**目前无法解决的坑**

在使用scrapy-spalsh爬取 “贵州省人民政府网站”（<http://www.gzgov.gov.cn/xxgk/jbxxgk/fgwj/gfxwj/>）时，总是出现 504 Gateway Timeout 错误

经过排查，是由于某个文件的返回头中出现了 Content-Disposition:attachment; 这个返回头字段及其值表示该资源为附件资源。

点击查看源码，发现该网站第一个js就是附件加载



与一般的js标签不同，一般的js标签点击地址会直接把js文件内容直接显示在浏览器上。在点击该标签的地址时，浏览器会直接进行文件下载（因为它是附件）。

之所以出现该原因是因为splash的渲染机制和浏览器不同，浏览器遇到这种js标签会直接进行解析，但splash不会，因此就出现了504的错误。splash不会解析的根本原因在于，splash是一种基于QT5实现的浏览器，对网页的解析也主要依赖于

SplashQWebPage( 继承QWebPage ）这个类来作解析，出现这种情况时，它会直接抛出一个unsupportedContent的信号，而不是去请求该资源来解析它。

不仅如此，在返回头的Content-Type中也会出现不支持的解析格式，我们知道，通常返回头的Content-Type主要为 text/html application/json text/xml等等常用的值，当出现其他不被QWebPage支持的值时，同样会报504错误.

**目前尚未找到较好的解决方案，望后续深入理解了能够解决！**