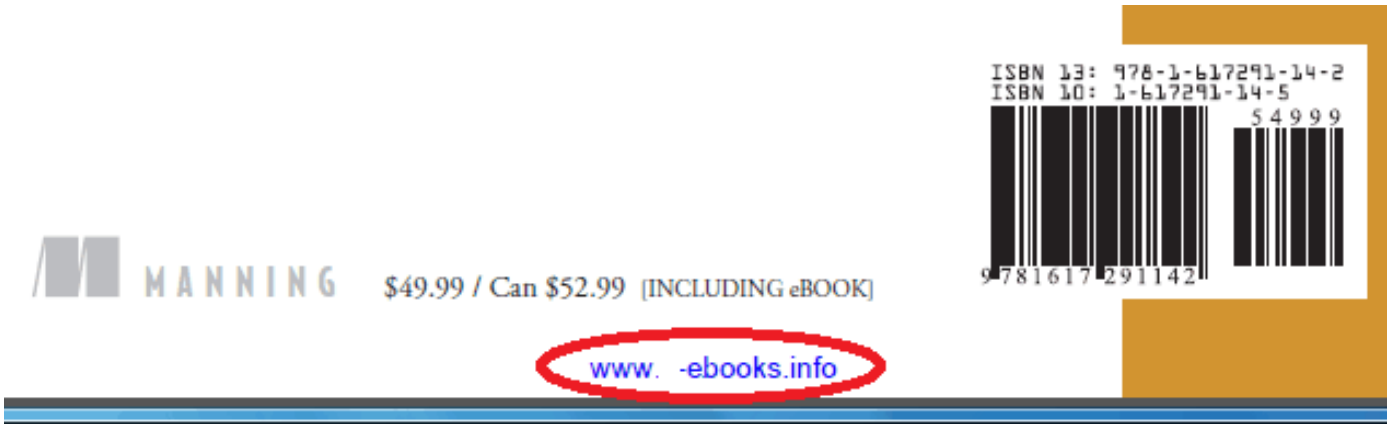
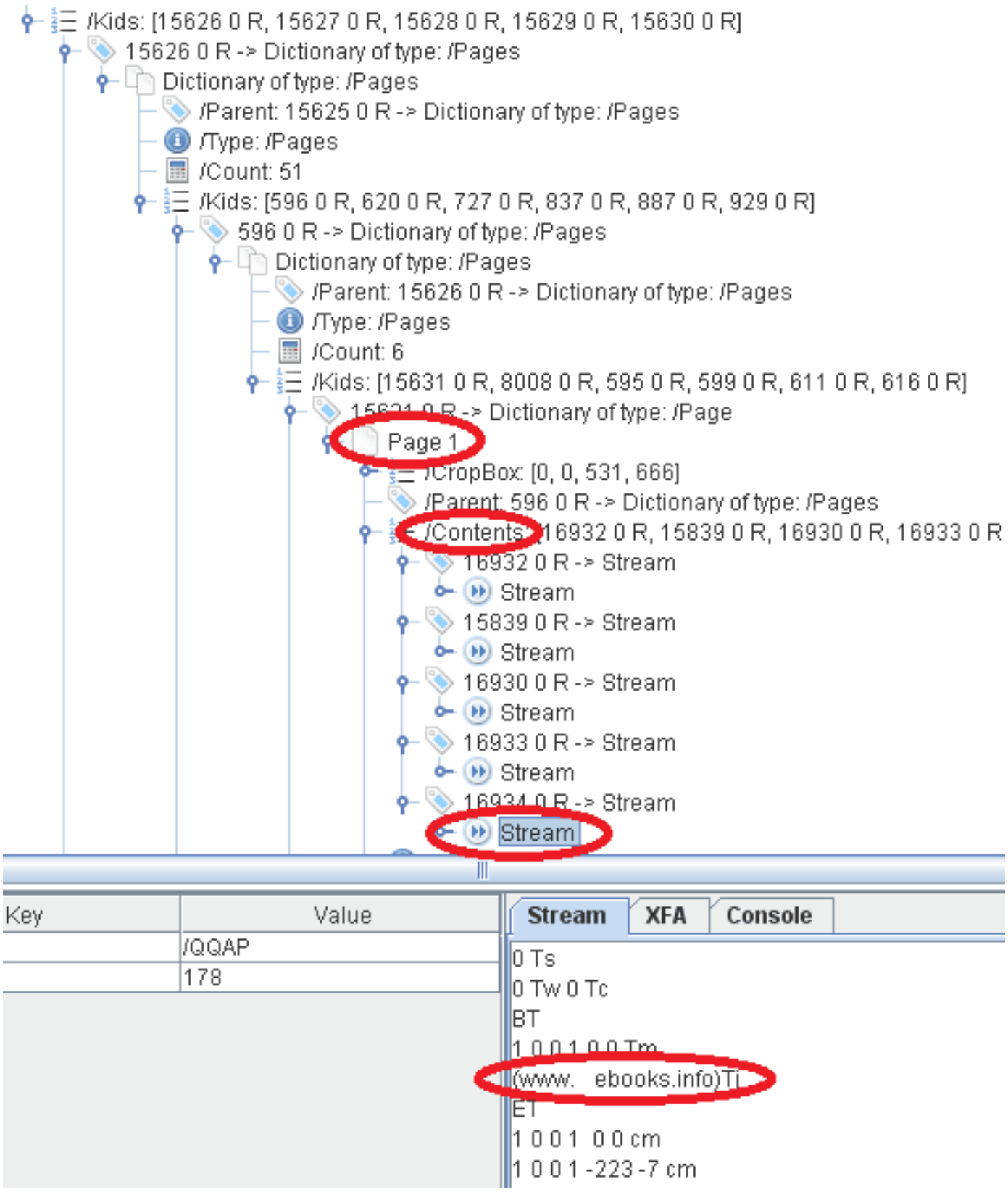
عنوان: **حذف لایه‌های جدید اضافه شده به فایل‌های PDF توسط iTextSharp**
نویسنده: وحید نصیری
تاریخ: ۱۳۹۲/۰۴/۳۱ ۸:۰
آدرس: www.dotnettips.info
برچسب‌ها: iTextSharp, Watermark

شاید یک سری از Ebookهای PDF ایی را دیده باشید که سایت‌های ثالث، آن‌ها را پس از افزودن لایه‌ای متنی، مثلا در ذیل تمام صفحات به همراه آدرس وب سایت خودشان، باز انتشار می‌دهند. در مطلب جاری قصد داریم، نحوه حذف این لایه‌های اضافی را توسط iTextSharp بررسی کنیم.



یافتن و حذف لایه‌های اضافه شده به صفحات یک فایل PDF

برای آشنایی با ساختار سطح پایین لایه‌های اضافه شده نیاز است به برنامه iText Rups مراجعه کنیم.



همانطور که مشاهده می‌کنید، برای رسیدن به لایه‌ای که حاوی متن اضافه شده به ذیل تمام صفحات است، نیاز است ابتدا صفحات را گشوده و سپس CONTENTS آن‌ها را استخراج کنیم. در این CONTENTS کلیه streamهای موجود را بررسی و هر کدام که حاوی متن مورد نظر ما بودند، یافته و سپس آن استریم را با مقدار دهی طول آن به صفر، حذف کنیم. روش کار را در متد ذیل مشاهده می‌کنید:

```
private static void removeWatermarkLayer(string watermarkedFile, string text,  string
unwatermarkedFile)
        {
            PdfReader.unethicalreading = true;
            PdfReader reader = new PdfReader(watermarkedFile);
            reader.RemoveUnusedObjects();
            int pageCount = reader.NumberOfPages;
            for (int i = 1; i <= pageCount; i++)
            {
                var page = reader.GetPageN(i);
                var contentarray = page.GetAsArray(PdfName.CONTENTS);
                if (contentarray == null)
                    continue;

                for (int j = 0; j < contentarray.Size; j++)
                {
                    var stream = (PRStream)contentarray.GetAsStream(j);
                    //دریافت محتوای خام صفحه
                    var content =
System.Text.Encoding.ASCII.GetString(PdfReader.GetStreamBytes(stream));
                    if (content.Contains(text))
                    {
                        //حذف کامل محتوا از فایل
                        stream.Put(PdfName.LENGTH, new PdfNumber(0));
                        stream.SetData(new byte[0]);
                    }
                }
            }

            using (var fileStream = new FileStream(unwatermarkedFile, FileMode.Create,
FileAccess.Write, FileShare.None))
            {
                using (var stamper = new PdfStamper(reader, fileStream))
                {
                    stamper.SetFullCompression();
                    stamper.Close();
                }
            }
        }
```

در این متد watermarkedFile همان فایل اصلی دارای لایه‌های اضافی است. Text متنی است که در استریم‌های صفحات به دنبال آن خواهیم گشت و unwatermarkedFile نام و مسیر فایل تصحیح شده نهایی است که قرار است تولید شود.