



(12) 发明专利申请

(10) 申请公布号 CN 116050377 A

(43) 申请公布日 2023. 05. 02

(21) 申请号 202211726571.4

(22) 申请日 2022.12.30

(71) 申请人 江西金格科技有限公司

地址 330000 江西省南昌市高新技术产业
开发区火炬大街579号绿悦大厦15层

(72) 发明人 余清留 金宏洲 程亮

(74) 专利代理机构 南昌恒桥知识产权代理事务
所(普通合伙) 36125

专利代理师 许明亮

(51) Int. Cl.

G06F 40/194 (2020.01)

G06F 16/538 (2019.01)

G06F 16/583 (2019.01)

权利要求书1页 说明书3页 附图1页

(54) 发明名称

一种查找和提示OFD文档间差异的方法

(57) 摘要

一种查找和提示OFD文档间差异的方法,其中:其步骤如下:S1读取、S2渲染、S3对比、S4提取、S5指示、S6标识。本专利提出使将OFD页面渲染成图像使用成熟的图像比对技术查找页面的主要差异区域,然后在这个缩小的区域内来解析OFD图元对象信息,比较这个小区内图元对象的差异详情。



1. 一种查找和提示OFD文档间差异的方法,其特征在于:其步骤如下:

S1:分别读取两份OFD文件;

S2:通过OFD渲染引擎拿取两份文档相同页面的渲染图像;

S3:计算两个页面图像之间的结构相似度指数,找到不同点的轮廓;

S4:根据轮廓的坐标值,提取对应坐标OFD图元对象的详细信息;

S5:对比两份图元对象的详情,将相同部分隐藏,分别指示出与相对图元对象不同点;

S6:在用户界面上使用图形标识出不同点,并将差异详细信息展现在不同点周围。

2. 根据权利要求1所述的一种查找和提示OFD文档间差异的方法,其特征在于:所述步骤二中,解析OFD的页面元素,根据OFD的成像模型将OFD页面绘制成图像。

3. 根据权利要求1所述的一种查找和提示OFD文档间差异的方法,其特征在于:所述步骤三利用OpenCV的图像处理能力获取对比文档两个页面的两幅图像的差异,具体步骤:

3.1 调整两幅图像大小到一致;

3.2 两幅图像相减得到图像差异;

3.3 对得到的差异矢量做绝对值处理然后改变图像位深为8位;

运用二值化阈值处理将图像处理成二值图;

3.5 对图像继续进行中值滤波和形态处理;

3.6 最后查找并绘制出差异的矩形框。

4. 根据权利要求1所述的一种查找和提示OFD文档间差异的方法,其特征在于:所述步骤四中,所述图元对象信息可包括文字对象和图形对象,所述文字对象为字体、字号、颜色;所述图形对象是线宽、颜色、线型;

根据步骤三获取的图像差异,通过差异的矩形坐标值,去获取坐标范围内的OFD图元的细节信息;

对于是差异范围内是文本对象,则可以逐字符比较字符差异;除字符差异外,还可以逐一比较文本属性如字体、字号、颜色的差异;

对于差异范围内是图形对象,则比较两个图形对象的坐标值,及图形属性;

对于差异范围内是图像对象的,则直接使用步骤三的差异信息。

一种查找和提示OFD文档间差异的方法

技术领域

[0001] 本发明属于OFD应用领域,具体涉及一种查找和提示OFD文档间差异的方法。

背景技术

[0002] OFD 文件,就是版式文件的一种,对标的是 PDF,它最大的特点在于,这是我国自主研发、自主制定的版式文件格式标准。

[0003] 相比 PDF 等其他版式文件,OFD 有一些技术上的优势。

[0004] 第一,OFD 文档内部采用可扩展标记语言 XML 来描述数据和结构,体积精简,安全开放,易于扩展。

[0005] 第二,OFD 支持国产加密算法,具有全面的安全保障体系,可防止信息被窃取,并且和数字签名技术结合,可防篡改抵赖,更加安全。

[0006] 第三,永久可读可用,可对文件长久保存,且可以精准呈现,文件的版式内容在不同场景、设备下都能保持一致性。

[0007] 第四,支持直接进行文件归档的一系列处理。

[0008] 从 OFD 诞生到今天,我国对 OFD 标准格式的推广也在逐步推进过程中。

[0009] 2016 年 10 月 14 日世界标准日,国家标准化管理委员会正式批准发布了基于自主技术的 OFD 国家标准(GB/T 33190-2016),

OFD采用GB/T 18793--2002中的XML技术描述文件数据。这些数据文件使用zip格式做容器存储。

[0010] OFD的成像模型:OFD采用二维矢量成像模型描述任何经过排版的图元对象,包括文字、图形、图像等。OFD成像模型可根据页面描述生成一个与设备无关的输出结果。

[0011] 同时,目前市场上并未发现OFD文档对比类的产品。

[0012] 如果要对比两份OFD文件一种可行方法是直接对OFD文件做哈希摘要然后比较两个哈希值是否相等,但这种方法只能提示文件是否有差异,并不比指出文件哪有差异。

[0013] 另一种可行的方法是展开OFD文件遍历OFD文件中的所有XML数据文件,但这种方法比较低效而且成功率不高,因为OFD中的图元是使用坐标值定位的因此XML中的各种对象的位置及表达方式不一定一致。

[0014] 因此,本发明提出一种针对OFD文档对比的新方法。

发明内容

[0015] 本发明的上述发明目的是通过以下技术方案得以实现的:

一种查找和提示OFD文档间差异的方法,其中:其步骤如下:

S1:分别读取两份OFD文件;

S2:通过OFD渲染引擎拿取两份文档相同页面的渲染图像;

S3:计算两个页面图像之间的结构相似度指数,找到不同点的轮廓;

S4:根据轮廓的坐标值,提取对应坐标OFD图元对象的详细信息;

S5:对比两份图元对象的详情,将相同部分隐藏,分别指示出与相对图元对象不同点;

S6:在用户界面上使用图形标识出不同点,并将差异详细信息展现在不同点周围。

[0016] 优选的,所述步骤二中,解析OFD的页面元素,根据OFD的成像模型将OFD页面绘制成图像。

[0017] 优选的,所述步骤三利用OpenCV的图像处理能力获取对比文档两个页面的两幅图像的差异,具体步骤:

3.1 调整两幅图像大小到一致;

3.2 两幅图像相减得到图像差异;

3.3 对得到的差异矢量做绝对值处理然后改变图像位深为8位;

3.4. 运用二值化阈值处理将图像处理成二值图;

3.5 对图像继续进行中值滤波和形态处理;

3.6 最后查找并绘制出差异的矩形框。

[0018] 优选的,所述步骤四中,所述图元对象信息可包括文字对象和图形对象,所述文字对象为字体、字号、颜色;所述图形对象是线宽、颜色、线型;

根据步骤三获取的图像差异,通过差异的矩形坐标值,去获取坐标范围内的OFD图元的细节信息;

对于是差异范围内是文本对象,则可以逐字符比较字符差异;除字符差异外,还可以逐一比较文本属性如字体、字号、颜色的差异;

对于差异范围内是图形对象,则比较两个图形对象的坐标值,及图形属性;

对于差异范围内是图像对象的,则直接使用步骤三的差异信息。

[0019] 综上所述,本发明包括以下至少一种有益技术效果:

本专利提出使将OFD页面渲染成图像使用成熟的图像比对技术查找到页面的主要差异区域,然后在这个缩小的区域内来解析OFD图元对象信息,比较这个小区域内图元对象的差异详情。

[0020] 本发明相对于直接比较OFD的XML数据文件,本专利的方法成功率高,能应对数据文件元素各种无序排列,同时方法精确,差异提示详细,用户体验好。

附图说明

[0021] 图1是本发明的对比差异流程示意图。

具体实施方式

[0022] 本发明的上述发明目的是通过以下技术方案得以实现的:一种查找和提示OFD文档间差异的方法,其中:其步骤如下:

S1:分别读取两份OFD文件;

S2:通过OFD渲染引擎拿取两份文档相同页面的渲染图像;

S3:计算两个页面图像之间的结构相似度指数,找到不同点的轮廓;

S4:根据轮廓的坐标值,提取对应坐标OFD图元对象的详细信息;

S5:对比两份图元对象的详情,将相同部分隐藏,分别指示出与相对图元对象不同

点；

S6:在用户界面上使用图形标识出不同点,并将差异详细信息展现在不同点周围。

[0023] 具体的,所述步骤二中,解析OFD的页面元素,根据OFD的成像模型将OFD页面绘制成图像。

[0024] 具体的,所述步骤三利用OpenCV的图像处理能力获取对比文档两个页面的两幅图像的差异,具体步骤:

- 3.1 调整两幅图像大小到一致;
- 3.2 两幅图像相减得到图像差异;
- 3.3 对得到的差异矢量做绝对值处理然后改变图像位深为8位;
- 3.4. 运用二值化阈值处理将图像处理成二值图;
- 3.5 对图像继续进行中值滤波和形态处理;
- 3.6 最后查找并绘制出差异的矩形框。

[0025] 具体的,所述步骤四中,所述图元对象信息可包括文字对象和图形对象,所述文字对象为字体、字号、颜色;所述图形对象是线宽、颜色、线型;

根据步骤三获取的图像差异,通过差异的矩形坐标值,去获取坐标范围内的OFD图元的细节信息;

对于是差异范围内是文本对象,则可以逐字符比较字符差异;除字符差异外,还可以逐一比较文本属性如字体、字号、颜色的差异;

对于差异范围内是图形对象,则比较两个图形对象的坐标值,及图形属性;

对于差异范围内是图像对象的,则直接使用步骤三的差异信息。

[0026] 最后应说明的是:以上所述仅为本发明的优选实施例而已,并不用于限制本发明,尽管参照前述实施例对本发明进行了详细的说明,对于本领域的技术人员来说,其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

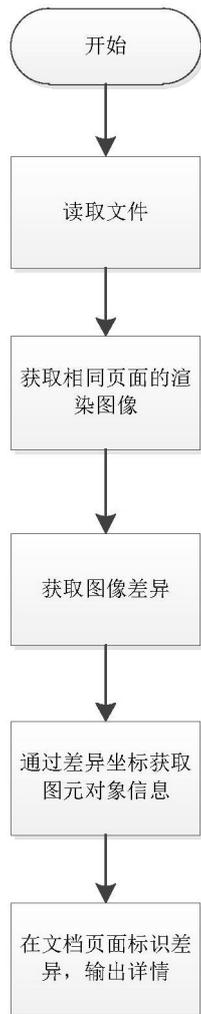


图1