# 张先生, 您好:

如上午电话沟通,经您的确认,该 case 将暂做关闭处理,以下为案例总结,请您知悉:

Case No: CAS-03900-Z6K3J1

## 问题描述:

用户反馈 OCR 在 IE 下存在不能识别的问题,经查因发现,系 CMGE 下,英文字符点阵颜色不纯造成,原 win7 正常纯色

## 问题分析:

| 系统环境       | www.k | paidu.com | OCR 页面 |     | OCR 的静态 html<br>页面 |     |
|------------|-------|-----------|--------|-----|--------------------|-----|
|            | 汉字    | 非汉字       | 汉字     | 非汉字 | 汉字                 | 非汉字 |
| Win7+IE8   | 纯色    | 杂色        | 纯色     | 纯色  | 纯色                 | 杂色  |
| Win7+IE11  | 杂色    | 杂色        | 纯色     | 杂色  | 纯色                 | 杂色  |
| Win10+IE11 | 杂色    | 杂色        | 纯色     | 杂色  | 杂色                 | 杂色  |

,在 IE8 中可以看到数字和字母都是柔化显示的,只有中文的渲染没有柔化显示。这一点在之后的某个 IE 版本中进行了改进,导致 IE11 呈现为现在这样的柔化结果。接下来又分别测试了 Chrome 浏览器和 Word,中文显示均有杂色,可以说柔化的文字渲染已经在业界达成了一种共识,IE8 是为数不多的锐化显示的案例。基于这样的一种优化,这不应该被认为是一个浏览器方面的问题,而是 OCR 识别应该更加智能化,如果只能通过某一种固定的颜色来识别具体是什么文字,其使用范围过于狭窄,因此建议客户通过升级 OCR 或寻求 OCR 技术支持看看是否有解决方案。

文字渲染引擎的工作机制是很复杂的,它有一套复杂的算法机制来决定文字的渲染结果,比如背景色是什么颜色?背景是纯色还是渐变色还是图片?文字本身是什么颜色?文字用的什么字体?有没有粗体?文字附近有没有其他文字或其它几何图形?文字本身是否覆盖于其它图层之上或者有没有其他图层覆盖在文字之上等等。文字渲染引擎通过这些因素最终决定渲染的时候哪里需要加入杂色来柔化显示效果,哪里是不需要的。举例来说黑色文字显示在白色背景之上,由于颜色反差度大,则需要柔化。相反如果在一个粉红色的背景上显示红色的文字,因为两种颜色比较接近,这时候为了让用户更容易看清楚文字,可能就不需要过度柔化文字显示,反而需要锐化文字显示,使文字的几何边界与背景更容易区分开来。因此,在IE11和CMGE都确定的情况下,不同的站点页面也会基于代码编写,CSS样式,图层间的色差甚至网络等诸多因素而呈现不同的字体渲染结果。从这个角度理解,如果该问题有解决方法,这个方法也是

更倾向于针对特定页面的解决,并不能通过固定的 IE 或 Windows 设置或代码改造来解决所有页面的字体渲染问题。

应该说有杂色和没有杂色都是一种正常的渲染结果,从这一层面来说,OCR 的文字识别应当进行相应的适配。通过文字边界、形状来正确识别文字,而不应该局限在固定某一种颜色上。

# 问题总结:

用户表示理解, 经用户的确认, 该 case 做关闭处理。

以上,如您后续有任何问题,可随时与我们联系,谢谢。

李琦 Li Qi

神州网信技术有限公司

C&M Information Technologies Co., Ltd.

服务支持电话: 4008180055

电子邮箱 Email: liqi@cmgos.com



发件人: Li Oi

**发送时间:** 2021 年 3 月 24 日 10:08 **收件人:** '张杰' < <u>hnhelp@163.com</u>>

**抄送:** tam\_sup < tam\_sup@cmgos.com >; 'win10 升级支持' < win10sup@sdc.icbc.com.cn >; Liu Wei < liuwei@cmgos.com >; Wang Dan < wangdan@cmgos.com >; CRM Case Email < casemail@cmgos.com >

**主题:** 回复: Re:回复: Re:回复: 回复: [案例号: CAS-03900-Z6K3J1] % |P2||CBC|OCR 无法在 IE 识别 % 初次响应 CMIT:0001898

# 张先生, 您好:

感谢您的回复。我十分理解您的疑问,结合我们这段时间的沟通与测试,我整理了如下表格方便您的理解。

| 系统环境      | www.baidu.com |     | OCR 内网页面(客<br>户提供的截图) |     | OCR 的静态 html<br>页面 |     |
|-----------|---------------|-----|-----------------------|-----|--------------------|-----|
|           | 汉字            | 非汉字 | 汉字                    | 非汉字 | 汉字                 | 非汉字 |
| Win7+IE8  | 纯色            | 杂色  | 纯色                    | 纯色  | 纯色                 | 杂色  |
| Win7+IE11 | 杂色            | 杂色  | 纯色                    | 杂色  | 纯色                 | 杂色  |

IE 浏览器版本针对不同的站点有不同的表现

形式,并且也没有找到对应的规律。对于字体渲染方式在 IE 浏览器的升级演变过程中,随着 W3C 标准的变化,我们已经确认是有变化的。但这种变化我没有在内外部的文档资料上找到 会引发当前问题的关联性。通过您提供的截图信息,也确实证明与外网站点的默认渲染结果是 不同的。正如之前电话中所述,IE 的文字渲染其实是调用了系统底层的绘图 API 进行的,举例来说,IE 告诉操作系统需要在屏幕的(x,y)坐标处画一个颜色为#000000 的大小为 12px 字体为 宋体的阿拉伯数字 2,而 IE 本身并不会自己去计算这个数字 2 的每个像素点应该怎么陈列。这就导致渲染结果有随机性。这中间牵涉的变量包括:

- 1, 不同操作系统和不同 IE 浏览器的区别(包括安装方式,版本差异,定制的组策略和服务内容)
- 2, 企业内网的网络环境
- 3. 安装的第三方软件的影响
- 4. 访问站点的源代码表现

我十分理解对于您来说,您希望的效果是纯色显示 OCR 页面的字体内容,而从我们已进行的测试结果来看,纯色显示其实并不是 Win10 和 IE11 的默认渲染结果,应是 OCR 页面加载后的编译结果。因此日志收集的建议是为了主要针对 OCR 页面加载后的渲染过程。即在内网环境下,使用 CMGE 系统抓取 OCR 页面的加载。通过 fiddler 的收取和本地编译,分析在加载页面的过程中该页面是如何渲染字体的。请您知悉。谢谢。

PS: fiddler 日志的分析可以对当前问题进行本地复现,因此请按照之前的步骤一步步按顺序进行,不要跳过任何一步(如 clear cache 的步骤)。

# 李琦 Li Qi

神州网信技术有限公司

C&M Information Technologies Co., Ltd.

服务支持电话: 4008180055

电子邮箱 Email: liqi@cmgos.com



发件人: 张杰 < hnhelp@163.com > 发送时间: 2021 年 3 月 24 日 7:13 收件人: Li Qi < liqi@cmgos.com >

主题: Re:回复: Re:回复: 回复: [案例号: CAS-03900-Z6K3J1 ] % |P2|ICBC|OCR 无法在 IE 识

别 % 初次响应 CMIT:0001898

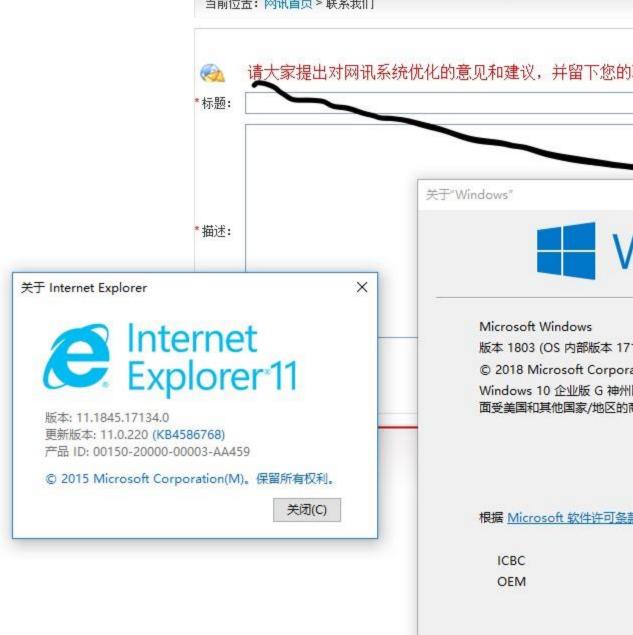
#### 李工你好!

为了方便您分析因查,我特地在内网用 WIN10 看了其他网站,您可以此为参照对象查因,谢谢!

张杰 2021.3.24



当前位置:网讯首页>联系我们



在 2021-03-23 16:50:03, "Li Qi" < liqi@cmgos.com > 写道:

张先生, 您好:

# 为进一步排查该问题,请在 CMGE 系统上按照如下步骤收取 fiddler 日志和步骤记录器内容

- 1. 从 https://www.telerik.com/download/fiddler 下载并安装 Fiddler。
- 2. 关闭所有 IE 浏览器进程,确保在任务管理器的进程列表中看不到 iexplore.exe

| Processes Performance App history Sta  | rtup User | S Details | Services  |     |           |            |
|--|-----------|-----------|-----------|-----|-----------|------------|
|  |           |           |           |     |           |            |
| Name   | PID       | Status    | User name | CPU | Memory (a | UAC virtua |
| explorer.exe   | 9016      | Running   | sheng     | 02  | 30,368 K  | Disabled   |
| operation of the second of the |           | Running   | sheng     | 00  | 10,640 K  | Enabled    |
| FileCoAuth.exe   | 9760      | Running   | sheng     | 00  | 2,908 K   | Disabled   |
| fontdrvhost.exe  | 508       | Running   | UMFD-0    | 00  | 60 K      | Disabled   |
| ■ GamingServices.exe   | 6948      | Running   | SYSTEM    | 00  | 2,964 K   | Not allowe |
| ■ GamingServicesNet.exe  | 6924      | Running   | LOCAL SE  | 00  | 24 K      | Not allowe |
| <b>☼</b> Google Crash Handler. exe   | 9868      | Running   | SYSTEM    | 00  | 272 K     | Not allowe |
| ₿ Google Crash Handler 64. exe   | 10016     | Running   | SYSTEM    | 00  | 272 K     | Not allowe |
| ■ HarmonyUserStartup.exe   | 15700     | Running   | sheng     | 00  | 252 K     | Disabled   |
| ■ HxTsr.exe  |           | Suspended | sheng     | 00  | 0 K       | Disabled   |
| ■ IGCC.exe   | 6792      | Running   | sheng     | 00  | 7,560 K   | Disabled   |
| ■ IGCCTray.exe   | 1852      | Running   | sheng     | 00  | 5,108 K   | Disabled   |
| ■ IntelCpHDCPSvc.exe   | 4492      | Running   | SYSTEM    | 00  | 24 K      | Not allowe |
| ■ IpOverUsbSvc.exe   | 4752      | Running   | SYSTEM    | 00  | 28 K      | Not allowe |
| LMIGuardianSvc.exe   | 4772      | Running   | SYSTEM    | 00  | 24 K      | Not allowe |
| ■ LockApp.exe  | 2664      | Suspended | sheng     | 00  | 0 K       | Disabled   |
| ■ Lsalso.exe   | 888       | Running   | SYSTEM    | 00  | 508 K     | Not allowe |
| ■ Isass.exe  | 896       | Running   | SYSTEM    | 00  | 8,352 K   | Not allowe |
| ■ Microsoft.Management.Services.Intune   | 20668     | Running   | SYSTEM    | 00  | 11,236 K  | Not allowe |
| ■ Microsoft.Photos.exe   | 4540      | Suspended | sheng     | 00  | 0 K       | Disabled   |
| ■ Microsoft.ServiceHub.Controller.exe  | 22520     | Running   | sheng     | 00  | 28 K      | Disabled   |
| ■ MoUsoCoreWorker.exe  | 2472      | Runnina   | SYSTEM    | 00  | 60 K      | Not allowe |
| <  |           |           |           |     |           | >          |

End task

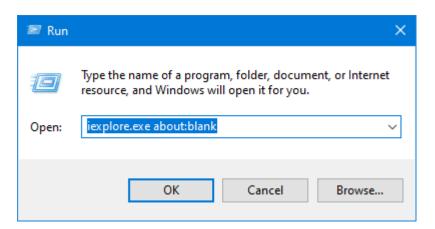
3. 打开步骤记录器(点击开始-windows 附件-步骤记录器)并开始录制。

4. 打开 Fiddler, 点击 Clear Cache 按钮清除 IE 缓存。

Fewer details



5. 然后运行 iexplore.exe about:blank 打开一个空白的 IE 窗口。



6. 在这个空白 IE 窗口中以最短的步骤打开到那个事件列表的页面。



- 7. 回到 Fiddler, 保存日志成 saz 文件。
- 8. 保存步骤记录器日志。

将 7.8 两步的收集日志按如下方法上传

# 日志上传:

您可以登陆 <a href="https://cduc.cmgos.com">https://cduc.cmgos.com</a>,通过数据上传系统上传您所收集的日志信息。如果登录不成功,点击后退即可

用户名: zhangjie

密码: zhangjie



文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)

{"status":"success", "message":"

×<\/a>Login success. Redirecting...<\/div>","location":"https:\/\/cduc.cmgos.com\/manage-files.php"}

# 注意,添加文件后点击上传文件

\_\_\_\_\_

=========

在向 CMIT 提供日志和数据前,请阅读并接受邮件下方隐私声明。

隐私声明

为向您提供本产品的相关技术支持及相关服务,您需要在同意本声明的前提下向神州网信提供为解决系统故障必要的数据和信息,包括但不限于与您相关的个人数据和隐私信息。通常情况下,我们仅需要如下数据以使我们的服务能够更好地满足您的需求:内存转储文件、注册表信息、组策略信息、事件日志、安装应用列表、驱动信息、网络日志或系统补丁等。

神州网信承诺将采取商业上可行及必要措施保护用户提供给神州网信的数据和信息,且仅在解决问题过程中使用。

神州网信保证不对外公开或向第三方提供、公开或共享您提供的数据和信息。

在以下情况下,神州网信对您的数据和信息的披露将不视为违约,具体包括:

- (1) 神州网信己获得您的明确授权;
- (2) 根据适用法律的要求,神州网信负有披露义务的;
- (3) 司法机关或行政机关基于法定程序要求神州网信提供的;
- (4) 为维护社会公共利益及神州网信合法权益,在合理范围内进行披露的。
- (5) 为了解决您的系统故障问题,神州网信可能会与神州网信的分包商披露您提供的数据和信息。在此情况下,第三方会承担与神州网信同等的隐私保护责任的,神州网信会在合理范围内对您的信息进行披露。

如果您欲提供的数据和信息中包含对您重要的保密信息以及商业秘密,在您向神州网信提供上述数据和信息前,务必对上述数据和信息进行脱敏处理,否则请不要提供该信息给神州网

信。作为一家商业软件公司,神州网信在商业可行的前提下,已为用户的数据和信息保护做了极大的努力,但是仍然不能保证现有的安全技术措施使您的数据和信息等不受任何形式的损失。您对上述情况充分知情,且不会因此追究神州网信的法律责任。

# 李琦 Li Qi

神州网信技术有限公司

C&M Information Technologies Co.,Ltd.

服务支持电话: 4008180055

电子邮箱 Email: <u>liqi@cmgos.com</u>



发件人: Li Qi

**发送时间:** 2021 年 3 月 22 日 11:11 **收件人:** '张杰' < hnhelp@163.com>

**抄送:** tam\_sup < tam\_sup@cmgos.com >; 'win10 升级支持' < win10sup@sdc.icbc.com.cn >; Liu Wei < liuwei@cmgos.com >; Wang Dan < wangdan@cmgos.com >; CRM Case Email < casemail@cmgos.com >

% 初次响应 CMIT:0001898

张先生, 您好:

收到您的邮件, 打开后无法显示如您截图的页面:



根据您的截图,OCR基于内网页面控件的渲染结果与外网的测试结果不同。

- 1. 外网为汉字内容表现不同, OCR 为非汉字内容表现不同。
- 2. 在 IE11 中, 外网为汉字显示杂色, OCR 为非汉字显示杂色。

您的需求目前我已了解,接下来我们首先针对第一点即 IE8 和 IE11 的不同渲染结果进行排查 (无论是汉字内容或非汉字内容的不同表现结果),请您知悉,谢谢。

# 李琦 Li Qi

神州网信技术有限公司

C&M Information Technologies Co.,Ltd.

服务支持电话: 4008180055

电子邮箱 Email: <u>liqi@cmgos.com</u>



**发件人:** 张杰 < <u>hnhelp@163.com</u>> **发送时间:** 2021 年 3 月 21 日 17:13

收件人: Li Qi liqi@cmgos.com>

**主题:** Re:回复: 回复: [案例号: CAS-03900-Z6K3J1 ] % |P2||CBC|OCR 无法在 IE 识别 % 初次响应 CMIT:0001898

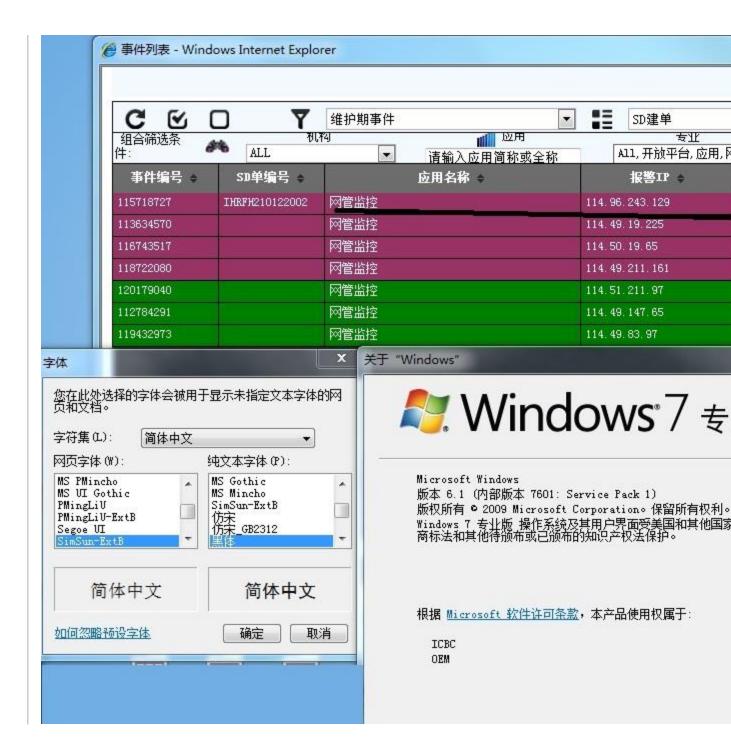
# 李工: 你好!

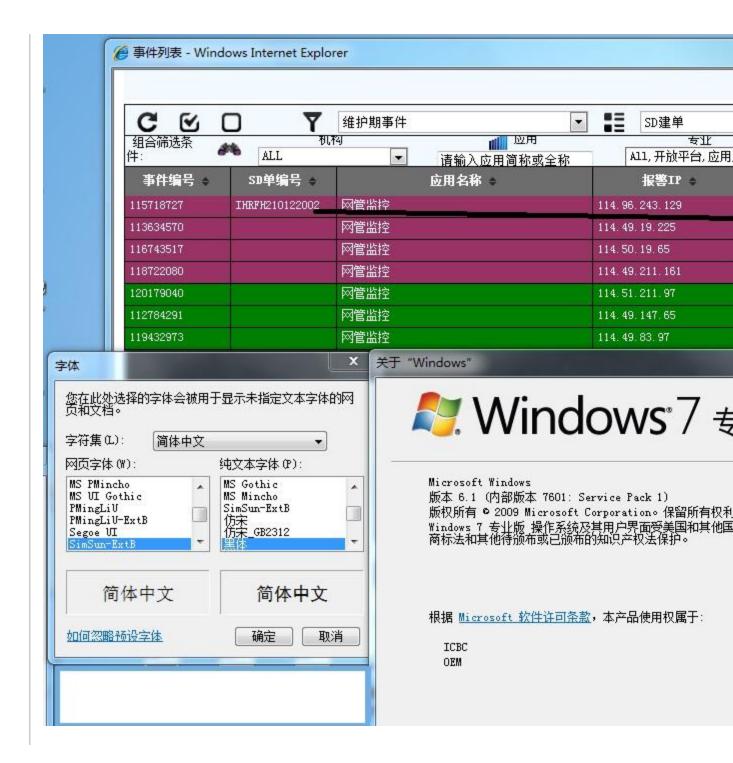
上班,如您要求,将测试后的页面发您分析,目标文件保存后只有菜单栏,无内容栏,但菜单栏的汉字也能正常纯色显示(您可以自己测试)。

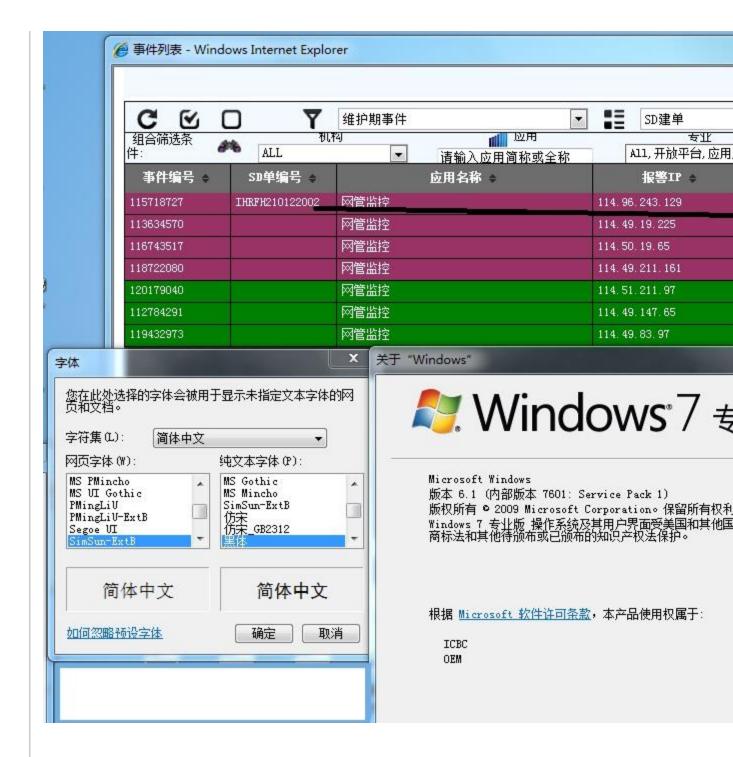
现将单位的 WIN10、Win7 的中英文测试图发你分析处理,谢谢!。

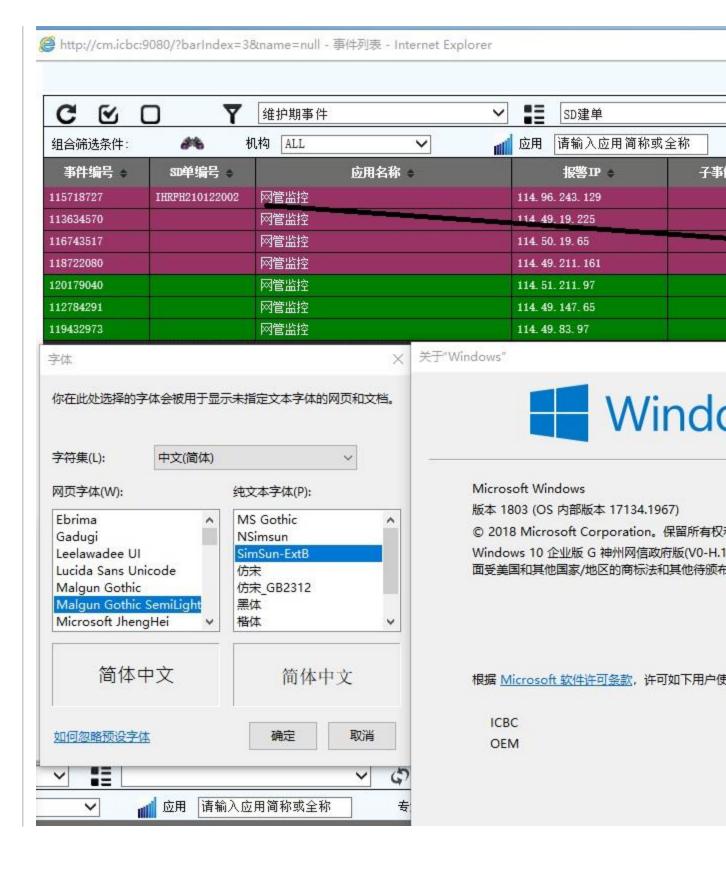
张杰

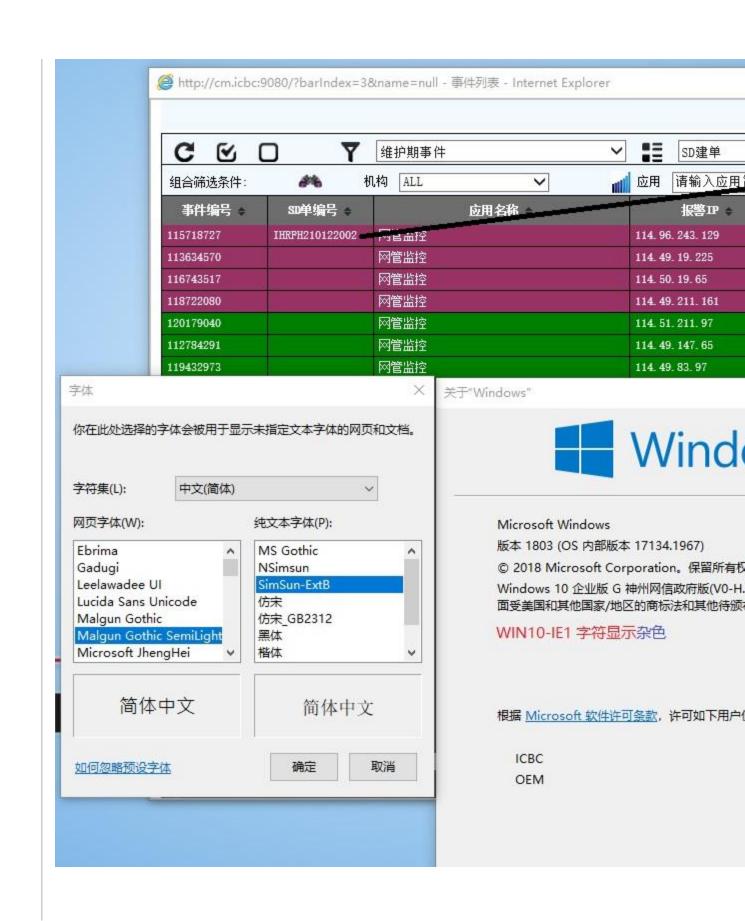
2021.3.21











在 2021-03-19 15:45:01, "Li Qi" < liqi@cmgos.com > 写道:

张先生, 您好:

好的。感谢您的回复。如可以将应用 OCR 控件的内网 HTML 页面发送给我,我将进行同样的测试,即

- a) 加载 OCR 控件后的内网 HTML 在 Win7-IE8 的汉字显示:
- b) 加载 OCR 控件后的内网 HTML 在 Win10-IE11 的汉字显示:

如无法发送给我的话,请您进行上述两个场景的测试并截图。

如上述两个测试<mark>结果不同</mark>,我将按照这组对比结果进行排查。请知晓,谢谢。

李琦 Li Qi

神州网信技术有限公司

C&M Information Technologies Co.,Ltd.

服务支持电话: 4008180055

电子邮箱 Email: <u>liqi@cmgos.com</u>



发件人: 张杰 < hnhelp@163.com > 发送时间: 2021 年 3 月 19 日 14:38 收件人: Li Qi < liqi@cmgos.com >

**抄送:** tam\_sup < tam\_sup@cmgos.com >; win10 升级支持 < win10sup@sdc.icbc.com.cn >; Liu Wei ; Wang Dan < wangdan@cmgos.com >; CRM Case Email < casemail@cmgos.com >

主题: 回复: [案例号: CAS-03900-Z6K3J1] % |P2|ICBC|OCR 无法在 IE 识别 % 初次响应 CMIT:0001898

谢谢李工支持 ,不过百度不目标网站,而且 win10 对汉字的内外网结果又不同(内网汉字全纯色),我还是申请下,考虑将内网的 HTML 发您分析吧,毕竟有针对性解决问题是需求。

张杰 2021.3.19

发自 网易邮箱大师

在 2021 年 03 月 19 日 14:25, Li Qi 写道:

张先生, 您好:

感谢您的大力支持。收到您的截图后,我在本地对此问题也进行了复现。其问题与您最初的问题描述有些不符,主要表现为 IE8 和 IE11 在非汉字部分的渲染均有杂色或灰度像素点;而汉字部分的渲染结果不一致,IE8 纯色显示,IE11 有杂色或灰度像素点。

目前的测试步骤与结果如下:

在 Win7-IE8 和 Win10-IE11 上设置相同的显示字体

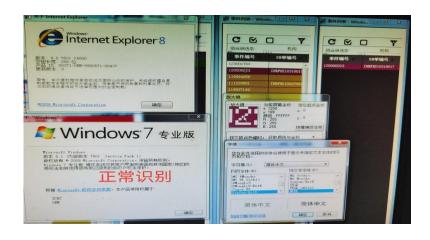


同样访问 www.baidu.com,并在首页内容中找到非汉字和汉字的内容,使用放大镜工具进行观察。可以看到非汉字部分均有杂色或灰度像素点。而汉字部分的渲染结果不同,Win7-IE8 是纯色显示;Win10-IE11 有杂色或灰度像素点。具体可参见下图:

Win7-IE8 的非汉字显示:



Win10-IE11 的非汉字显示:



# Win7-IE8 的汉字显示:



Win10-IE11 的汉字显示:



基于上述结果,接下来对此问题的处理范围定义更新如下:

# 问题定义:

在 IE11 使用"网页字体: SimSun-ExtB; 纯文体字体: 黑体"的设置下, 网页 渲染的汉字字体为杂色或有灰度像素点, 与 IE8 的表现不符。非汉字字体与 IE8 的表现一致, 均为杂色或有灰度像素点。

# 问题范围:

我们将协助您分析处理汉字字体部分的问题,并对定义的问题给予最大的技术支持。

如果能及时解决问题,或问题属于产品设计的行为,或问题涉及到三方,我们将考虑关闭案例。如果存在多个问题,则我们考虑拆分案例进行分析。

接下来,我们将开始合作解决这个问题。如果您对以上的问题范围界定有任何异议,请尽快告知。如果您有其他任何疑问。也欢迎随时与我联系。

# 目前已测试的内容为:

将 <u>www.baidu.com 加入本地 intranet</u>站点,尝试使用 IE8 的文字渲染方式进行页面渲染,问题依旧

通过 F12 修改文档模式至 IE8, 问题依旧

因此如之前电话所述,目前大概率判断为 IE11 对汉字渲染部分进行的变更导致,接下来,我将在内部进行查询相关变更内容的影响,如有更新,我会第一时间与您联系,感谢您的大力配合。

### 李琦 Li Qi

神州网信技术有限公司

C&M Information Technologies Co.,Ltd.

服务支持电话: 4008180055

电子邮箱 Email: <u>liqi@cmgos.com</u>



发件人: 张杰 < hnhelp@163.com > 发送时间: 2021 年 3 月 19 日 10:59 收件人: Li Qi < liqi@cmgos.com >

主题: 截图

# 李工你好:

现将家与单位截图与你,方便查因。

家里有 WIN7 截图,没有 WIN10 系统(本本是 WIN8)

结果如下:

| 单位 WIN10,IE1<br>1 | 单位 WIN7,IE11 | 单位 WIN7,IE8 | 家 WIN7,IE8 |
|-------------------|--------------|-------------|------------|

单位同一网址

字符杂色,汉字正常

字符杂色,汉字正常

字符正常, 汉字正常

百度网址

字符杂色,汉字正常

由于单位与家访问不同网址,个人觉得还是说明不了什么问题,但有一个共同点,汉字点阵都正常显示,能否设置调整与汉字模式一样?

张杰

2021.3.19

单位 WIN7 IE8, 能正常识别。

Internet 选项

| 字体                      |         |             |
|-------------------------|---------|-------------|
| 你在此处选择的字体会被             | 用于显示    | 示未指定文本字体的   |
|                         |         |             |
| 字符集(L): 中文(             | (简体)    |             |
| 7-19 <del>3.</del> (E). | (IBIVA) |             |
| 网页字体( <u>W</u> ):       |         | 纯文本字体(P):   |
| PMingLiU-ExtB           | ^       | MS Gothic   |
| Segoe UI                |         | NSimsun     |
| Segoe UI Light          |         | SimSun-ExtB |
| Segoe UI Semibold       |         | 仿宋          |
| Segoe UI Semilight      |         | 黑体          |
| Simsun                  |         | 楷体          |
| SimSun-ExtB             | ~       | 新宋体         |
|                         |         |             |
| A*** (-1                |         | A-6- L-1    |
| 简体中文                    |         | 简体。         |
|                         |         |             |
|                         |         |             |
| 如何忽略预设字体                |         | 确定          |
|                         |         |             |
|                         |         | 删除(         |
|                         |         | Wild Kate ( |
| 外观 ————                 |         |             |
|                         |         |             |
| 颜色(O)                   | 语言(     | (L) 字体      |
|                         |         |             |

单位 WIN7 IE11,不能正常识别。









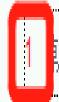




🔐 ▼ 🔝 ▼ 📑 🖶 ▼ 页面(P) ▼ 安全(S) ▼ 🦠

# Baid的百度

# 百度热榜



卵鲜宣布与马来西亚断交 热



王毅说中:

2 人民日报评暂停网贷进校园

变异新毒

3 中美对话现场:杨洁篪严正表态

多地公务

# 家 WIN7 IE8,不能正常识别。

| https://www.bai | du.com/      |         | w | <u> </u> |   | _ |
|-----------------|--------------|---------|---|----------|---|---|
| 就知道             | X 🥝 31省份新增本: | 土确诊病例1例 |   |          |   |   |
|                 |              |         |   | 9        | C |   |
|                 |              |         |   |          |   |   |
|                 |              |         |   |          |   |   |
|                 |              |         |   |          |   |   |

# 百度热榜

- 1 朝鲜宣布与马来西亚断交 热
  - 2 人民日报评暂停网贷进校园
  - 3 中美对话现场:杨洁篪严正表态

关于百度 About Baidu 百度营销 使用百度前必读 意见反馈 帮助