



# [12] 发明专利说明书

[21] ZL 专利号 98115571.5

[45] 授权公告日 2004 年 10 月 6 日

[11] 授权公告号 CN 1170240C

[22] 申请日 1998.7.1 [21] 申请号 98115571.5

[30] 优先权

[32] 1997. 7. 1 [33] JP [31] 190716/1997

[32] 1997. 7. 22 [33] JP [31] 195408/1997

[71] 专利权人 株式会社日立制作所

地址 日本东京

[72] 发明人 冈本卓哉 高桥亨 青山由纪

山崎纪之 村田英子

审查员 陈汝岩

[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

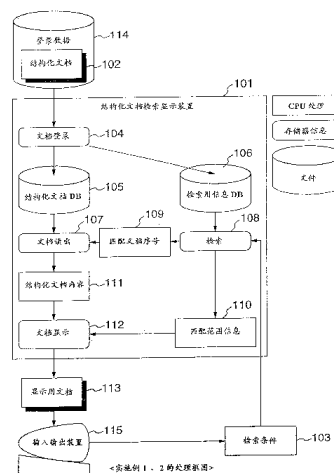
代理人 于 静

权利要求书 6 页 说明书 44 页 附图 42 页

[54] 发明名称 结构化文档检索显示方法和装置

[57] 摘要

将文件的结构化文档作为输入进行文档登录，生成进行过结构分析的结构化文档和用于文档检索的信息，分别存储到 DB105 和 DB106 中。然后，输入检索条件并分析，进行检索输出匹配的文档序号信息和匹配范围信息。从 DB105 中读出对应的结构分析过的结构化文档。在文档根据匹配范围信息，对结构化文档插入匹配信息，生成附加了高亮度信息的显示用的结构化文档，并显示该结构化文档。



1. 一种用于具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，其特征在于：

分析输入的结构化文档，生成分析过的结构化文档，将该分析过的结构化文档存储到上述文件装置中；

从该分析过的结构化文档中取得各结构内的内容字符串信息，生成文档检索用信息，并存储到上述文件装置中；

根据输入的检索条件检索该文件装置中存储的文档检索用信息，判断是否有满足该检索条件的内容字符串信息，取得具有被认为满足该检索条件的内容字符串信息的文档的分析过的结构化文档，并且取得满足该文档的检索条件的范围的信息；

根据满足上述文档的检索条件的范围的信息生成在结构化文档中附加了用于进行高亮度显示的信息的显示用结构化文档。

2. 按权利要求1所述的结构化文档检索显示方法，其特征在于：对于检索条件中的2个检索项，当满足关于各检索项的相对的出现位置的条件时，对构成检索条件的各检索项的高亮度显示和对包含该2个检索项的最小的字符串范围的高亮度显示、分别使用不同的高亮度显示形式、进行高亮度显示。

3. 按权利要求1所述的结构化文档检索显示方法，其特征在于：对于检索条件的多个检索项，对构成检索条件的各检索项的高亮度显示和对包含该检索项的整个结构的高亮度显示分别使用不同的高亮度显示形式进行高亮度显示。

4. 按权利要求1所述的结构化文档检索显示方法，其特征在于：在生成附加了基于上述显示用文档类型定义的用于进行高亮度显示的信息的显示用结构化文档时，使用在检索条件中指定的方法附加进行高亮度显示的信息。

5. 按权利要求1所述的结构化文档检索显示方法，其特征在于：检索结果的高亮度显示，对各检索项使用多个高亮度显示中的某一个

进行高亮度显示。

6. 按权利要求5所述的结构化文档检索显示方法，其特征在于：各检索项的高亮度显示的高亮度显示形式，根据各检索项的出现频度的信息而决定。

7. 按权利要求5所述的结构化文档检索显示方法，其特征在于：各检索项的高亮度显示的高亮度显示形式，根据对各检索项预先给定的加权的的信息而决定。

8. 一种具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，其特征在于：

分析输入的结构化文档，生成分析过的结构化文档，将该分析过的结构化文档存储到上述文件装置中；

生成从上述输入的结构化文档中去除预先给定的检索对象外的结构信息后的文档检索用的规格化处理过的结构化文档，并且生成用于将该被去除的结构信息复原的复原信息，存储到上述文件装置中；

根据输入的检索条件检索该文件装置中存储的规格化处理过的结构化文档，判断是否有满足该检索条件的规格化处理过的结构化文档，取得被认为满足该检索条件的文档的规格化处理过的结构化文档，并且取得满足该文档的检索条件的范围的信息；

利用上述复原信息将通过上述检索而取得的规格化处理过的结构化文档复原为具有被去除的结构信息的结构化文档，根据满足上述文档的检索条件的范围的信息，生成在该复原后的结构化文档中附加了用于进行高亮度显示的信息的显示用结构化文档。

9. 按权利要求8所述的结构化文档检索显示方法，其特征在于：对于检索条件中的2个检索项，当满足关于各检索项的相对的出现位置的条件时，对构成检索条件的各检索项的高亮度显示和对包含该2个检索项的最小的字符串范围的高亮度显示、分别使用不同的高亮度显示形式、进行高亮度显示。

10. 按权利要求8所述的结构化文档检索显示方法，其特征在于：对于检索条件的多个检索项，对构成检索条件的各检索项的高亮

度显示和对包含该检索项的整个结构的高亮度显示、分别使用不同的高亮度显示形式、进行高亮度显示。

1 1. 按权利要求 8 所述的结构化文档检索显示方法, 其特征在于: 在生成附加了基于上述显示用文档类型定义的用于进行高亮度显示的信息的显示用结构化文档时, 使用在检索条件中指定的方法附加进行高亮度显示的信息。

1 2. 按权利要求 8 所述的结构化文档检索显示方法, 其特征在于: 检索结果的高亮度显示, 对各检索项使用多个高亮度显示中的某一个进行高亮度显示。

1 3. 按权利要求 1 2 所述的结构化文档检索显示方法, 其特征在于: 各检索项的高亮度显示的高亮度显示形式, 根据各检索项的出现频度的信息而决定。

1 4. 按权利要求 1 2 所述的结构化文档检索显示方法, 其特征在于: 各检索项的高亮度显示的高亮度显示形式, 根据对各检索项预先给定的加权的的信息而决定。

1 5. 一种具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法, 其特征在于:

分析输入的结构化文档, 生成分析过的结构化文档, 将该分析过的结构化文档存储到上述文件装置中;

生成从上述输入的结构化文档中去除预先给定的检索对象外的结构信息后的文档检索用的规格化处理过的结构化文档, 并且生成用于将该被去除的结构信息复原的复原信息, 存储到上述文件装置中;

根据输入的检索条件检索该文件装置中存储的规格化处理过的结构化文档, 判断是否有满足该检索条件的规格化处理过的结构化文档, 取得被认为满足该检索条件的文档的规格化处理过的结构化文档, 并且取得满足该文档的检索条件的范围的信息;

取得输入的显示对象的部分结构;

对该显示对象的部分结构, 利用上述复原信息将通过上述检索而取得的规格化处理过的结构化文档复原为具有被去除的结构信息的结构

化文档，根据满足上述文档的检索条件的范围的信息，生成在该复原后的结构化文档中附加了用于进行高亮度显示的信息的部分结构显示用结构化文档。

1 6. 按权利要求 1 5 所述的结构化文档检索显示方法，其特征在于：对于检索条件中的 2 个检索项，当满足关于各检索项的相对的出现位置的条件时，对构成检索条件的各检索项的高亮度显示和对包含该 2 个检索项的最小的字符串范围的高亮度显示、分别使用不同的高亮度显示形式、进行高亮度显示。

1 7. 按权利要求 1 5 所述的结构化文档检索显示方法，其特征在于：对于检索条件的多个检索项，对构成检索条件的各检索项的高亮度显示和对包含该检索项的整个结构的高亮度显示、分别使用不同的高亮度显示形式、进行高亮度显示。

1 8. 按权利要求 1 5 所述的结构化文档检索显示方法，其特征在于：根据上述显示用文档类型定义生成附加了用于进行高亮度显示的信息的显示用结构化文档时，使用在检索条件中指定的方法附加用于进行高亮度显示的信息。

1 9. 按权利要求 1 5 所述的结构化文档检索显示方法，其特征在于：检索结果的高亮度显示，对各检索项使用多个高亮度显示中的某一个进行高亮度显示。

2 0. 按权利要求 1 9 所述的结构化文档检索显示方法，其特征在于：各检索项的高亮度显示的高亮度显示形式，根据各检索项的出现频度的信息而决定。

2 1. 按权利要求 1 9 所述的结构化文档检索显示方法，其特征在于：各检索项的高亮度显示的高亮度显示形式，根据对各检索项预先给定的加权的的信息而决定。

2 2. 一种具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，其特征在于：

将符合输入的特定的文档类型定义的结构化文档作为保留着标记的明文存储到上述文件装置中；

根据输入的检索条件检索该文件装置中存储的明文,判断是否有满足该检索条件的范围,取得具有满足该检索条件的范围的文档作为明文,并且取得满足该文档的检索条件的范围的信息;

将上述特定的文档类型定义作为显示用文档类型定义,对上述输入的结构化文档,针对满足上述检索条件的范围生成附加了基于该显示用文档类型定义的用于进行高亮度显示的信息的显示用结构化文档。

2 3. 按权利要求 2 2 所述的结构化文档检索显示方法,其特征在于:判断满足检索条件的范围在结构化文档中是否在表示文档结构的标记的属性信息中存在;

满足该检索条件的范围在标记的属性信息中存在时,在结构化文档的内容字符串中追加包含满足该检索条件的范围的字符串的字符串,对在该字符串中满足该检索条件的范围,生成附加了基于上述特定的文档类型定义的用于进行高亮度显示的信息的显示用结构化文档。

2 4. 按权利要求 2 2 所述的结构化文档检索显示方法,其特征在于:根据输入的检索条件,检索在该文件装置中作为仍然保留标记的明文而存储的结构化文档时,从检索对象中去除构成预先指定的特定的标记的字符串,对于对连结了构成该特定的标记的字符串的前后的字符串进行检索而得到的满足检索条件的范围,生成附加了基于上述特定的文档类型定义的用于进行高亮度显示的信息的显示用结构化文档。

2 5. 按权利要求 2 2 所述的结构化文档检索显示方法,其特征在于:根据输入的检索条件,检索在该文件装置中作为明文而存储的结构化文档时,判断满足检索条件的范围是否夹在表示预先指定的文档结构的开始的特定的标记和表示文档结构的结尾的特定的标记之间,

当该范围夹在其间时就在表示文档结构的开始的特定的标记之前或表示文档结构的结尾的标记之后的内容字符串中,追加包含满足该检索条件的范围的字符串的字符串,对在该字符串中满足该检索条件的范围,生成附加了基于上述特定的文档类型定义的用于进行高亮度显示的信息的显示用结构化文档。

---

2 6. 按权利要求 2 2 所述的结构化文档检索显示方法, 其特征在于: 在生成附加了基于上述显示用文档类型定义的用于进行高亮度显示的信息的显示用结构化文档时, 使用在检索条件中指定的方法附加用于进行高亮度显示的信息。

## 结构化文档检索显示方法和装置

### 技术领域

本发明涉及根据 S G M L、H T M L 等而生成的结构化文档的检索显示技术，特别是对结构化文档进行检索并对检索结果以高亮度进行显示的结构化文档检索显示方法和装置。

### 背景技术

随着文字处理器等的普及，所生成的文档信息的电子化获得了很大发展。这些电子化文档根据生成的机器、软件不同而具有各自的格式，别的机器或软件不能利用，或者需要准备某种变换手段。

作为这种文档交换用的共同格式，已提出了各种结构化文档。这些结构化文档，不仅可以定义作为文档的基本结构的章、节、项等层次结构，而且可以包含格式信息。

作为结构化文档的记述语言，现在已标准化了，这就是 S G M L (Standard Generalized Markup Language) = 「标准通用置标语言」。

S G M L 使用通过把称为标记的特定结构化文档的结构信息的字符串插入到文本中来表现文档的结构的方法。在 S G M L 中，可以根据 D T D (Document Type Definition) = 「文档类型定义」规定标记的名称、内容以及由标记所示的文档结构。

对于上述 S G M L、D T D，在「实践 S G M L」(S G M L 恳谈会实用化 W G 监译 1992 年 4 月 20 日 财团法人 日本规格协会发行) 中已详细说明。

设想了将这些结构化文档登录到检索系统的 D B 中，指定结构名进行检索的情况。想登录的各文档的 D T D 不同时，作为处理方法，提出了对各文档分析文档结构，在分析了所指定的结构名与哪一部分相当的基础上，取得检索对象的字符串进行检索的方法。

但是，该方法需要很长的处理时间。另外，使用对各结构名用表就各文档的对应的地方的等方法时，必须全部统一管理在各文档中出现的结构名，必须对各结构名登录各文档的对应的部分，从而需要庞大



的管理表。

此外，即使登录了不同的 D T D 混合存在的文档，也不限于所有的文档具有检索对象的结构，另外，例如像「摘要」、「大意」那样，即使是相同的内容也起了不同的结构名时，则必须全部指定这些不同的结构名进行检索，所以，不能认为是现实的结构化文档的检索。

因此，在结构化文档的检索中，必须运用的是只登录按相同的文档类型定义生成的文档。对于预先指定的结构名，管理各文档的对应的部分。

在进行检索时，如果指定了检索对象的结构名和检索条件，在与各文档的指定的结构对应的部分包含符合检索条件的字符串时，就判定与检索条件匹配。

作为结构化文档的检索结果，下面介绍用于显示文档的内容的功能的先有技术。

首先，作为第 1 先有技术，有特开平 8—3 3 9 3 6 9「文档显示装置和文档显示方法」。

本先有技术，介绍了 S G M L 文档的结构分析和向结构显示用的格式的变换以及进行指定结构的内容的显示的方法，通过使用本技术可以按结构单位显示结构化文档。此外，在本先有技术中，提供了指定结构的高亮度显示（就是强调的显示，改变颜色、字体、字的大小等，附加下划线）的手段。

但是，这里所说的所谓的高亮度显示手段，就是对各结构控制显示方法的手段，按结构单位进行有无显示、高亮度显示等的指定。因此，在本先有技术中，当然没有给出实现结构化文档的检索结果的显示时所需要的实现对已匹配的检索项的高亮度显示的方法。

另外，作为第 2 先有技术，有在特开平 8—2 1 2 2 3 0「文档检索方法和文档检索装置」中给出的对结构化文档以外的文档的检索结果的高亮度显示方法。

但是，本先有技术只实现了对用于进行显示的文本的匹配范围的取得和高亮度信息的附加，对于作为结构化文档的检索结果而得到的文档则不具有附加高亮度信息的功能。

只将上述 2 个先有技术组合，对于作为对结构化文档的检索结果而输出的文档不能实现对已匹配的项的高亮度信息的附加。

即，在结构化文档中，为了实现高亮度显示，需要在显示对象的文档生成时的 D T D 中生成追加高亮度用的结构信息的 D T D 的手段。

关于对结构化文档附加高亮度信息时的文档类型定义的变更方法，就是第 3 先有技术。在特愿平 8—1 5 9 2 0 2 「结构化文档的版管理方法和装置」中给出了对原来的 D T D 生成追加新的结构的 D T D 的方法。

通过使用本先有技术，可以生成附加高亮度信息的文档类型定义。

利用第 1、第 2 先有技术可知，进行显示可以知道结构化文档的结构，此外在未结构化的文档中，可以进行匹配范围的高亮度显示。

另外，通过使用第 3 先有技术，可以指定附加了对各结构取得的高亮度信息的文档类型定义。

通过将这些技术组合，输出对结构化文档的特定的结构的检索结果附加了高亮度信息的结构化文档，可以实现高亮度显示。

另外，作为获得最新的信息的方法，近年来国际互网络获得了突飞猛进的发展。作为从国际互网络上的大量的信息中最快地知道自己所需要的信息的手段，也充实了网上的信息的检索功能。

H T M L (Hyper Text Markup Language)是在 W W W (World Wide Web)上记述文档内容、用于表现向其他资源的链接信息、文档的格式的语言。H T M L 可以视为按照特定的 D T D 记述的 S G M L。作为生成、加工该 H T M L 文档的手段，有 H T M L 编辑器。另外，还有分析并显示所生成的 H T M L 文档的 H T M L 浏览器。

有的具有将进行检索的字符串(以下，称为「检索项」)输入 H T M L 浏览器，对显示中的 H T M L 文档进行检索并将匹配的地方进行反相显示等强调显示的功能。

对于 S G M L，也存在具有进行格式显示并加工的功能的 S G M L 浏览器。在 S G M L 浏览器上，对浏览器上显示中的 S G M L 文档进行全文检索并对与检索条件匹配的地方进行高亮度显示。在这些浏览器中，在显示文档时进行文档的分析并生成显示用的数据。检索是对该浏览器上的显示用的数据进行检索，并在画面上将匹配位置进行高亮度显示。

利用上述先有技术的组合，对所给定的结构化文档，可以将对各结构检索的结果分别进行高亮度显示。

但是,在结构信息中不仅表示了像章、节、项那样的文档结构本身,而且有时也包含下划线的附加等格式用的信息。这些结构信息不一定按文章的段落而插入。在进行文档检索时,如果未去除这种结构信息,尽管是包含在文档中的语句,也不能进行检索。这样,在以下的说明中将去除检索时不需要的结构信息的处理称为「规格化处理」。

为了以进行了规格化处理的结构化文档为检索对象、对原来的结构化文档实现附加了高亮度信息的显示,对进行了规格化处理的结构化文档只利用使用上述先有技术的方法是不能实现的。即,使用该方法在进行检索时只保留原来的文档的结构信息的一部分,所以,对该结构信息只附加高亮度信息,对原来的结构化文档不能实现与检索项匹配的高亮度显示。

### 发明内容

本发明的目的在于,为了根据对规格化处理过的文档的检索结果实现对原来的文档的高亮度信息的附加,实现从检索用的文档向原来的文档的高亮度范围信息的变换。

本发明的其他目的在于,规格化后的匹配项在原来的文档与多个结构匹配时,就按各结构对匹配的范围附加高亮度信息,进行高亮度显示。

本发明的另一个目的在于,为了将包括满足包含匹配的项的整个结构的高亮度显示或者出现位置的距离条件的2个检索项的整个区域进行高亮度显示等处理,附加层次性的高亮度信息,利用不同的高亮度显示形式进行高亮度显示。

本发明的另一个目的在于,在只抽出结构化文档的部分结构进行显示时,对这部分结构的内容也附加高亮度信息进行高亮度显示。

另一方面,HTML文档存在通过浏览器的单独的扩展根据多个DTD生成的HTML文档,但是,不知道是根据哪个DTD记述的。此外,由于存在大量的未根据SGML的语法正确地记述的文档,所以,难于使用和SGML相同的方法进行结构分析。

另外,(1)对明文(plain text)文档进行检索处理,通过生成在检索匹配位置的前后插入高亮度用的标记的HTML文档,在HTML浏览器上就可以强调显示检索匹配的字符串。但是,标记内的字符

串与检索项一致时，如果对该检索匹配位置的前后插入高亮度用的标记，由于改变了原来的HTML的标记的内容，所以，将不能正确地进行显示。

此外，(2)在HTML浏览器上连续地显示的字符串的中途，有时插入表现格式的标记，对HTML文档进行检索时，如果未除去标记进行检索，就不能正确地进行检索。例如，在HTML文档中写入「本月的〈FONT SIZE = + 1〉专集〈/FONT〉记事」，将检索项定为「专集记事」时，在HTML文档中，由于在「专集」和「记事」之间记述了用于放大字符进行显示的「〈/FONT〉」的标记，所以，如果不跳过标记进行检索，就不能正确地进行检索。

为了解决上述问题，本发明是具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，上述处理装置分析输入的结构化文档，生成分析过的结构化文档，将该分析过的结构化文档存储到上述文件装置中，从该分析过的结构化文档中取得各结构内的内容字符串信息、生成文档检索用信息，并存储到上述文件装置中，根据输入的检索条件检索该文件装置存储的文档检索用信息，判断是否有满足该检索条件的内容字符串信息，取得具有认为满足该检索条件的内容字符串信息的文档的分析过的结构化文档，并且取得满足该文档的检索条件的范围的信息，根据满足上述文档的检索条件的范围的信息生成在结构化文档中附加了用于进行高亮度显示的信息的显示用结构化文档。

另外，本发明是具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，上述处理装置分析输入的结构化文档，生成分析过的结构化文档，将该分析过的结构化文档存储到上述文件装置中，生成从上述输入的结构化文档中去除预先给定的检索对象外的结构信息的文档检索用的规格化处理过的结构化文档，并且生成用于将该被去除的结构信息复原的复原信息，并存储到上述文件装置中，根据输入的检索条件检索该文件装置存储的规格化处理过的结构化文档，判断是否有满足该检索条件的规格化处理过的结构化文档，取得认为满足该检索条件的文档的规格化处理过的结构化文档，并且取得满足该文档的检索条件的范围的信息，利用上述复原信息将通过上述检索而取得的规格化处理过的结构化文档复原

为具有被去除的结构信息的结构化文档，根据满足上述文档的检索条件的范围的信息，生成在该复原后的结构化文档中附加了用于进行高亮度显示的信息的显示用结构化文档。

另外，本发明是一种具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，其特征在于：分析输入的结构化文档，生成分析过的结构化文档，将该分析过的结构化文档存储到上述文件装置中；生成从上述输入的结构化文档中被去除预先给定的检索对象外的结构信息后的文档检索用的规格化处理过的结构化文档，并且生成用于将该被去除的结构信息复原的复原信息，存储到上述文件装置中；根据输入的检索条件检索该文件装置中存储的规格化处理过的结构化文档，判断是否有满足该检索条件的规格化处理过的结构化文档，取得被认为满足该检索条件的文档的规格化处理过的结构化文档，并且取得满足该文档的检索条件的范围的信息；取得输入的显示对象的部分结构；对该显示对象的部分结构，利用上述复原信息将通过上述检索而取得的规格化处理过的结构化文档复原为具有去除的结构信息的结构化文档，根据满足上述文档的检索条件的范围的信息，生成在该复原后的结构化文档中附加了用于进行高亮度显示的信息的部分结构显示用结构化文档。

另外，本发明是具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示装置，上述处理装置包括分析输入的结构化文档生成分析过的结构化文档并将该分析过的结构化文档存储到上述文件装置中的单元；生成从上述输入的结构化文档中去除预先给定的检索对象外的结构信息的文档检索用的规格化处理过的结构化文档并存储到上述文件装置中的单元；生成用于将该去除的结构信息复原的复原信息并存储到上述文件装置中的单元；根据输入的检索条件检索该文件装置存储的规格化处理过的结构化文档、判断是否有满足该检索条件的规格化处理过的结构化文档、取得认为满足该检索条件的规格化处理过的结构化文档的信息并且取得满足该文档的检索条件的范围的信息的单元；生成用于将满足该文档的检索条件的范围进行高亮度显示的显示用文档类型定义的单元；根据上述复原信息将通过上述检索而取得的规格化处理过的结构化文档复原为具有去除的结构信息的结构化文档的单元；和根据满足上述文档的检索条

件的范围的信息和显示用文档类型定义生成附加了用于在结构化文档中进行高亮度显示的信息的显示用结构化文档的单元。

本发明是具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，上述处理装置将符合输入的特定的文档类型定义的结构化文档作为保留着标记的明文存储到上述文件装置中，根据输入的检索条件检索该文件装置存储的明文，判断是否有满足该检索条件的范围，取得具有满足该检索条件的范围的文档作为明文，并且取得满足该文档的检索条件的范围的信息，将上述特定的文档类型定义作为显示用文档类型定义，对上述输入的结构化文档，根据该显示用文档类型定义对满足上述检索条件的范围生成附加了用于进行高亮度显示的信息的显示用结构化文档。

另外，本发明是具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，上述处理装置将符合输入的特定的文档类型定义的结构化文档作为保留着标记的明文存储到上述文件装置中，根据输入的检索条件检索该文件装置存储的明文，判断是否有满足该检索条件的范围，取得具有满足该检索条件的范围的文档作为明文，并且取得满足该文档的检索条件的范围的信息，判断满足检索条件的范围在结构化文档中表示文档结构的标记的属性信息中是否存在，满足该检索条件的范围在标记的属性信息中存在时，就在结构化文档的内容字符串中追加包含满足该检索条件的范围的字符串的字符串，对该字符串中满足该检索条件的范围根据上述特定的文档类型定义生成附加了用于进行高亮度显示的信息的显示用结构化文档。

另外，本发明是具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，上述处理装置将符合输入的特定的文档类型定义的结构化文档作为保留着标记的明文存储到上述文件装置中，从检索对象中去除构成预先指定的特定的标记的字符串，对于满足通过对连结构成该特定的标记的字符串的前后的字符串进行检索而得到的检索条件的范围，根据上述特定的文档类型定义生成附加了用于进行高亮度显示的信息的显示用结构化文档。

另外，本发明是具有处理装置、存储装置、文件装置和输入输出装置的信息处理系统的结构化文档检索显示方法，上述处理装置将符合

输入的特定的文档类型定义的结构化文档作为保留着标记的明文存储到上述文件装置中，根据输入的检索条件检索该文件装置作为明文存储的结构化文档时，判断满足检索条件的范围是否夹在表示预先给定的文档结构的开始的特定的标记和表示文档结构的结尾的特定的标记之间，夹在上述两标记之间时，就在表示文档结构的开始的特定的标记之前或表示文档结构的结尾的标记之后的内容字符串中追加包含满足该检索条件的范围的字符串的字符串，对该字符串中满足该检索条件的范围，根据上述特定的文档类型定义生成附加了用于进行高亮度显示的信息的显示用结构化文档。

#### 附图说明

- 图 1 是实施例 1、2 的结构化文档检索显示装置的处理框图。
- 图 2 是表示结构化文档检索显示处理的流程图。
- 图 3 是表示结构化文档登录的内容的图。
- 图 4 是表示结构化文档登录处理的流程图。
- 图 5 是表示检索用的文本的图。
- 图 6 是表示更新处理的流程图。
- 图 7 是表示指定结构的抽出处理的流程图。
- 图 8 是表示作为结构指定的分析结果而输出的信息的图。
- 图 9 是表示文档显示处理的流程图。
- 图 10 是表示结构化文档和高亮度处理结果的例子的图。
- 图 11 是表示文档显示用 DTD 生成处理的流程图。
- 图 12 是表示结构化文档检索用的规格化处理的内容的图。
- 图 13 是表示进行了规格化处理的结果的存储内容的图。
- 图 14 是表示进行了规格化处理时的匹配范围信息的变换处理内容的图。
- 图 15 是表示进行了规格化处理时的匹配范围信息的变换处理的流程图。
- 图 16 是表示附加高亮度信息的处理的流程图。
- 图 17 是表示实施例 2 的匹配范围信息的图。
- 图 18 是表示实施例 2 的各匹配信息的高亮度方法的定义的图。
- 图 19 是表示实施例 2 的向高亮度显示用 DTD 的变换内容的图。

图 2 0 是表示实施例 2 的高亮度处理的流程图。

图 2 1 是表示根据实施例 2 附加了高亮度信息的 S G M L 文档的例子图。

图 2 2 是表示高亮度显示的例子图。

图 2 3 是实施例 3 的结构化文档检索显示装置的概略处理框图。

图 2 4 是表示实施例 3 的处理内容的流程图。

图 2 5 是表示向部分结构显示用的 D T D 的变换处理的图。

图 2 6 是表示部分结构显示用的 D T D 生成处理的流程图。

图 2 7 是实施例 4 的系统结构图。

图 2 8 是表示数据控制部的流程图。

图 2 9 是表示实施例 4 的字符检索处理和高亮度位置信息的生成处理的流程图。

图 3 0 是高亮度位置信息存储区域的结构。

图 3 1 是高亮度数存储区域的结构。

图 3 2 是高亮度标记字符存储区域的结构。

图 3 3 是表示实施例 4 的带高亮度标记的 H T M L 文档的生成处理的流程图。

图 3 4 是高亮度插入例子图。

图 3 5 是高亮度插入后的例子。

图 3 6 是实施例 5 的系统结构图。

图 3 7 是表示实施例 5 的检索处理和高亮度位置信息生成处理的流程图。

图 3 8 是表示实施例 5 的标记内检索和标记外检索处理的流程图。

图 3 9 是表示实施例 5 的标记外用检索处理的流程图。

图 4 0 是表示实施例 5 的高亮度用标记的插入 H T M L 文档的生成处理的流程图。

图 4 1 是表示实施例 5 的高亮度标记插入处理的流程图。

图 4 2 是表示实施例 5 的标记外高亮度标记插入处理的流程图。

图 4 3 是实施例 6 的检索条件的例子图。

图 4 4 是实施例 6 的匹配范围信息的例子图。

图 4 5 是表示实施例 6 的向高亮度显示用 D T D 的变换处理的图。

图 4 6 是表示实施例 6 的高亮度显示用的 S G M L 文档的例子图。



图。

图 4 7 是实施例 6 的高亮度显示的例子的图。

### 具体实施方式

实施例 1 的概略的处理框图示于图 1。

1 0 1 是结构化文档检索显示装置。通过将登录数据文件( 1 1 4 ) 中存储的结构化文档( 1 0 2 ) 作为输入进行文档登录的处理, 生成进行结构分析过的结构化文档( 后面根据图 3 进行说明) 和用于文档检索的文档检索用信息( 后面根据图 5 进行说明)。

进行结构分析过的结构化文档存储到结构化文档数据库( 以后, 将数据库记为 D B ) ( 1 0 5 ) 中, 检索用信息存储到检索用信息 D B ( 1 0 6 ) 中。

其次, 从输入输出装置( 1 1 5 ) 输入检索条件( 1 0 3 ) 时, 就分析检索条件, 得出文档检索用信息, 进行检索处理( 1 0 8 )。作为检索结果, 输出匹配的文档序号的信息( 1 0 9 ) 和匹配范围的信息( 1 1 0 )。

显示处理, 首先在文档得出的处理( 1 0 7 ) 中, 根据匹配的文档序号的信息( 1 0 9 ), 从结构化文档 D B ( 1 0 5 ) 中得出指定的结构分析过的结构化文档( 1 1 1 )。在文档显示( 1 1 2 ) 的处理中, 根据匹配范围信息( 1 1 0 ), 对结构分析过的结构化文档( 1 1 1 ) 生成插入了匹配信息的显示用的结构化文档( 1 1 3 )。生成的显示用的结构化文档在输入输出装置( 1 1 5 ) 上进行显示。

图 2 表示结构化文档检索显示的处理的流程图。

首先, 进行结构化文档的登录处理( 2 0 1 )。对于登录处理的内容, 后面使用图 4 的流程图进行说明。

其次, 使用指定的检索条件检索结构化文档( 2 0 2 )。检索处理的详细情况, 后面使用图 6 的流程图进行说明。

作为检索结果, 有识别匹配文档数和匹配文档的序号和各文档的检索项的匹配范围。匹配范围的信息是用于识别包含匹配的检索项的结构的结构 I D ( 结构标识符) 和结构内的匹配开始位置, 输出文本长度的信息。

在结构化文档检索的处理中, 如果匹配文档数大于 1 ( 2 0 3 ), 就

顺序读出匹配的文档的内容(204),取得读出的文档的匹配范围信息(205),实现高亮度显示(206)。对于显示处理的详细情况,后面使用图9进行说明。

此外,如果有匹配的文档,就反复进行204~206的处理。

显示处理结束时,就确认有无下一个检索处理(208),如果没有检索条件,就结束处理,如果有检索条件,就返回到202的处理,反复进行结构化文档的检索显示处理。

图3是表示结构化文档登录处理的概要的图。

首先,分析SGML文档(301)的结构,生成树结构(302)。将生成的树结构的各项目的内容以表形式的数据(303)输出,并将其作为分析过的结构化文档进行登录。这里,所谓CDATA,就是字符串数据。

图4是结构化文档登录处理的流程图。

首先,分析结构化文档(401)。将分析过的结构化文档作为已完成分析的结构化文档进行登录(402)。对于结构化文档的分析,可以利用DTD,通过使用分析SGML文档的SGML检查器而实现。

其次,对分析过的结构化文档,进行用于去除检索所不需要的结构的规格化处理(403)。

对于规格化处理的顺序,后面使用图12进行说明。并且,将规格化处理过的结构化文档登录到文档数据库中(404)。

此外,从登录到数据库中的分析过的结构化文档中,取出结构信息、结构内的文本的信息作为结构化文档的检索所需要的检索用信息(405)。将这里得到的检索用信息登录到检索用信息DB(106)中(406)。这里,登录的检索用信息去除了SGML文档中的结构信息(标记),对各结构存储表示结构信息及其内容的文本串。

图5表示由上述检索用信息和已规格化的结构化文档构成的检索用的文本的存储例子。对登录文档反复进行上述处理,在没有了登录文档时将结束处理(407)。登录内容用于登录文档的全文检索。

图5是作为检索用的文本而输出的内容的例子。这样,就将由使文档结构的结构ID与文本串对应的表和字符串信息构成的信息作为检索用的

文本进行登录。在检索时，根据结构 I D 抽出所需要的字符串进行检索。

图 6 是图 2 的结构化文档检索显示处理的 2 0 2 步骤的结构化文档检索的处理流程。

检索条件按「检索对象的结构指定：检索条件式」那样给定。

检索对象的结构，例如像「〈文档. 标题〉」那样，用「〈」和「〉」包围，上位结构（例中为「文档」）和下位结构（例中为「标题」）用「.」进行区分，指定对层次结构中的哪个结构进行检索。

检索条件式，例如用 and（“检索”、“文档”）表示“检索”和“文档”两个都出现的条件，用 C < = 1 0（“检索”、“文档”）表示“检索”和“文档”夹着小于 1 0 字符的字符而出现的条件。

结构化文档检索中，首先清除匹配文档数的计数器（6 0 1），其次，进行检索条件中的检索对象的结构指定的部分的分析（6 0 2）。在 6 0 2，从如〈文档. 标题〉那样指定结构的字符串中取得可以唯一地特定分析过的结构化文档对应的结构的结构 I D（结构标识符）。结构 I D 取得的处理内容，后面使用图 7 的流程图进行说明。

其次，读出作为检索对象而登录的文档（检索用的文本），取得与在 6 0 2 取得的指定结构 I D 对应的文本部分（6 0 3）。

根据检索条件分析由检索项以及多个检索项出现的“与”、距离条件等的逻辑条件构成的检索条件式（6 0 4），根据得到的检索项进行在 6 0 3 取得的文本部分的全文检索，判断是否满足检索条件式的逻辑条件，即判断是否与检索条件匹配（6 0 5）。

与检索条件匹配时（6 0 6），作为检索结果，输出文档序号、包含检索项的结构的 I D 和结构中的检索项匹配的范围的信息（6 0 7）。

此外，计数匹配的文档的数（6 0 8），对全部文档进行本处理后（6 0 9），输出匹配文档数（6 1 0）。

图 7 是图 6 的检索条件的分析中结构指定内容的分析处理的流程图。

首先，取得文档的最上位结构（7 0 1）。其次，从最上位结构开始顺序取得下位结构。如果取得的结构是指定结构的下位结构（7 0 3），就将该结构作为检索对象的结构而输出结构 I D（7 0 4）。

如果有下位结构（7 0 5），则进而对该下位结构同样判断是否为指

定的结构的下位结构，如果是下位结构，就反复进行输出结构 I D 的处理（7 0 6）直至不再有下位结构为止（7 0 7），如果对全部结构的处理结束，便可获得检索对象的结构 I D 的一览表。

图 8 表示作为检索对象的结构 I D 一览表的输出形式。

输出作为检索对象的结构 I D 的数（8 0 1）和作为检索对象的 I D（8 0 2）。

图 9 是表示显示处理的内容的流程图。下面，使用本流程图说明显示处理的内容。

首先，由于检索对象的结构化文档是进行清除检索所不需要的结构的规格化处理后的文档，所以，根据检索而匹配的结构和匹配范围信息与登录的未规格化的结构化文档的结构和范围不一定一致（参见图 3 的树结构 3 0 2 和图 1 2 的树结构 1 2 0 3）。

显示所使用的文档，是对登录的未规格化的结构化文档将高亮度信息附加到匹配的范围上的文档。

因此，首先，根据登录文档的 D T D 进行显示所使用的文档用的显示用 D T D 的生成处理（9 0 1）。对于显示用 D T D 生成处理的内容，后面使用图 1 1 进行说明。

此外，对于对规格化后的结构化文档而得到的匹配范围，变换为规格化前登录的结构化文档的结构和高亮度范围信息（9 0 2）。对于规格化后的文档的匹配范围信息向规格化前的文档的高亮度范围信息的变换处理的内容，后面使用图 1 5 进行说明。

其次，读出显示所使用的分析过的文档的最上位结构的信息，通过顺序反复进行 9 0 3 ~ 9 1 1 的处理，进行显示用的文档的输出处理。

首先，读出结构信息（9 0 3），开始，输出结构的开始标记（9 0 4）。在本结构中如果进而存在下位结构（9 0 5），就对下位结构再次进行显示处理（9 0 3 ~ 9 1 1 的处理）（9 0 6）。如果没有下位结构，就转移到输出表示结构的结尾的标记的处理（9 1 1）。

这里，所谓下位结构，包括字符串。因此，对于

〈文档〉

〈标题〉

## 结构化文档

〈正文〉

〈强调〉结构化文档〈/强调〉的检索、 . . .

〈/正文〉

〈/文档〉

等的结构化文档，作为〈标题〉的下位结构，存在字符串（在SGML中，表现为CDATA）这样的结构。CDATA不具有下位结构，作为字符串信息，上述例子的情况就是具有「结构化文档」这样的内容。

对于〈正文〉，同样，作为下位结构也存在〈强调〉这样的结构和具有「的检索、 . . .」这样的内容的字符串。

在905的步骤，判定不存在下位结构时，由于是字符串的结构，所以，对本结构的内容，与匹配范围信息进行比较（908），如果是包含匹配范围的结构，就进行高亮度处理（909）。对于高亮度处理，后面使用图16进行说明。

如果不包含匹配范围的字符串，就直接将内容作为文本而输出（910）。输出内容为字符串时，在904、911的步骤，就不输出开始标记和结束标记。

通过上述处理，实现各结构的高亮度显示。如果有应进一步处理的结构，就反复进行从903开始的处理（912）。

图10是登录用DTD（1001）、登录的SGML文档（文档实例）的例子（1002）、高亮度显示使用的显示用DTD（1003）和变换为显示用的SGML文档（文档实例）的例子（1004）。所谓DTD（Document Type Definition），如在先有的技术一项所述的那样，是规定标记的名称、内容以及由标记表示的文档结构的文档类型定义。

在DTD中，表现结构时，在“〈!ELEMENT标记名”之后，并列2个“-”或“0”。

开始的“-”或“0”表示可否省略结构开始标记，为“-”时，就不能省略。为“0”时，可以省略。第2个“-”或“0”表示可否省略结束标记。

下面，对于内容模型，描述在下位结构可以出现的结构。

为图10的DTD1001的(标题,正文)时,标题为下位结构1,正文为下位结构2。

描述为“(下位结构1,下位结构2?)”那样时,表示在下位结构1之后分别只出现1次下位结构2,“?”表示也可以不出现下位结构2。

为“(下位结构1|下位结构2)\*”时,表示下位结构1、2顺序不同地多次(包括0次)出现。

这里,在内容模型中记述为“CDATA”时,表示在该结构中只存在1个字符串。

#PCDATA也表示字符串,但是,可以反复出现。字符串和结构混合存在时,就必须使用#PCDATA。

在内容模型中,有时指定“RCDATA”取代“CDATA”。CDATA和RCDATA的不同,是在结构内出现实体参照(如“&XXXX,”那样记述时,就利用于向外字的置换等。)时,CDATA不进行向实体(外字等)的变换,所出现的字符串直接作为字符串进行处理。指定“RCDATA”时,就将向进行向实体的变换的字符串作为字符串进行处理。

为了进行高亮度显示,必须改变文档结构,用以对字符串附加高亮度信息。如在1003用下划线所示的变更点那样,对各结构的字符串部分全部追加高亮度显示用的结构信息,此外,还必须附加高亮度显示用的结构信息(〈!ELEMENT高亮度--(#PCDATA)〉)。

在原来的DTD中成为内容模型的“CDATA”的部分变更为“( #PCDATA | 高亮度)\*”,表示CDATA在该结构中只存在1个字符串,不能作为反复的要素出现。由于附加了高亮度用的标记,所以,即使原来的结构是CDATA,变更为#PCDATA后,就采用“( #PCDATA | 高亮度)\*”,以使高亮度可以反复出现。

图11是表示用于根据登录用的DTD生成高亮度显示用的DTD的处理内容的流程图。

首先,读出登录用DTD(1101),分析DTD的内容,取得ELEMENT项目(1102)。在ELEMENT项目的内容模型中指

定了CDATA、RCDATA、#PCDATA等时，就全部改变内容模型，使之成为可以附加高亮度用的结构（1103~1106）。

内容模型的变更，首先将“CDATA”、“RCDATA”变更为“#PCDATA”，然后将“#PCDATA”定义为如“（#PCDATA | 高亮度）\*”那样反复出现用高亮度标记包围的字符串和不包围的字符串。

原来的内容模型如“（#PCDATA | 下划线）\*”那样记述为多个结构反复出现时，如“（#PCDATA | 下划线 | 高亮度）\*”那样，可以只记述出现高亮度结构。

对于所有的ELEMENT声明变更处理结束时（1107），作为高亮度用的结构的定义，追加“（!ELEMENT 高亮度--CDATA）”（1108）。通过以上的处理，生成图10的1003所示的高亮度用的DTD。

图12是表示结构化文档的规格化处理的内容的图。

将图10的1001所示的结构化文档表示为树结构时，就成为1201那样。

作为不需要的结构，指定了“下划线”时，作为规格化处理的最初的处理，如1202所示的那样，就是删除下划线这样的结构，包含在下划线的下位结构中的字符串直接作为上位结构即“正文”的要素。

此外，作为“正文”的下位结构，由于并列了2个字符串（CDATA），所以，如1203那样，将字符串连结，作为1个字符串数据。

图13是分析规格化处理前的结构化文档（1301）、规格化处理后的结构化文档（1302）的内容变换为表形式而输出的内容。1303是存储结构信息的表，附加了0~6的结构ID的结构是规格化前的结构的信息。0是最上位结构，通过追寻下位结构的信息，就知道了文档结构。

附加了7~9的结构ID（结构标识符）的结构是规格化后变更、追加的结构。

7是最上位结构，通过追寻下位结构，便可知道规格化后的文档结构。这里，未变更的结构即“标题”以下的结构即结构ID1、2的结构信息

仍然保留。

此外，对于在规格化处理中追加的结构 I D 7 ~ 9 的结构，根据 1 3 0 4 的规格化对应表，存储与规格化前的结构的对应关系。

图 1 4 表示对规格化后的结构化文档将检索时的匹配范围的信息变换为规格化前的结构化文档的范围信息的结果。

利用图 1 3 的 1 3 0 4 的规格化对应表的信息，将根据 1 4 0 1 的规格化后的结构信息而得到的匹配范围的信息变换为规格化前的结构化文档的范围信息（1 4 0 2）。

在本图的例子中，由于在规格化前的文档中分为结构 I D 5 和 6，所以，规格化后的结构 I D 9 的匹配范围变更为 2 个结构中的高亮度对象的范围信息。

图 1 5 表示将对图 9 的 9 0 2 步骤的规格化处理后的结构化文档的匹配范围信息变换为对规格化处理前的结构化文档的匹配范围信息的处理内容的流程图。

首先，顺序读出规格化后的匹配范围信息（1 5 0 1），判断匹配范围信息的结构 I D 是规格化后追加的还是规格化前就存在的（1 5 0 2）。

如果是规格化前就存在的结构 I D，就不变更，所以，直接作为规格化前的匹配范围信息而输出（1 5 0 3）。

如果是规格化后生成的结构 I D，就追寻图 1 4 的规格化对应表的规格化后结构 I D，根据字符范围的信息获得对应的规格化前的结构 I D 和匹配范围（1 5 0 4）。如果得到了规格化处理前的结构的匹配范围，就将其作为规格化前的匹配范围而输出（1 5 0 5）。

对所有的匹配范围信息处理结束时（1 5 0 6），就可以获得显示用的高亮度范围信息。

图 1 6 是图 9 的 9 0 9 步骤的高亮度处理的流程图。首先，输出从文档的开头到高亮度开始的字符串（1 6 0 1）。其次，输出高亮度显示使用的结构的开始标记（1 6 0 2）。

此外，输出高亮度范围的字符串（1 6 0 3），输出高亮度显示使用的结构的结束标记（1 6 0 4）。

所有的高亮度处理结束时（1 6 0 5），就输出其余的文本，并结束



高亮度处理(1606)。

下面,对实施例2说明根据匹配条件改变高亮度显示方法的处理以及按层次进行多个高亮度处理时的处理。概略处理框图和图1相同。

图17是在本实施例中使用的匹配范围信息(1701)。

对图14所示的匹配范围信息追加的信息,就是追加存储各匹配的条件区域(1702)。

此外,在图14中,只输出匹配的检索项的范围,但是,根据检索条件不同,除了匹配的检索项外,还可以指定对包含该检索项的整个结构的高亮度等包含检索项的区域。

这些匹配条件的信息,在结构化文档的检索处理时附加。这里,附加了检索条件使用的距离条件和各检索项的出现频度等信息,但是,也可以使用对各检索项预先进行加权等方法。

图18是定义匹配条件和高亮度方法(高亮度显示形式)的对应的表(1801)。

记述与匹配条件(1802)对应的高亮度方法(1803)。按照各匹配条件而匹配的范围根据本表的内容进行高亮度显示。

此外,给定层次信息(1804),层次信息的值越大,则越是整个结构的高亮度等上位的高亮度结构。

图19是表示用于实现上述高亮度处理的显示用DTD生成的处理内容。

对于登录所使用的原来的DTD(1901),在上位的高亮度结构内可以按层次指定下位的高亮度结构,进而为了也可以省略而生成变更或追加定义的高亮度显示用的DTD(1902)。

DTD的生成方法,对于使用图11进行的上述处理,可以在1106步骤的高亮度信息附加时全部附加上多个存在的高亮度信息(1903),此外,在1108步骤的高亮度用ELEMENT声明追加时,可以根据图18的层次信息(1804),对于各高亮度结构的下位结构,将下位的高亮度结构和字符串作为内容模型进行保持。

如果没有下位的高亮度结构,则作为内容模型只出现字符串(1904)。

图 2 0 是实施例 2 的高亮度处理的流程图。

首先，将高亮度信息进行分类，将开始位置顺序作为第 1 键，将从层次信息的上位到下位的顺序作为第 2 键（2 0 0 1）。其次，输出高亮度开始之前的文本（2 0 0 2），并输出高亮度开始标记（2 0 0 3）。

此外，如果在高亮度范围的结尾之前已开始了下一个高亮度，由于存在下位的结构信息（2 0 0 4），所以，在输出该位置之前的文本之后（2 0 0 5），进行下位的高亮度结构的高亮度处理（2 0 0 6）。下位结构的高亮度处理和 2 0 0 3 - 2 0 0 9 的处理相同。

对下位的高亮度结构的处理结束之后，如果进而还有下位的高亮度结构（2 0 0 7），就返回到 2 0 0 5 步骤的处理，输出下一个高亮度结构之前的文本，进行下位的高亮度结构的处理。

如果已没有下位的高亮度结构，就输出到结构的结尾为止的文本（2 0 0 8），并输出高亮度结束标记（2 0 0 9）。

如果还有高亮度的信息，就返回到 2 0 0 2 的步骤，反复进行处理。如果高亮度的信息结束了（2 0 1 0），就输出其余的文本，并结束处理（2 0 1 1）。

图 2 1 是通过上述处理而生成的 S G M L 文档的例子。

图 2 2 是图 2 1 的 S G M L 文档的正文的显示例子。对于重叠的高亮度范围，重复进行多个高亮度用的显示方法。作为实施例 3，就是只分离出结构化文档的部分结构，进行高亮度显示时的处理内容。

图 2 3 是表示本实施例的概略处理框图。

图 1 的变更点，是指定显示对象的结构（2 3 0 1），根据显示对象的结构的内容进行部分结构显示的处理（2 3 0 2）取代文档显示（1 1 2）的处理。

图 2 4 是表示抽出部分结构进行显示时的处理顺序的流程图。

首先，生成部分结构显示用的 D T D（2 4 0 1）。对于部分结构显示用的 D T D 的生成处理，后面使用图 2 6 进行说明。

此外，将对规格化后的结构化文档而得到匹配范围变换为规格化前的登录时的文档的结构 I D 和匹配范围信息（2 4 0 2）。关于规格化后的文档的信息向规格化前的文档的范围信息的变换处理的内容，可以使用上

述使用图 1 6 说明的方法。

其次，读出作为显示对象的分析过的文档的结构的信息，通过反复进行 2 4 0 3 ~ 2 4 1 1 的处理，进行显示用的文档的输出处理。

首先，读出作为显示对象的结构信息 ( 2 4 0 3 )。这里，是否为显示对象的结构判断，使用上述使用图 7 说明的方法来实现。

如果是显示对象的结构信息，首先输出结构的开始标记 ( 2 4 0 4 )。进而如果存在下位结构 ( 2 4 0 5 )，就对下位结构进行显示处理 ( 2 4 0 3 ~ 2 4 1 1 的处理 ) ( 2 4 0 6 )。如果没有下位结构，就转移到输出表示结构的结尾的标记的处理 ( 2 4 1 1 )。

在 2 4 0 5 的步骤中判定不存在下位结构时，由于是字符串的结构，所以，对本结构的内容与匹配范围信息进行比较 ( 2 4 0 8 )，如果是匹配范围包含的结构，就进行高亮度处理 ( 2 4 0 9 )。对于高亮度处理，使用上述使用图 1 5 说明的方法。

如果是高亮度范围不包含的字符串，就将内容直接作为文本而输出 ( 2 4 1 0 )。输出内容为字符串时，在 2 4 0 4、2 4 1 1 的步骤中，就不输出开始标记和结束标记。

通过上述处理，实现各结构的高亮度显示。如果进而有应处理的结构，就反复进行从 2 4 0 3 开始的处理 ( 2 4 1 2 )。

图 2 5 是部分结构显示用的 D T D 的生成内容。

根据部分结构的输出，有时不输出定义为在原来的 D T D ( 2 5 0 1 ) 一定出现的结构。此外，也不一定必须输出上位结构。

因此，即使对于不是必须出现上位结构的开始标记和结束标记的结构本身，也必须变更为可以不一定出现。生成的部分结构显示用的 D T D 如 2 5 0 2 所示。

使用该 D T D 生成的 S G M L 文档，如 2 5 0 3 所示。在该例中，只抽出了标题。

图 2 6 是表示部分结构显示用的 D T D 生成顺序的流程图。首先，取得登录用的 D T D ( 2 6 0 1 )。

其次，取出 D T D 中的 E L E M E N T 项目 ( 2 6 0 2 )。在内容模型中包含 C D A T A、R C D A T A、# P C D A T A 时，附加高亮度信

息(2603~2606)。

高亮度信息的附加,和图11的1103~1106步骤的处理相同。

其次,检查内容模型中的出现指示符(\*、+、?、无),如果是“+”(2607),就变更为“\*” (2608),如果没有指示符(2609),就附加“?”(2610)。

对所有的ELEMENT声明的处理结束时(2611),就追加高亮度用的结构的ELEMENT声明(2612),进而如果必须出现存在下位结构的结构的标记,就变更为不需要(0)。

下面,参照附图说明使用本发明的实施例4。

图27是本实施例的系统结构图。

WWW(World Wide Web)检索系统(2700)使用网络(2702)与客户(2701)连接。

客户(2701)是PC、WC等,在客户(2701)上工作的网络浏览器(2703)上的检索项设定画面上,输入检索项。在WWW检索系统(2700)中,使用该检索项进行检索,并将该检索结果向网络浏览器(2703)输出。

WWW检索系统(2700)由接收客户(2701)的检索项的HTTP服务器(2704)、插入检索处理和高亮度用标记的数据控制部(2705)和预先存储高亮度标记的位置信息等存储器(2706)构成,连接预先存储作为检索对象的HTML文档的磁盘装置(2707)。

在数据控制部(2705)中,将HTTP服务器(2704)接收的检索项对在磁盘(2707)中存在的HTML文档进行检索处理,将高亮度标记插入到与检索项匹配的HTML文档的检索匹配位置。

存储器(2706)由存储各文档的检索匹配数的高亮度数存储区域(2708)、存储检索结果位置信息的高亮度位置信息存储区域(2709)、预先存储插入的高亮度用标记的内容的高亮度用标记字符存储区域(2710)、存储插入高亮度用标记的HTML文档的HTML文档

暂时存储区域(2711)和由WWW检索系统(2700)的HTTP服务器(2704)取得在客户(2701)输入的检索项并暂时存储的检索项存储区域(2712)构成。

由WWW检索系统(2700)插入高亮度用标记的HTML文档,从HTTP服务器(2704)经过网络(2702)在客户(2701)的网络浏览器(2703)上进行显示。

下面,使用图28说明数据控制部(2705)的处理内容。

这里,取得由客户(2701)设定的检索项,进行检索处理,检测检索匹配位置,生成高亮度位置信息(2709),将高亮度用的标记插入到与检索条件匹配的HTML文档的检索项匹配的HTML文档的检索匹配位置,并在客户(2701)的网络浏览器(2703)上进行显示。

步骤2800:

在WWW检索系统(2700)中,使用HTTP服务器(2704)取得由客户(2701)设定的检索项。取得的检索项存储到存储器(2706)的检索项存储区域(2712)中。

步骤2801:

使用在步骤2800存储到检索项存储区域(2712)中的检索项,对磁盘装置(2707)存储的HTML文档进行全文检索。检索匹配时,取得HTML文档中的检索匹配位置及检索匹配数等,并将该信息存储到高亮度位置信息存储区域(2709)和高亮度数存储区域(2708)中。对于该处理,使用图29详细说明。

步骤2802:

根据在步骤2801生成的存储在高亮度位置信息存储区域(2709)中的信息,将存储在高亮度标记字符存储区域(2710)中的高亮度用标记插入到HTML文档的检索匹配的位置,并存储到HTML文档暂时存储区域(2711)中。详细情况,使用图33进行说明。

步骤2803:

使用HTTP服务器(2704)将由步骤2802生成的存储到HTML文档暂时存储区域(2711)中的高亮度用HTML文档在客户

(2701)的网络浏览器(2703)上进行显示。

通过反复进行步骤2800~2803的处理,使用由客户(2701)输入的检索条件检索磁盘(2707)存储的HTML文档,对与检索条件匹配的文档可以采用多个地方的检索匹配位置的高亮度显示。

下面,使用图29说明图28的步骤2801的高亮度位置信息的生成处理。

步骤2900:

读出磁盘(2707)存储的HTML文档。图34的HTML文档(3400),就是读出的HTML文档的例子。

用网络浏览器显示该HTML文档时,则显示为3401所示的画面。

步骤2901:

将存储高亮度位置信息的区域即高亮度位置信息存储区域(2709)确保 $\alpha$ 件数。 $\alpha$ 是任意的正整数。另外,确保存储高亮度数的高亮度数存储区域(2708)。

高亮度位置信息存储区域(2709)和高亮度数存储区域(2708)的数据形式示于图30和图31。

如图30所示,高亮度位置信息存储区域(2709)由HTML文档序号(3000)、从开头开始的高亮度位置序号(3001)、高亮度字节数(3002)和高亮度插入标记序号(3003)构成。

HTML文档序号(3000),是在步骤2900读出的HTML文档的序号。存储着存储HTML文档时附加的系列序号等。

从开头开始的高亮度位置序号(3001)在由步骤2900读出的HTML文档中与在步骤2800取得的检索示于匹配时,从文档开头开始按字节数存储HTML文档中的检索匹配位置。

高亮度字节数(3002),按字节数存储高亮度的长度。即,存储检索项的字符串长度。

高亮度插入标记序号(3003),在按多个检索项进行高亮度显示时,可以按各检索项区别高亮度有标记而进行显示。根据这里存储的信息,

区别高亮度用标记。即，这里存储判断高亮度显示所利用的标记的种类的数据。

步骤 2 9 0 2 :

将表示高亮度位置信息存储区域 ( 2 7 0 9 ) 存储的计数的 i.cnt 初始设定为 0。

步骤 2 9 0 3 :

检查在步骤 2 8 0 0 读出的检索项与在步骤 2 9 0 0 读出的 HTML 文档是否一致。存在检索匹配地方时，就进入步骤 2 9 0 4。另外，不存在时，就进入步骤 2 9 0 8。

步骤 2 9 0 4 :

检查在步骤 2 9 0 1 或 2 9 0 5 确保的高亮度位置信息存储区域 ( 2 7 0 9 ) 是否大于表示高亮度存储数的 i.cnt。还存在存储数据的区域时，就进入步骤 2 9 0 6。另外，不存在存储的区域时，就进入步骤 2 9 0 5。

步骤 2 9 0 5 :

将高亮度位置信息存储区域 ( 2 7 0 9 ) 放大一定值，重新进行确保，并进入步骤 2 9 0 6。

步骤 2 9 0 6 :

将 HTML 文档序号 ( 3 0 0 0 )、从 HTML 文档的开头开始的位置 ( 3 0 0 1 )、高亮度字符数 ( 3 0 0 2 ) 和高亮度标记插入序号 ( 3 0 0 3 ) 存储到在步骤 2 9 0 1 或 2 9 0 5 确保的高亮度位置信息存储区域 ( 2 7 0 9 ) 的第 i.cnt 个的位置。由于 i.cnt 初始化为 0，所以，i.cnt 为 0 时，第 0 个就存储数据。

在 1 个 HTML 文档中存储多个高亮度信息时，由于更新 i.cnt，所以，存储到 i.cnt 所示的位置。

将在步骤 2 9 0 0 读出的 HTML 文档 ( 3 4 0 0 ) 定为 HTML 文档序号「 0 0 1」。此外，将在步骤 2 8 0 0 抽出的检索项定为「专集」。

在该 HTML 文档 ( 3 4 0 0 ) 中检索该检索项「专集」时，就可以从 HTML 文档 ( 3 4 0 0 ) 的开头开始到第 1 2 2 字节 ( 3 4 0 3 ) 找到「专集」的字符。

这时，对于 HTML 文档序号 ( 3 0 0 0 )，就存储作为 HTML 文

档序号的「001」(3404), 对于从HTML文档的开头开始的位置(3001), 存储「122」(3405), 对于高亮度字符数(3002), 存储「专集」的字节数「4」(3406)。最后, 对于高亮度标记插入序号(3003), 存储表示用于强调检索结果的标记的序号。这里, 存储「1」(3407)。

高亮度插入标记与实际存储的高亮度标记对应的结构示于图32。在图32的(1)中, 示出了高亮度标记字符存储区域(2710)存储的高亮度插入标记用的结构体3200。

高亮度插入标记用的结构体(3200)由存储系列序号的标记序号1(3202)、存储高亮度开始标记名的开始标记1(3203)、存储高亮度结束标记名的结束标记1(3204)和存储标记的个数的高亮度标记数(3201)构成。存在高亮度标记数存储个数的标记序号、开始标记、结束标记。

下面, 说明高亮度标记字符存储区域的使用例(2)。

这里, 给出了存储3种高亮度标记的情况。因此, 在存储高亮度标记数的地方存储「3」(3205)。对于标记序号「0」(3206)的地方的开始标记, 存储表示红色的标记「<FONT COLOR = "RED">」(3207), 对于结束标记, 存储「</FONT>」(3208)。同样, 对于标记序号「1」(3209), 存储表示闪烁的标记「<BLINK>」, 对于标记序号「2」(3210), 存储表示字符大的「<H1>」。

高亮度标记字符存储区域(2710), 在高亮度位置信息存储区域(2709)生成之前生成。另外, 该高亮度标记字符存储区域(2710), 也可以使用用户界面而生成。

通过准备多个高亮度用标记, 在进行不同标记或同义词的检索处理时, 对于用不同表记检索的字符, 可以用标记序号「1」表示, 对于用同义词检索的字符, 可以用标记序号「2」表示, 对各检索条件可以进行不同的高亮度显示。

对于高亮度用标记, 使用「<BLINK>」时, 对于高亮度位置信息存储区域(3402)的高亮度标记插入序号(3407), 存储「1」。



**步骤 2 9 0 7 :**

由于在步骤 2 9 0 6 将数据存储到了高亮度位置信息存储区域 ( 2 7 0 9 ) 中, 所以, 将 `i.cnt` 增加 1, 并返回到步骤 2 9 0 3。

**步骤 2 9 0 8 :**

取得在步骤 2 9 0 0 取得的 HTML 文档中的高亮度数, 并存储到高亮度数存储区域 ( 2 7 0 8 ) 中。下面, 使用图 3 1 说明高亮度数存储区域 ( 2 7 0 8 ) 的结构体的内容。

图 3 1 是高亮度数存储区域 ( 2 7 0 8 ) 的结构体的内容。3 1 0 0 是在步骤 2 9 0 0 读出的 HTML 文档的文档序号。另外, 3 1 0 1 是预先存储取得的高亮度数的地方。这里, 将文档序号「0 0 1」存储到文档序号 3 1 0 0 处, 将 `i.cnt` 存储到高亮度数存储区域 ( 3 1 0 1 ) 处, 并结束处理。

下面, 使用图 3 3 说明带高亮度用标记的 HTML 文档生成处理。

**步骤 3 3 0 0 :**

检查在步骤 2 9 0 0 读出的 HTML 文档中是否必须插入高亮度标记。

存在 ( 2 7 0 9 ) HTML 文档序号 ( 3 0 0 0 ) 时, 就进入步骤 3 3 0 1。不存在时, 在步骤 3 3 0 9 就输出所有的文本, 并结束处理。

**步骤 3 3 0 1 :**

将表示处理计数的 `i.cnt` 初始化为 0。

**步骤 3 3 0 2 :**

确保存储插入了高亮度标记的 HTML 文档的 HTML 文档暂时存储区域 ( 2 7 1 1 )。

HTML 文档暂时存储区域 ( 2 7 1 1 ) 确保 HTML 原档字节数中将高亮度用开始标记和结束标记的字节数总和乘以高亮度插入数的字节数的区域。

高亮度的开始标记和结束标记, 根据高亮度位置信息存储区域 ( 2 7 0 9 ) 的高亮度插入标记序号 ( 3 0 0 3 ) 计算高亮度用标记的标记字符串长度。高亮度数取得在步骤 2 9 0 8 存储到高亮度数存储区域 ( 2 7 0 8 ) 中的高亮度数 ( 3 1 0 1 )。

**步骤 3 3 0 3 :**

检查高亮度数 ( 3 1 0 1 ) 是否小于 icnt.

小于时, 由于存在未处理的高亮度地方, 所以, 就进入步骤 3 3 0 4. 除此之外, 由于应处理的高亮度地方全部结束了, 所以, 就进入步骤 3 3 0 9.

**步骤 3 3 0 4 :**

将高亮度位置之前的 HTML 文档存储到在步骤 3 3 0 2 确保的 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中.

**步骤 3 3 0 5 :**

将高亮度开始标记存储 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中. 高亮度开始标记, 是根据从高亮度插入标记序号 ( 3 0 0 3 ) 抽出的序号而得到的高亮度标记字符存储区域 ( 2 7 1 0 ) 中的标记名.

图 3 4 ( 3 ) 的情况, 对于高亮度插入标记序号 ( 3 0 0 3 ), 就是存储「 1 」。将图 3 2 ( 2 ) 所示的高亮度标记字符存储区域 ( 2 7 1 0 ) 存储为标记序号「 1 」 ( 3 2 0 9 ) 的「 < B L I N K > 」存储到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中.

**步骤 3 3 0 6 :**

将检索项存储到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中.

图 3 4 的情况, 就是将「 专集 」存储到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中.

**步骤 3 3 0 7 :**

将刮亮度结束标记存储到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中. 高亮度结束标记, 和在步骤 3 3 0 5 处理的高亮度开始标记一样, 存储根据用高亮度插入标记序号 ( 3 0 0 3 ) 存储的序号而得到的高亮度标记字符存储区域 ( 2 7 1 0 ) 存储的标记名.

图 3 4 ( 3 ) 的情况, 就是存储「 1 」。因此, 将存储为图 3 2 ( 2 ) 的标记序号「 1 」的「 < / B L I N K > 」存储到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中.

**步骤 3 3 0 8 :**

在步骤 3 3 0 5 ~ 步骤 3 3 0 7, 将数据存储到 HTML 文档暂时存

储区域(2711)中后,将icnt增加1,并返回到步骤3303。

步骤3309:

从高亮度插入位置到HTML文档最后,将文本存储到HTML文档暂时存储区域(2711)中,结束带高亮度标记的HTML文档的生成处理。

通过使用上述处理,根据客户(2701)设定的检索项检索HTML文档,对与检索项一致的文档,可以生成存储高亮度数的高亮度数存储区域(2708)和存储高亮度位置的高亮度位置信息存储区域(2709)的内容。

上述处理结果的例子示于图9。3500是插入高亮度用标记的HTML文档。高亮度用标记插入到检索匹配的「专集」的前后(3501、3502)。

将该HTML文档显示在画面上时,如3503那样,检索匹配的「专集」(3504)进行闪烁显示。

以上,作为本发明的实施例1,说明了从HTML文档(2707)中检索客户(2701)插入的检索项并生成将高亮度用标记插入到匹配位置的带高亮度用标记的HTML文档的方法。

下面,使用图36~图42说明本发明的实施例5。

图36是用HTML文档的标记分割检索项时和在标记内存在检索项时的高亮度显示方法的系统结构图。

和图27一样,在客户(2701)的网络浏览器(2703)上设定检索项。

WWW检索系统(2700)由取得检索项的HTTP服务器(2704)、进行检索处理的数据的控制部(2705)和确保区域的存储器(2706)构成。

存储器(2706),除了在图27的说明中所述的以外,由用设计显示等使用的标记存储跳跃读取的标记名的跳跃读取标记名存储区域(3600)、预先暂时存储客户(2701)输入的检索项和HTML文档(2707)一致的开始位置的开始位置存储区域(3601)、在HT

ML 标记的开始字符「〈」和结束字符「〉」之间存在与检索项匹配的位置时预先存储作为记号的标志的再次记述标志存储区域(3602)、检索项与HTML 标记的开始标记和结束标记之间匹配时预先存储在检索匹配的地方的前后不能输入高亮度用标记的高亮度用标记不能插入标记名存储区域(3603)构成。

检索项用HTML 文档的标记跨越时或在标记内存在检索项时的检索项的取得、高亮度位置信息的生成和高亮度用标记插入方法按图28所示的处理顺序进行。另外,使用图37~图42说明各个处理内容。

使用在步骤2800取得的检索项,在步骤2801的处理中,进行检索处理和高亮度位置信息生成处理。处理内容示于图37的流程图。

步骤3700:

从磁盘(2707)中读出作为处理对象的HTML 文档。

步骤3701:

在存储器(2706)中确保持存储高亮度位置信息的高亮度位置信息存储区域(2709)和高亮度数存储区域(2708)。

步骤3702:

读出插入到检索匹配位置的前后的高亮度标记。

在图32(2)的使用例中示出了具体例,从高亮度标记字符存储区域(2710)中读出高亮度用标记。这时,根据「3」(3205)可知高亮度插入标记序号的个数为「3个」。对于第1个的「0」(3206),存储「〈FONT COLOR = "RED"〉」(3207)和「〈/FONT〉」(3208)。因此,高亮度插入标记序号第0个的开始标记为「〈FONT COLOR = "RED"〉」,结束标记为「〈/FONT〉」。同样,高亮度插入标记序号第1个的开始标记为「〈BLINK〉」,结束标记为「〈/BLINK〉」,高亮度插入标记序号第2个的开始标记为「〈H1〉」,结束标记为「〈/H1〉」。

步骤3703:

将表示HTML 文档的处理过的字符数的计数的icnt和存储高亮度数的区域的内容初始设定为0。

步骤3704:

检查检索项和HTML文档的字符串是否一致。作为检查方法，根据HTML文档的第i.cnt个字节检索与检索项的开头字符一致的字符。在步骤3703，由于初始设定是设定为0，所以，开始时从HTML文档的第0个字节开始检索一致的字符。一致时，就进入步骤3705。不一致时，就结束处理。

另外，这里，作为抽出检索项的方法，使用跳过指定的标记进行检索的方法。具体而言，就是在HTML文档中存在跳跃读取标记名存储区域(3600)存储的标记名时，就跳跃读取该标记名，进行检索处理。

将「IMG」预先存储到跳跃读取标记名存储区域(3600)中，检索图34的HTML文档(3400)时，从HTML文档(3400)中的开头开始扫描数据，在抽出「IMG」(3413)的时刻，跳过标记内的字符。即，跳跃到标记的结束字符「>」(3414)之前。

通过将该跳跃读取标记名存储区域(3600)预先设定到检索处理之前，就可以进行跳跃读取处理。

步骤3705：

将在步骤3704从HTML文档的开头到与检索项的开头字符一致的字符的字符数暂时存储到开始位置存储区域(3601)中。

步骤3706：

检查检索项的字符串与书写在HTML文档中的字符是否一致，一致时就检查一致的地方是否存在于HTML标记内或HTML标记外。此外，用从HTML文档的开头开始的字符数确保检索匹配的字符串的最后的字符的位置。详细情况，使用图38说明。

步骤3707：

检查步骤3706的结果是否已检索匹配了。在HTML文档中存在检索项时，就进入步骤3708。不存在检索项时，就进入步骤3712。

步骤3708：

将在步骤3701确保的高亮度数存储区域(3708)与高亮度存储数进行比较，如果确保的区域多于存储的高亮度数，就进入步骤3709。如果少于存储的高亮度数，就进入步骤3710。

步骤3709：

在高亮度位置信息存储区域(2709)中,由于存储数据的区域不足,所以,再次进行区域设定,并进入步骤3710。

步骤3710:

将高亮度显示的字符数和高亮度的位置信息存储到高亮度位置信息存储区域(3600)中。具体而言,对于用图30说明的高亮度位置信息存储区域(3600)的HTML文档序号(3000),存储在步骤3700读出的HTML文档的文档序号,对于从开头开始的高亮度位置信息(3001),存储在步骤3705取得的开始位置。另外,对于高亮度的字节数(3002),存储检索项的字符串长度,对于高亮度插入标记序号(3003),存储在步骤3702读出的标记的序号。

高亮度插入标记序号(3003),作为缺省,设定为「0」。

步骤3711:

对于检索项存在多个匹配的字符串时,从检索匹配的位置的下一个字符开始,再次进行检查检索项与HTML文档中的一致的地方的处理。因此,将从记述在步骤3706确保的检索匹配的最后的字符的位置的HTML文档的开头开始的字符数加上1的值赋值于iCnt。如果更新了处理位置,就返回到步骤3704。

步骤3712:

在步骤3705取得的开始位置存储区域(3600)存储的从开始位置开始的字符串与检索项不一致时,就从开始位置的下一个字符开始再次进行检查检索项与HTML文档中的一致的地方的处理。因此,将开始位置存储区域(3600)存储的开始位置增加1的值赋值于iCnt。如果更新了处理位置,就返回到步骤3704。

以上,说明了包括标记内和标记外的检查的检索处理和高亮度位置信息生成处理。

下面,使用图38说明步骤3706的标记内的检索和标记外的检索处理。这里,检查在步骤3705取得的检索匹配的开始位置是存在于表示文档结构的标记的属性中或存在于标记之外,并检查从检索匹配的开始位置开始的字符串与检索项是否一致。

步骤3800:

在步骤3705中开始位置存储区域(3600)存储的检索匹配的  
开始位置,检查是HTML标记内或标记外。

从步骤3706时刻的HTML文档的第i.cnt个字节开始,检查到检索匹配的  
开始位置之前的数据。检查与标记的开始字符「<」对应的标记  
的结束字符「>」,检查在标记内是否存在检索匹配的  
开始位置。有标记的开始字符「<」并在标记的结束字符「>」之前存在检索匹配的  
开始位置时,就假定开始位置存在于标记内,并进入步骤3801。在未由标记  
的开始字符「<」和标记的结束字符「>」包围的范围内存在检索匹配的  
开始位置时,就假定检索匹配的  
开始位置存在于标记外,并进入步骤3804。

#### 步骤3801:

检查检索项与从检索匹配的  
开始位置开始的字符串是否一致。检索项的字符串由多个字节构成时,就检查每个字节的字符串。检索项的字符串与从检索匹配的  
位置开始的字符串一致时,就进入步骤3802。不一致时,就进入步骤3803。

#### 步骤3802:

在步骤3801,与检索项一致时,就假定已「检索匹配」,并结束处理。

另外,求检索匹配的字符串的终端位置。假定终端位置是在检索匹配的  
开始字符位置加上检索项的字符串长度的字节数。这里所求的终端位置,在步骤3711中使用。

#### 步骤3803:

在步骤3801检索项不一致时,就假定「检索不匹配」,不结束处理。

#### 步骤3804:

在步骤3800,检索匹配的  
开始位置存在于标记外时,就进行标记外用的检索处理。标记外用的检索处理,使用图39进行说明。

#### 步骤3805:

检查在步骤3804检索项匹配的地方在HTML文档中是否存在。存在时,就进入步骤3807。不存在时,就进入3806。

步骤 3 8 0 6 :

在步骤 3 8 0 5 检索项不匹配时, 就结束处理。

步骤 3 8 0 7 :

在步骤 3 8 0 5 与检索项一致时, 就假定「检索匹配」, 并结束处理。

另外, 求检索匹配的字符串的终端位置。假定终端位置是在检索匹配的  
的开始字符位置加上记述在步骤 3 8 0 4 检测的检索匹配的最后的字符的  
位置的值。这里所求的终端位置, 在步骤 3 7 1 1 中使用。

以上, 说明了标记内检索和标记外检索处理。

下面, 使用图 3 9 说明步骤 3 8 0 4 的标记外用的检索处理。

步骤 3 9 0 0 :

检查在 HTML 文档中是否存在检索项。检查检索项的字符串与在 HTML  
文档中存在字符串是否一致, 但是, 从检索匹配的开始位置, 通过  
跳过途中存在的标记, 有一致的情况, 所以, 这里检查从检索匹配的开始  
位置开始是否每 1 个字符与检索项一致。

下面, 具体地使用图 3 4 进行说明。

假定检索项为「专集记事」时, 就在 (2) 的显示画面上, 在 3 4 0  
8 处显示「专集记事」。但是, 在 HTML 文档 (3 4 0 0) 中, 在「专  
集」(3 4 0 3) 和「记事」(3 4 1 6) 之间有「〈/H1〉」(3 4  
1 7) 的标记。这样, 在检索项的途中存在 HTML 标记时, 就跳跃读取  
该 HTML 标记, 抽出与检索项一致的字符串。

这里, 将检索项按每 1 字符与 HTML 文档的字符进行对照检查。

检索项的第 1 个字符与 HTML 文档中的字符一致时, 就对检索项的  
下一个字符和 HTML 文档的下一个字符反复进行相同的处理。

检索项的字符串的所有的字符一致时, 具体地就进行「特」(3 4 0  
3)、「集」、与字符的比较, 跳跃读取「〈/H1〉」(3 4 1 7),  
进而进行「记」(3 4 1 6)、「事」与字符比较。将所有的检索项全部  
抽出时, 就进入步骤 3 9 0 1。

检索项与 HTML 文档中的字符串完全不一致时, 就进入步骤 3 9 0  
2。

步骤 3 9 0 1 :



由于在HTML文档中存在检索项，所以，设定已「检索匹配」，并结束处理。

另外，求检索匹配的终端位置。终端位置就是在步骤3900最后抽出的字符的位置。

步骤3902：

在步骤3900检索项的字符与HTML文档的字符不一致时，就检查HTML文档的字符是否为标记的开始字符「〈」。

是标记的开始字符「〈」时，就进入步骤3903。是除此以外的字符时，就进入步骤3904。

步骤3903：

在步骤3902，HTML文档中的字符是标记的开始字符「〈」时，就跳过标记的内容，并返回到步骤3900。

具体而言，就是抽出标记的结束字符「〉」，跳跃读取到抽出的字符之前。在图34的HTML文档中(3400)，假定检索项为「专集记事」时，就跳跃读取从处于「专集」(3403)的下一个字符的「〈」(3417)到标记的结束字符「〉」(3418)。即，跳跃读取「〈/H1〉」。

步骤3904：

在步骤3902，检索项不一致时，就假定「检索不匹配」，并结束处理。

以上，说明了在HTML文档中存在检索项的字符串时抽出检索匹配位置并将高亮度位置信息存储到高亮度位置信息存储区域中的处理。

下面，使用图40说明根据存储到高亮度位置信息存储区域中的高亮度位置信息为了强调显示HTML文档的检索匹配的字符串而插入高亮度用标记的方法。

步骤4000：

读出在步骤3710存储到高亮度信息存储区域(2709)中的高亮度位置信息。

步骤4001：

确保用于存储插入了高亮度标记的HTML文档的HTML文档暂时

存储区域(2711)。

确保的区域的大小,是HTML文档的原文档的数据乘以高亮度标记数的高亮度开始标记和结束标记的长度之和的值的字节数。

从高亮度标记数存储区域(2708)读出高亮度标记数。另外,根据高亮度位置信息存储区域(2709)的高亮度插入标记序号(3003)和高亮度标记字符存储区域(2710)检测高亮度的开始标记和结束标记,并求检测的标记的字符串长度。

步骤4002:

将表示HTML文档中的处理过的位置的`icnt`和高亮度处理数初始化为0。

步骤4003:

将处理过的高亮度地方的数即高亮度处理数与应处理的高亮度数进行比较。高亮度处理数少时,用于进行插入高亮度用标记的处理,所以,就进入步骤4004。在除此以外的情况时,就进入步骤4007。

步骤4004:

将从表示处理过的位置的`icnt`到检索匹配的起始位置的数据存储到HTML文档暂时存储区域(2711)中。

具体而言,就是在图34的HTML文档(3400)中,将检索项作为「专集记事」时,就将从HTML文档开头到「专集记事」(3403)前的字符「本月的」的数据存储到HTML文档暂时存储区域(2711)中。

步骤4005:

将高亮度用标记存储到检索匹配位置。使用图41说明高亮度用标记的插入处理。

步骤4006:

将从插入了高亮度结束标记的位置的开头开始的字节数赋值于表示HTML文档的处理过的位置的`icnt`,并返回到步骤4003。

步骤4007:

将从表示HTML文档的处理过的位置的`icnt`到HTML文档的最后数据存储到HTML文档暂时存储区域(2711)中,并结束处理。

下面，使用图 4 1 说明在步骤 4 0 0 5 中处理的高亮度标记的插入处理。

这里，检查检索匹配的位置是在标记内或标记外，并进行将高亮度用标记插入到检索匹配位置的前后的处理。

步骤 4 1 0 0：

检查在 HTML 文档中匹配的位置是在 HTML 标记内还是标记外。

检查方法，是在到检索匹配的开始位置之前的 HTML 文档中，找 HTML 标记的开始字符「〈」和标记的结束字符「〉」的对应，来判断在标记内还是标记外。检索匹配的开始位置处于标记的开始字符「〈」与标记的结束字符「〉」之间时，就视为存在于标记内，并进入步骤 4 1 0 1。是除此以外的情况时，就视为检索匹配位置的开始位置存在于标记外，并进入步骤 4 1 1 0。

步骤 4 1 0 1：

从标记的开始字符「〈」的下一个字符开始抽出字符，取得标记的种类。

例如，在图 3 4 的 HTML 文档（3 4 0 0）的情况时，假定检索项为「hitachi」时，就可以在 HTML 文档（3 4 0 0）中的 3 4 0 9 处取得「hitachi」。取得该 HTML 标记的种类时，就可以知道写在标记的开始字符「〈」之后的「A」（3 4 1 0）。

步骤 4 1 0 2：

检查在步骤 4 1 0 1 取得的标记是开始用的标记还是结束用的标记。是结束用的标记时，则标记的开始字符「〈」的下一个字符就是「/」（斜杠）。因此，就检查、判断标记的开始字符「〈」的下一个字符。该标记的开始字符「〈」的下一个字符是「/」时，就判定为结束用的标记，并进入步骤 4 1 0 5。除此以外的情况时，就判定为开始标记，并进入步骤 4 1 0 3。

步骤 4 1 0 3：

检查是否可以将高亮度用的标记插入到开始用标记与结束用标记之间。

可以插入时，就进入步骤 4 1 0 5。另外，不能插入时，就进入步骤

4 1 0 6。

具体而言，在图 3 4 的 HTML 文档 ( 3 4 0 0 ) 中，检索项为「 hitachi 」时，HTML 标记的种类就是在「 hitachi 」 ( 3 4 0 9 ) 之前存在的「 A 」 ( 3 4 1 0 )。所谓开始用的标记，就是夹在包围 HTML 标记的种类的「 < 」和「 > 」之间的部分的「 < A ~ > 」 ( 3 4 1 0 )，所谓结束用的标记，就是「 < / A > 」 ( 3 4 1 2 )。

该 HTML 标记是写为开始用标记和结束用标记的字符串在画面上显示并选择该字符串时与在开始与标记内指定的 URL 链接的标记。将高亮度用的标记插入到该「 < A ~ > 」 ( 3 4 1 0 ) 和「 < / A > 」 ( 3 4 1 2 ) 之间时，就发生不能正确地进行强调显示的现象。因此，视为不能将高亮度用的标记插入到该「 < A ~ > 」 ( 3 4 1 0 ) 和「 < / A > 」 ( 3 4 1 2 ) 之间，并进入步骤 4 1 0 4。

高亮度标记的插入可否的判断，是根据 HTML 标记的种类存储在高亮度标记插入不可能标记名存储区域 ( 3 6 0 3 ) 中时，如果与标记的某一个一致，就判定为不可能，如果与哪一个都不一致，就判定为可以插入。

存储在高亮度标记插入不可能标记名存储区域 ( 3 6 0 3 ) 中的标记，使用用户接口在进行数据控制 ( 2 7 0 5 ) 的处理之前预先生成。

步骤 4 1 0 4：

不能将检索匹配的字符串插入到开始用标记和结束用标记之间时，跳跃到可以插入的地方读取。

这里，跳跃到结束用标记的结尾的字符「 > 」进行读取，取得跳跃读取的数据。

具体而言，在图 3 4 的 HTML 文档 ( 3 4 0 0 ) 中，设检索项为「 hitachi 」时，HTML 标记就是「 A 」 ( 3 4 1 0 )。其结束标记就是「 < / A > 」 ( 3 4 1 2 )。

因此，直至「 < / A > 」取得数据。

具体而言，设图 3 4 的 HTML 文档 ( 3 4 0 0 ) 的检索项为「 imagefile.gif 」 ( 3 4 1 1 ) 时，HTML 标记的种类就是「 IMG 」 ( 3 4 1 3 )，其结尾的标记字符就是「 > 」 ( 3 4 1 4 )。因此，跳跃读取到「 > 」 ( 3 4 1 4 )，取得跳跃读取的数据。

**步骤 4 1 0 6 :**

将在步骤 4 1 0 4、步骤 4 1 0 5 取得的数据插入到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中。另外, 求检索匹配的字符的终端位置。终端位置就是在步骤 4 1 0 4 或步骤 4 1 0 5 跳跃读取的位置的字节数。

**步骤 4 1 0 7 :**

将高亮度用的开始标记插入到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中。将写为图 3 2 的开始标记 1 ( 3 2 0 3 ) 的标记名插入到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中。

设高亮度位置信息存储区域存储的数据为 3 4 0 2、高亮度标记字符存储区域为图 3 2 的 ( 2 ) 时, 由于对高亮度位置信息存储区域的高亮度插入标记序号 ( 3 4 0 7 ) 记述为「 1 」, 所以, 从图 3 2 的 ( 2 ) 可以取得「 < B L I N K > 」。因此, 这里, 就将「 < B L I N K > 」插入到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中。

**步骤 4 1 0 8 :**

将检索匹配字符插入到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中。但是, 由于记述再次检索匹配字符, 所以, 存储再次记述的标志。这里, 插入再次记述标志的存储区域 ( 3 6 0 2 ) 存储的数据。

再次记述标志的存储区域 ( 3 6 0 2 ) 在进行数据控制之前预先存储数据。例如, 预先存储「 @ 」标志。

**步骤 4 1 0 9 :**

将高亮度用的结束标记插入到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中。将写为图 3 2 的结束标记 1 ( 3 2 0 4 ) 的标记名插入到 HTML 文档暂时存储区域 ( 2 7 1 1 ) 中。这里, 是插入「 < / B L I N K > 」。

**步骤 4 1 1 0 :**

检索匹配的位置存在标记外时, 进行标记外用高亮度标记插入处理。使用图 4 2 说明该处理。

下面, 使用图 4 2 说明在步骤 4 1 1 0 进行处理的标记外用高亮度插入处理。

**步骤 4 2 0 0 :**

是「 < A ~ > 」标记时, 将高亮度用的标记插入到夹在开始用的标记

「〈A~〉」和结束用的标记「〈/A〉」之间的地方时，不能反映强调显示。这里，检查是否夹在开始标记和结束标记之间，判断是否为在步骤4201可以插入高亮度标记的标记。但是，在HTML文档中，存在「HTML」和「〈BODY〉」的标记，由于这些标记作为包围整个文档的标记使用，所以，这里不作为检查的对象。

检查存储在从图30的开头开始的高亮度位置信息(3001)中的位置是否被HTML文档的开始用标记和结束用标记所包围。被包围时，就进入步骤4201。不被包围时，就进入步骤4209。

具体而言，在图34的HTML文档(3400)中，检索项为「日立」时，就抽出HTML文档(3400)的「日立」(3415)。由于它被HTML标记「〈A~〉」(3410)和「〈/A〉」(3412)所包围，所以，这里就进入步骤4201。

步骤4201：

检查是否可以将高亮度用标记插入到检索匹配的字符串的前后。

检查方法，就是抽出被检索匹配的检索字符串所包围的HTML标记。将抽出的标记的种类与高亮度标记插入不可能标记名存储区域(3603)存储的标记进行比较。如果一致，就进入步骤4202，不一致时，就进入步骤4209。

在高亮度标记插入不可能标记名存储区域(3603)中记述的标记名，在开始用标记和结束用标记之间不能插入高亮度用标记。

因此，与高亮度标记插入不可能标记名存储区域(3603)存储的HTML标记进行比较，如果一致，就进入步骤4202，不一致时，就进入步骤4209。

该高亮度插入不可能标记名存储区域(3603)，使用用户界面在数据控制(2705)之前预先生成。

步骤4202：

不能插入高亮度用标记时，就跳跃读取结束用标记的结尾的字符「〉」之前的HTML文档。

在图34中，假定为检索项「日立」时，就跳跃读取「〈/A〉」(3412)之前的HTML文档。

**步骤4 2 0 3 :**

将在步骤4 2 0 2 跳跃读取的HTML文档存储到HTML文档暂时存储区域(2 7 1 1)中。

在图3 4中,假定为检索项「日立」时,将从位于在步骤4 0 0 2 设定的第i.cnt个或在步骤4 0 0 6 更新的第i.cnt个HTML文档的字符开始到「</A>」(3 4 1 2)的数据存储到HTML文档暂时存储区域(2 7 1 1)中。

**步骤4 2 0 4 :**

将高亮度用标记的开始标记插入到HTML文档暂时存储区域(2 7 1 1)中。设高亮度位置信息存储区域为3 4 0 2、高亮度标记字符存储区域为图3 2的(2)时,就抽出「<BLINK>」。因此,这里,就插入「<BLINK>」。

**步骤4 2 0 5 :**

存储再显示用标志。和步骤4 1 0 8一样,读出存储在再次记述标志的存储区域(3 6 0 2)中的HTML文档,存储到HTML文档暂时存储区域(2 7 1 1)中。

**步骤4 2 0 6 :**

再一次将检索匹配的字符串插入到HTML文档暂时存储区域(2 7 1 1)中。

但是,在检索匹配的字符串的途中存在标记时,就插入存在标记的地方之前的字符串。

**步骤4 2 0 7 :**

将高亮度用的结束标记插入到HTML文档暂时存储区域(2 7 1 1)中。这里,就是插入「</BLINK>」。

**步骤4 2 0 8 :**

检查在步骤4 2 0 6 是否存储了全部检索项的字符串。

在检索匹配的字符串中存在标记、未全部存储检索匹配的字符串时,就返回到步骤4 2 0 0。另外,存储了全部字符串时,就结束处理。

**步骤4 2 0 9 :**

跳跃到检索匹配的起始位置读取数据,将跳跃读取的HTML文档存

储到HTML文档暂时存储区域(2711)中。

具体而言,在图34的HTML文档(3400)中,假定检索项为「专集记事」时,就将在检索匹配的「专集」(3403)之前存在的「本月的」之前的HTML文档插入到HTML文档暂时存储区域(2711)中。

步骤4210:

和步骤4205一样,将高亮度用标记的开始标记存储到HTML文档暂时存储区域(2711)中。这里,就是插入「<BLINK>」。

步骤4211:

将检索匹配的字符串插入到HTML文档暂时存储区域(2711)中。

但是,在检索匹配的字符串的途中存在标记时,就插入存在标记的地方之前的字符串。

例如,在HTML文档(着400)中,假定检索项为「专集记事」时,在「专集」(3403)和「记事」(3416)之间存在「</H1>」。因此,这里,就存储「专集」。

步骤4212:

将高亮度用标记的结束标记插入到HTML文档暂时存储区域(2711)中。这里,就是插入「</BLINK>」。

步骤4213:

检查在步骤4211是否已将检索项的全部字符串插入到HTML文档暂时存储区域(2711)中。通过去除HTML的标记,使检索项与HTML文档的字符串匹配,在从检索匹配的开头位置开始的检索项长度的字符串之间存在HTML的标记时,在步骤4211,就将HTML标记之前的HTML文档插入到HTML文档暂时存储区域(2711)中。这时,必须从HTML标记开始处理其余的检索匹配的字符。

将所有的检索项插入到HTML文档暂时存储区域中时,就结束处理。另外,从HTML标记开始处理其余的检索匹配的字符时,就返回到步骤4200。

在图34的HTML文档(3400)中,假定检索项为「专集记事」



时，在「专集」(3403)和「记事」(3416)之间，存在「〈/H1〉」。在步骤4206只插入了「专集」的状态下，未插入「记事」，所以，返回到步骤4200。

通过进行这样的处理，使用客户(2701)设定的检索项，将高亮度标记插入到与检索项一致的HTML文档中，就可以在网络浏览器(2703)上显示高亮度匹配地方。

这里，对1个检索项检查HTML文档，如果在HTML文档中存在检索项的字符串，就在客户(2701)的网络浏览器上显示表示处理结果的处理方法，但是，也可以对1个检索项从多个HTML文档中进行检索，存储检索匹配的HTML文档数的高亮度位置信息，一起生成存储了高亮度用标记的多个HTML文档。

另外，也可以对多个检索项从多个HTML文档中进行检索，存储检索匹配的HTML文档数的高亮度位置信息，一起生成存储了高亮度用标记的多个HTML文档。

下面，说明使用本发明的实施例6。

本实施例与实施例2的不同点在于，在检索条件中可以与检索项等一起定义与检索条件匹配时的高亮度方法。这样，对任意的检索条件，就可以按各检索条件指定高亮度方法。

本实施例的系统结构和图1相同。但是，检索条件103的记载方法不同。下面，使用图43说明本实施例的检索条件103的记述方法的例子。

图43表示本实施例的检索条件的例子。如图所示，在各检索项或检索条件等之后，如「{下划线}」那样指定高亮度方法。实施例2的检索条件，是「检索对象的结构指定：检索条件式」，但是，此处则为「检索对象的结构条件{高亮度方法}：带高亮度方法的检索条件式」。高亮度方法的指定，可以省略。省略时，用实施例2所示的方法进行高亮度显示。即，对于未在检索条件中记载高亮度方法的地方，读出图18所示的高亮度方法定义1801，使用本定义信息记载的高亮度方法进行高亮度显示。

图44表示本实施例的匹配范围信息4401的存储内容。与实施例

2的图17所示的匹配范围信息的不同点在于,对各匹配范围不仅存储匹配条件4402,而且存储高亮度方法4403。本信息可以通过使用图43分析上述检索条件、读出检索条件中记载的高亮度方法的信息而取得。

图45表示本实施例的高亮度显示用DTD的生成方法。在本例中,每次检索时可以重新改变高亮度方法,所以,每次进行高亮度显示时,生成只追加所需要的结构的高亮度显示用DTD。这时,在DTD中,不仅记述检索条件,而且直接记述高亮度方法。

如图所示,对于登录使用的原来的DTD(1901),在上位的高亮度结构内,可以按层次指定下位的高亮度结构,进而也可以省略而生成改变、追加定义的高亮度显示用的DTD(4501)。

DTD的生成方法,首先在图44的匹配范围信息中未记载高亮度方法4403时,根据图18所示的高亮度方法定义取得与匹配条件对应的高亮度方法。首先,对于原来的DTD的各结构,变更结构信息(4502),用以可以在内容模型中具有在下位结构中出现的的高亮度方法。此外,从匹配范围信息4401的匹配范围的层次关系中获得出现的的高亮度用结构的层次关系。根据这里得到的高亮度显示用的层次关系,作为各高亮度结构的下位结构,则具有下位的高亮度结构和字符串作为内容模型。如果没有下位的高亮度结构,则作为内容模型,就只出现字符串(4503)。

根据本实施例的高亮度处理,不是将检索条件作为高亮度结构,而是生成记述所记载的高亮度方法的高亮度显示用结构化文档和高亮度显示用的DTD。因此,按照本实施例,是生成图46所示的显示用结构化文档。将本显示用结构化文档进行高亮度显示时,就进行图47所示的高亮度显示。

按照本发明,作为结构化文档的检索结果,显示匹配的文档的内容时,可以向各结构的检索项匹配的范围输出附加了高亮度信息的高亮度结构化文档。通过向结构化文档中插入高亮度信息而不是与浏览器有关的高亮度信息,则不论在哪个浏览器上都可以进行高亮度显示。

可以根据检索时的条件或各检索项的重要程度、出现频度等条件而进行不同的高亮度处理,对于重要的检索项,可以进行表明是高加权的高亮

度处理。此外，通过在检索条件中记述高亮度方法，可以对各用户进行任意的高亮度显示。

此外，只抽出部分结构就可以输出附加了高亮度信息的结构化文档。

另外，从存在表示文档结构的HTML标记的文档中检索字符串时，在HTML标记内存在与设定的检索项一致的字符串时，或检索项跨越HTML标记记述时，也可以很容易地进行检索匹配。另外，可以将检索匹配的字符串进行高亮度显示。

图 1

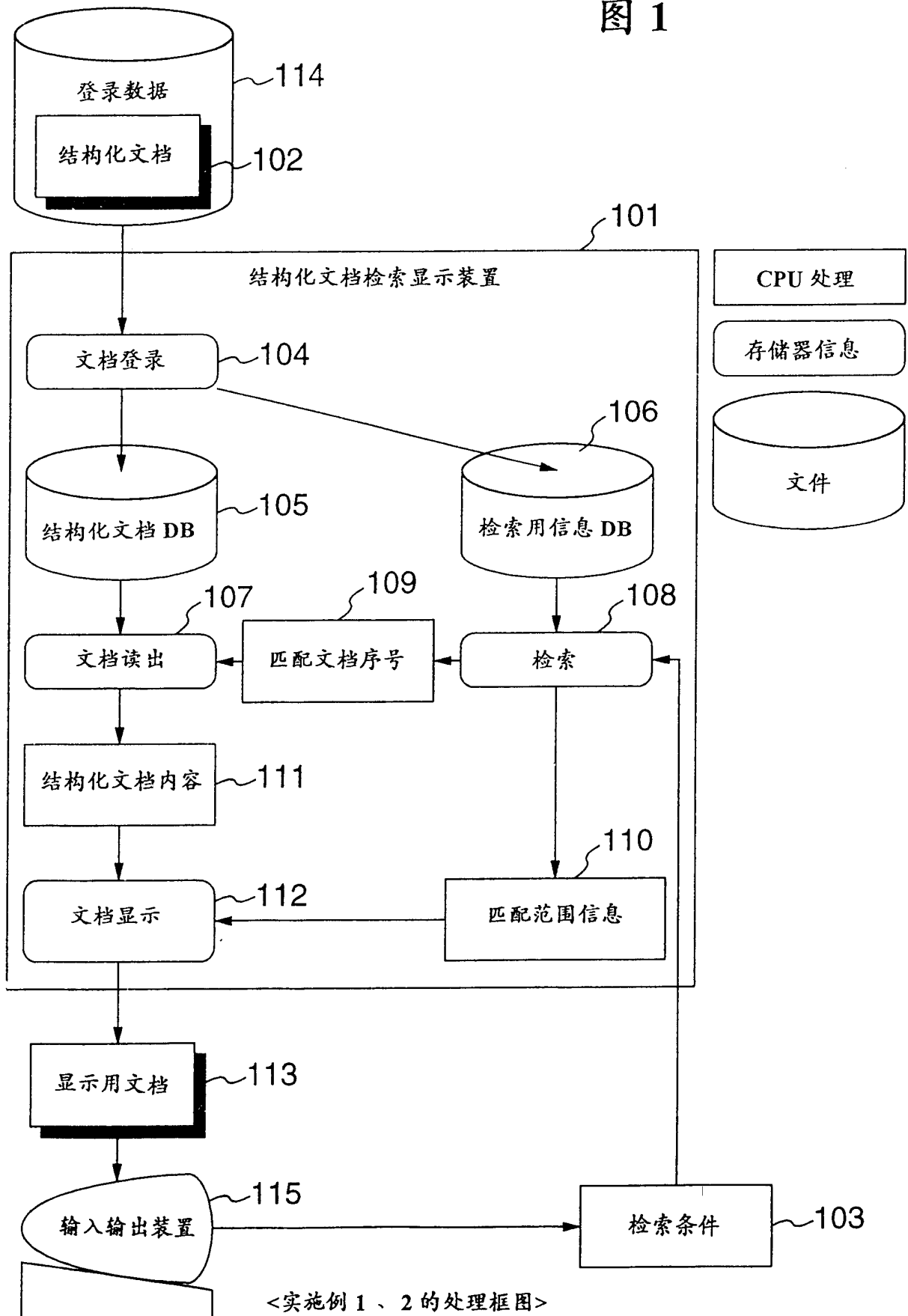
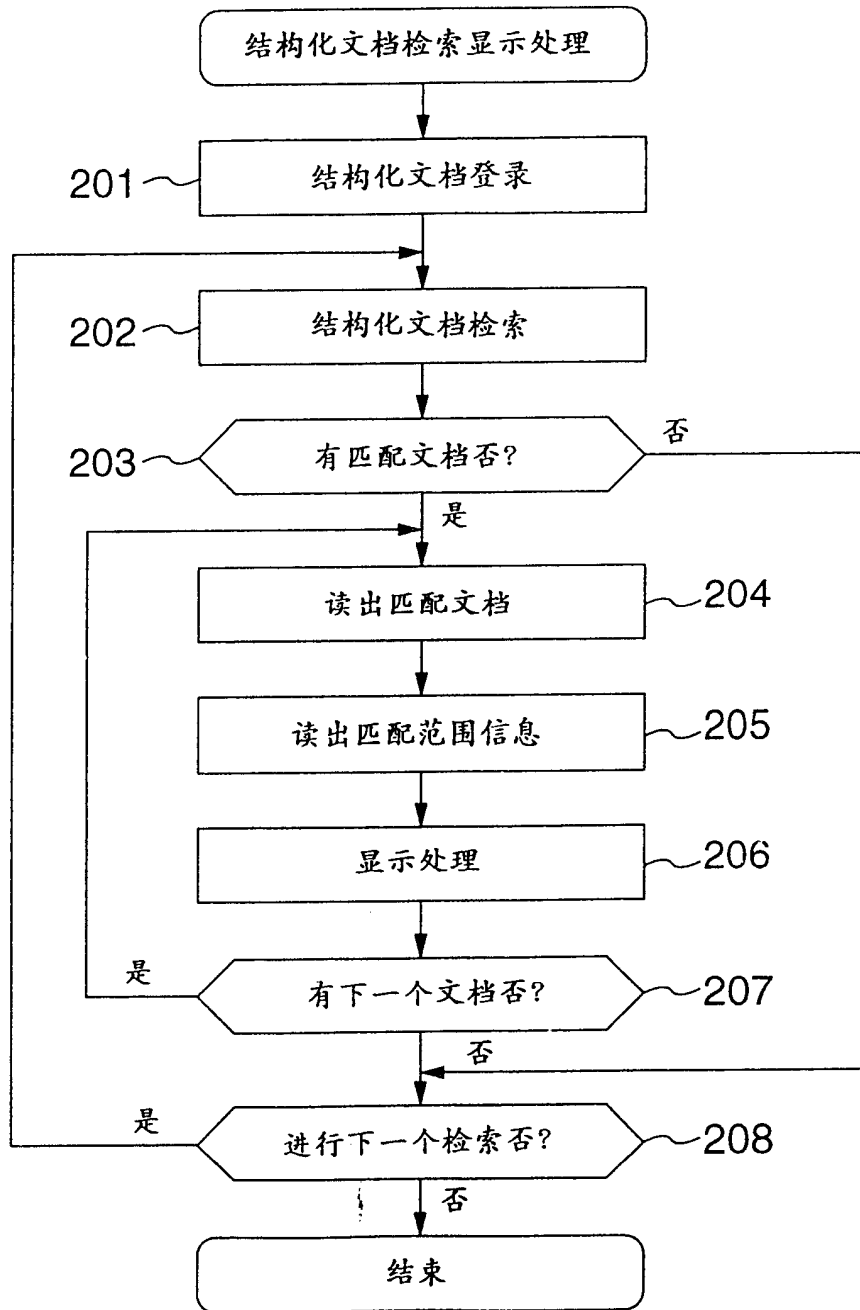


图 2



<实施例 1 的文档显示处理的流程图>

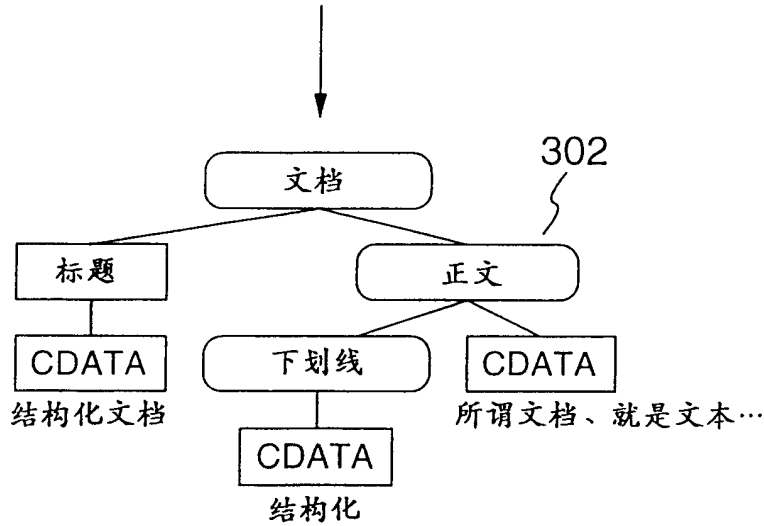
图 3

<结构化文档的登录>

301

```

<!DOCTYPE 文档 SYSTEM " 文档 .dtd">
<文档>
<标题>结构化文档</标题>
<正文>
<下划线>结构化</下划线>文档、就是文本...
</正文>
</文档>
    
```



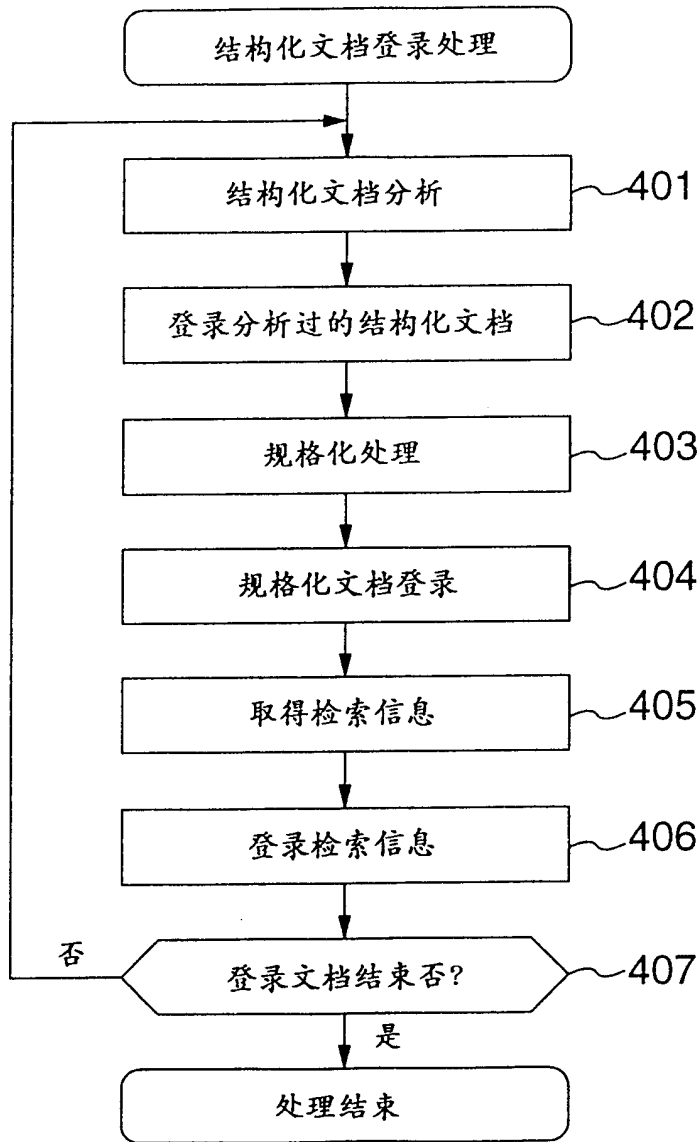
302

向列表形式的数据的变换

303

	结构种类	标记	内容
0	文档	文档	下位结构：1，3
1	要素	标题	下位结构：2
2	CDATA	-	字符串：“结构化文档”
3	要素	正文	下位结构：4，6
4	要素	下划线	下位结构：5
5	CDATA	-	字符串：“结构化”
6	CDATA	-	字符串：“文档和”

图 4



<结构化文档登录处理的流程图>

## 图 5

<检索用的文本>

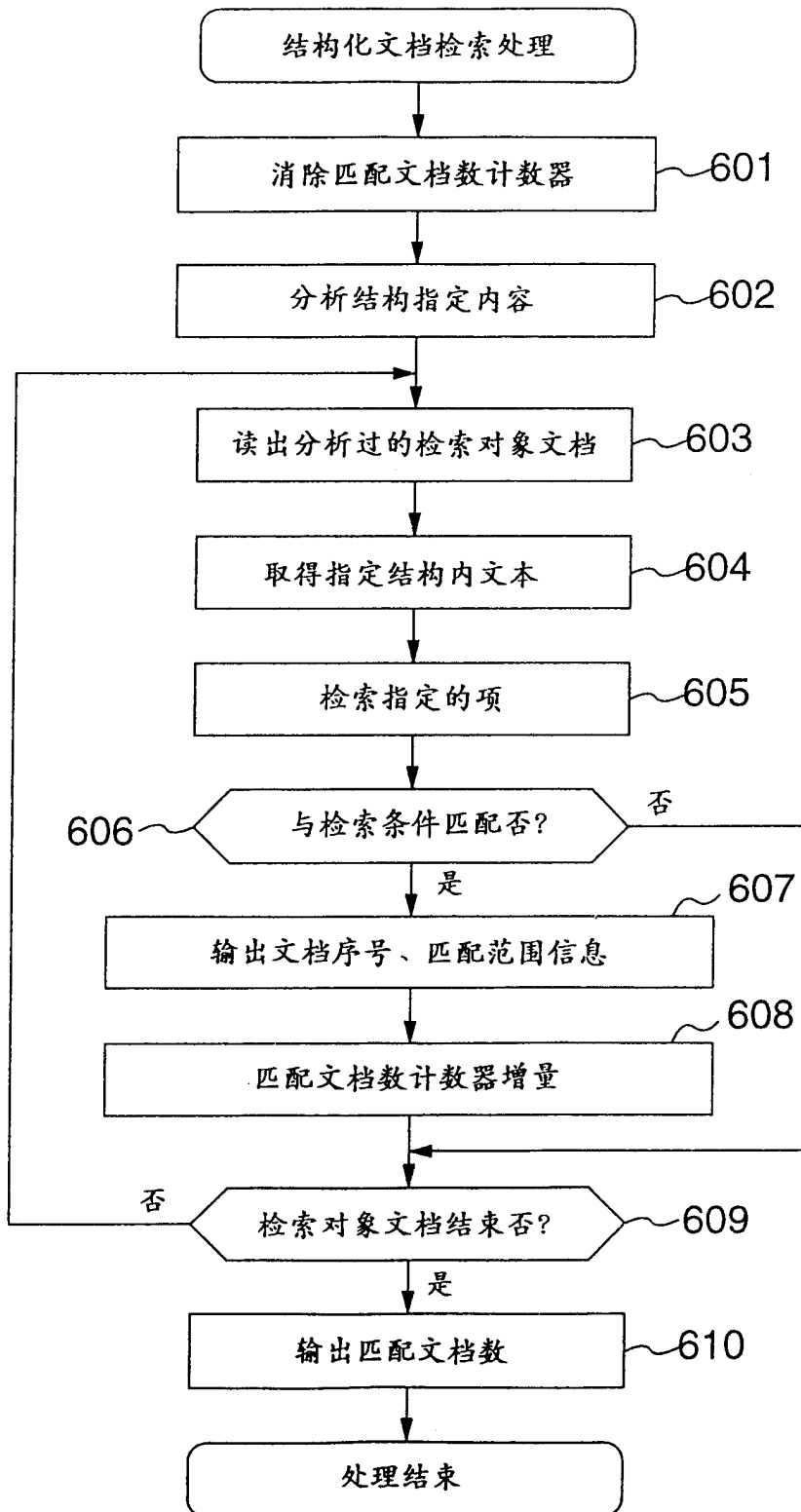
结构 ID	开始文字位置	字符串长度
2	0	10
5	10	6
6	16	44
⋮	⋮	⋮

<字符串>

所谓结构化文档、…

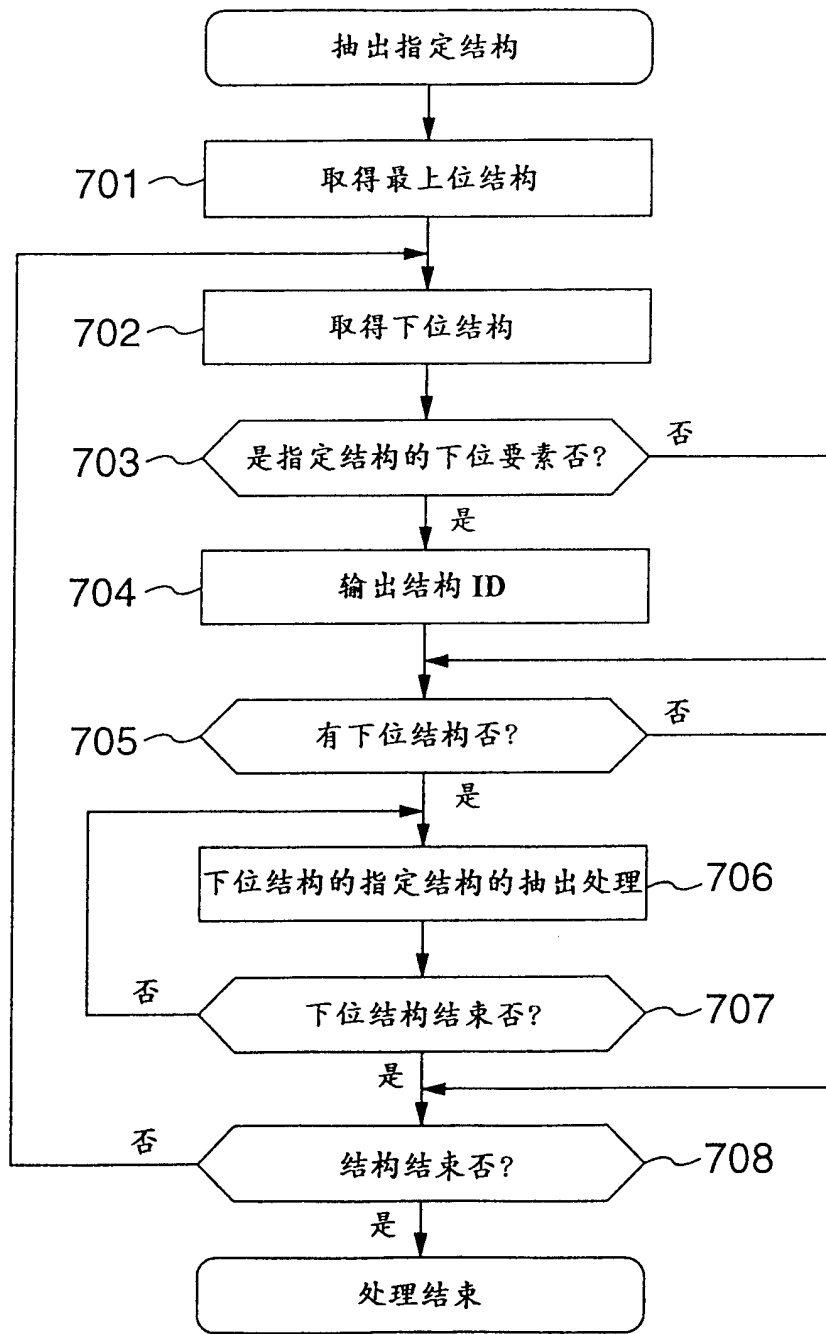


图 6



<结构化文档检索处理的流程图>

图 7



<指定结构抽出处理的流程图>

# 图 8

<结构指定的分析结果>

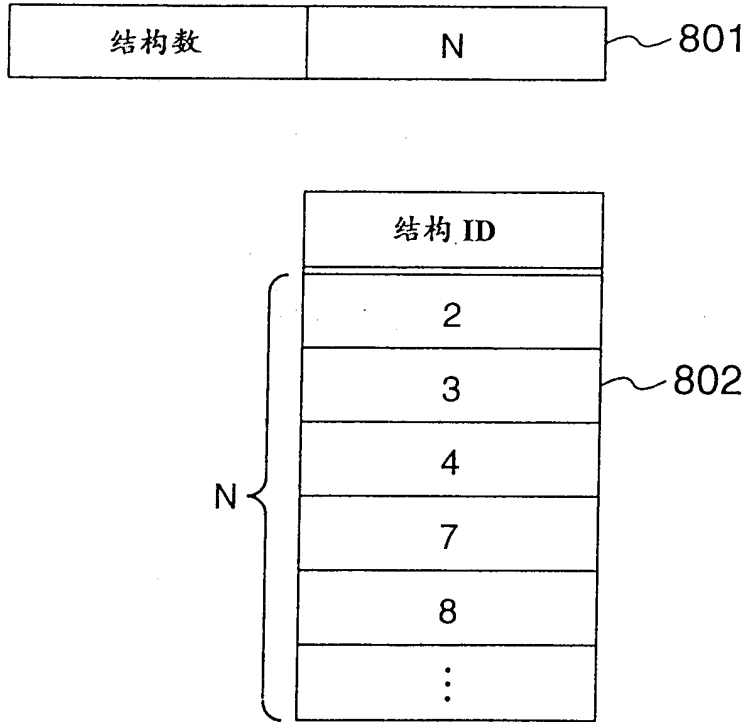
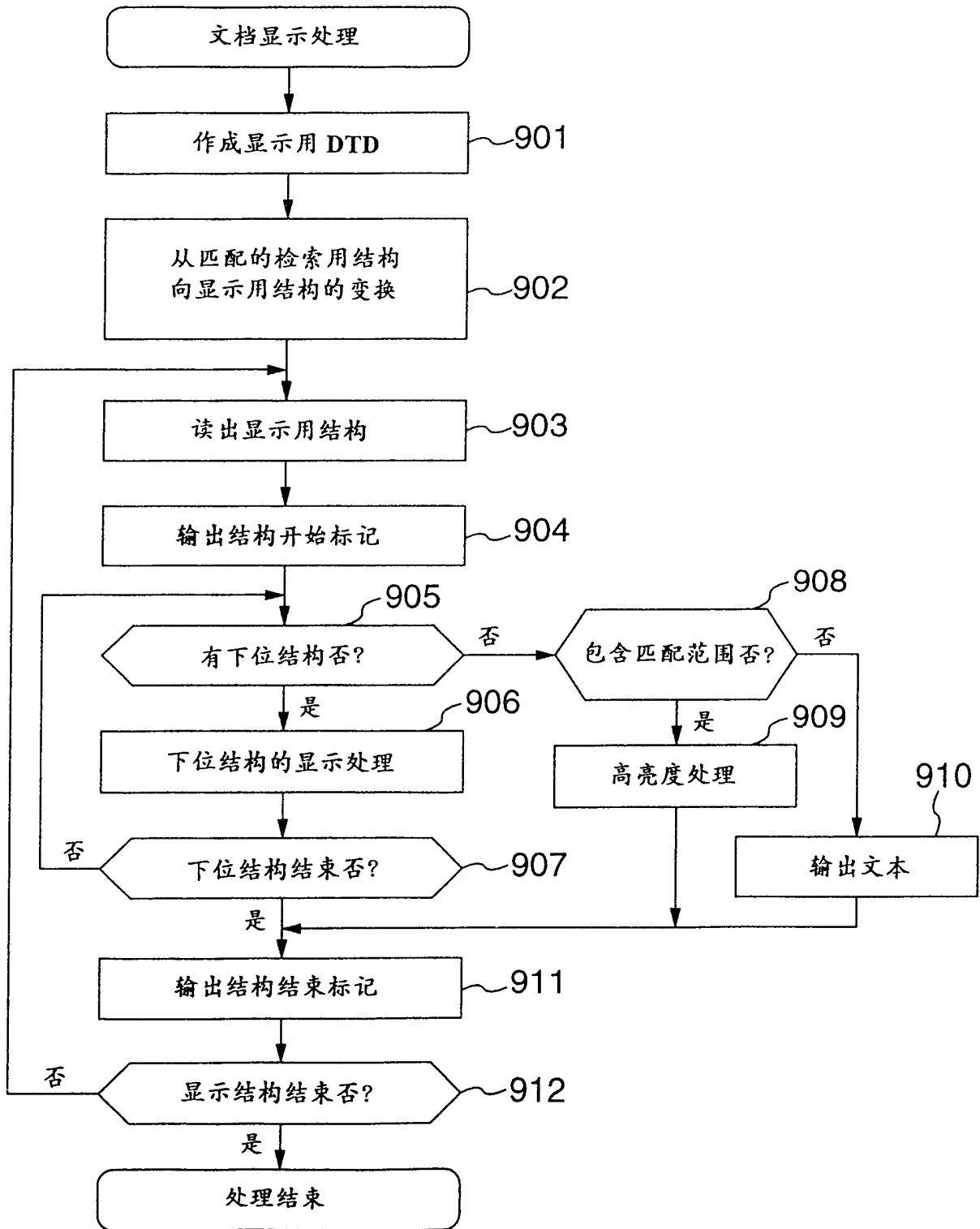


图 9



<结构化文档显示处理的流程图>

# 图 10

<结构化文档的例子>

[DTD] (文档.dtd)

1001

```

<!ELEMENT 文档      ... (标题, 正文) >
<!ELEMENT 标题      ... CDATA>
<!ELEMENT 正文      ... (#PCDATA | 下划线  )*>
<!ELEMENT 下划线    ... CDATA>

```

[文档下划线]

1002

```

<!DOCTYPE 文档      SYSTEM "文档.dtd">
<文档>JMENT>
<标题>结构化文档</标题>
<正文>`
<下划线>结构化</下划线>文档、就是文本...

</正文>
</文档>

```

检索“结构化文档”，显示结果

<高亮度处理后的结构化文档的例子>

[DTD] (显示.dtd)

1003

```

<!ELEMENT 文档      ... (标题, 正文) >
<!ELEMENT 标题      ... (#PCDATA | 高亮度  )*>
<!ELEMENT 正文      ... (#PCDATA | 下划线 | 高亮度  )*>
<!ELEMENT 下划线    ... (#PCDATA | 高亮度  )*>
<!ELEMENT 高亮度    >... (#PCDATA)

```

[文档实例]

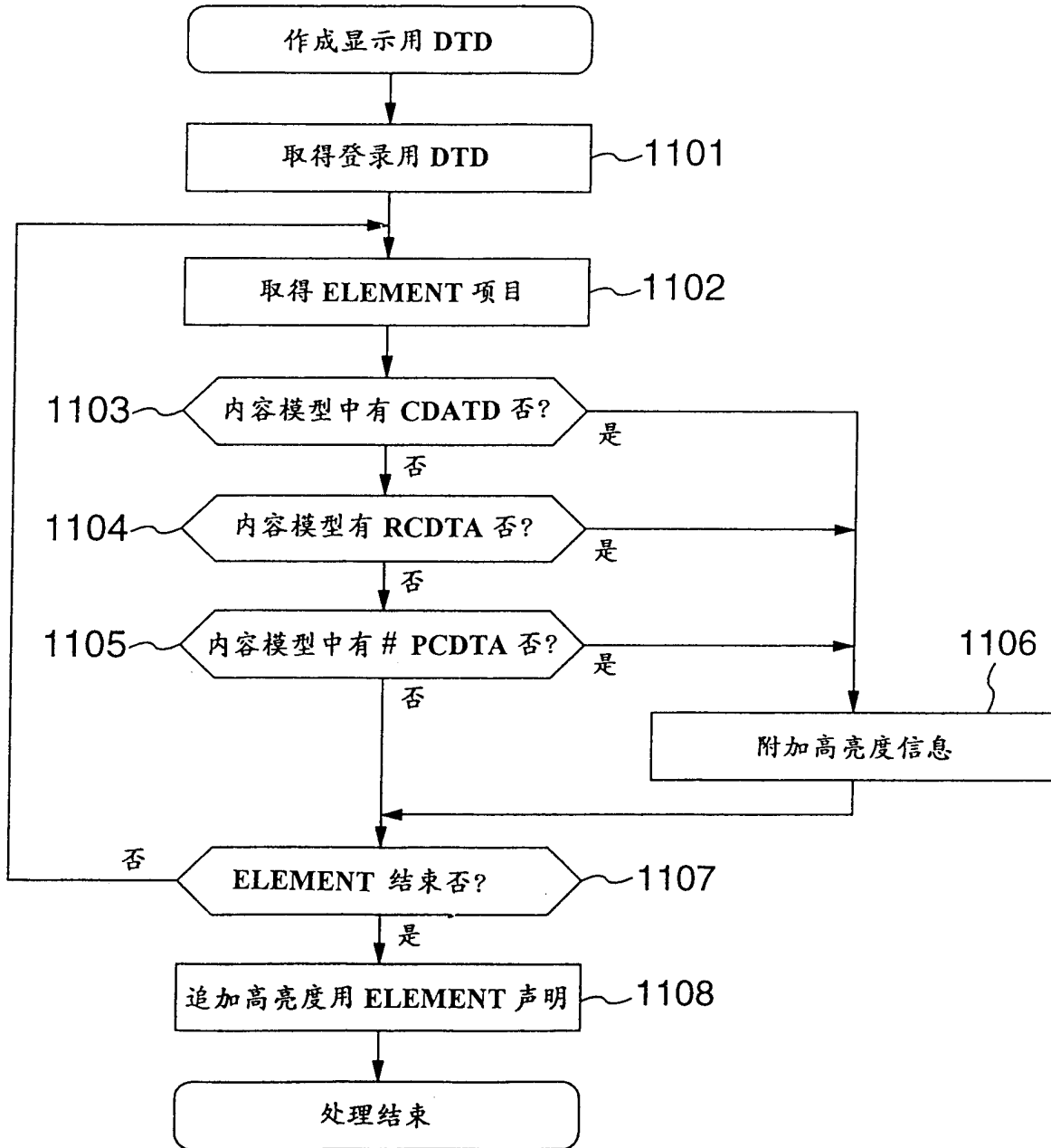
1004

```

<!DOCTYPE 文档      SYSTEM "显示.dtd">
<文档>
<标题><高亮度>结构化文档</高亮度></标题>
<正文>
<下划线><高亮度>结构化</高亮度></下划线>
<高亮度>文档</高亮度>就是文档...
</正文>
</文档>

```

图 11



<显示用 DTD 作成的流程图>

### 图 12

<结构化文档检索用规格化处理>

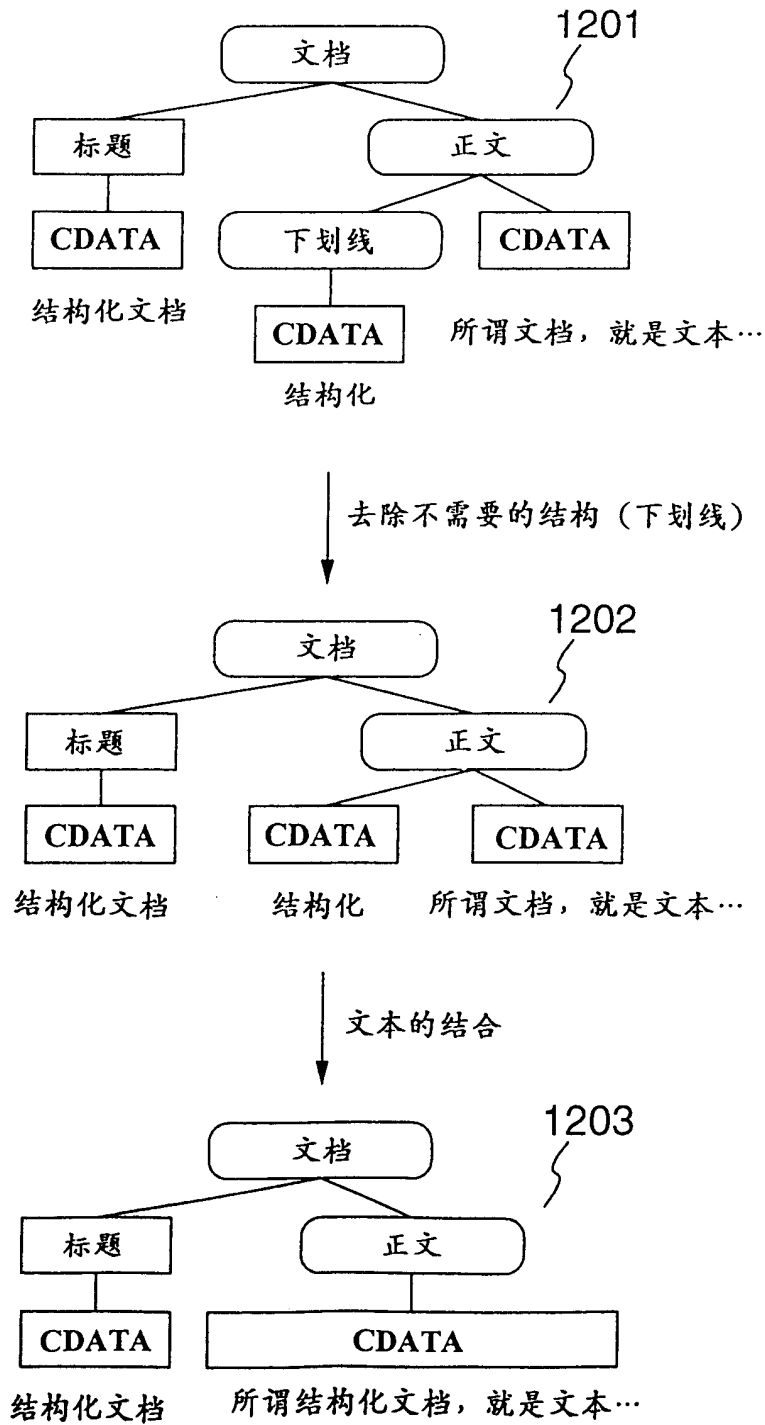
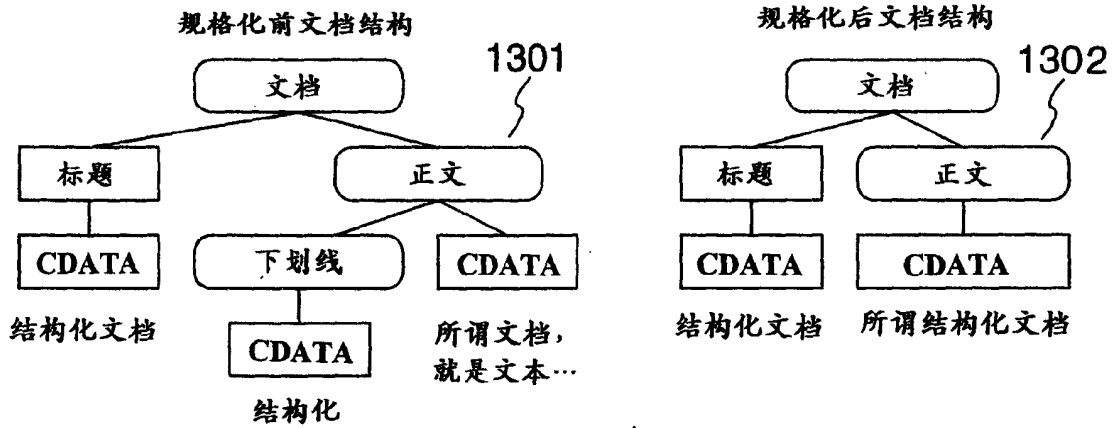


图 13

<向规格化文档的表形式的数据的变换>



结构信息表

向表形式的数据的变换 1303

结构标识符	结构种类	标记	内容
0	文档	文档	下位结构: 1, 3
1	要素	标题	下位结构: 2
2	CDATA	-	字符串: "结构化文档"
3	要素	正文	下位结构: 4, 6
4	要素	下划线	下位结构: 5
5	CDATA	-	字符串: "结构化"
6	CDATA	-	字符串: "所谓文档..."
7	文档	文档	下位结构: 1, 8
8	要素	正文	下位结构: 9
9	CDATA	-	字符串: "所谓结构化文档,..."

规格化后追加

规格化对应表

1304

规格化后结构标识符	结构种类	规格化前结构标识符	字符串范围	
			规格化后	规格化前
7	文档	0	-	
8	要素	3	-	
9	CDATA	5	0-5	0-5
		6	6-49	0-43



图 14

&lt;匹配范围信息&gt;

1401

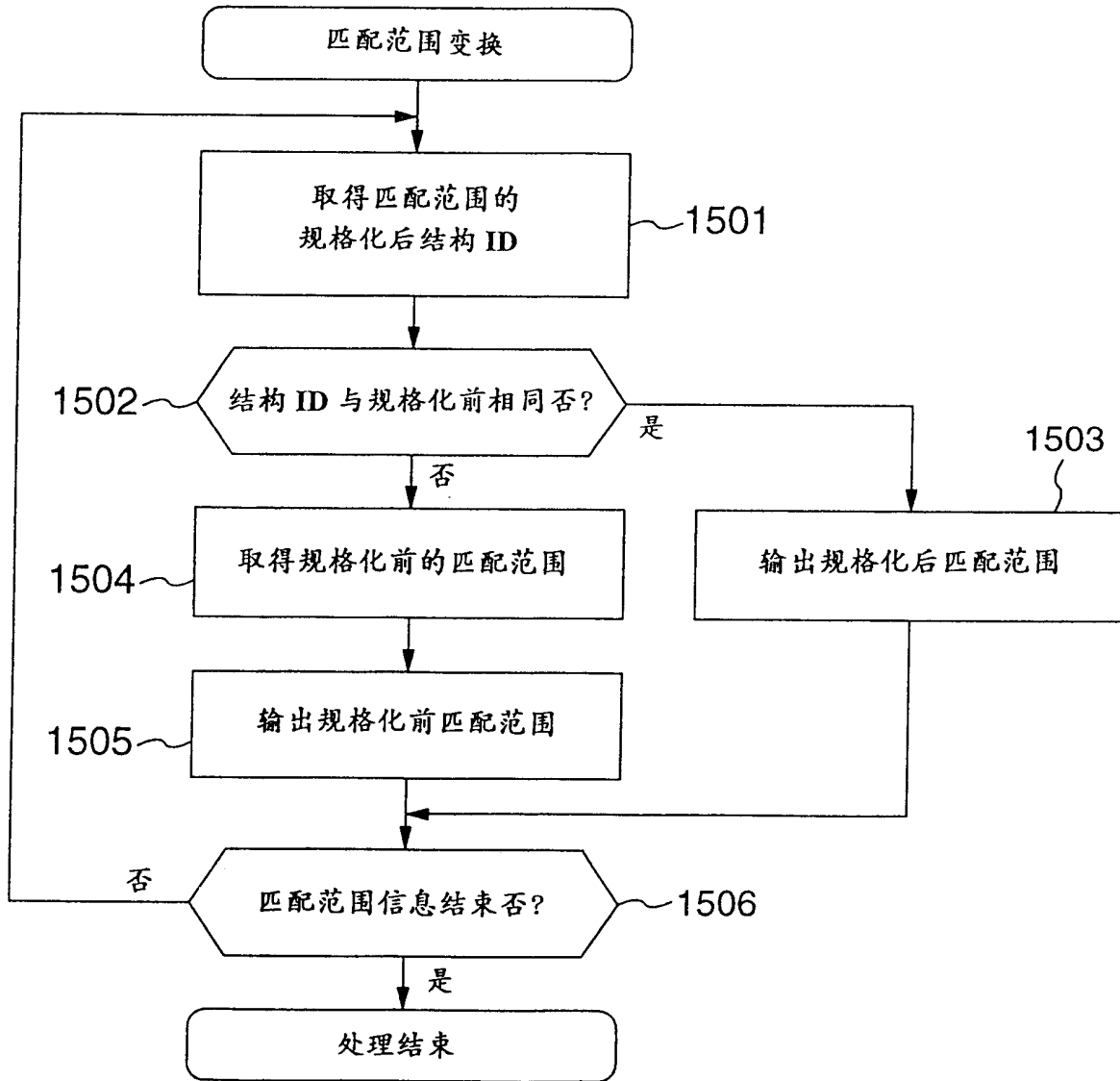
规格化后结构 ID	匹配范围	
	开始位置	长度
2	0	10
9	0	10

↓  
向规格化前的范围的变换

1402

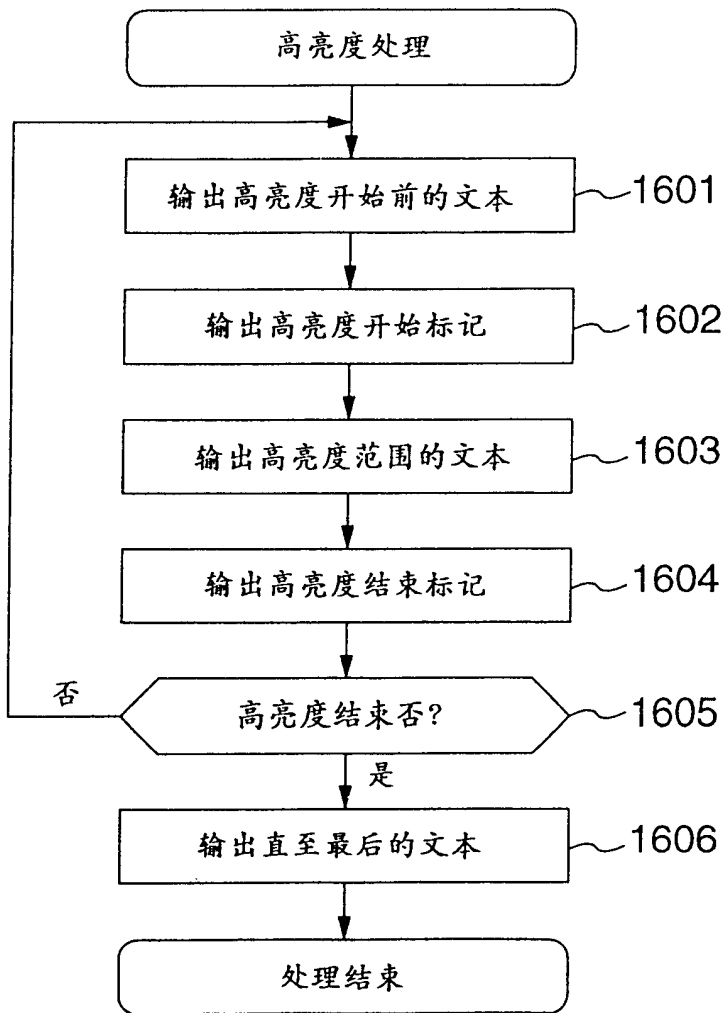
规格化前结构 ID	匹配范围	
	开始位置	长度
2	0	10
5	0	6

图 15



<匹配范围取得处理的流程图>

图 16



<高亮度处理的流程图>

图 17

&lt;匹配范围信息&gt;

结构 ID	匹配条件	匹配范围	
		开始位置	长度
2	项	0	10
9	频度	0	10
9	结构	0	50
9	项	30	4
9	项	36	4
9	距离	30	10

图 18

&lt;高亮度方法定义&gt;

匹配条件	高亮度方法	层次信息
项	红色	1
频度	字体大	1
距离	下划线	2
结构	斜体	3

图 19

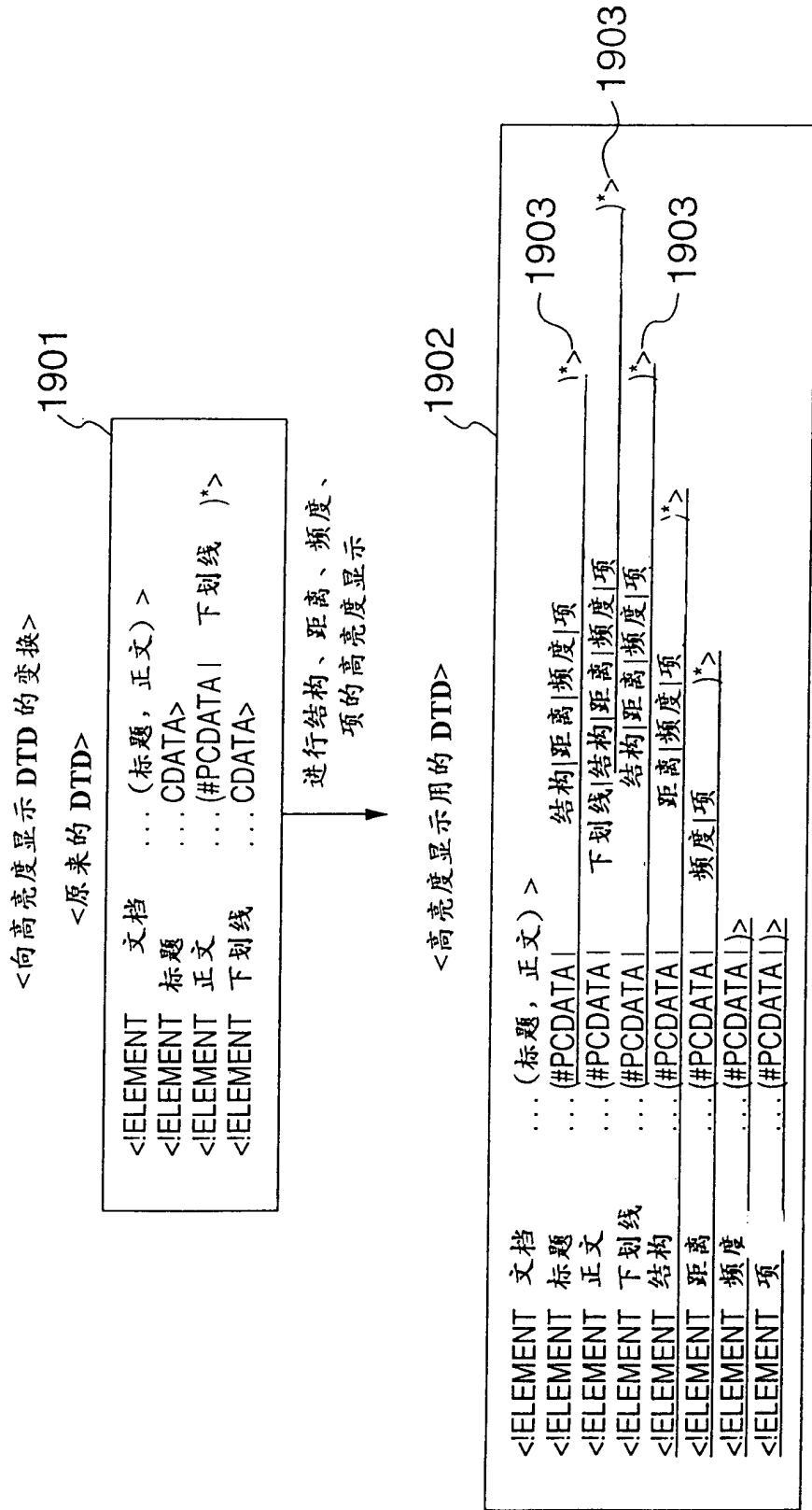
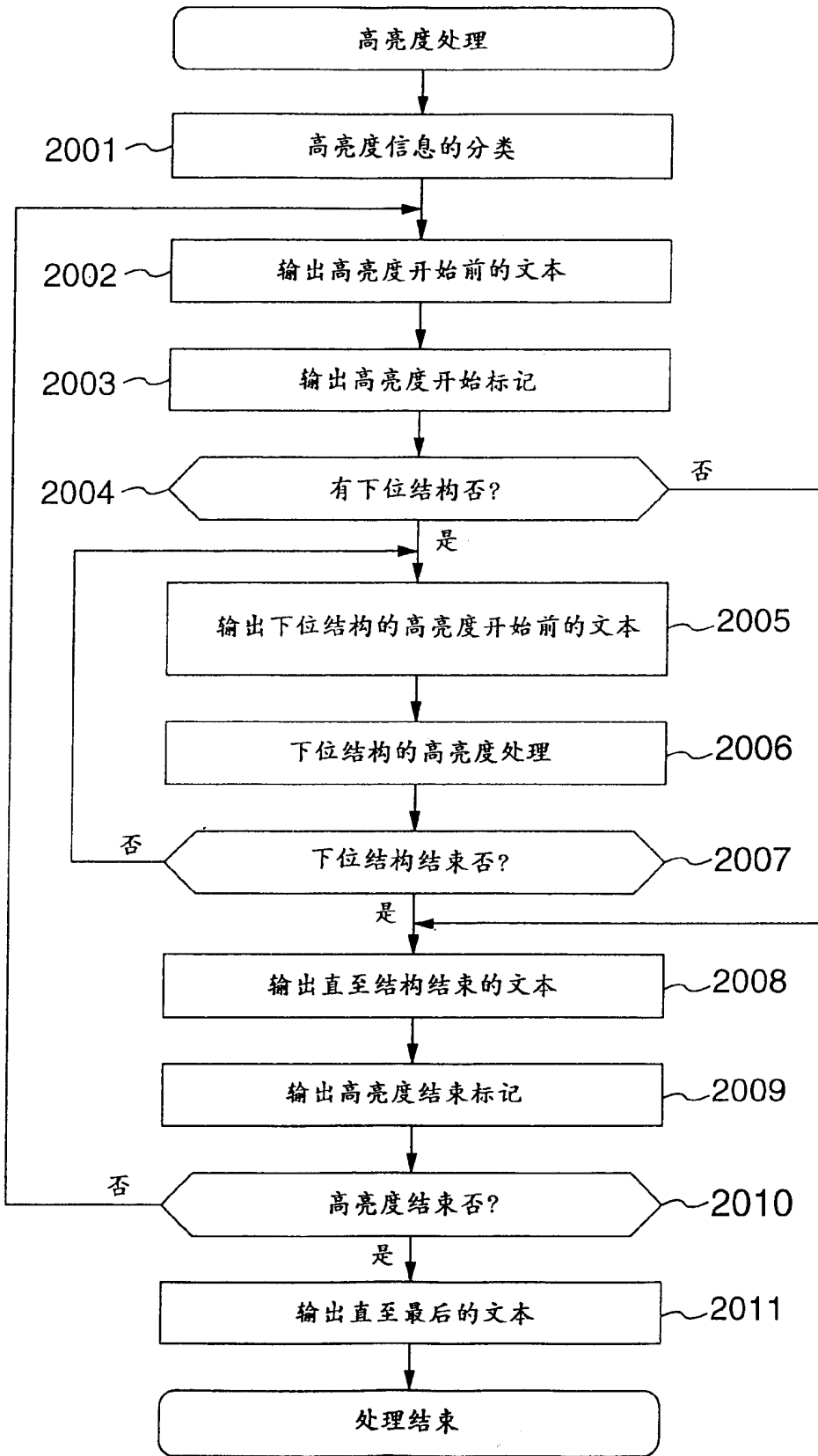


图 20



<匹配范围取得处理的流程图>

图 21

```

<文档>
<标题>><项>结构化文档</项></标题>> <正文>
<文档>
<下划线>><结构>><频度>结构化</频度></结构>></下划线>><结构>><频度>
文档</频度>>就是将<距离>><项>>标记</项>>向文本中进行<项>>插入
</项>></距离>>的文档.....</结构>

</正文>
</文档>

```

图 22

<高亮度显示的例子>

所谓结构化文档，就是将标记插入到文本中的文档.....

图 23

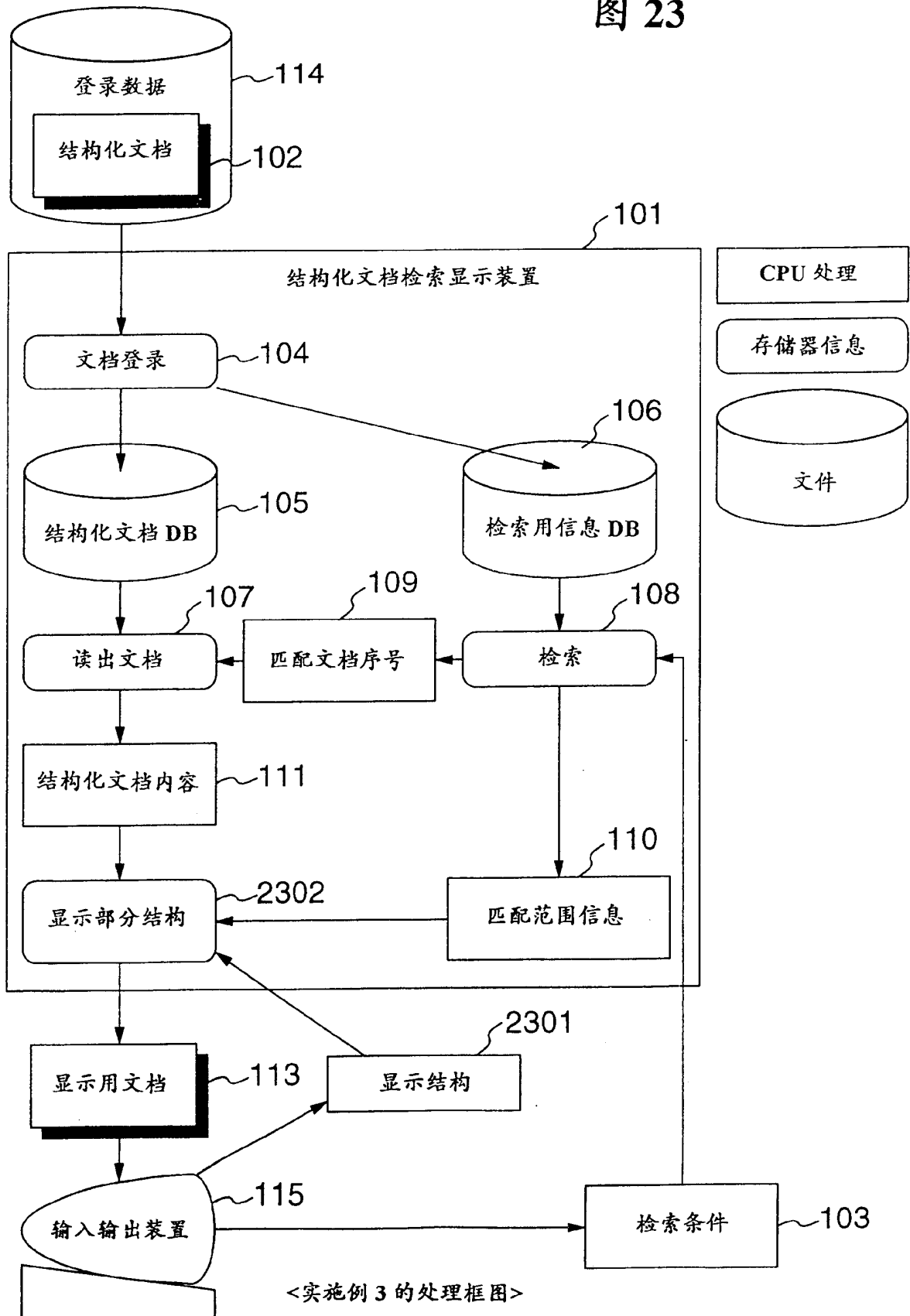
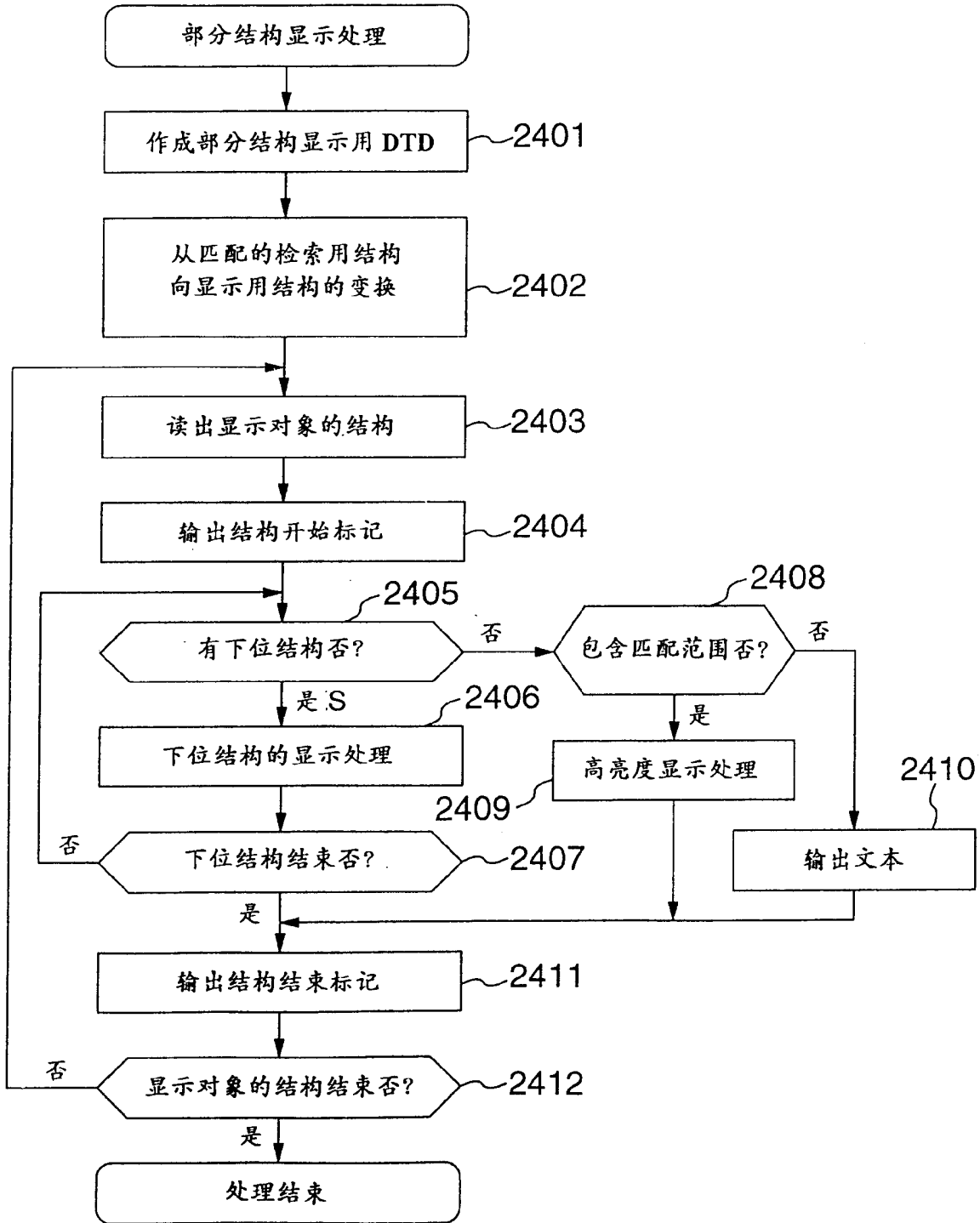




图 24



<结构化文档显示处理的流程图>

# 图 25

<向部分结构显示用 DTD 的变换>

<原来的 DTD>

2501

```

<!ELEMENT 文档      ... (标题, 正文) >
<!ELEMENT 标题      ... CDATA>
<!ELEMENT 正文      ... (#PCDATA | 下划线  )*>
<!ELEMENT 下划线    ... CDATA>

```

设可以抽出部分结构

附加 高亮度

<部分结构显示用的>

2502

```

<!ELEMENT 文档      00 (标题?, 正文?) >
<!ELEMENT 标题      ... (#PCDATA | 高亮度  )*>
<!ELEMENT 正文      00(#PCDATA | 下划线 | 高亮度  )*>
<!ELEMENT 下划线    ... (#PCDATA | 高亮度  )*>
<!ELEMENT 高亮度    ... (#PCDATA)>

```

<抽出了部分结构的 SGML 的例子>

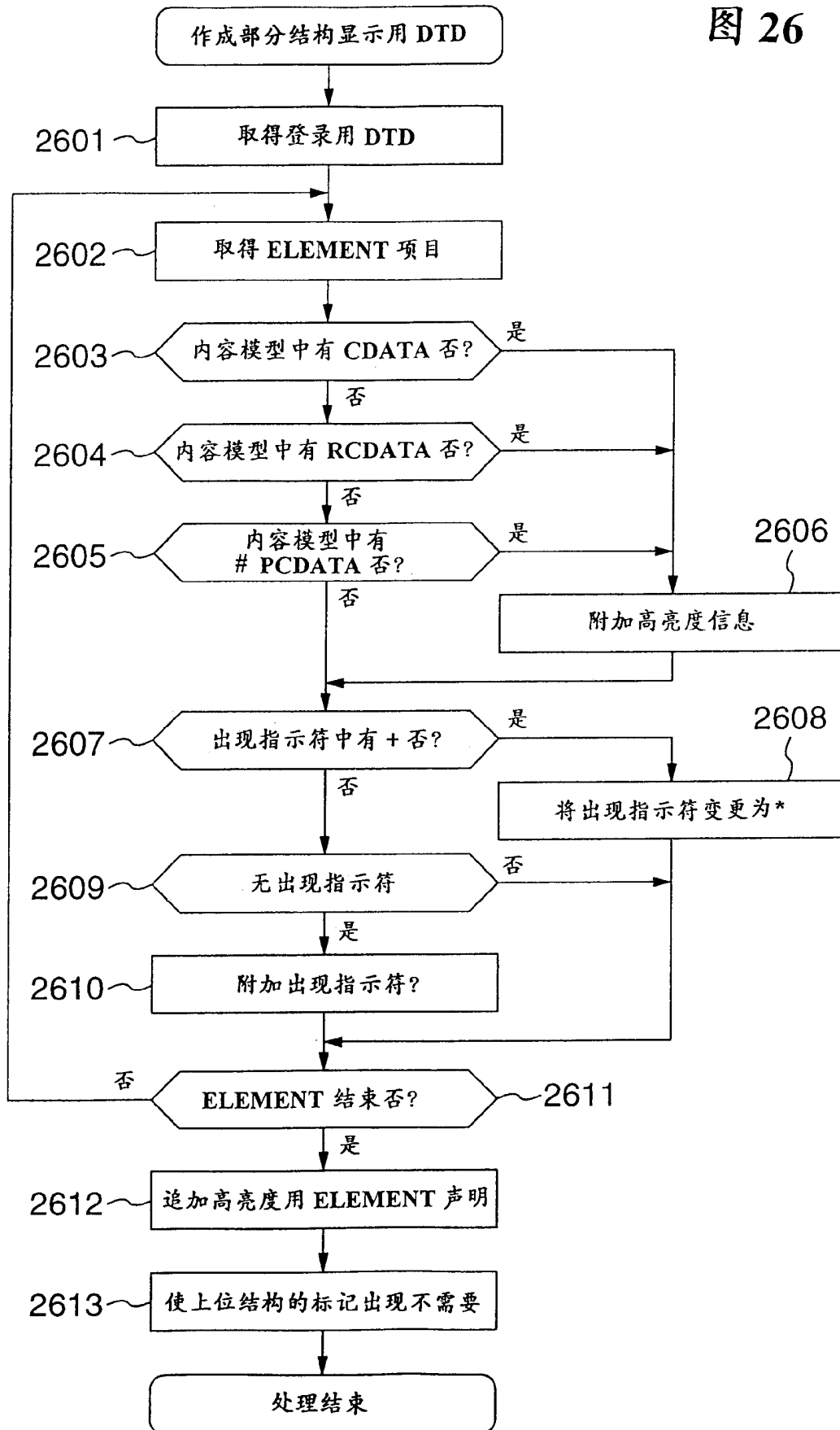
2503

```

<标题>
<高亮度>结构化文档</高亮度>
</标题>

```

图 26



<部分结构显示用 DTD 作成的流程图>

图 27

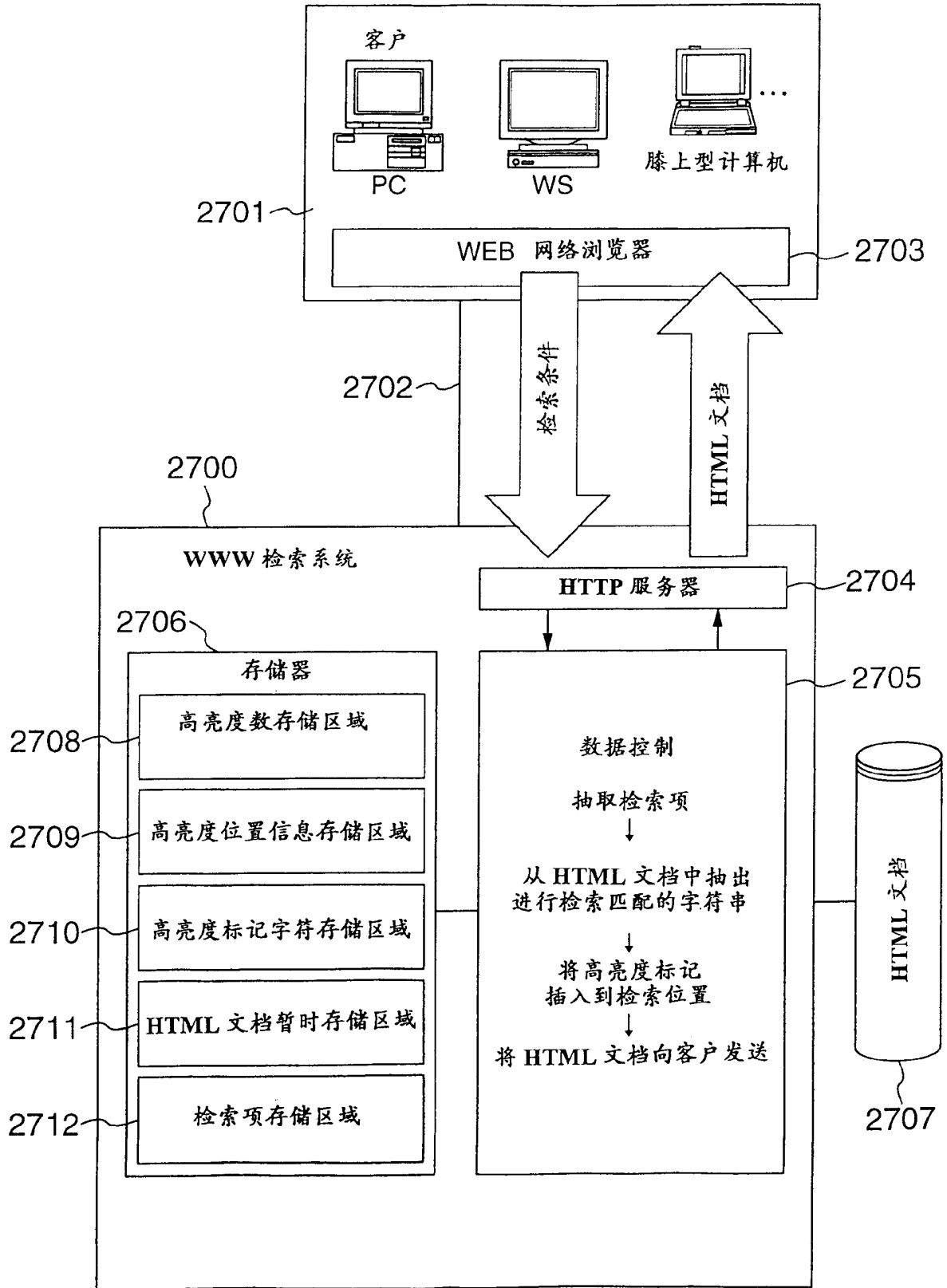


图 28

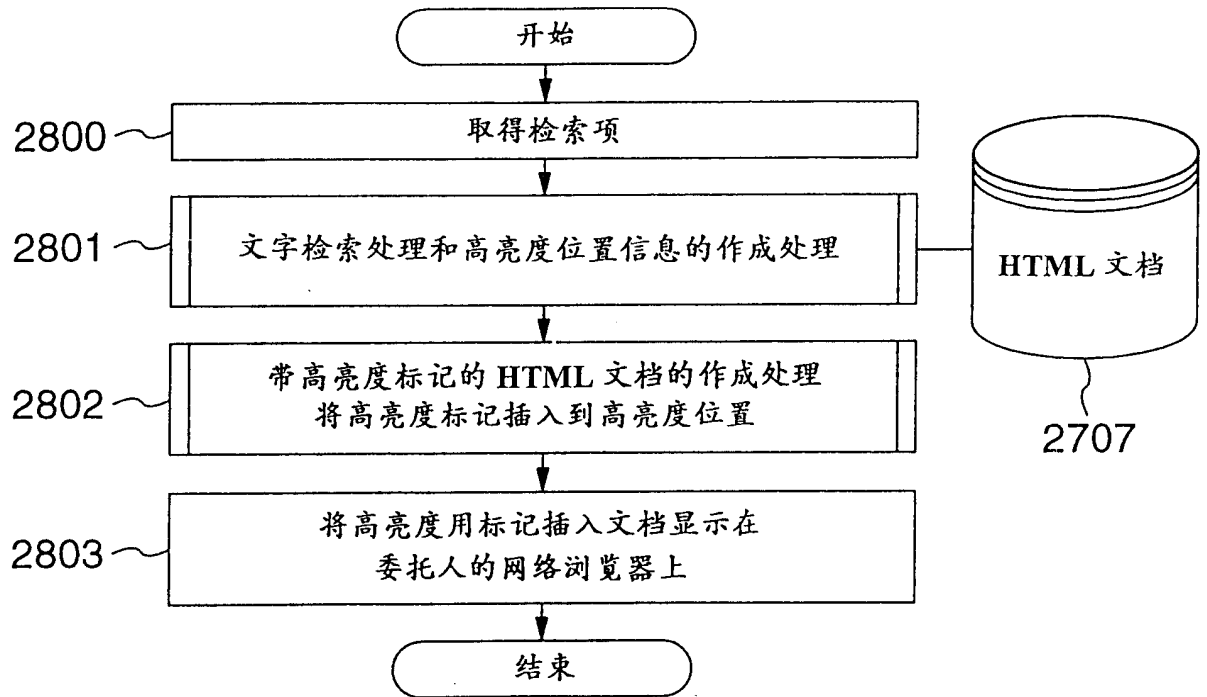


图 29

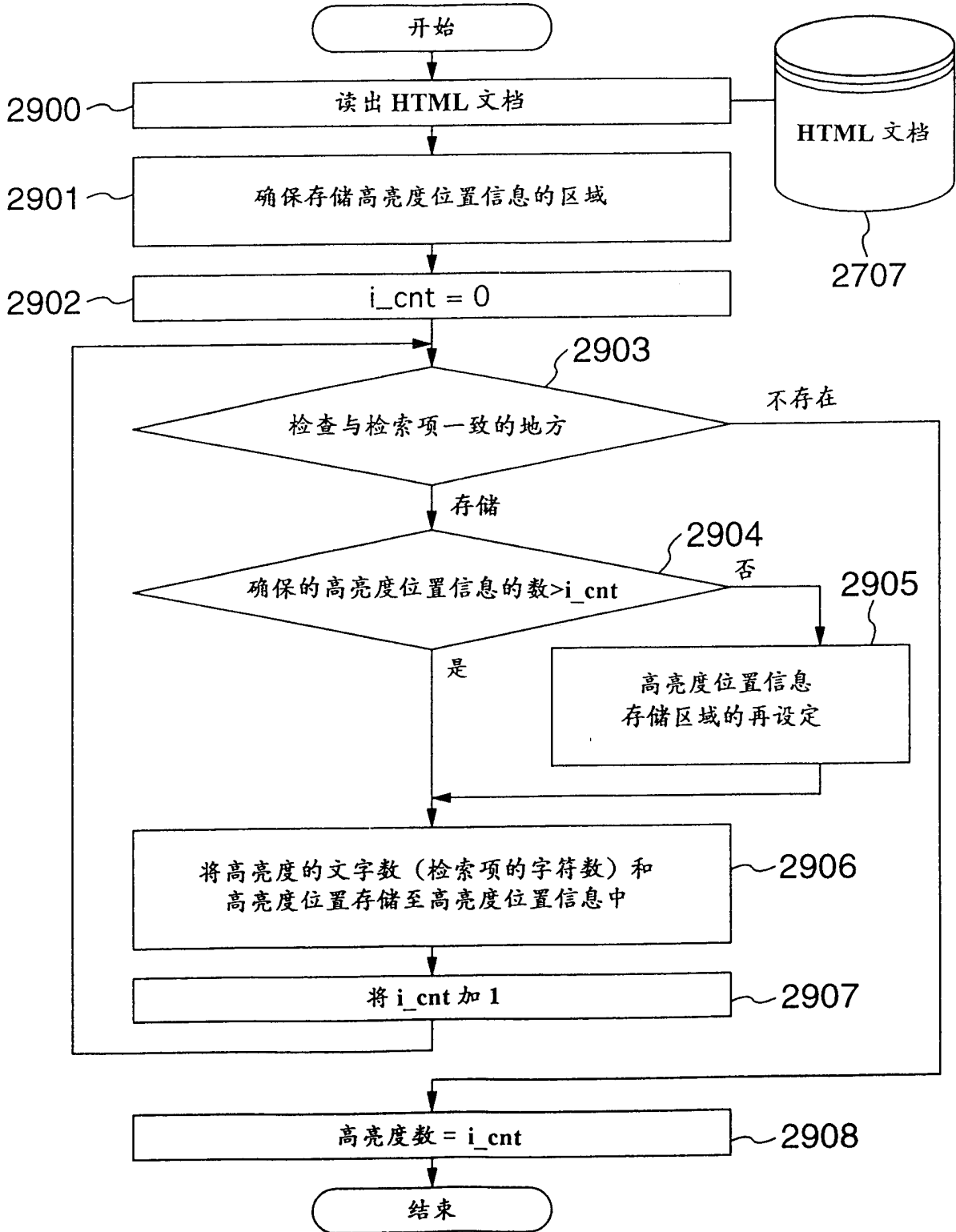


图 30

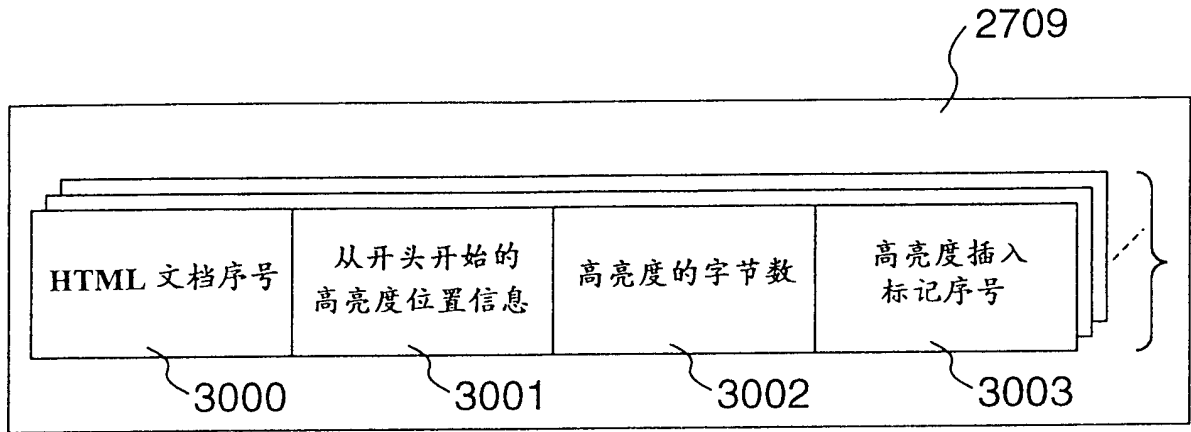


图 31

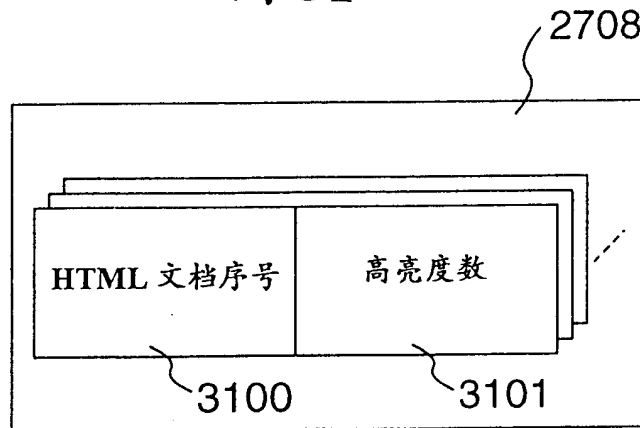
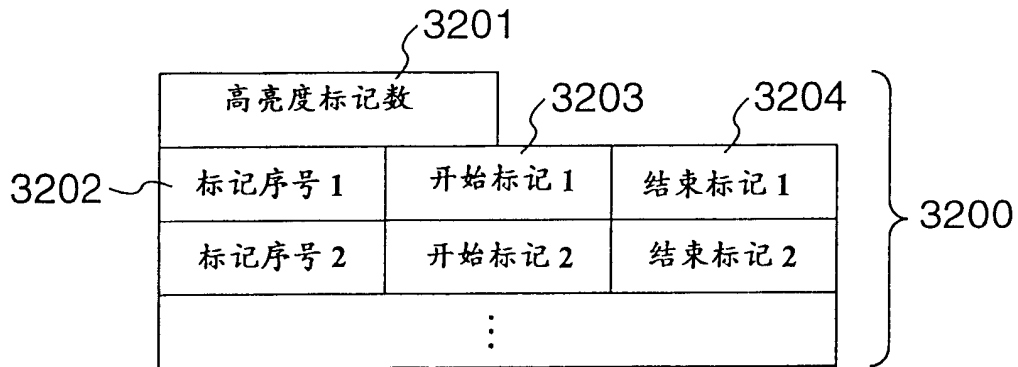


图 32

(1) 结构体



(2) 使用例

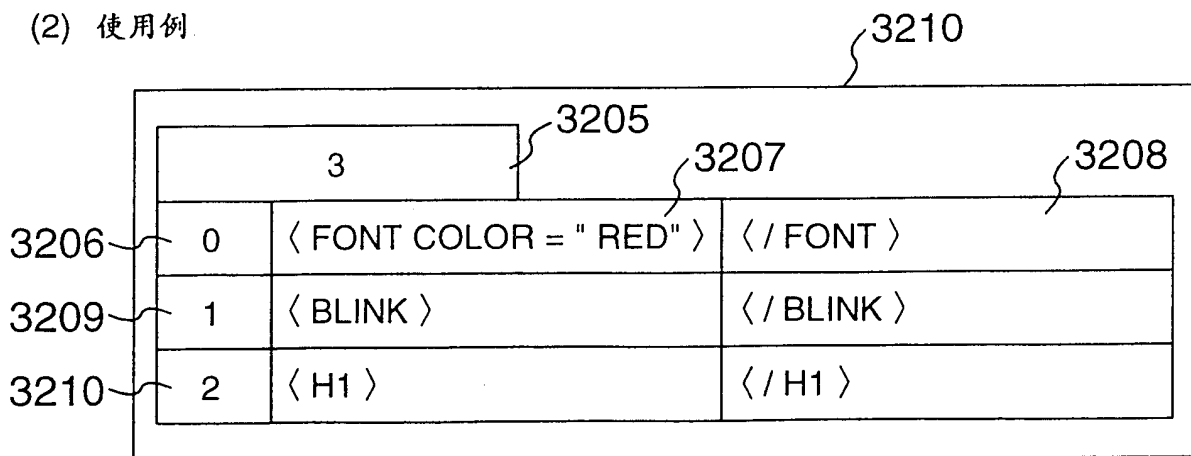




图 33

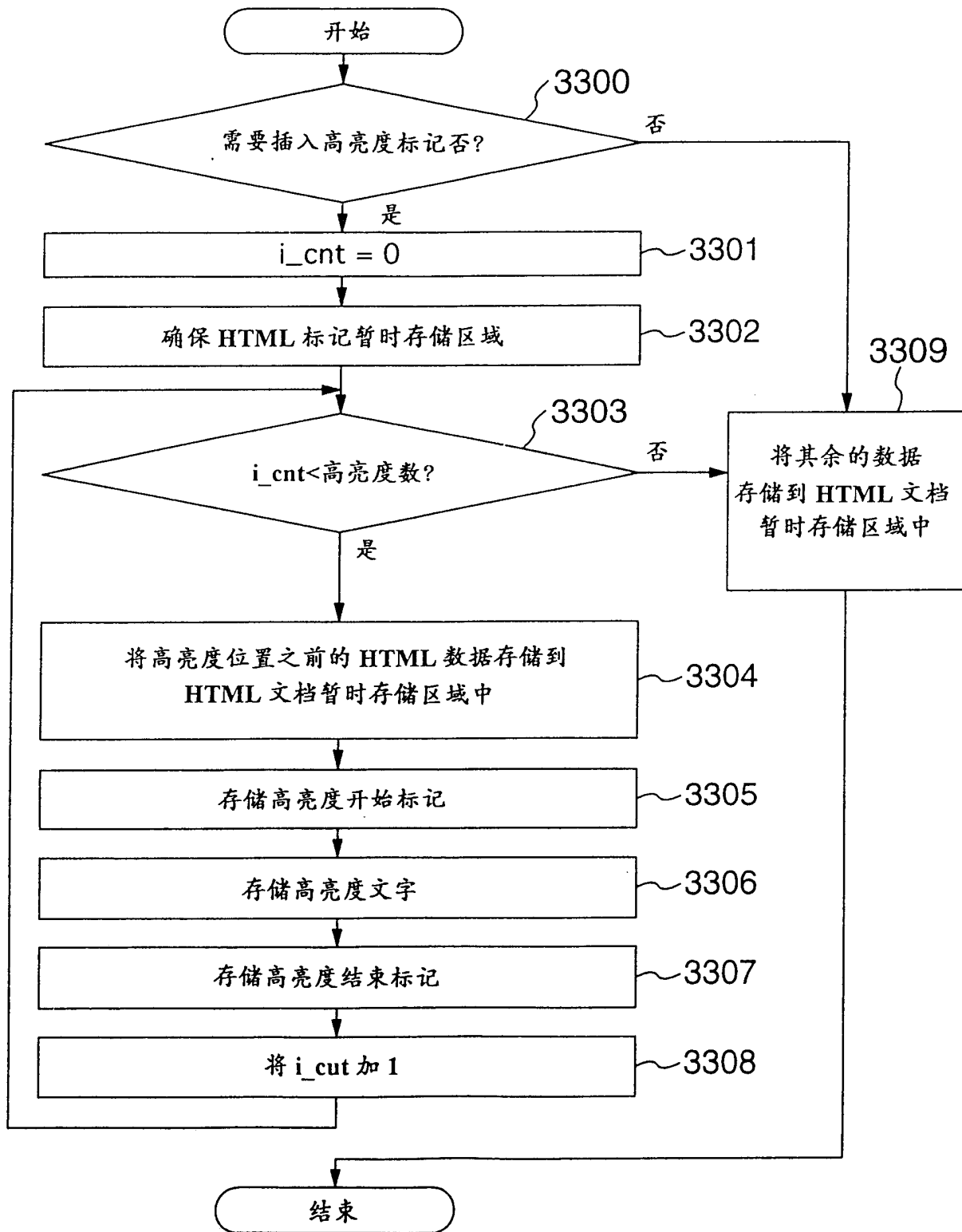
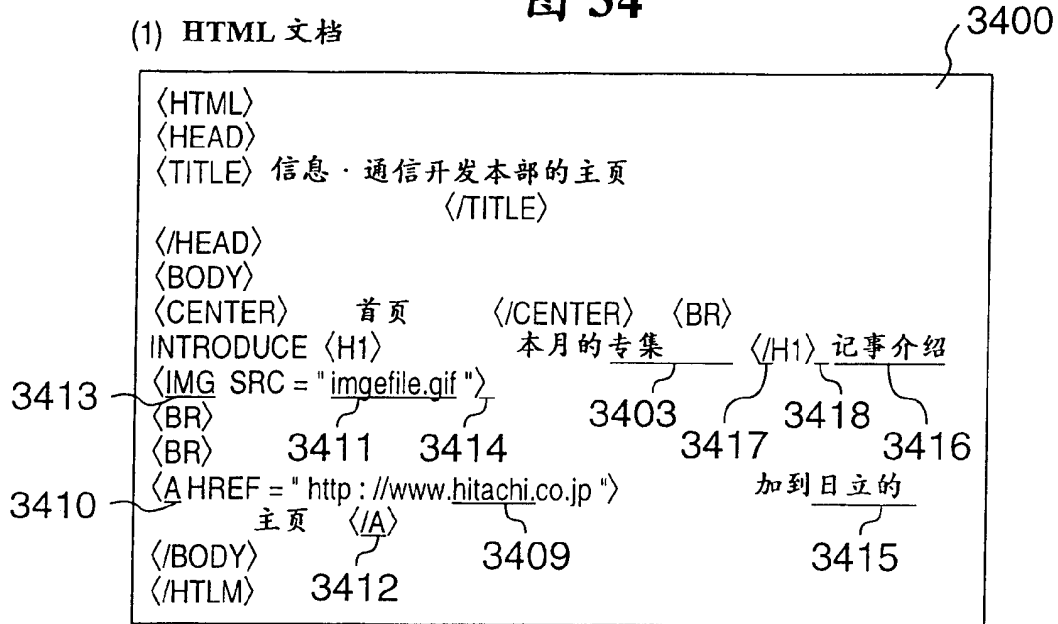
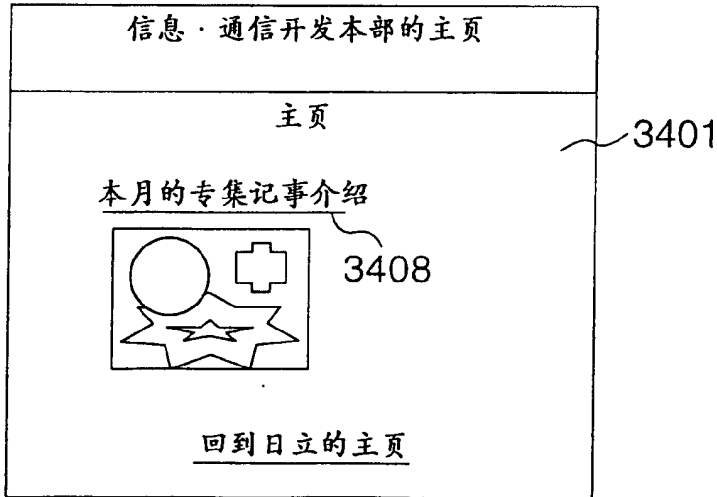


图 34



(2) DISPLAY SCREEN



(3) 高亮度位置信息存储区域

HTML 文档序号	从开头开始的高亮度位置信息	高亮度的字节数	高亮度插入标记序号
001	122	4	1

图 35

(1) 高亮度标记插入后 HTML 文档

3500

```

<HTML>
<HEAD>
<TITLE> 信息·通信开发本部的主页
      </TITLE>
</HEAD>
<BODY>
<CENTER>    主页    </CENTER> <BR>
              <H1>    本月的<BLINK>专集    </BLINK>
</H1> 记事介绍
<IMG SRC = " imgefile.gif ">
<BR>
<BR>
<A HREF = " http ://www.hitachi.co.jp "> 回到日立的主页
</A>
</BODY>
</HTLM>

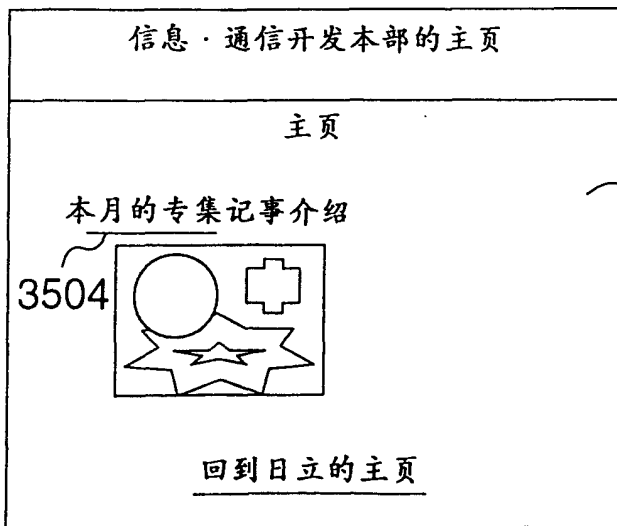
```

3501

3502

(2) 高亮度标记插入后显示画面

3503



3504

图 36

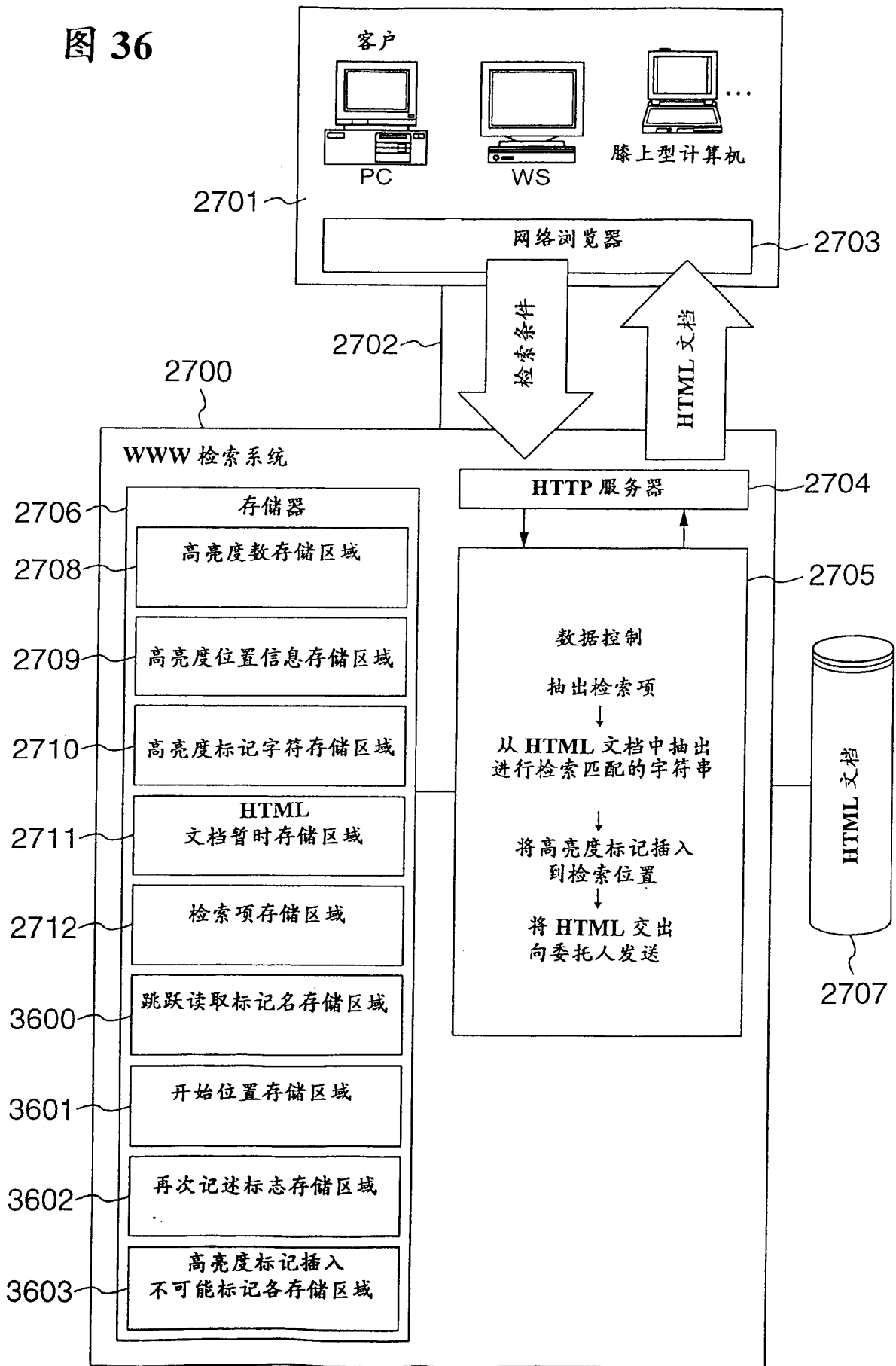


图 37

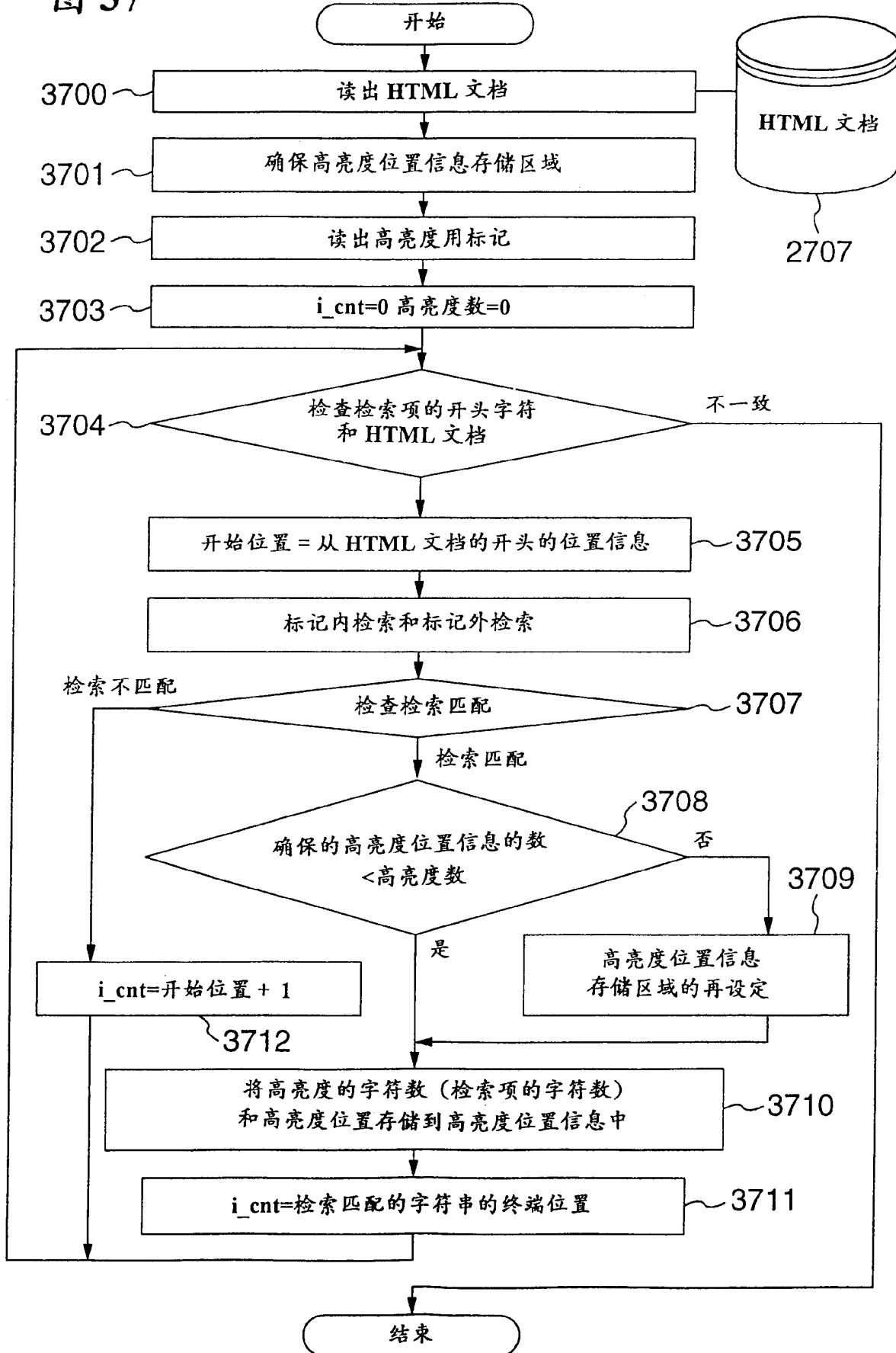


图 38

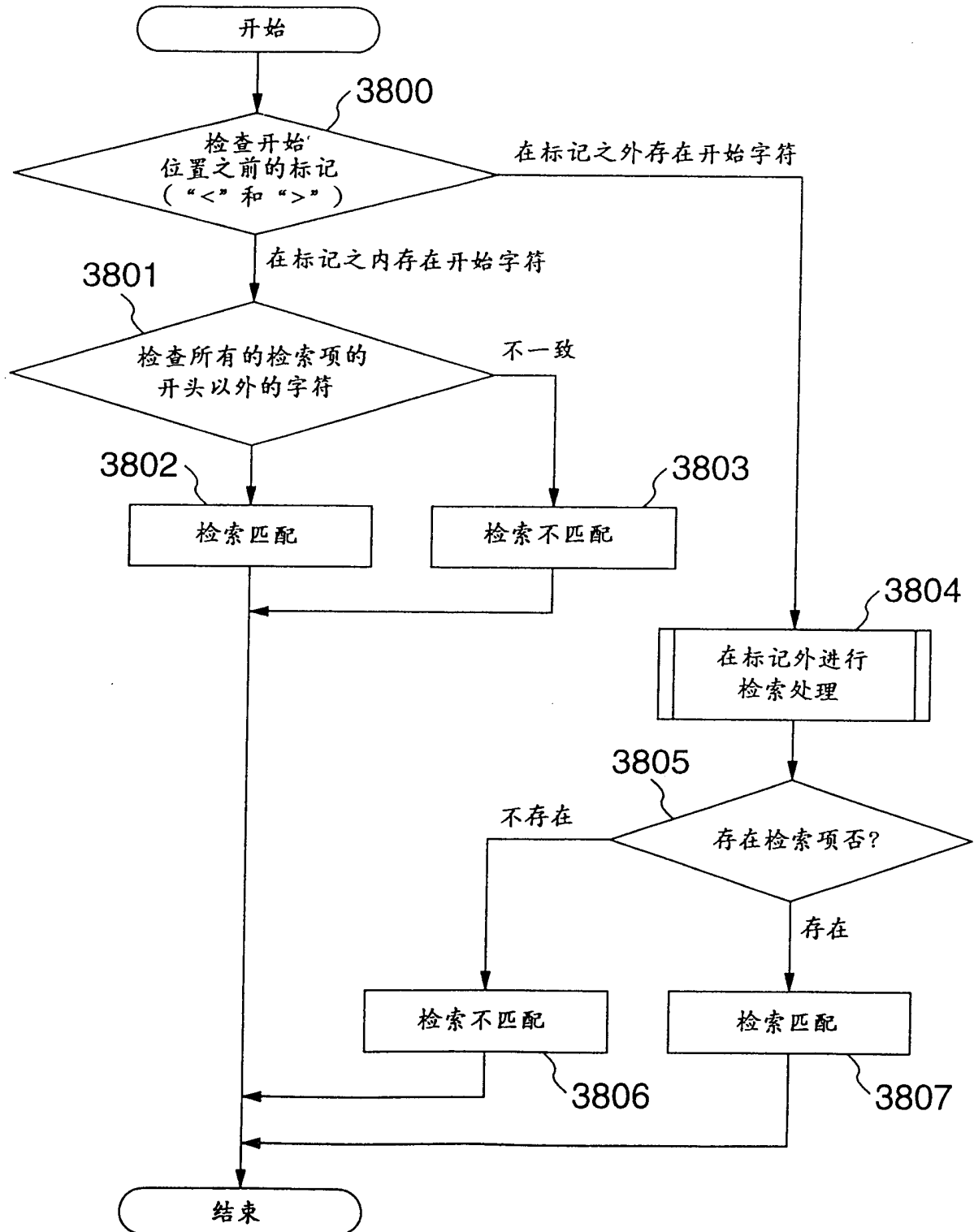


图 39

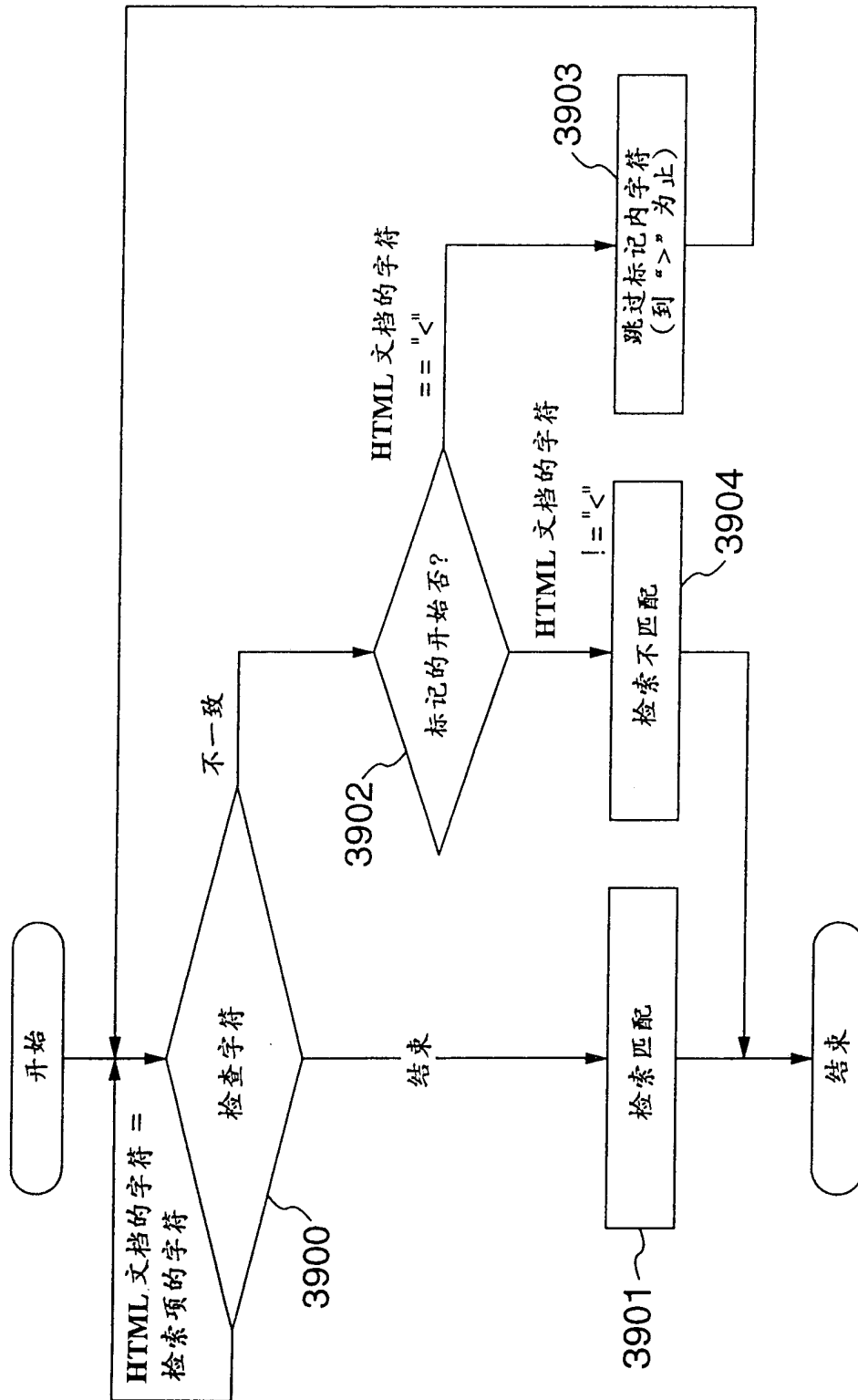


图 40

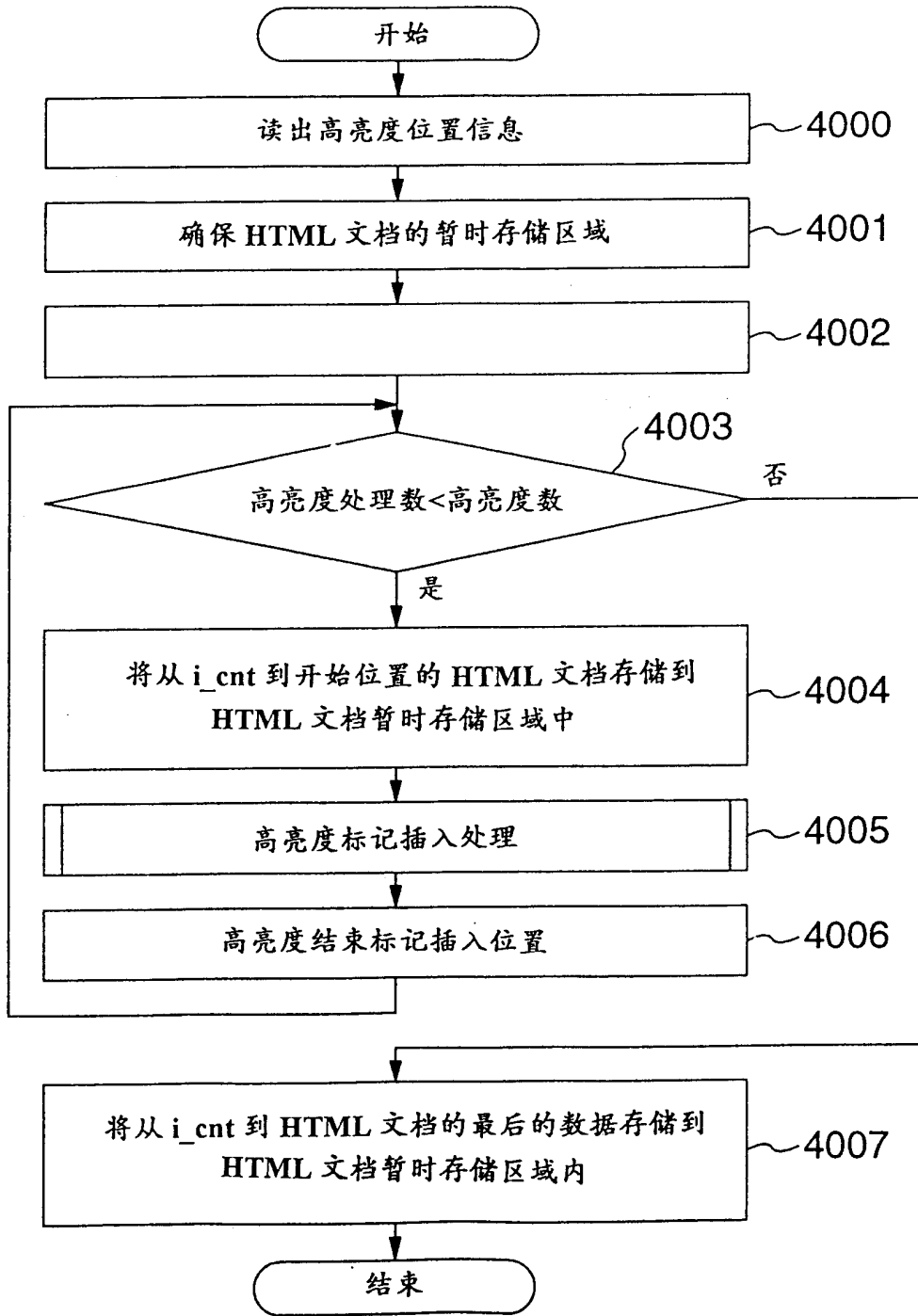




图 41

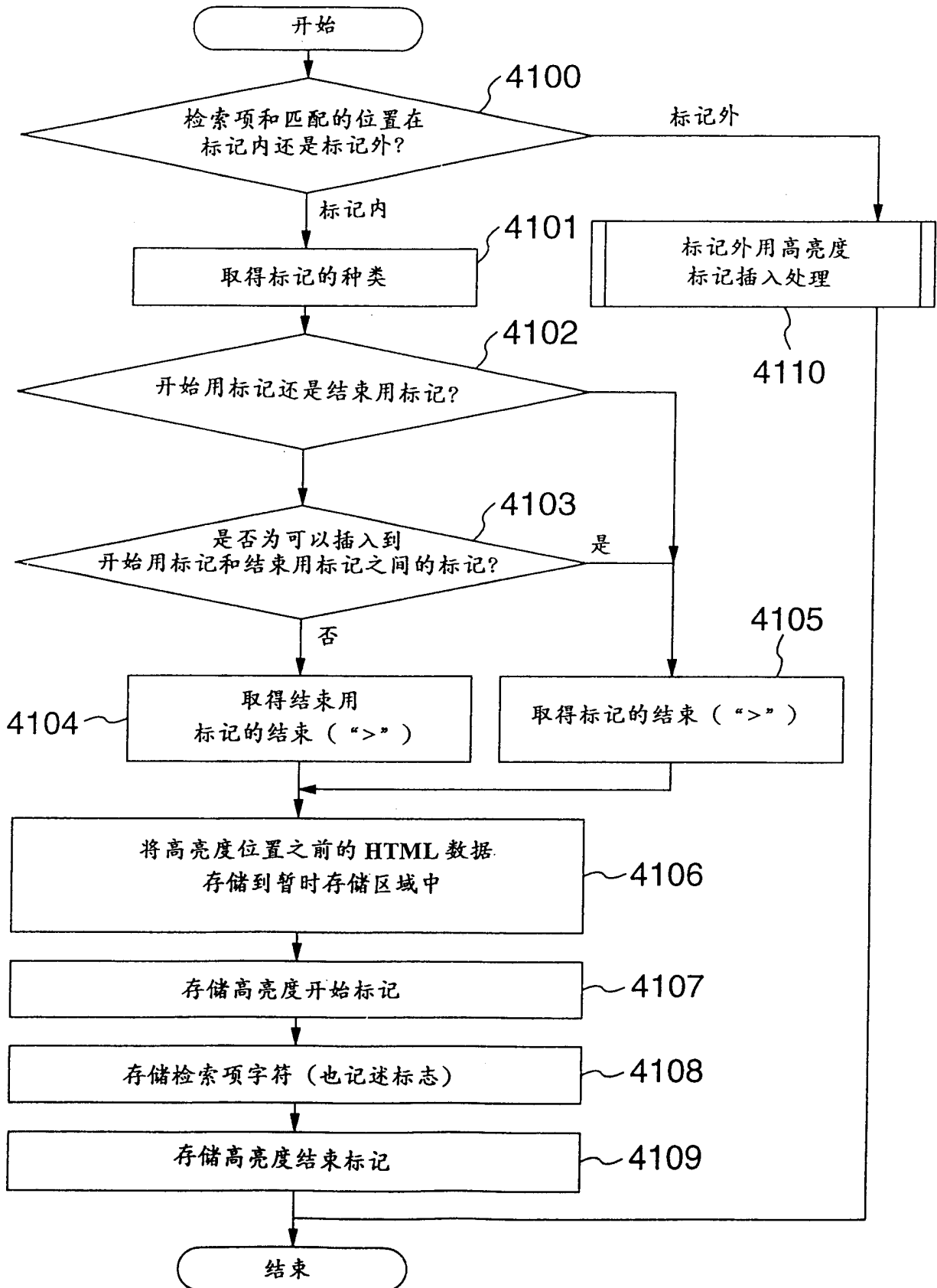


图 42

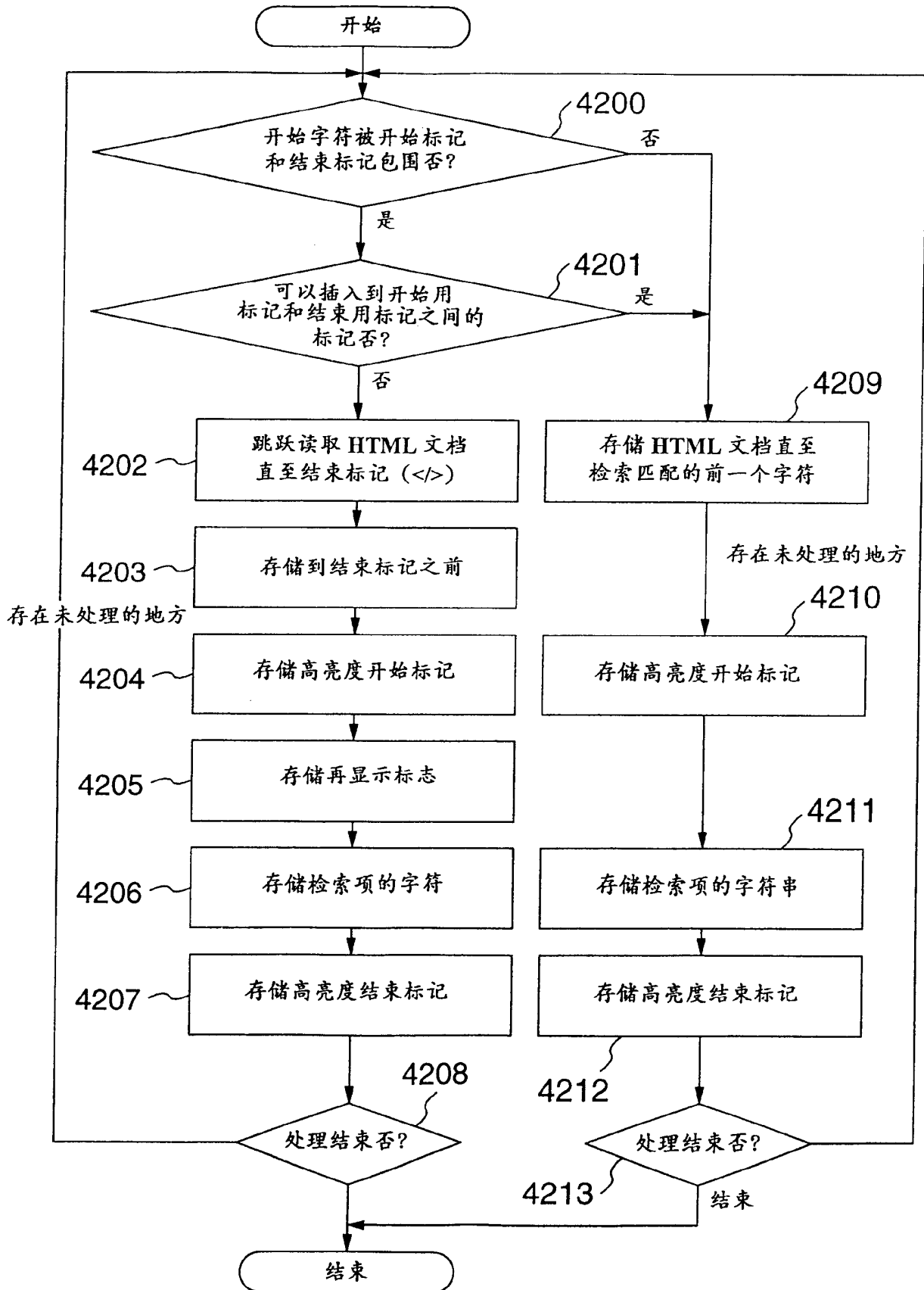


图 43

<文档.标题>: “结构化文档” {兰色}  
 <文档.正文>{波浪下划线}: “结构化文档”  
 OR C<10 {“标记”}, (“插入”))

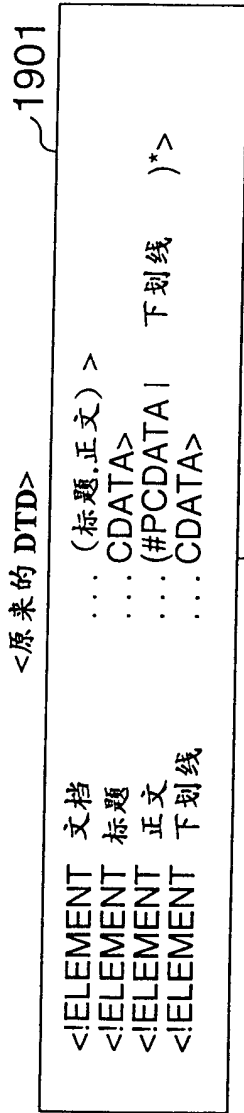
图 44

<匹配范围信息>

结构 ID	高亮度方法	匹配条件	匹配范围	
			开始位置	长度
2	<u>兰色</u>	项	0	10
9	—	频度	0	10
9	<u>波浪下划线</u>	结构	0	50
9	—	项	30	4
9	—	项	36	4
9	<u>反相</u>	距离	30	10

图 45

<实施例 6 的向高亮度显示用 DTD 的变换>



附加显示文档用的高亮度信息

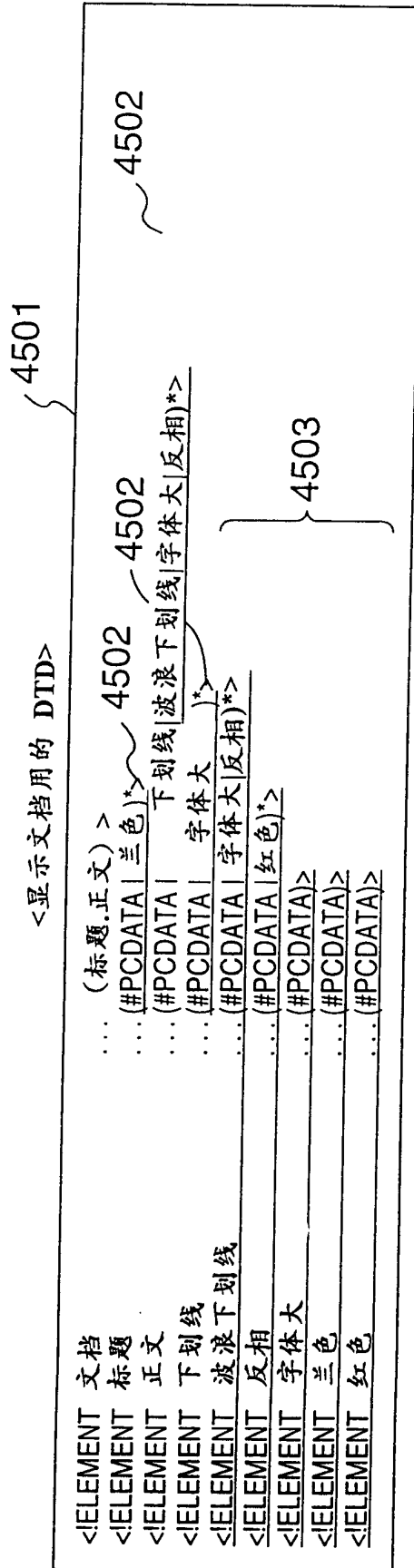


图 46

```

<文档>
<标题><<三色>结构化文档</三色></标题>
<正文>
<下划线><<波浪下划线><字体大>结构化</波浪下划线></下划线><波浪下划线><字体大>文档</波浪下划线><反相><红色>标记</红色>采用<红色>插入</红色></反相>的文档……</波浪下划线>
<正文>
<文档>

```

图 47

<高亮度显示的例子>

```

所谓结构化文档 就是在文本中 插入标记 的文档……

```