400 进行说明。在本实施方式中，来源信息记述和用户信息记述作为输入文本的一部分被输入，输入文本的特定的块中分别记述来源信息记述、用户信息记述。

从输入单元 102 输入文本。

属性赋予单元 202 利用含义属性赋予规则对包含被输入的来源信息记述的文本赋予含义属性，将附带含义属性的文本输出到附带含义属性文本存储单元 206。图 24 (a) 表示除了来源信息记述和用户信息记述的块以外的文本例 1~4。到此为止的处理的流程由于与实施方式 2 至 3 一样，所以省略详细的说明。图 19 (a) 中表示来源信息记述的例子，图 19 (b) 中表示用户信息

记述的例子，图 19 (c) 中表示附带含义属性的例子，图 19 (d) 中表示附带含义属性用户信息记述的例子。而且，图 20 (a) 中表示来源含义属性赋予规则的例子，图 20 (b) 中表示用户含义属性赋予规则的例子。

接着，观点/记述提取单元 120 利用存储在观点/记述提取规则存储单元 122 中的观点/记述提取规则、利用观点/记述提取规则、以及用户观点/记述提取规则，从存储在附带含义属性文本存储单元 206 中的附带含义属性文本、附带含义属性来源信息以及附带含义属性用户信息中分别提取至少由观点和记述的组构成的要素元数据、来源元数据以及用户元数据。

首先，对观点/记述提取单元 120 从前述附带含义属性文本的来源信息记述和用户信息记述的块提取来源元数据和用户元数据的情况进行说明。在提取来源元数据和用户元数据时，如图 22 (a) 所示，对各来源元数据赋予来源元数据 ID，如图 22 (b) 所示，对用户元数据赋予用户元数据 ID。而且，在本实施方式中，设为以各个<文本 ID>-S<观点/记述对的来源信息内的号码>、<文本 ID>-U<观点/记述对的用户信息内的号码>的形式赋予来源元数据 ID 和用户元数据 ID，但是来源元数据 ID 的形式和用户元数据 ID 的形式不限于此。

在图 21 (a) 中表示来源观点/记述提取规则的例子，在图 21 (b) 中表示用户观点/记述提取规则的例子。图 21 (a) 的来源观点/记述提取规则和图 21 (b) 的用户观点/记述提取规则中，与实施方式 1 的观点/记述提取规则一样，在规则的规范中，相当于观点/记述的字符串及其周边的字符串的句法的属性以及含义的属性被指定。而且，作为字符串的句法的属性的指定方法，在图 21 (a) (b) 中使用列表，作为含义属性的指定方法使用含义属性的含义分类和详细信息，但是本发明不限于此，也可以仅指定句法的属性和含义的属性的任何一个，例如作为句法的属性也可以使用词类分类等。

以下，对利用图 21 (a) 的利用观点/记述提取规则或者图 21 (b) 的用户观点/记述提取规则，从图 19 (c) 的附带含义属性来源信息记述和图 19 (d) 的附带含义属性用户信息记述提取来源元数据和用户元数据的情况进行说明。例如，在对图 19 (c) 的附带含义属性来源信息记述应用图 19 (a) 的来源观点/记述提取规则 1 的情况下，图 19 (c) 的字符串<URL type = 公司网页 s>[</URL>](http://aaa.co.jp/article1)相当于前述规则 1 的规范，其中，相当于规范中的最初的‘( )’中被括住的部分的 <http://aaa.co.jp/article1>，相当于与在规则

中被指定的观点“文本的取得元”对应的记述。

利用图 21 (a) 的利用观点/记述提取规则或者图 21 (b) 的用户观点/记述提取规则, 从图 19 (c) 的附带含义属性来源信息记述和图 19 (d) 的附带含义属性用户信息记述提取的来源元数据提取结果和用户元数据提取结果的例子分别表示在图 22 (a)、图 22 (b) 中。

接着, 对观点/记述提取单元 120 从前述附带含义属性文本的来源信息记述、用户信息记述的块以外提取要素元数据, 并且话题事物推定单元 310 推定话题事物为止的流程进行说明。图 24 (b) 中表示属性赋予单元 202 对图 24 (a) 的各文本赋予含义属性的例子, 图 25 中表示观点/记述提取规则的例子。利用图 25 的观点/记述提取规则, 与实施方式 2 或 3 一样从图 24 (b) 的附带含义属性文本中提取观点/记述。例如, 在对图 24 (b) 的附带含义属性文本 1 应用图 25 的规则 1 时, 对于观点“容量”, 提取两个记述“20 升”“大的”。同样, 通过图 25 的规则, 从图 24 (b) 的附带含义属性文本 1~4 中提取图 26 所示的观点和记述。进而, 利用图 14 的话题事物推定规则与实施方式 3 一样从图 24 (b) 的附带含义属性文本 1~4 中进行推定。

在图 26 中表示将观点/记述提取单元 120 从图 24 (b) 的附带含义属性文本 1~4 提取的观点/记述、以及它们的含义的属性、话题事物推定单元 310 推定的话题事物进行集中而作为要素元数据的例子。而且, 在图 26 中, 仅表示要素元数据的一部分。而且, 在实施方式 3 中, 关于话题事物的推定, 说明了仅利用从文本得到的信息来进行推定的方法, 但是此外也可以利用从来源信息或用户信息得到的元数据。

接着, 元数据核对单元 106 的客观性/可靠性判定单元 412 在观点/记述提取单元 120 中从附带含义属性文本中提取的要素元数据、来源元数据和用户元数据中至少利用其中一个, 按照存储在客观性/可靠性判定规则存储单元 414 中的客观性/可靠性判定规则来判定前述要素元数据的客观性和可靠性。

这里, 所谓要素元数据的客观性表示要素元数据是否被客观地记述, 例如, 如果作为事实被记述的话则认为客观性高, 如果作为意见被记述的话则认为客观性低。客观性可以作为数值表现, 也可以通过阈值或者判定条件用“事实”“意见”等的分类来表现。

而且, 所谓要素元数据的可靠性, 表示要素元数据是否可靠, 例如, 个人的主页上作为意见被书写的记述的可靠性被认为较低, 在报纸报道中作为

事实被书写的记述的可靠性被认为较高等。而且，可靠性可以作为数值表现，或者也可以通过阈值或者判定条件用“可靠性高”“可靠性低”的分类来表现。

要素元数据的客观性/可靠性的判定中，假设至少使用要素元数据、来源元数据、用户元数据的其中一个，但是，除此之外也可以将字符串的句法的属性、含义的属性或者统计的信息等组合来使用。

在图 23 中表示客观性/可靠性判定规则的例子。这里，将客观性设为 1~0（设 1 的客观性高，0 的客观性低），将可靠性设为 1~0（设 1 的可靠性高、0 的可靠性低）来表现。例如，规则 4 是对于观点为“用途”而记述的含义分类为“USAGE”那样的要素元数据，如果来源元数据的文本的取得元为“公司网页”，则客观性判定为 1，可靠性也判断为 1 的规则。

接着，对于图 26 的要素元数据，说明通过文本的要素元数据、来源元数据、句法的属性，利用可客观性/可靠性判定规则，进行了客观性/可靠性的判定处理的例子。

当前，假设通过观点/记述提取单元 120 从输入文本中，图 26 的要素元数据的提取元的对应于文本 1~4 的来源信息记述意见用户信息记述的块中，分别提取以下那样的来源元数据和用户元数据。

文本 1

来源元数据

观点：文本的取得元

假设的含义属性：公司网页

文本 2

来源元数据

观点：文本的取得元

记述的含义属性：个人网页

用户元数据

观点：性别

记述：男性

文本 3

来源元数据

观点：文本的取得元

记述的含义属性：个人网页

### 用户元数据

观点：性别

记述：女性

文本 4

### 来源元数据

观点：文本的取得元

记述的含义属性：个人网页

### 用户元数据

观点：性别

记述：男性

利用上述的来源元数据和用户元数据，利用图 23 的客观性/可靠性判定规则判定图 26 的要素元数据的客观性和可靠性。例如，在图 26 的要素元数据 ID 为 1-3a 的要素元数据的情况下，要素元数据的观点为“容量”、记述的含义分类为“QUANT”，提取元的文本 1 为公司网页，所以应用图 23 的规则 6，客观性、可靠性都被判定为 1。另一方面，在图 26 的要素元数据 ID 为 1-3b 的要素元数据的情况下，要素元数据的观点为“容量”、记述的含义分类为“无”，提取元的文本 1 为公司网页，进而包含要素元数据的句子的“句尾为不确定表现 1 以外”，所以应用图 23 的规则 9，客观性被判定为 0、可靠性被判定为 0.5。同样，图 27 表示利用上述来源元数据和用户元数据，对图 26 的要素元数据，客观性/可靠性判定单元 412 利用图 23 的客观性/可靠性判定规则判定的客观性/可靠性判定结果例。而且，对于规则的记法和构成要素定义与图 3、图 7 等一样，省略说明。

而且，作为客观性/可靠性判定规则的条件，在上述说明中利用了文本的要素元数据和来源元数据和句法的属性，但是，只要是包含要素元数据、来源元数据和用户元数据的至少其中一个，则本发明不限于此。而且，在图 23 的客观性/可靠性判定规则中，将对应于来源元数据的观点“文本的取得元”的记述的含义属性用于规则的条件的一部分，但是也可以利用其它的观点和记述的组。例如可以利用“作成日”而作成日早的要素元数据判定为可靠性低，或者利用“作成日”而将特点的人写的文本的可靠度提高或降低。而且，将要素元数据和其它的信息组合的情况下，例如，与统计的信息组合，将对于同样的观点具有多个类似的内容的记述的要素元数据的可靠度提高。或者，也

可以将具有与多数人的记述不同的内容的记述的要素元数据的可靠度降低。而且，在图 23 的客观性/可靠性判定规则中，在一个规则中同时判定客观性和可靠性，但是也可以区分客观性的判定规则和可靠性的判定规则，从而在一个规则中判定其中一个。

接着，元数据核对单元 106 分别比较/核对被提取的要素元数据的观点间、记述间，推定关联性。元数据核对单元 106 的观点/记述的核对方法没有特别限定。这里，设为与实施方式 1、2 或 3 相同，但是，也可以还利用客观性/可靠性，在由观点/记述间的核对结果推定为关联性高的要素元数据中，客观性或可靠性的值接近者进一步推定为关联性高。

而且，在上述说明中，来源元数据和用户元数据仅在可靠性或可靠性的判定中使用，但是有利于将它们在元数据核对单元 106 进行要素元数据的比较/核对时直接使用。例如，也可以在存在关于从多个个人网页中提取的要素元数据的某个产品的容量的记述的情况下，用户元数据的“性别”的记述相同，或者“年龄”的记述在一定范围内时，使关联性高。

接着，元数据综合单元 108 综合包含要素元数据、来源元数据、用户元数据和评价的要素元数据，将综合结果存储在元数据存储单元 110 中。

综合的方法虽然没有特别限定，但是这里作为例子，设为以下的(1)~(4)

#### (要素元数据)

(1) 综合具有相同话题的元数据

(2) 在相同的话题中综合具有同义的观点的要素元数据

(3) 在相同的话题中在具有同义的观点的元数据中如果存在同义的记述则进行综合。

(4) 在相同的话题中在具有同义的观点和同义的记述的元数据中如果含义属性相同则进行综合。

在将综合的方法设为(1)~(4)的情况下，对元数据综合单元 108 综合图 27 的要素元数据的情况进行说明。首先，图 27 的元数据由于全部具有相同的话题“A200”，所以按照上述(1)，以共同的话题进行综合。接着，与实施方式 1 一样，判定具有相同话题的各要素元数据的观点是否同义。在图 27 的例子中，观点仅为“产品分类”、“产品名”、“容量”、“用途”四种，它们不同义，所以如果各自综合具有这四种观点的要素元数据，则以观点“产品分

类”综合要素元数据 1-1、2-1、3-1、4-1，以观点“产品名”综合要素元数据 1-2、2-2、3-2、4-2，以观点“容量”综合要素元数据 1-3a、1-3b、2-3、3-3、4-3。

接着，与实施方式 1 一样，判定在相同话题中具有同义的观点的元数据的记述是否同义。在图 27 的例子中，例如，由于在话题“A200”中具有同义的观点“产品分类”的要素元数据的记述都是“包”，所以它们被视为同义，按照上述(3)，要素元数据 1-1、2-1、3-1、4-1 的记述被综合。同义，在话题“A200”中具有同义的观点“产品名”的要素元数据 1-2、2-2、3-2、4-2 的记述，意见具有同义的观点“用途”的要素元数据 3-4、4-4 被分别综合。另一方面，例如，在话题“A200”中具有同义的观点“容量”要素元数据的记述“20 升”、“大的”、“海外出差用 - 不足够”、“国内出差用 - 太大”、“国内出差用 - 足够”不被判定为同义，所以不被综合。

接着，在话题“A200”中具有同义的观点“产品分类”和同义的记述“包”的要素元数据的含义分类都为“PROD\_TYPE”，所以它们被视为同义，按照上述(4)，要素元数据 1-1、2-1、3-1、4-1 的含义分类被综合。同样，在话题“A200”中，具有同义的观点“产品名”和同义的记述“A200”的要素元数据 1-2、2-2、3-2、4-2 的含义分类、以及具有与同义的观点“用途”同义的记述“国内出差用”的要素元数据 3-4、4-4 的含义分类也被分别综合。

如上所述，在图 28 中表示元数据综合单元 108 综合了图 27 的元数据的结果、被存储在元数据存储单元 110 中的元数据的综合结果的例子。而且，在图 28 中，要素元数据的一部分省略记述。

在图 28 的例子中，作为客观性、可靠性都高的信息，即作为事实的可能性高的信息，有“A200”的“包”的“容量”为“20 升”的信息。而且，可知作为与其相关的客观性低的信息，即被认为意见的信息，虽然在公司的网页中，其容量被评价为“大的”，但是，在个人网页中，在“海外出差用”中被一名男性评价为“不够”，在“国内出差用”中被一名女性评价为“太大”，被一名男性评价为“足够”。

接着，如果从用户请求处理单元 302 有用户请求的指定，则元数据输出形式生成单元 304 按照用户请求的指定生成元数据输出形式，并通过元数据输出单元 306 对用户提示，但是，此前的流程与实施方式 3 一样。而且，在本实施方式中，也可以将要素元数据的评价数据指定作为用户请求。这里，以以下情况为例进行说明，即从图 28 的元数据的综合结果，接受包含如下那

样的要素元数据的评价数据的用户请求的指定，元数据输出形式生成单元 304 对与用户指定的条件一致的元数据生成元数据表的情况。

#### 用户请求的指定

话题事物：A200

客观性：0

文本的种类：个人网页

该指定是作为被写在个人网页的文本中的关于“A200”的事物的评价数据，求“客观性为 0”的记述，即意见。而且，上述是用户请求的指定方法的一例，指定方法不限于上述情况。

图 29 表示通过上述的用户请求的指定，与实施方式 3 一样生成的元数据表的例子。从图 29 的元数据表可知，作为被写在与 A200 的事物有关的个人网页的文本中的意见，容量和用途的观点被采纳，作为用途对于海外出差用、国内出差用两个用途进行评价，容量对于海外出差用被评价为不够（男性 1 名），对于国内出差用被评价为太大（女性 1 名）、足够（男性 1 名）。

这样，按照本实施方式，可以将文本中被表现的与事物有关的事实和意见的记述内容与被推定的话题的事物一起，将事物与意见附带对应来提取。而且，以容易进行被提取的事物和意见的关联性的比较的形式提取，并将事实和意见对于某个话题事物附带对应后，还包含客观性和可靠性的评价结果来对用户提示。由此，可以适当地解释用于提示用户的信息，并且仅确切地选择用户需要的信息。

本发明虽然根据附图所示的优选实施例进行了说明，但是只要是本技术领域的技术人员就明白，可以容易地变更和改变本发明，这样的变更部分也被包含在发明的范围内。

#### 本发明在产业上的可利用性

与本发明有关的信息提取系统具有观点/记述提取单元、观点/记述提取规则存储单元、元数据存储单元，并作为信息提取系统、信息检索系统有用。而且，也可以应用在信息分析/评价系统、信息发信系统等用途中。

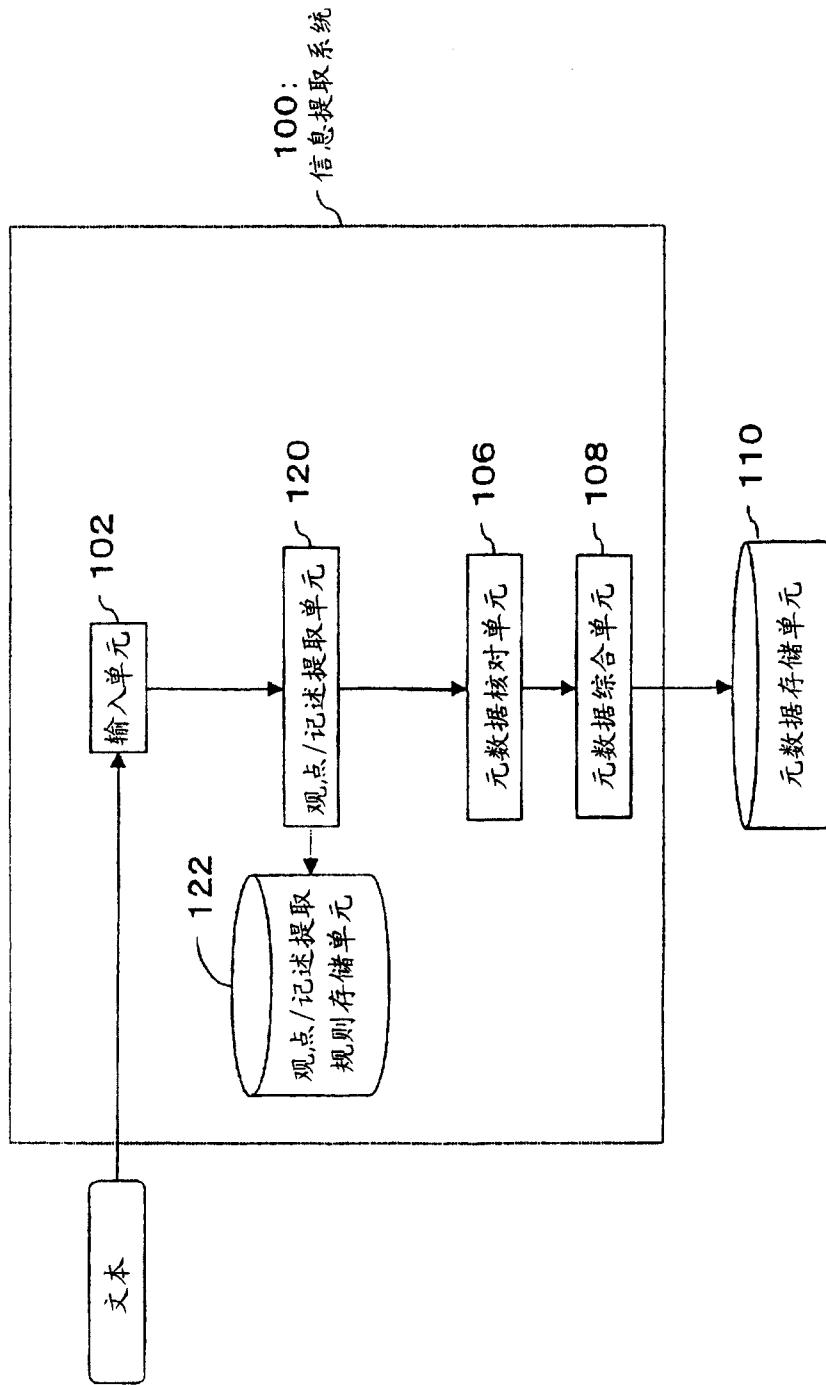


图 1

## (a) 输入文本例

|     |  |
|-----|--|
| 文本1 | A公司的小型包A20开口部为30cm相当大。拉链的开闭也顺滑。<br>皮的触感温和湿润。 |
| 文本2 | A20的皮的手感惊人的光滑。色调也温和湿润。                       |

## (b) 观点/记述认定例



|     |   |
|-----|---|
| 文本1 | A公司的小型包A20<VIEW1>开口部</VIEW1>为<DESC1a>30cm</DESC1a>&<DESC1b>相当大</DESC1b>。<VIEW2>拉链的开闭</VIEW2>也<DESC2>顺滑</DESC2>。<VIEW3>皮的触感</VIEW3>!<br><DESC3>温和湿润</DESC3>。 |
| 文本2 | A20!<VIEW1>皮的手感</VIEW1>为<DESC1>惊人的光滑</DESC1>。<VIEW2>也<DESC2>温和湿润</DESC2>。   |

## (c) 要素元数据提取结果



| 要素元数据 ID | 要素元数据 | 记述   | 观点    |
|----------|-------|------|-------|
| 1-1a     |       | 30cm | 开口部   |
| 1-1b     |       | 相当大  |       |
| 1-2      | 顺滑    |      | 拉链的开闭 |
| 1-3      | 温和湿润  |      | 皮的触感  |
| 2-1      | 惊人的光滑 |      | 皮的手感  |
| 2-2      | 温和湿润  |      | 色调    |

图 2

## (a) 观点/记述提取规则例

| 规则 | 规范   | 观点  | 记述       |
|----|--|-----|----------|
| 1  | [は( { 汉字 / 平假名连续1 } ) [がも] ( { 英数字连续1 } ) ] + ( { 汉字 / 平假名连续1 } ) \$1 \$2    \$3<br>1 } { 形容词词尾1 } | \$1 | \$2  \$3 |
| 2  | [は( { 汉字 / 平假名连续1 } ) [がも] ( { 汉字 / 平假名连续1 } ) { 形容动词词尾1 } ) \$1 \$2                               | \$1 | \$2      |
| 3  | [は( { 汉字 / 平假名连续1 } ) [がも] ( { 汉字 / 平假名连续1 } ) { 形容词词尾1 } ) \$1 \$2                                | \$1 | \$2      |

观点/记述提取规则的构成要素定义例

[]:[] 内的其中之一

+; 之前的规范要素的一次以上的重复

(): 参照后面(依次以"\$ (整数)"被参照)

\$ { 整数 }: 变量(在规范中与在(整数)号码中用"()"括住的部分匹配的字符串)

## (b) 观点/记述提取规则的构成要素定义例

| 构成要素名       | 定义            |
|-------------|---------------|
| 汉字 / 平假名连续1 | [あ-ん垂-瑠]+     |
| 英数字连续1      | [0-9A-Z]+     |
| 形容动词词尾1     | だろ,た,で,だつ,に,な |
| 形容词词尾1      | く,かつ,う,ゆう,い   |

图 3

| 元数据综合结果 |       | 记述    | 观点/记述对ID |
|---------|-------|-------|----------|
| 观,点     |       |       |          |
| 开 口 部   | 30cm  | 1-1 a |          |
|         | 相当大   | 1-1 b |          |
| 拉链的开闭   | 顺滑    | 1-2   |          |
|         | 温和湿润  | 1-3   |          |
| 皮的触感    | 惊人的光滑 | 2-1   |          |
|         | 温和湿润  | 2-2   |          |
| 色 调     |       |       |          |

图 4

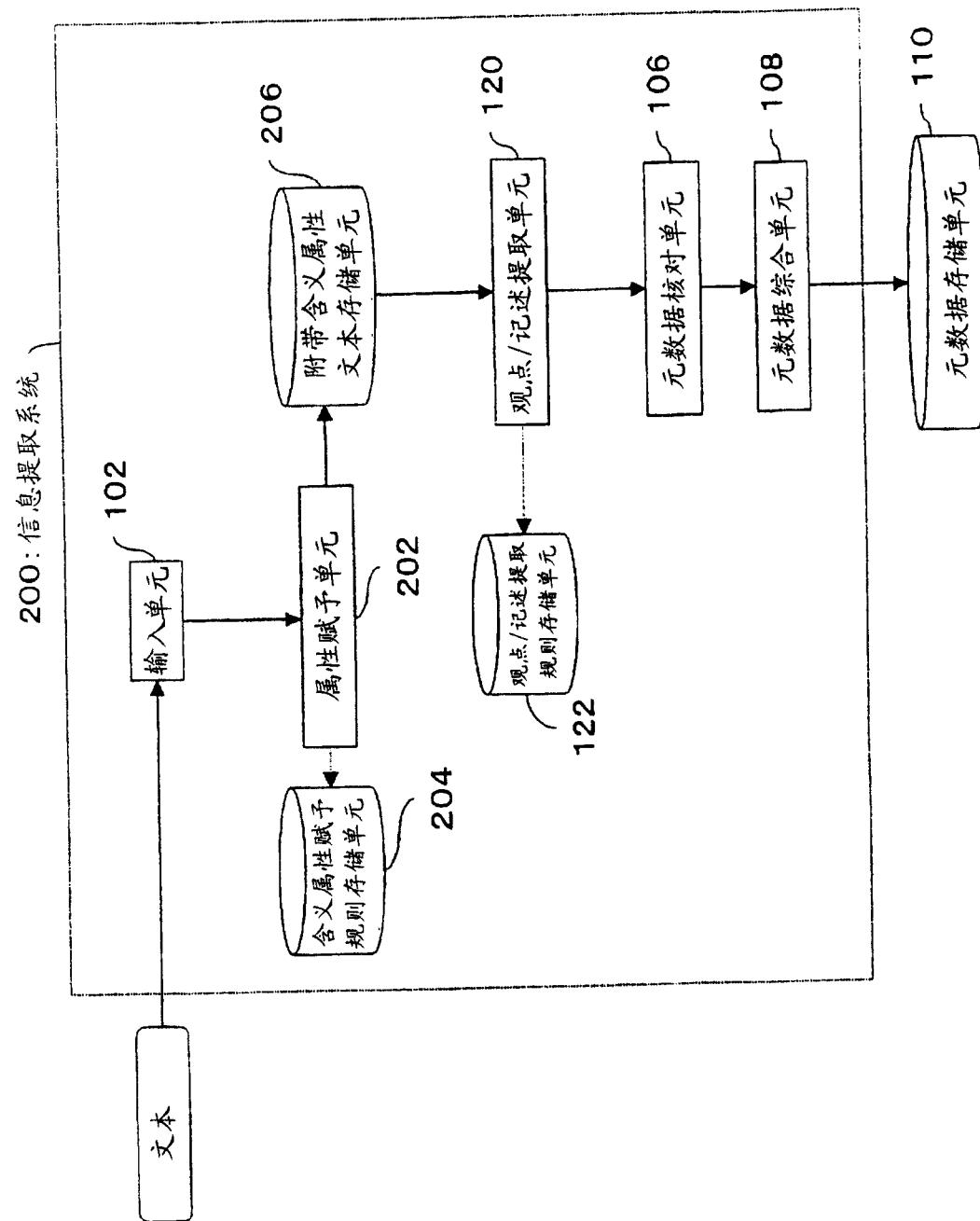


图 5

(a) 文本例

|     |                  |
|-----|------------------|
| 文本1 | A公司的包A200容量大至20升 |
| 文本2 | 认为包A-200容量不足够    |



(b) 含义属性赋予例

|     |   |
|-----|---|
| 文本1 | <pre>&lt;ORGANIZATION type=company&gt; A公司 &lt;/ORGANIZATION&gt; 的&lt;PROD_TYPE&gt; 包 &lt;/PROD_TYPE&gt;&lt;PROD_NAME&gt;A200&lt;/PROD_NAME&gt;[<br/> &lt;QUANT_TYPE&gt;容量 &lt;/QUANT_TYPE&gt;が&lt;QUANT unit=1, val=20&gt;20升<br/> &lt;/QUANT&gt; 大。</pre> |
| 文本2 | <pre>&lt;PROD_TYPE&gt; 包 &lt;/PROD_TYPE&gt;&lt;PROD_NAME&gt;A-200&lt;/PROD_NAME&gt;<br/> は&lt;QUANT_TYPE&gt;容量 &lt;/QUANT_TYPE&gt;认为不足够。</pre>  |

图 6

(a) 含义属性赋予规则例

| 规则 | 规范                        | 对象部分    | 含义属性         |                                      |
|----|---------------------------|---------|--------------|--------------------------------------|
|    |                           |         | 含义分类         | 详细信息                                 |
| 1  | ({{ 数字连续 }{ 数量单元 }})      | \$1 \$2 | QUANT        | 'unit': 数量单位的表现<br>'val': 将数值连续正规化的值 |
| 2  | ({{ 数量分类 }})              | \$1     | QUANT_TYPE   |                                      |
| 3  | ({{英字母连续}}{ 公司名后接表现 })    | \$1     | ORGANIZATION | 'type:company'                       |
| 4  | ({{ 产品分类名 }})             | \$1     | PROD_TYPE    |                                      |
| 5  | ({{ 产品分类名 }}{ 英数字记号连续1 }) | \$1     | PROD_NAME    |                                      |

(b) 含义属性赋予规则构成要素的定义例

| 构成要素名    | 定义                 |
|----------|--------------------|
| 公司名后接表现  | '公司'               |
| 数量分类     | '容量'               |
| 数量单元     | '升'、'卡'、'克' . . .  |
| 产品分类名    | '包'、'鞋'、'帽子' . . . |
| 数字连续     | [0-9]+             |
| 英字母连续    | [A-Z]+             |
| 英数字记号连续1 | [-9-zA-Z]+         |

图 7

## (a) 附带含义属性文本例

|     |   |
|-----|---|
| 文本1 | <ORGANIZATION type=company> A公司 </ORGANIZATION> 的 <PROD_TYPE><br>包 </PROD_TYPE><PROD_NAME>A200</PROD_NAME>[足<br><QUANT_TYPE>容量 </QUANT_TYPE> が <QUANT unit=l, val=20>20升<br></QUANT> 大。 |
| 文本2 | <PROD_TYPE> 包 </PROD_TYPE><PROD_NAME>A-200</PROD_NAME><br>[足<QUANT_TYPE>容量 </QUANT_TYPE> 认为不够。  |



## (b) 观点 / 记述认定例

|     |   |
|-----|---|
| 文本1 | <DESC1><ORGANIZATION type=company> A公司 </ORGANIZATION><DESC1><br>的 <DESC2><PROD_TYPE> 包 </PROD_TYPE></DESC2> <DESC3><br><PROD_NAME>A200</PROD_NAME>[足<VIEW4><QUANT_TYPE><br>容量 </QUANT_TYPE><VIEW4>が <DESC4a><QUANT unit=l, val=20>20升<br>ル</QUANT><DESC4a>と<DESC4b> 大 </DESC4b>。 |
| 文本2 | <DESC1><PROD_TYPE> 包 </PROD_TYPE></DESC1><DESC2><br><PROD_NAME>A-200</PROD_NAME></DESC2><VIEW3>[足<QUANT_TYPE><br>容量 </QUANT_TYPE><VIEW3>が <DESC3> 不够 </DESC3> 认为。   |

图 8

(a) 观点 / 记述提取规则例

| 规则 | 规范   | 观点               | 记述         |
|----|--|------------------|------------|
| 1  | <QUANT_TYPE>({ 标记开始记号以外的任意的字符串 })<QUANT_TYPE>[が は]<QUANT>({ 记开始记号以外的任意的字符串 })<QUANT_TYPE>[が は]<{ 汉字/平假名连续 } { 形容词词尾 }> | \$1              | \$2    \$3 |
| 2  | <QUANT_TYPE>({ 标记开始记号以外的任意的字符串 })<QUANT_TYPE>[が は]<{ 汉字/平假名连续 } { 形容动词词尾 }>  | \$1              | \$2        |
| 3  | { 标记开始记号以外的任意的字符串 }*<{ 含义属性1 }>({ 标记开始记号以外的任意的字符串 }*)<{ 含义属性1 }>   | \$1 的含义属性<br>的别名 | \$2        |
| 4  | { 结束标记 }<{ 含义属性1 }><{ 标记开始记号以外的任意的字符串 }><{ 含义属性1 }>  | \$1 的含义属性<br>的别名 | \$2        |

+ : 之前的规范要素1次以上的重复  
 \* : 之前的规范要素0次以上的重复

(b) 观点 / 记述提取规则构成要素定义例

| 构成要素名            | 定义   |
|------------------|--|
| 形容动词词尾1          | 'たろ'、'だ'、'飞'、'たづ'、'に'、'な'  |
| 形容词词尾1           | 'く'、'かづ'、'う'、'ゆう'、'い'  |
| 含义属性1            | 'ORGANIZATION'、'ORGANIZATION type=company'、<br>'PROD_TYPE'、'PROD_NAME'、<br>'PERSON'、'DATE'、'TIME'、'PERIOD'、... |
| 标记开始记号以外的任意的字符串1 | [`<`]+   |
| 标记结束记号以外的任意的字符串1 | [`>`]  |
| 结束标记             | </[`>`]+>  |

图 9

| 要素元数据提取结果<br>观点 | 记述        | 含义属性         |                | 要素元<br>数据ID  |
|-----------------|-----------|--------------|----------------|--------------|
|                 |           | 含义分类         | 详细信息           |              |
| 公司名             | A公司       | ORGANIZATION | type=company   | 1—1          |
| 产品分类            | 包         | PROD_TYPE    |                | 1—2          |
| 产品名             | A200      | PROD_NAME    |                | 1—3          |
|                 |           | QUANT_TYPE   | -              |              |
| 容量              | 20升<br>大的 | QUANT        | unit=1, val=20 | 1—4a<br>1—4b |
| 产品分类            | 包         | PROD_TYPE    |                | 2—1          |
| 产品名             | A-200     | PROD_NAME    |                | 2—2          |
| 容量              | 不够        | QUANT_TYPE   |                | 2—3          |

图 10

元数据综合结果

| 观 点  | 记述              | 含义属性         |                | 要素元<br>数据 ID        |
|------|-----------------|--------------|----------------|---------------------|
|      |                 | 含义分类         | 详细信息           |                     |
| 公司名  | A公司             | ORGANIZATION | type=company   | 1-1                 |
| 产品分类 | 包               | PROD_TYPE    |                | 1-2<br>2-1          |
| 产品名  | A200            | PROD_NAME    |                | 1-3<br>2-2          |
|      |                 | QUANT_TYPE   |                | -                   |
| 容量   | 20升<br>大的<br>不够 | QUANT        | unit=l, val=20 | 1-4a<br>1-4b<br>2-3 |

图 11

300：信息提取系统

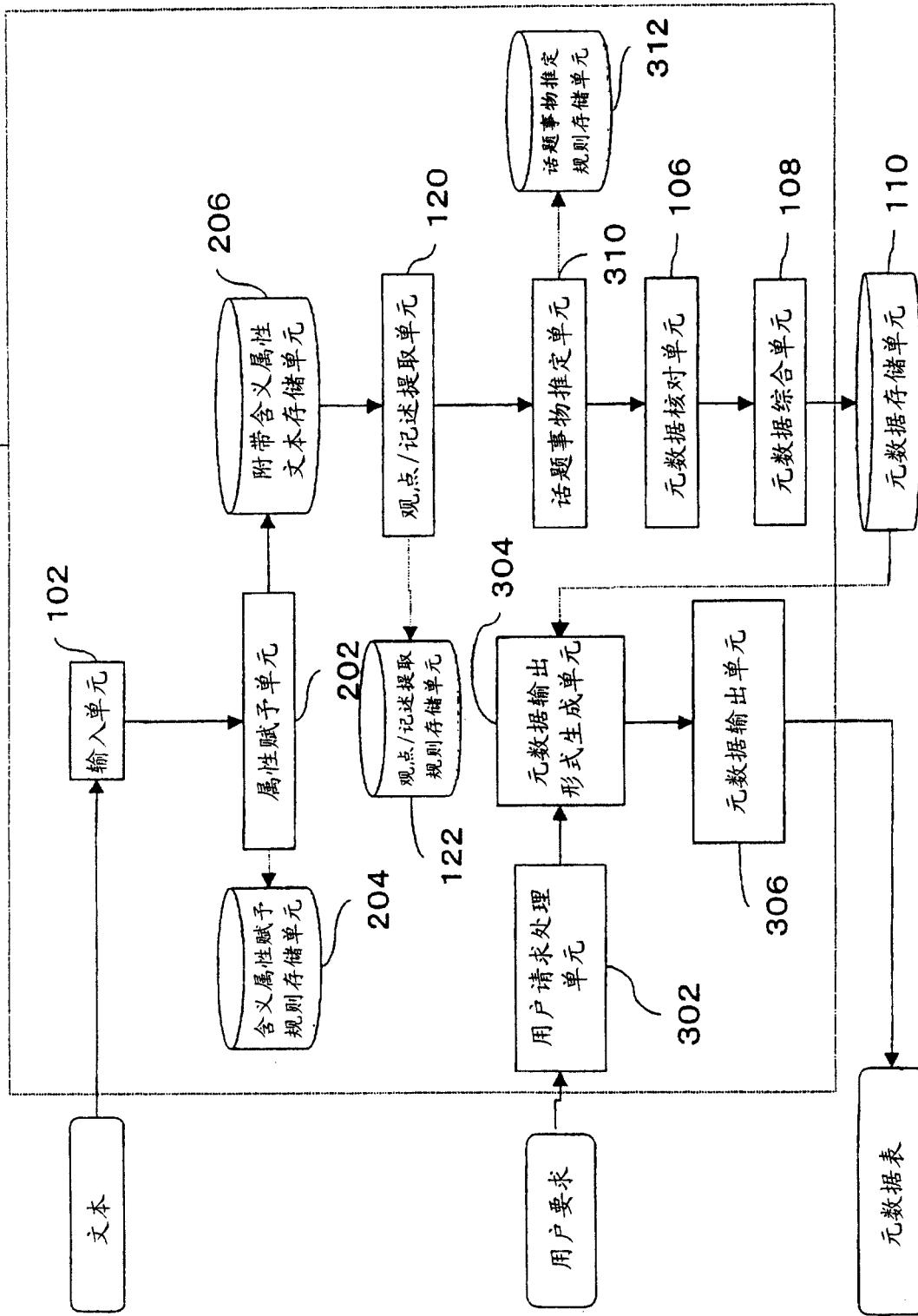


图 12

## (a) 观点/记述认定例

|     |  |
|-----|--|
| 文本1 | <DESC1><PROD_TYPE> 包 </PROD_TYPE></DESC1><DESC2><PROD_NAME>A200</PROD_NAME></DESC2>は<VIEW3><QUANT_TYPE>容量 </QUANT_TYPE></VIEW3>が<DESC3> 不足够 </DESC3>し、<DESC4><PROD_TYPE> 包 </PROD_TYPE></DESC4><DESC5><PROD_NAME>A300</PROD_NAME></DESC5>は<VIEW6><QUANT_TYPE> 容量 </QUANT_TYPE></VIEW6>が<VIEW7> 太大 </VIEW7>。  |
| 文本2 | <DESC1><PROD_TYPE> 包 </PROD_TYPE></DESC1><DESC2><PROD_NAME>A200</PROD_NAME></DESC2>は<VIEW3><QUANT_TYPE> 容量 </QUANT_TYPE></VIEW3>が<DESC3><QUANT unit=l val=20> 20升 </QUANT></DESC3>で、<DESC4><PROD_TYPE> 包 </PROD_TYPE></DESC4><DESC5><PROD_NAME>A300</PROD_NAME></DESC5>の<VIEW6><QUANT_TYPE> 容量 </QUANT_TYPE></VIEW6>は<VIEW7><QUANT unit=l val=30>30升 </QUANT></VIEW7>。 |

## (b) 要素元数据提取结果例

| 观点   | 记述   | 含义属性       |                   | 要素的数据ID |
|------|------|------------|-------------------|---------|
|      |      | 含义分类       | 详细信息              |         |
| 产品分类 | 包    | PROD_TYPE  |                   | 1-1     |
| 产品名  | A200 | PROD_NAME  |                   | 1-2     |
|      |      | QUANT_TYPE |                   | 1-3     |
| 容量   | 不够   |            |                   |         |
| 产品分类 | 包    | PROD_TYPE  |                   | 1-4     |
| 产品名  | A300 | PROD_NAME  |                   | 1-5     |
|      |      | QUANT_TYPE |                   | 1-6     |
| 容量   | 太大   |            |                   |         |
| 产品分类 | 包    | PROD_TYPE  |                   | 2-1     |
| 产品名  | A200 | PROD_NAME  |                   | 2-2     |
|      |      | QUANT_TYPE |                   | 2-3     |
| 容量   | 20升  | QUANT      | unit=l,<br>val=20 |         |
| 产品分类 | 包    | PROD_TYPE  |                   | 2-4     |
| 产品名  | A300 | PROD_NAME  |                   | 2-5     |
|      |      | QUANT_TYPE |                   | 2-6     |
| 容量   | 30升  | QUANT      | unit=l,<br>val=30 |         |

图 13

## (a) 话题事物推定规则的例子

| 规则 | 条件   | 被推定的话题事物                       |
|----|--|--------------------------------|
| 1  | <DESC[0-9]+><{PROD_TYPE PERSON}>{<br>标记开始记号以外的任意的字符串<[<br>PROD_TYPE PERSON]></DESC[0-9]+>  | 记述\$1的要素<br>元数据的话题事物是\$1       |
| 2  | ({ 标记开始记号以外的任意的字符串<br>})</{PROD_TYPE PERSON}></DESC[0-9]+>[<br>(<VIEW[0-9]+>{ 标记开始记号以外的任意的字符串<br>></VIEW[0-9]+>{ 标记开始记号<br>以外的任意的字符串 }</DESC[0-9]+><br>这里\$1和\$4, \$5和\$7的字符串相同) | 观点\$3记述\$6的要素的数据<br>的话题中事物是\$1 |
| 3  | ({ 标记开始记号以外的任意的字符串<br>})</{PROD_TYPE PERSON}></DESC[0-<br>9]+><{PROD_TYPE PERSON}><br><DESC[0-9]+><{PROD_TYPE PERSON}><br><{ 标记开始记号以外的任意的字符串<br>></PROD_TYPE PERSON>             | 记述\$2的要素元数据的<br>话题事物是\$4       |

规则的记法的说明

(A|B) : A和B的其中一个

(b) 话题事物推定规则构成要素定义的例子

|         |               |
|---------|---------------|
| 构成要素名   | 定义            |
| 标记构成字符串 | [0-9A-Za-z_]+ |
| 任意字符串   | +             |

图 14

| 观点   | 记述   | 含义属性       |                   | 要素元数据ID | 话题事物 |
|------|------|------------|-------------------|---------|------|
|      |      | 含义分类       | 详细信息              |         |      |
| 产品分类 | 包    | PROD_TYPE  |                   | 1-1     | A200 |
| 产品名  | A200 | PROD_NAME  |                   | 1-2     | A200 |
|      |      | QUANT_TYPE |                   | 1-3     | A200 |
| 容量   | 不足够  |            |                   |         |      |
| 产品分类 | 包    | PROD_TYPE  |                   | 1-4     | A300 |
| 产品名  | A300 | PROD_NAME  |                   | 1-5     | A300 |
|      |      | QUANT_TYPE |                   | 1-6     | A300 |
| 容量   | 太大   |            |                   |         |      |
| 产品分类 | 包    | PROD_TYPE  |                   | 2-1     | A200 |
| 产品名  | A200 | PROD_NAME  |                   | 2-2     | A200 |
|      |      | QUANT_TYPE |                   | 2-3     | A200 |
| 容量   | 20升  | QUANT      | unit=1,<br>val=20 |         |      |
| 产品分类 | 包    | PROD_TYPE  |                   | 2-4     | A300 |
| 产品名  | A300 | PROD_NAME  |                   | 2-5     | A300 |
|      |      | QUANT_TYPE |                   | 2-6     | A300 |
| 容量   | 30升  | QUANT      | unit=1,<br>val=30 |         |      |

图 15

| 综合结果例<br>话题事物 | 观点点  | 记述   | 含义属性       |                   | 要素元<br>数据ID |
|---------------|------|------|------------|-------------------|-------------|
|               |      |      | 含义分类       | 详细信息              |             |
| A200          | 产品分类 | 包    | PROD_TYPE  |                   | 1-1<br>2-1  |
|               | 产品名  | A200 | PROD_NAME  |                   | 1-2<br>2-2  |
|               |      |      | QUANT_TYPE |                   | -           |
|               | 容量   | 20升  | 不足够        |                   | 1-3         |
|               |      |      | QUANT      | unit=1,<br>val=20 | 2-3         |
|               |      |      | PROD_TYPE  |                   | 1-4<br>2-4  |
| A300          | 产品分类 | 包    | PROD_NAME  |                   | 1-5<br>2-5  |
|               | 产品名  | A300 | QUANT_TYPE |                   | -           |
|               |      |      | 太大         |                   | 1-6         |
|               | 容量   | 30升  | QUANT      | unit=1,<br>val=30 | 2-6         |
|               |      |      | PROD_TYPE  |                   | -           |
|               |      |      | QUANT      |                   | -           |

图 16

| 话题事物 | 观点  | 记述 | 含义属性       |                  | 要素元数据ID |
|------|-----|----|------------|------------------|---------|
|      |     |    | 含义分类       | 详细信息             |         |
| A200 |     |    | QUANT_TYPE | -                |         |
|      | 不够  |    |            | 1-3              |         |
|      | 20升 |    | QUANT      | unit=1<br>val=20 | 2-3     |
|      | 容量  |    |            |                  |         |

图 17

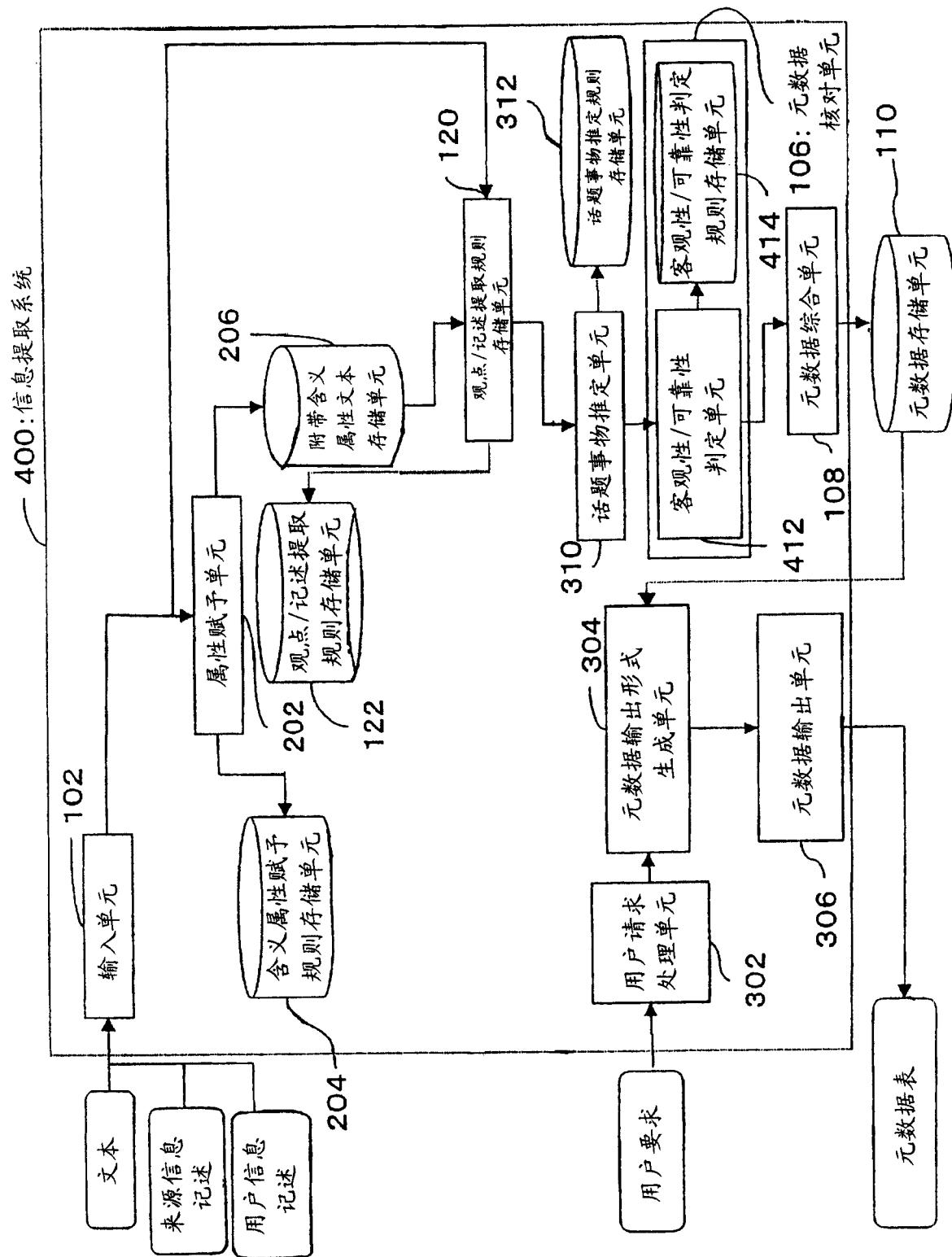


图 18

(a) 来源信息记述的例子

|         |   |
|---------|---|
| 来源信息记述1 | <a href="http://www.aaa.co.jp/article1">http://www.aaa.co.jp/article1</a><br>作成日：2003年10月1日 |
| 来源信息记述2 | <a href="http://www.ddd.fff.jp/~zzzz">http://www.ddd.fff.jp/~zzzz</a><br>作成日：2003年5月1日      |

(b) 用户信息记述的例子

|         |               |
|---------|---------------|
| 用户信息记述1 | 公司名 :aaa      |
| 来源信息记述2 | 作者zzzz。20多岁男性 |

(c) 附带含义属性的来源信息记述的例子

|         |  |
|---------|--|
| 用户信息记述1 | <URL type=公司网页> <a href="http://www.aaa.co.jp/article1">http://www.aaa.co.jp/article1</a> </URL><br><DATE value=2003:10:01> 2003年10月1日 </DATE> |
|---------|--|

(d) 附带含义属性的用户信息的例子

|         |  |
|---------|--|
| 用户信息记述1 | 作者<AUTHOR>zzzz</AUTHOR>。<AGE value=20:29> 20多岁<br></AGE><GENDER type=M> 男性</GENDER>。 |
|---------|--|

图 19

## (a) 附带来源含义属性的规则的例子

| 规则 | 规范                            | 含义分类      | 含义属性                             |
|----|-------------------------------|-----------|----------------------------------|
| 1  | (http://.*.co*.jp.*)          | URL( 网页 ) | type=company                     |
| 2  | (数字4行) 年(数字1~2行)<br>月(数字1~2行) | 日期        | 这里在\$1、\$2、\$3为1行的情况下<br>将0附加在前面 |

¥：“.” (句号)

## (b) 附带用户含义属性的规则的例子

| 规则 | 规范            | 含义分类   | 含义属性                           |
|----|---------------|--------|--------------------------------|
| 1  | 作者[:]*([. ]+) | AUTHOR |                                |
| 2  | (数字1行) 0代     | AGE    | value=(\\$1*10+1): (\\$1*10+9) |
| 3  | 男性            | GENDER | value=M                        |

[. ]: 与□内的字符不一致的字符

图 20

(a) 来源观点/记述提取规则的例子

| 规则 | 规范   | 观点    | 记述  |
|----|--|-------|-----|
| 1  | <URL type="公司网页">{{ 标记开始记号以外<br>的标志的字符串 }}</URL> | 文本取得元 | \$1 |
| 2  | ( 作成日 ):< 日期 >{{ 标记开始记号以外的标志的<br>字符串 }}</ 日期 >   | \$1   | \$2 |

(b) 用户观点/记述提取规则的例子

| 规则 | 规范   | 观点 | 记述  |
|----|--|----|-----|
| 1  | <AUTHOR>{{ 标记开始记号以外的标志的字符串<br>}}</AUTHOR>          | 作者 | \$1 |
| 2  | <AGE value="20:29">{{ 标记开始记号以外的标志<br>的字符串 }}</AGE> | 年龄 | \$1 |
| 3  | <GENDER type=M>{{ 标记开始记号以外的标志<br>的字符串 }}</GENDER>  | 性别 | 男性  |

图 21

(a) 来源元数据提取结果的例子

| 观点              | 记述   | 含义属性    |                  | 来源元数据 ID |
|-----------------|--|---------|------------------|----------|
|                 |  | 含义分类    | 详细信息             |          |
| 文本的取得元<br>文本的分类 | <a href="http://www.aaa.co.jp/article1">http://www.aaa.co.jp/article1</a><br>公司网页<br>( <a href="http://www.aaa.co.jp/article1">http://www.aaa.co.jp/article1</a> ) | URL(网页) | type=company     | 1—S1     |
| 作成日             | 2003年10月1日   | 日期      | value=2003:10:01 | 1—S3     |

(b) 用户观点/记述提取规则的例子

| 观点 | 记述    | 含义属性   |             | 用户元数据 ID |
|----|-------|--------|-------------|----------|
|    |       | 含义分类   | 详细信息        |          |
| 作者 | zzz   | AUTHOR |             | 2—U1     |
| 年龄 | 20-29 | AGE    | value=20:29 | 2—U2     |
| 性别 | 男性    | GENDER | value=M     | 2—U3     |

图 22

| 规则     | 文本的要素的数据      |         |             | 条件 |                 | 可靠性   |
|--------|---------------|---------|-------------|----|-----------------|-------|
|        | 观点            | 记述的含义分类 | 观点          | 记述 | 来源元数据           |       |
| 1 组织名  |               |         |             |    |                 | 1 1   |
| 2 产品分组 |               |         |             |    |                 | 1 1   |
| 3 产品名  |               |         |             |    |                 | 1 1   |
| 4 用途   | USAGE         | 文本的分类   | 公司网页 / 报纸报道 |    |                 | 1 1   |
| 5 用途   | USAGE         | 文本的分类   | 个人网页        |    |                 | 0 0.5 |
| 6 容量   | QUANT         | 文本的分类   | 公司网页 / 报纸报道 |    |                 | 1 1   |
| 7 容量   | QUANT         | 文本的分类   | 个人网页        |    |                 | 1 0.9 |
| 8 容量   | 无, 或 QUANT 以外 |         |             |    | 句未为 {不确定表现1}    | 0 0.2 |
| 9 容量   | 无, 或 QUANT 以外 | 文本的分类   | 公司网页 / 报纸报道 |    | 句未为 {不确定表现1} 以外 | 0 0.5 |
| 10 容量  | 无, 或 QUANT 以外 | 文本的分类   | 个人网页        |    | 句未为 {不确定表现1} 以外 | 0 0.3 |

客观性/可靠性判定规则例  
不确定表现1：と思う,と思える,と思われる,かも知れない,らしい

图 23

## (a) 文本例

- |     |                    |
|-----|--------------------|
| 文本1 | 包A200的容量大至20升。     |
| 文本2 | 包A200的容量对于海外出差不够。  |
| 文本3 | 包A200的容量对于国内出差太大。  |
| 文本4 | 包A200的容量对于国内出差用足够。 |

## (b) 附带含义属性的文本例

- |     |   |
|-----|---|
| 文本1 | <pre>&lt;ORGANIZATION type=company&gt; A公司 &lt;/ORGANIZATION&gt; 的 &lt;PROD_TYPE&gt; 包 &lt;PROD_TYPE&gt;&lt;PROD_NAME&gt;A200&lt;/PROD_NAME&gt;は &lt;QUANT_TYPE&gt;容量 &lt;/QUANT_TYPE&gt;が&lt;QUANT unit=l, val=20&gt; 20升 &lt;/QUANT&gt; 大至。</pre> |
| 文本2 | <pre>&lt;PROD_TYPE&gt; 包 &lt;/PROD_TYPE&gt;&lt;PROD_NAME&gt;A200&lt;PROD_NAME&gt; 的 &lt;QUANT_TYPE&gt;容量 &lt;/QUANT_TYPE&gt;は&lt;USAGE&gt; 海外出差用 &lt;/USAGE&gt; 认为不够。</pre>   |
| 文本3 | <pre>&lt;PROD_TYPE&gt; 包 &lt;/PROD_TYPE&gt;&lt;PROD_NAME&gt;A200&lt;PROD_NAME&gt; 的 &lt;QUANT_TYPE&gt;容量 &lt;/QUANT_TYPE&gt;は&lt;USAGE&gt; 国内出差用 &lt;/USAGE&gt; 太大。</pre>   |
| 文本4 | <pre>&lt;PROD_TYPE&gt; 包 &lt;/PROD_TYPE&gt;&lt;PROD_NAME&gt;A200&lt;PROD_NAME&gt; 的 &lt;QUANT_TYPE&gt;容量 &lt;/QUANT_TYPE&gt;は&lt;USAGE&gt; 国内出差用 &lt;/USAGE&gt; 足够。</pre>   |

图 24

(a) 观点 / 记述提取规则例

| 规则 | 规范  | 观点              | 记述       |
|----|---|-----------------|----------|
| 1  | <QUANT_TYPE>(< 标记开始记号以外的任意的字符串 ><br>(</QUANT_TYPE>)[が は]<QUANT>(< 标记开始记号以外的任意的字符串 >)<br></QUANT><{ 汉字 / 平假名连续 1 }> { 形容词词尾 1 } )      | \$1             | \$2  \$3 |
| 2  | <QUANT_TYPE>(< 标记开始记号以外的任意的字符串 ><br>(</QUANT_TYPE>)[が は]<USAGE>(< 标记开始记号以外的任意的字符串 ><br><用>)</USAGE>(<{ 汉字 / 平假名连续 1 }> { 形容动词词尾 1 } ) | \$1             | \$2&&\$3 |
| 3  | <QUANT_TYPE>(< 标记开始记号以外的任意的字符串 ><br></QUANT_TYPE>[が は]<{ 汉字 / 平假名连续 1 }> { 形容动词词尾 1 } )   | \$1             | \$2      |
| 4  | { 标记开始记号以外的任意的字符串 } * <{ 含义属性 }><br>({ 标记开始记号以外的任意的字符串 })<{ 含义属性 1 }>   | \$1的含义<br>属性的别名 | \$2      |
| 5  | { 结束标记 } <{ 含义属性 1 }> <{ 标记开始记号以外的任意的字符串 }><br><{ 含义属性 1 }>   | \$1的含义<br>属性的别名 | \$2      |

(b) 观点 / 记述提取规则构成要素定义例

|                 |          |
|-----------------|----------|
| 构成要素名           | 定义       |
| 标记结束记号以外的任意的字符串 | [^>]     |
| 结束标记            | </[^>]+> |

图 25

要素元数据提取结果例

| 观 点     | 记 述       | 含 义 屬 性    |                | 要 素 元<br>数 据 ID | 话 题 事 物 |
|---------|-----------|------------|----------------|-----------------|---------|
|         |           | 含 义 分 类    | 详 细 信 息        |                 |         |
| 产品分 类   | 包         | PROD_TYPE  |                | 1-1             | A-200   |
| 产 品 名   | A200      | PROD_NAME  |                | 1-2             | A-200   |
|         |           | QUANT_TYPE |                | -               | A-200   |
|         |           | QUANT      | unit=1, val=20 | 1-3a            |         |
| 容 量     | 20升       |            |                | 1-3b            |         |
|         | 大 き い     |            |                |                 |         |
| 产 品 分 类 | 包         | PROD_TYPE  |                | 2-1             | A-200   |
| 产 品 名   | A200      | PROD_NAME  |                | 2-2             | A-200   |
| 容 量     |           | QUANT_TYPE |                | 2-3             | A-200   |
|         | 海 外 出 差 用 | USAGE      |                |                 |         |
|         | 不 足 够     |            |                |                 |         |
| 用 途     | 海 外 出 差 用 | USAGE      |                | 2-4             | A-200   |
| 产 品 分 类 | 包         | PROD_TYPE  |                | 3-1             | A-200   |
| 产 品 名   | A200      | PROD_NAME  |                | 3-2             | A-200   |
| 容 量     |           | QUANT_TYPE |                | 3-3             | A-200   |
|         | 国 内 出 差 用 | USAGE      |                |                 |         |
|         | 太 大       |            |                |                 |         |
| 用 途     | 国 内 出 差 用 | USAGE      |                | 3-4             | A-200   |
| 产 品 分 类 | 包         | PROD_TYPE  |                | 4-1             | A-200   |
| 产 品 名   | A200      | PROD_NAME  |                | 4-2             | A-200   |
| 容 量     |           | QUANT_TYPE |                | 4-3             | A-200   |
|         | 国 内 出 差 用 | USAGE      |                |                 |         |
|         | 足 够       |            |                |                 |         |
| 用 途     | 国 内 出 差 用 | USAGE      |                | 4-4             | A-200   |

图 26

## 客观性/可靠性判定结果例

|      |             | 要素元数据      |         |     |     | 来源元数据 |    |         |      | 用户元数据 |         |
|------|-------------|------------|---------|-----|-----|-------|----|---------|------|-------|---------|
| 观点   | 记述          | 含义分类       | 要素元数据ID | 客观性 | 可靠性 | 观点    | 记述 | 来源元数据ID | 观点   | 记述    | 用户元数据ID |
| 产品分类 | 包           | PROD_TYPE  | 1-1     | 1.0 | 1.0 |       |    |         | 1-S2 | 性别    | -       |
| 产品名  | A200        | PROD_NAME  | 1-2     | 1.0 | 1.0 |       |    |         |      |       |         |
| 容量   | 20升<br>大的   | QUANT_TYPE | -       | -   | -   | 文本的分类 |    |         |      |       |         |
|      |             | QUANT      | 1-3a    | 1.0 | 1.0 |       |    |         |      |       |         |
| 产品分类 | 包           | PROD_TYPE  | 2-1     | 1.0 | 1.0 |       |    |         | 2-S2 | 男性    | 2-U3    |
| 产品名  | A200        | PROD_NAME  | 2-2     | 1.0 | 1.0 |       |    |         |      |       |         |
| 容量   | 海外出差用<br>不够 | QUANT_TYPE | -       | -   | -   | 个人网页  |    |         |      |       |         |
|      |             | USAGE      | 2-3     | 0   | 0.2 |       |    |         |      |       |         |
| 用途   | 海外出差用       | USAGE      | 2-4     | 0   | 0.5 |       |    |         |      |       |         |
| 产品分类 | 包           | PROD_TYPE  | 3-1     | 1.0 | 1.0 |       |    |         | 3-S2 | 女性    | 3-U3    |
| 产品名  | A200        | PROD_NAME  | 3-2     | 1.0 | 1.0 |       |    |         |      |       |         |
| 容量   | 国内出差用<br>太大 | QUANT_TYPE | -       | -   | -   | 个人网页  |    |         |      |       |         |
|      |             | USAGE      | 3-3     | 0   | 0.3 |       |    |         |      |       |         |
| 用途   | 国内出差用       | USAGE      | 3-4     | 0   | 0.5 |       |    |         |      |       |         |
| 产品分类 | 包           | PROD_TYPE  | 4-1     | 1.0 | 1.0 |       |    |         | 4-S2 | 男性    | 4-U3    |
| 产品名  | A200        | PROD_NAME  | 4-2     | 1.0 | 1.0 |       |    |         |      |       |         |
| 容量   | 国内出差用<br>足够 | QUANT_TYPE | -       | -   | -   | 个人网页  |    |         |      |       |         |
|      |             | USAGE      | 4-3     | 0   | 0.3 |       |    |         |      |       |         |
| 用途   | 国内出差用       | USAGE      | 4-4     | 0   | 0.5 |       |    |         |      |       |         |

图 27

元数据综合结果例

| 话题事物 | 观点           | 描述    | 含义分类      | 要素元数据      |      |     | 来源元数据 |      |         | 用户元数据 |      |         |
|------|--------------|-------|-----------|------------|------|-----|-------|------|---------|-------|------|---------|
|      |              |       |           | 要素元数据ID    | 客观性  | 可靠性 | 观点点   | 记述   | 来源元数据ID | 观点点   | 记述   | 用户元数据ID |
| A200 | 产品分类         | 包     | PROD_TYPE | 1-1        | 1.0  | 1.0 | 公司网页  | 1-S2 | 性别      | -     | -    | -       |
|      |              |       |           | 2-1        |      |     |       | 2-S2 |         | 男性    | 2-U3 |         |
|      |              |       |           | 3-1        |      |     |       | 3-S2 |         | 女性    | 3-U3 |         |
|      |              |       |           | 4-1        |      |     |       | 4-S2 |         | 男性    | 4-U3 |         |
|      | 产品名          | A200  | PROD_NAME | 1-2        | 1.0  | 1.0 | 公司网页  | 1-S2 | 性别      | -     | -    | -       |
|      |              |       |           | 2-2        |      |     |       | 2-S2 |         | 男性    | 2-U3 |         |
|      |              |       |           | 3-2        |      |     |       | 3-S2 |         | 女性    | 3-U3 |         |
|      |              |       |           | 4-2        |      |     |       | 4-S2 |         | 男性    | 4-U3 |         |
|      |              |       |           | QUANT_TYPE | -    |     |       |      |         | -     | -    | -       |
|      |              |       |           | QUANT      | 1-3a | 1.0 | 1.0   | 公司网页 | 1-S2    | -     | -    | -       |
| 容量   | 大的           | 20升   | USAGE     | 1-3b       | 0    | 0.5 | 2-S2  | 性别   | 男性      | 2-U3  |      |         |
|      |              |       |           | 2-3        | 0    | 0.2 | 3-S2  |      | 女性      | 3-U3  |      |         |
|      |              |       |           | -          |      |     | 4-S2  |      | 男性      | 4-U3  |      |         |
|      |              |       |           | 3-3        | 0    | 0.3 |       |      | 女性      | 3-U3  |      |         |
|      |              |       |           | -          |      |     |       |      | 男性      | 4-U3  |      |         |
|      | 不足以<br>国外出差用 | 国内出差用 | USAGE     | 4-3        | 0    | 0.3 | 个人网页  | 2-S2 | 性别      | 男性    | 2-U3 |         |
|      |              |       |           | 2-4        | 0    | 0.5 |       | 3-S2 |         | 女性    | 3-U3 |         |
|      |              |       |           | 3-4        |      |     |       | 4-S2 |         | 男性    | 4-U3 |         |
|      |              |       |           | 4-4        |      |     |       |      |         |       |      |         |
|      |              |       |           |            |      |     |       |      |         |       |      |         |
| 用途   | 海外出差用        | 国内出差用 | USAGE     |            |      |     |       |      |         |       |      |         |

|      |    | 要素元数据        |            |         |     |     | 来源元数据         |      |         | 用户元数据 |      |         |  |
|------|----|--------------|------------|---------|-----|-----|---------------|------|---------|-------|------|---------|--|
| 话题事物 | 观点 | 记述           | 含义分类       | 要素元数据ID | 客观性 | 可靠性 | 观点            | 记述   | 来源元数据ID | 观点    | 记述   | 用户元数据ID |  |
| A200 | 容量 | 海外出差用<br>不够  | QUANT_TYPE | -       | -   | -   | 文本的种类<br>个人网页 | 2-S2 | 性别      | -     | 男性   | 2-U3    |  |
|      |    |              | USAGE      | 2-3     | 0   | 0.2 |               |      |         |       |      |         |  |
|      |    | 国内出差用<br>太多  | USAGE      | 3-3     | 0   | 0.3 |               | 3-S2 |         | 女性    | 3-U3 |         |  |
|      |    |              | -          | -       | -   | -   |               | 4-S2 |         |       |      |         |  |
|      |    |              | USAGE      | 4-3     | 0   | 0.3 |               | 4-S2 |         |       |      |         |  |
|      | 用途 | 海外出差用<br>足够  | USAGE      | 2-4     | 0   | 0.5 | 4-S2          | 2-S2 | 男性      | 4-U3  | 男性   | 2-U3    |  |
|      |    |              | -          | -       | -   | -   |               |      |         |       |      |         |  |
|      |    | 国内出差用<br>4-4 | USAGE      | 3-4     | 0   | 0.5 | 3-S2          | 3-S2 | 女性      | 3-U3  | 女性   | 4-U3    |  |
|      |    |              | -          | -       | -   | -   |               |      |         |       |      |         |  |
|      |    |              | USAGE      | 4-4     | 0   | 0.5 |               |      |         |       |      |         |  |

图 29