



(12) 发明专利申请

(10) 申请公布号 CN 102737012 A

(43) 申请公布日 2012. 10. 17

(21) 申请号 201110084821. 4

(22) 申请日 2011. 04. 06

(71) 申请人 鸿富锦精密工业(深圳)有限公司

地址 518109 广东省深圳市宝安区龙华镇油
松第十工业区东环二路 2 号

申请人 鸿海精密工业股份有限公司

(72) 发明人 李忠一 林海洪 谢德意 陶帅军
易志强 罗安胜 江威

(51) Int. Cl.

G06F 17/22 (2006. 01)

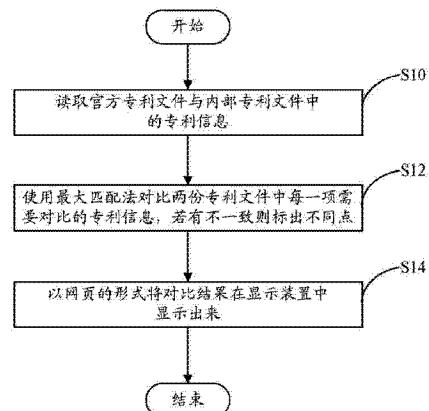
权利要求书 3 页 说明书 5 页 附图 3 页

(54) 发明名称

文本信息对比方法及系统

(57) 摘要

一种文本信息对比方法，该方法包括步骤：读取要对比的两份文本文件中的文本信息；使用最大匹配法对比两份文本文件中每一项需要对比的文本信息，若有不一致则标出不同点；将对比结果在显示装置中显示出来。本发明还提供一种文本信息对比系统。本发明可以对比文本信息并直观地标识出信息错误点。



1. 一种文本信息对比方法,其特征在于,该方法包括:

读取步骤:读取要对比的两份文本文件中的文本信息;

对比步骤:使用最大匹配法对比两份文本文件中每一项需要对比的文本信息,若有不一致则标出不同点;

显示步骤:将对比结果在显示装置中显示出来。

2. 如权利要求1所述的文本信息对比方法,其特征在于,所述对比步骤具体包括:

设置步骤:提取第一份文本文件中的要对比的一项文本信息,设为字符串A,提取第二份文件中相应的文本信息,设为字符串B,另外分别设字符串C及字符串D,均为空值;

判断步骤:判断所述字符串A及字符串B长度是否均大于0,若两字符串长度均大于0,则执行第一匹配步骤,若至少有一个字符串长度为0,则执行标识步骤;

第一匹配步骤:将字符串A中首字符与字符串B进行匹配,若该首字符在字符串B中出现,则继续将首字符和第二字符组成的串与字符串B进行匹配,依此类推,直到无法匹配为止,得到字符串A对字符串B的最大匹配长度和字符串B中的开始匹配位置,若该首字符在字符串B中未出现,开始匹配位置小于0,则匹配失败,执行第二匹配步骤,若该开始匹配位置不小于0,则将此开始匹配位置之前的字符串设置成不同点,执行截取步骤;

第二匹配步骤:继续将字符串A中第二字符与字符串B进行匹配,若该第二字符在字符串B中出现,则继续将第二字符和第三字符组成的串与字符串B进行匹配,若该第二字符在字符串B中未出现,则继续将第三字符与字符串B进行匹配,依此类推,直到无法匹配为止,得到字符串A对字符串B的最大匹配长度及两个字符串中的开始匹配位置,若字符串A中所有字符在字符串B中均未出现,两个字符串中的开始匹配位置均小于0,则匹配失败,执行标识步骤,若有一个字符串中的开始匹配位置不小于0,则将两字符串的开始匹配位置之前的字符串设置成不同点,执行截取步骤;

截取步骤:根据最大匹配长度、开始匹配位置及已经设置的不同点,分别截取新的字符串A、B、C、D,截取之后返回所述判断步骤;

标识步骤:若字符串A长度大于0,则将字符串A中的剩余字符设置成不同点,加入字符串C的字符后面,并清空字符串A,若字符串B长度大于0,则将字符串B中的剩余字符设置为不同点,加入字符串D的字符后面,并清空字符串B,若字符串A与B长度均等于0,则结束对比。

3. 如权利要求2所述的文本信息对比方法,其特征在于,所述截取步骤具体包括:

截取新的字符串A为原来的字符串A已经匹配的字符后面的剩余部分;

新的字符串B为原来的字符串B已经匹配的字符后面的剩余部分;

新的字符串C为原来的字符串C后面加上原来的字符串A中已经匹配的字符部分,已经设置的不同点用不同的字体或颜色标出;

新的字符串D为原来的字符串D后面加上原来的字符串B中已经匹配的字符部分,已经设置的不同点用不同的字体或颜色标出。

4. 如权利要求2所述的文本信息对比方法,其特征在于,所述对比结果为完成对比步骤后得到的字符串C与字符串D。

5. 如权利要求1所述的文本信息对比方法,其特征在于,所述显示步骤中以网页的形式在显示装置中显示对比结果。

6. 一种文本信息对比系统,其特征在于,该系统包括:

读取模块,用于读取要对比的两份文本文件中的文本信息;

对比模块,用于使用最大匹配法对比两份文本文件中每一项需要对比的文本信息,若有不一致则标出不同点;

显示模块,用于将对比结果在显示装置中显示出来。

7. 如权利要求 6 所述的文本信息对比系统,其特征在于,所述对比模块的对比过程具体包括:

设置步骤:提取第一份文本文件中的要对比的一项文本信息,设为字符串 A,提取第二份文件中相应的文本信息,设为字符串 B,另外分别设字符串 C 及字符串 D,均为空值;

判断步骤:判断所述字符串 A 及字符串 B 长度是否均大于 0,若两字符串长度均大于 0,则执行第一匹配步骤,若至少有一个字符串长度为 0,则执行标识步骤;

第一匹配步骤:将字符串 A 中首字符与字符串 B 进行匹配,若该首字符在字符串 B 中出现,则继续将首字符和第二字符组成的串与字符串 B 进行匹配,依此类推,直到无法匹配为止,得到字符串 A 对字符串 B 的最大匹配长度和字符串 B 中的开始匹配位置,若该首字符在字符串 B 中未出现,开始匹配位置小于 0,则匹配失败,执行第二匹配步骤,若该开始匹配位置不小于 0,则将此开始匹配位置之前的字符串设置成不同点,执行截取步骤;

第二匹配步骤:继续将字符串 A 中第二字符与字符串 B 进行匹配,若该第二字符在字符串 B 中出现,则继续将第二字符和第三字符组成的串与字符串 B 进行匹配,若该第二字符在字符串 B 中未出现,则继续将第三字符与字符串 B 进行匹配,依此类推,直到无法匹配为止,得到字符串 A 对字符串 B 的最大匹配长度及两个字符串中的开始匹配位置,若字符串 A 中所有字符在字符串 B 中均未出现,两个字符串中的开始匹配位置均小于 0,则匹配失败,执行标识步骤,若有一个字符串中的开始匹配位置不小于 0,则将两字符串的开始匹配位置之前的字符串设置成不同点,执行截取步骤;

截取步骤:根据最大匹配长度、开始匹配位置及已经设置的不同点,分别截取新的字符串 A、B、C、D,截取之后返回所述判断步骤;

标识步骤:若字符串 A 长度大于 0,则将字符串 A 中的剩余字符设置成不同点,加入字符串 C 的字符后面,并清空字符串 A,若字符串 B 长度大于 0,则将字符串 B 中的剩余字符设置为不同点,加入字符串 D 的字符后面,并清空字符串 B,若字符串 A 与 B 长度均等于 0,则结束对比。

8. 如权利要求 7 所述的文本信息对比系统,其特征在于,所述截取步骤具体包括:

截取新的字符串 A 为原来的字符串 A 已经匹配的字符后面的剩余部分;

新的字符串 B 为原来的字符串 B 已经匹配的字符后面的剩余部分;

新的字符串 C 为原来的字符串 C 后面加上原来的字符串 A 中已经匹配的字符部分,已经设置的不同点用不同的字体或颜色标出;

新的字符串 D 为原来的字符串 D 后面加上原来的字符串 B 中已经匹配的字符部分,已经设置的不同点用不同的字体或颜色标出。

9. 如权利要求 7 所述的文本信息对比系统,其特征在于,所述对比结果为完成对比过程后得到的字符串 C 与字符串 D。

10. 如权利要求 6 所述的文本信息对比系统,其特征在于,所述显示模块以网页的形式

在显示装置中显示对比结果。

文本信息对比方法及系统

技术领域

[0001] 本发明涉及一种文本信息对比方法及系统。

背景技术

[0002] 现有的文本信息对比方式,虽然可以对比出信息的不同,却无法直观的显示出来,特别是当信息量较大的时候,给用户带来了很大不便,而且还会耗费多余的时间去检查错误点。

发明内容

[0003] 鉴于以上内容,有必要提供一种文本信息对比方法,可以对比文本信息并直观地标识出信息错误点。

[0004] 鉴于以上内容,还有必要提供一种文本信息对比系统,可以对比文本信息并直观地标识出信息错误点。

[0005] 所述文本信息对比方法包括:读取步骤:读取要对比的两份文本文件中的文本信息;对比步骤:使用最大匹配法对比两份文本文件中每一项需要对比的文本信息,若有不一致则标出不同点;显示步骤:将对比结果在显示装置中显示出来。

[0006] 所述文本信息对比系统包括:读取模块,用于读取要对比的两份文本文件中的文本信息;对比模块,用于使用最大匹配法对比两份文本文件中每一项需要对比的文本信息,若有不一致则标出不同点;显示模块,用于将对比结果在显示装置中显示出来。

[0007] 相较于现有技术,所述的文本信息对比方法及系统,能够使用最大匹配法对比文本信息,并直观地标识出信息错误点,使用户第一时间发现错误的具体所在。

附图说明

[0008] 图1是本发明文本信息对比系统较佳实施例的架构图。

[0009] 图2是本发明文本信息对比系统较佳实施例的功能模块图。

[0010] 图3是本发明某实施例的对比结果网页示意图。

[0011] 图4是本发明文本信息对比方法较佳实施例的流程图。

[0012] 图5是图4中步骤S12的具体流程图。

[0013] 主要元件符号说明

对比服务器	1
FTP服务器	2
内部服务器	3
显示装置	4
文本信息对比系统	10
数据库	20
读取模块	100
对比模块	200
显示模块	300

如下具体实施方式将结合上述附图进一步说明本发明。

具体实施方式

[0014] 如图 1 所示,是本发明文本信息对比系统较佳实施例的架构图。本实施例以官方专利文件和企业内部专利文件的专利信息对比为例进行说明。所述文本信息对比系统 10 运行于对比服务器 1 中,所述对比服务器 1 与 FTP 服务器 2、内部服务器 3 进行数据通信,并连接于显示装置 4。所述对比服务器 1 中还包括数据库 20。

[0015] 所述对比服务器 1 用于对专利局官方来文中的专利文件(以下简称为官方专利文件)及企业内部存储的同一件专利文件(以下简称为内部专利文件)中需要进行对比的每一项专利信息依次进行对比,若有不一致则标出不同点,在所述显示装置 4 中以网页形式显示对比结果,以供用户查看。通过该对比结果,用户可以方便地找出官方专利文件中的专利信息出现的错误,及时进行处理。

[0016] 所述 FTP 服务器 2 用于下载所述官方专利文件。

[0017] 所述内部服务器 3 用于提供所述内部专利文件。

[0018] 所述数据库 20 用于存储对比过程中所使用的字符串等相关数据。

[0019] 如图 2 所示,是本发明文本信息对比系统较佳实施例的功能模块图。

[0020] 所述文本信息对比系统 10 包括读取模块 100、对比模块 200 及显示模块 300。

[0021] 所述读取模块 100 用于读取所述官方专利文件与内部专利文件中的专利信息。所述专利文件包括但不限于 Word、PDF、XML 等格式。

[0022] 所述对比模块 200 用于使用最大匹配法对比两份专利文件中每一项需要对比的专利信息,若有不一致则标出不同点。所述最大匹配法的具体对比过程包括:

设置步骤:所述对比模块 200 提取所述官方专利文件中的某项专利信息(如发明人信息),设为字符串 A;提取所述内部专利文件中相应的专利信息,设为字符串 B;另外分别设字符串 C 及字符串 D,均为空值。

[0023] 判断步骤:所述对比模块 200 判断所述字符串 A 及字符串 B 长度是否均大于 0。当两字符串长度均大于 0 时,执行第一匹配步骤;当至少有一个字符串长度为 0 时,执行标识步骤。

[0024] 第一匹配步骤:所述对比模块 200 将字符串 A 中首字符与字符串 B 进行匹配,若该首字符在字符串 B 中出现,则继续将首字符和第二字符组成的串与字符串 B 进行匹配,依此类推,直到无法匹配为止,得到字符串 A 对字符串 B 的最大匹配长度和字符串 B 中的开始匹配位置。若该首字符在字符串 B 中未出现,开始匹配位置小于 0,则匹配失败,执行第二匹配步骤。若该开始匹配位置不小于 0,则将此开始匹配位置之前的字符串设置成不同点(用不同的字体或颜色标出),执行截取步骤。所述开始匹配位置为字符串 B 中第一次出现的与字符串 A 中首字母相同的字符所在位置。在本实施例中,将字符串中第一个字符所在位置设为 0,第二个字符所在位置设为 1,依此类推。

[0025] 第二匹配步骤:所述对比模块 200 继续将字符串 A 中第二字符与字符串 B 进行匹配,若该第二字符在字符串 B 中出现,则继续将第二字符和第三字符组成的串与字符串 B 进行匹配;若该第二字符在字符串 B 中未出现,则继续将第三字符与字符串 B 进行匹配。依此类推,直到无法匹配为止,得到字符串 A 对字符串 B 的最大匹配长度及两个字符串中的开始

匹配位置。若字符串 A 中所有字符在字符串 B 中均未出现,两个字符串中的开始匹配位置均小于 0,则匹配失败,执行标识步骤。若有一个字符串中的开始匹配位置不小于 0,则将两字符串的开始匹配位置之前的字符串设置成不同点,执行截取步骤。字符串 A 中的开始匹配位置为字符串 A 中可以与字符串 B 进行匹配的第一个字符所在位置。字符串 B 中的开始匹配位置为字符串 B 中可以与字符串 A 进行匹配的第一个字符所在位置。

[0026] 截取步骤:所述对比模块 200 根据最大匹配长度、开始匹配位置及已经设置的不同点,分别截取新的字符串 A、B、C、D。其中,新的字符串 A 为原来的字符串 A 已经匹配的字符后面的剩余部分;新的字符串 B 为原来的字符串 B 已经匹配的字符后面的剩余部分;新的字符串 C 为原来的字符串 C 后面加上原来的字符串 A 中已经匹配的字符部分,已经设置的不同点用不同的字体或颜色标出;新的字符串 D 为原来的字符串 D 后面加上原来的字符串 B 中已经匹配的字符部分,已经设置的不同点用不同的字体或颜色标出。截取之后返回所述判断步骤。

[0027] 标识步骤:若字符串 A 长度大于 0,则将字符串 A 中的剩余字符设置成不同点,加入字符串 C 的字符后面,并清空字符串 A;若字符串 B 长度大于 0,则将字符串 B 中的剩余字符设置为不同点,加入字符串 D 的字符后面,并清空字符串 B;若字符串 A 与 B 长度均等于 0,则结束对比。

[0028] 下面以字符串“Lung-sheng Tai”与“sLTJng-sheng Ta”的对比过程为例进行具体说明:

(1) 首先设置字符串 A :Lung-sheng Tai

字符串 B :sLTJng-sheng Ta

字符串 C :空值

字符串 D :空值

(2) 判断得到字符串 A 及字符串 B 长度均大于 0,执行第一匹配步骤。

[0029] (3) 字符串 A 中首字符“L”在字符串 B 中出现,继续将首字符和第二字符“Lu”与字符串 B 进行匹配,在字符串 B 中未出现,匹配结束。得到字符串 A 对字符串 B 的最大匹配长度为 1,开始匹配位置为 1。开始匹配位置为 1 大于 0,将此位置之前的字符串“s”设置成不同点(此处用粗斜体、18 号字体标出)。

[0030] (4) 截取新的字符串 A :ung-sheng Tai

字符串 B :TJng-sheng Ta

字符串 C :L

字符串 D :sL

(5) 再次判断得到字符串 A 及字符串 B 长度均大于 0,执行第一匹配步骤。

[0031] (6) 字符串 A 中首字符“u”在字符串 B 未中出现,得到开始匹配位置小于 0,匹配失败,执行第二匹配步骤。

[0032] (7) 字符串 A 中首字符“u”在字符串 B 中未出现,继续将第二字符“n”与字符串 B 进行匹配,在字符串 B 中出现,可以匹配,最终得到字符串 A 对字符串 B 的最大匹配长度为 11,字符串 A 中的开始匹配位置为 1,将此位置之前的字符串“u”设置成不同点;字符串 B 中的开始匹配位置为 2,将此位置之前的字符串“TJ”设置成不同点。

[0033] (8) 截取新的字符串 A :i

字符串 B :空值

字符串 C :Lung-sheng Ta

字符串 D :sLTJng-sheng Ta

(9) 再次判断得到字符串 A 长度大于 0, 字符串 B 长度等于 0, 执行标识步骤。

[0034] (10) 将字符串 A 中的剩余字符“i”设置成不同点, 加入字符串 C 的字符后面, 并清空字符串 A。

[0035] 得到新的字符串 A :空值

字符串 B :空值

字符串 C :Lung-sheng Tai

字符串 D :sLTJng-sheng Ta

至此对字符串“Lung-sheng Tai”与“sLTJng-sheng Ta”的对比过程结束。

[0036] 所述对比模块 200 采用上述最大匹配法依次对所述官方专利文件及内部专利文件中每一项需要对比的专利信息进行对比, 得到每一项专利信息的对比结果。所述对比结果为完成对比过程后得到的字符串 C 与字符串 D。

[0037] 所述显示模块 300 用于以网页的形式将对比结果在所述显示装置 4 中显示出来, 以供用户查看。(参阅图 3 所示)

如图 3 所示, 是本发明某实施例的对比结果网页示意图。在对内部卷号为 2004A-7012 的专利文件进行官方专利文件和内部专利文件中申请号、申请日、第一发明人这三项专利信息的对比之后, 得到标出不同点的对比结果, 在网页中显示出来, 供用户查看。

[0038] 如图 4 所示, 是本发明文本信息对比方法较佳实施例的流程图。

[0039] 步骤 S10, 所述读取模块 100 读取所述官方专利文件与内部专利文件中的专利信息。

[0040] 步骤 S12, 所述对比模块 200 使用最大匹配法对比两份专利文件中每一项需要对比的专利信息, 若有不一致则标出不同点。(参阅图 5 中的描述)

步骤 S14, 所述显示模块 300 以网页的形式将对比结果在所述显示装置 4 中显示出来, 以供用户查看。

[0041] 如图 5 所示, 是图 4 中步骤 S12 的具体流程图。

[0042] 步骤 S200, 所述对比模块 200 提取所述官方专利文件中的某项专利信息, 设为字符串 A; 提取所述内部专利文件中相应的专利信息, 设为字符串 B; 另外分别设字符串 C 及字符串 D, 均为空值。

[0043] 步骤 S202, 所述对比模块 200 判断所述字符串 A 及字符串 B 长度是否均大于 0。若两字符串长度均大于 0, 则执行步骤 S204; 若至少有一个字符串长度为 0, 则执行步骤 S218。

[0044] 步骤 S204, 所述对比模块 200 将字符串 A 中首字符与字符串 B 进行匹配, 若该首字符在字符串 B 中出现, 则继续将首字符和第二字符组成的串与字符串 B 进行匹配, 依此类推, 直到无法匹配为止, 得到字符串 A 对字符串 B 的最大匹配长度和字符串 B 中的开始匹配位置。

[0045] 步骤 S206, 所述对比模块 200 判断所述开始匹配位置是否小于 0。若该首字符在字符串 B 中未出现, 开始匹配位置小于 0, 则匹配失败, 执行步骤 S210。若该开始匹配位置不小于 0, 则执行步骤 S208。

[0046] 步骤 S208,所述对比模块 200 将此开始匹配位置之前的字符串设置成不同点,执行步骤 S216。

[0047] 步骤 S210,所述对比模块 200 继续将字符串 A 中第二字符与字符串 B 进行匹配,若该第二字符在字符串 B 中出现,则继续将第二字符和第三字符组成的串与字符串 B 进行匹配;若该第二字符在字符串 B 中未出现,则继续将第三字符与字符串 B 进行匹配。依此类推,直到无法匹配为止,得到字符串 A 对字符串 B 的最大匹配长度及两个字符串中的开始匹配位置。

[0048] 步骤 S212,所述对比模块 200 判断两个字符串中的开始匹配位置是否均小于 0。若字符串 A 中所有字符在字符串 B 中均未出现,则两个字符串中的开始匹配位置均小于 0,则匹配失败,执行步骤 S218。若有一个字符串中的开始匹配位置不小于 0,则执行步骤 S214。

[0049] 步骤 S214,所述对比模块 200 将两字符串的开始匹配位置之前的字符串设置成不同点。

[0050] 步骤 S216,所述对比模块 200 根据最大匹配长度、开始匹配位置及已经设置的不同点,分别截取新的字符串 A、B、C、D。其中,新的字符串 A 为原来的字符串 A 已经匹配的字符后面的剩余部分;新的字符串 B 为原来的字符串 B 已经匹配的字符后面的剩余部分;新的字符串 C 为原来的字符串 C 后面加上原来的字符串 A 中已经匹配的字符部分,已经设置的不同点用不同的字体或颜色标出;新的字符串 D 为原来的字符串 D 后面加上原来的字符串 B 中已经匹配的字符部分,已经设置的不同点用不同的字体或颜色标出。截取之后返回步骤 S202。

[0051] 步骤 S218,若字符串 A 长度大于 0,则将字符串 A 中的剩余字符设置成不同点,加入字符串 C 的字符后面,并清空字符串 A;若字符串 B 长度大于 0,则将字符串 B 中的剩余字符设置为不同点,加入字符串 D 的字符后面,并清空字符串 B;若字符串 A 与 B 长度均等于 0,则结束对比。所述对比结果为完成对比过程后得到的字符串 C 与字符串 D。

[0052] 可以理解,本发明并不局限于对比官方专利文件和内部专利文件中的专利信息,本领域技术人员可以很容易利用本发明所述方法及系统对比其它文本信息。

[0053] 以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

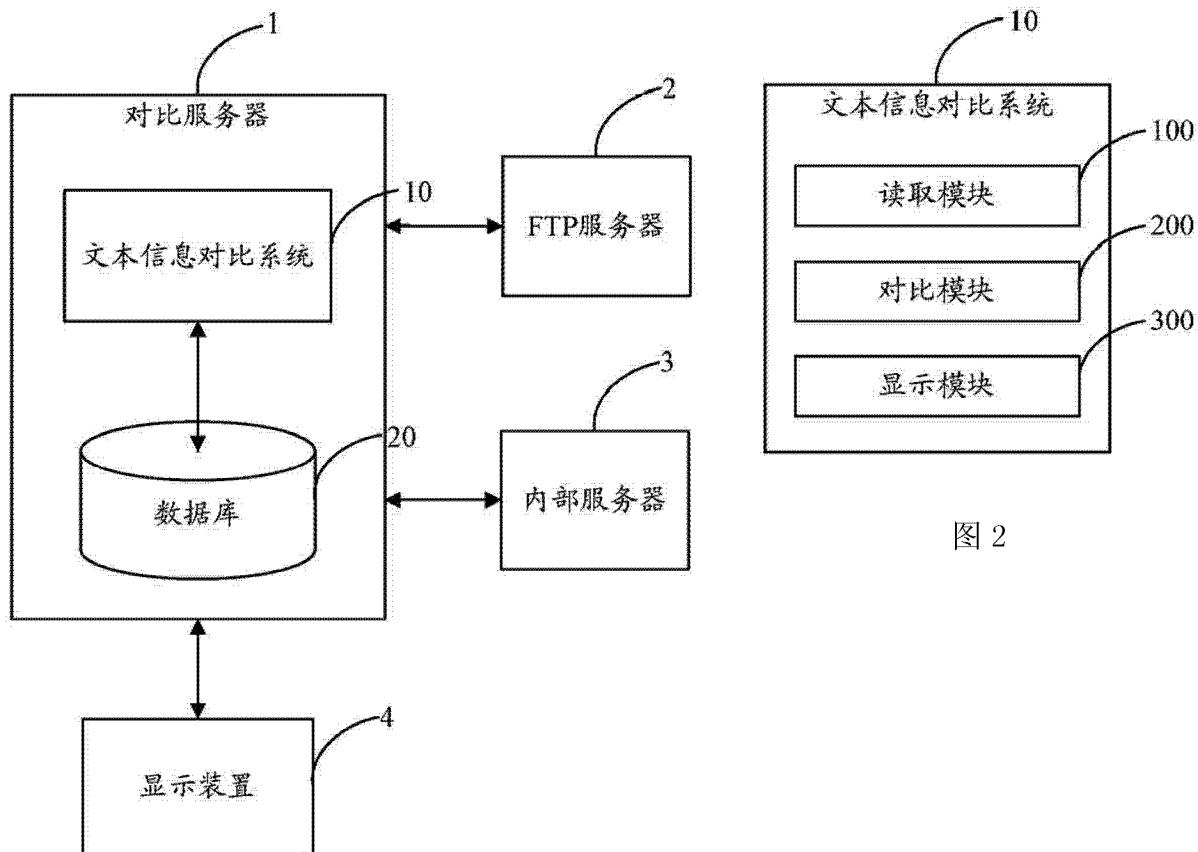


图 1

卷号: 2004A-7012
申请号 官方文: 11/765453 对比文: 11/795453
申请日 官方文: 2007-08-08 对比文: 2001-08-08
第一发明人 官方文: Liung-sheng Tai 对比文: SLTJng-sheng Ta

图 3

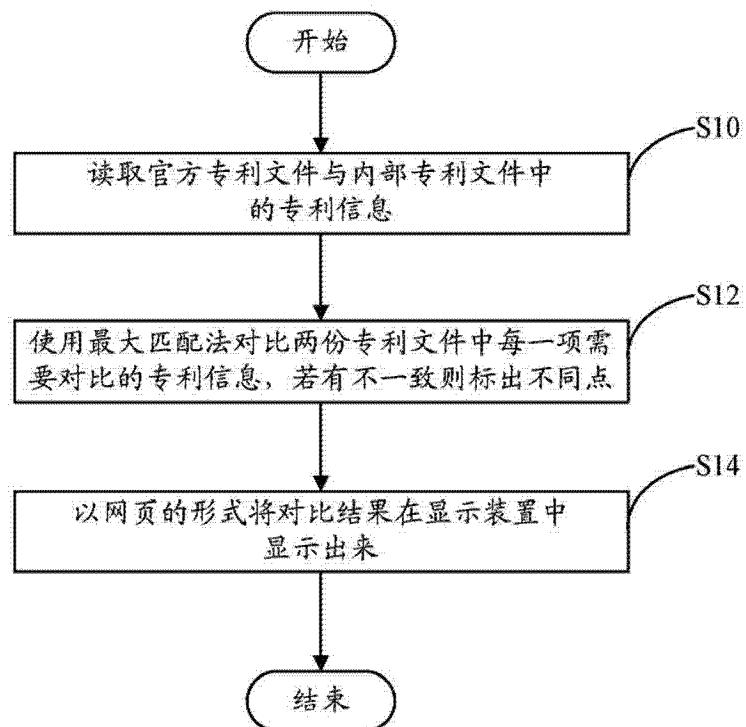


图 4

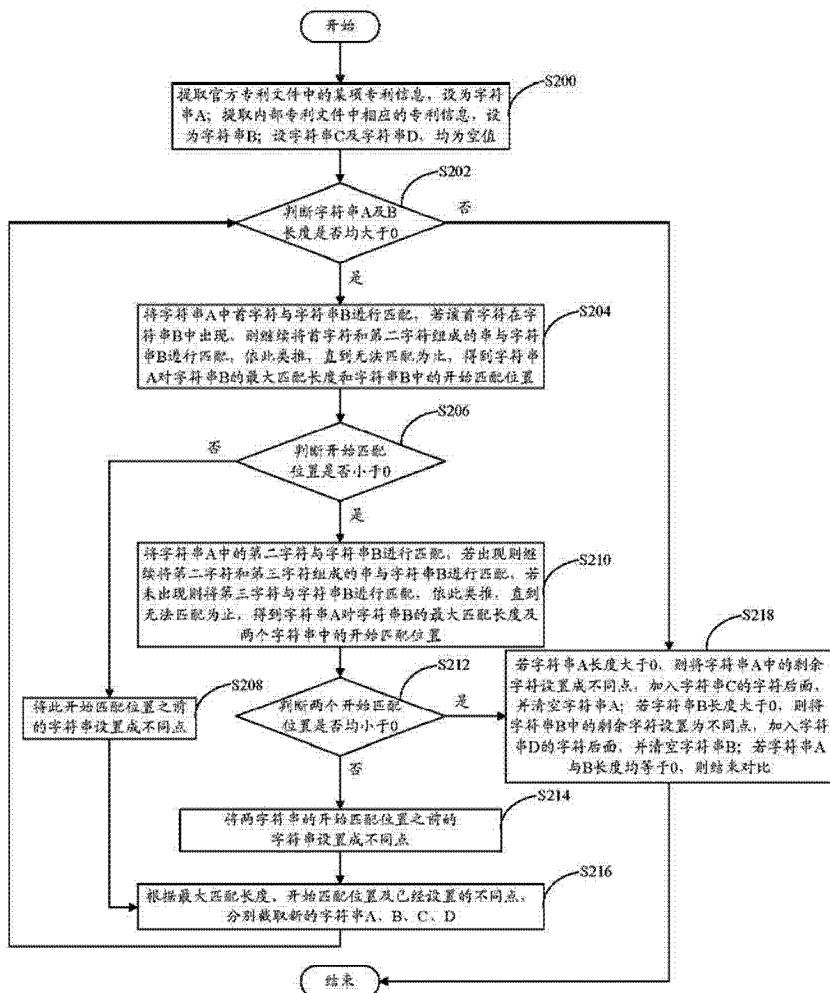


图 5