

یک سری از دوره‌های پلورال‌سایت دارای زیرنویس هستند که تحت عنوان [Transcript](#) در کنار آن‌ها قرار گرفته‌اند:



Building Applications with ASP.NET MVC 4

This course is a comprehensive introduction to ASP.NET MVC 4, and will give you the essentials you need to start building applications with Microsoft's MVC framework.



Table of Contents

Description

Transcript

Exercise Files

Assessment

Discussion

این زیرنویس‌ها فرمت ویژه‌ای دارند:

```
<li class="transcript-module">
  Introduction to ASP.NET MVC 4
  <ul>
    <li class="transcript-clip" data-p="author=scott-allen&name=mvc4-
building-m1-intro&mode=live&clip=0&course=mvc4-building"><a href="javascript:void(0)"
onclick="LaunchPlayerWindow('http://pluralsight.com/training', 'author=scott-allen&name=mvc4-
building-m1-intro&mode=live&clip=0&course=mvc4-building');">Introduction</a><br />
    <div>
      <a href="javascript:void(0)" onclick="p(this);" data-
s="1.636">Hi, this is Scott Allen and this is the first module in the course design</a>
    </div>
    </li>
    <li class="transcript-clip" data-p="author=scott-allen&name=mvc4-
building-m1-intro&mode=live&clip=1&course=mvc4-building"><a href="javascript:void(0)"
onclick="LaunchPlayerWindow('http://pluralsight.com/training', 'author=scott-allen&name=mvc4-
building-m1-intro&mode=live&clip=1&course=mvc4-building');">Web Platform Installer</a><br
/>
    <div>
      ...
    </div>
  </ul>
</li>
```

در آن، هر li که دارای کلاسی به نام transcript-clip است، حاوی یک div می‌باشد و این div دارای تعدادی لینک است. این لینک‌ها توسط ویژگی datas آن‌ها که بیانگر زمان شروع گفتگو است، مشخص می‌شوند و همین‌طور الی آخر. بنابراین اگر بخواهیم برای آن‌ها ساختاری را تهیه کنیم، به کلاس‌های ذیل خواهیم رسید:

```
public class TranscriptClip
{
    public string Title { set; get; }
    public IList<TranscriptItem> TranscriptItems { set; get; }
}

public class TranscriptItem
{
    public double StartTime { set; get; }
    public string Text { set; get; }
}
```

هر li دارای کلاس transcript-clip، یک شیء TranscriptClip را تشکیل می‌دهد. هر شیء TranscriptClip می‌تواند دارای چندین TranscriptItem باشد.

برای استخراج این اطلاعات، یکی از بهترین ابزارها، کتابخانه [HTML Agility pack](#) است که توسط آن می‌توان به liهای یاد شده دسترسی یافت:

```
var nodes = doc.DocumentNode.SelectNodes("//li[@class='transcript-clip']/div");
```

و سپس اطلاعات آن‌ها را استخراج نمود.

```
using System;
using System.Collections.Generic;
using System.Globalization;
using System.IO;
using System.Linq;
using System.Text;
using System.Text.RegularExpressions;
using System.Web;
using HtmlAgilityPack;

namespace PluralsightTranscripts
{
    public class TranscriptClip
    {
        public string Title { set; get; }
        public IList<TranscriptItem> TranscriptItems { set; get; }
    }

    public class TranscriptItem
    {
        public double StartTime { set; get; }
        public string Text { set; get; }
    }

    public class ExtractSubtitle
    {
        public static void ConvertToSrt(string fileName)
        {
            var transcriptClips = extractItems(fileName);
            var itemNumber = 1;
            foreach (var item in transcriptClips)
            {
                transcriptClipToSrt(item, itemNumber);
                itemNumber++;
            }
        }

        private static void transcriptClipToSrt(TranscriptClip item, int itemNumber)
        {
            var count = item.TranscriptItems.Count;
            var srtFileContent = transcriptItemsToSrt(item.TranscriptItems, count);
            var fileName = removeIllegalCharacters(string.Format("{0}-{1}.srt",
itemNumber.ToString("00"), item.Title));
            File.WriteAllText(fileName, srtFileContent);
        }

        private static string transcriptItemsToSrt(IList<TranscriptItem> items, int count)
        {
            var lineNumber = 1;
            var sb = new StringBuilder();
            for (int row = 0; row < count; row++)
            {
                sb.AppendLine(lineNumber.ToString(CultureInfo.InvariantCulture));
                sb.AppendLine(getTimeLine(items, count, row));
                sb.AppendLine(items[row].Text);
                sb.AppendLine(string.Empty);
                lineNumber++;
            }
            return sb.ToString();
        }

        private static string getTimeLine(IList<TranscriptItem> items, int count, int row)
        {
            var startTs = TimeSpan.FromSeconds(items[row].StartTime);
            var endTs = row + 1 < count ? TimeSpan.FromSeconds(items[row + 1].StartTime) :
TimeSpan.FromSeconds(items[row].StartTime + 5);
            return string.Format("{0} --> {1}", timeSpanToString(startTs), timeSpanToString(endTs));
        }

        private static string timeSpanToString(TimeSpan lineTs)
        {

```

```

    {
        return string.Format("{0}:{1}:{2},{3}", lineTs.Hours.ToString("D2"),
lineTs.Minutes.ToString("D2"), lineTs.Seconds.ToString("D2"), lineTs.Milliseconds.ToString("D3"));
    }

    private static string removeIllegalCharacters(string fileName)
    {
        string regexSearch = string.Format("{0}{1}",
                                                    new string(Path.GetInvalidFileNameChars()),
                                                    new string(Path.GetInvalidPathChars()));
        var r = new Regex(string.Format("[{0}]", Regex.Escape(regexSearch)));
        return r.Replace(fileName, ".");
    }

    private static IList<TranscriptClip> extractItems(string fileName)
    {
        var htmlContent = File.ReadAllText(fileName);
        var results = new List<TranscriptClip>();

        var doc = new HtmlDocument
        {
            OptionCheckSyntax = true,
            OptionFixNestedTags = true,
            OptionAutoCloseOnEnd = true,
            OptionDefaultStreamEncoding = Encoding.UTF8
        };
        doc.LoadHtml(htmlContent);

        var nodes = doc.DocumentNode.SelectNodes("//li[@class='transcript-clip']/div");
        foreach (var node in nodes)
        {
            var itemList = new List<TranscriptItem>();
            var title = node.ParentNode.ChildNodes.First(x => x.Name == "a").InnerText;

            foreach (var childNode in node.ChildNodes)
            {
                if (childNode.Name != "a") continue;

                var dataS = childNode.Attributes.First(x => x.Name == "data-s");
                itemList.Add(new TranscriptItem
                {
                    StartTime = double.Parse(dataS.Value),
                    Text = HttpUtility.HtmlDecode(childNode.InnerText.Trim())
                });
            }

            results.Add(new TranscriptClip { TranscriptItems = itemList, Title = title });
        }

        return results;
    }
}

```

اگر این اطلاعات را کنار هم قرار دهیم، به کلاس کمکی فوق خواهیم رسید. کار با گره‌های li شروع می‌شود. سپس در این گره‌ها، کلیه گره‌های a یا لینک‌ها، یافت شده و سپس dataS و متن آن‌ها استخراج می‌شوند. اگر این‌ها را نهایتاً کنار هم قرار دهیم، می‌توان به فرمت SRT متداول که اکثر پخش‌کننده‌های فایل‌های تصویری قادر به پردازش آن‌ها هستند، رسید. فرمت SRT ساختار ساده‌ای دارد. هر گفتگوی آن حداقل از سه سطر تشکیل می‌شود. سطر اول یک شماره خود افزایشی است. سطر دوم زمان شروع و پایان گفتگو را مشخص می‌کند و سطر سوم بیانگر متن گفتگو است. برای مثال:

```

1
00:00:01,636 --> 00:00:05,616
Hi, this is Scott Allen and this is the first module in the course design

```

دریافت پروژه کامل این مطلب

[PluralsightTranscripts.zip](#)

نظرات خوانندگان

نویسنده: الهام
تاریخ: ۲۲:۲۱ ۱۳۹۲/۰۱/۰۹

سلام آقای نصیری

آیا شما عضو سایت پلورال سایت هستید؟ چقدر پرداخت کردید و این مبلغ را چطور با توجه به وضع ایران واریز کردید؟ آیا ارزش عضو شدن رو داره؟

بخشید چون من دانشجو هستم و بدنبال یادگیری حرفه ای برنامه نویسی و زبانم هم خوبه میخوامم از منابع انگلیسی استفاده کنم.

با تشکر

نویسنده: وحید نصیری
تاریخ: ۲۲:۳۴ ۱۳۹۲/۰۱/۰۹

سلام! من عضو نیستم.

نویسنده: حسین
تاریخ: ۰:۱۳ ۱۳۹۲/۰۱/۱۰

این زیرنویس‌ها فقط برای اعضای اون سایت در دسترسه.از کجا میشه بهشون دسترسی پیدا کرد ؟

نویسنده: وحید نصیری
تاریخ: ۰:۱۶ ۱۳۹۲/۰۱/۱۰

لطفا روی لینک مطرح شده در سطر اول مطلب فوق [کلیک کنید](#) . کل بحث جاری در مورد استخراج اطلاعات و تبدیل فرمت خاص صفحه وبی بود که ملاحظه می‌کنید. این صفحه هم عمومی است (هر چند ظاهر ساده‌ای دارد، اما پشت صحنه و سورس آن، متن زمانبندی شده کل دوره است).

نویسنده: علیرضا جهانشاهلو
تاریخ: ۲:۴۰ ۱۳۹۲/۰۱/۱۴

اتفاقا سایت Lynda هم از همین روش استفاده میکنه و من با کمی تغییر موفق شدم که فایل‌های Transcript آموزشی هاشو استخراج کنم.

ممنون مهندس

نویسنده: سیروان عقیفی
تاریخ: ۱۵:۴۳ ۱۳۹۲/۰۷/۲۷

ظاهراً ساختار عوض شده به این شکل (البته در اینجا data-s حذف شده و مقدار آن به صورت رشته ایی در انتهای مقدار ng-click اضافه شده است به صورت start=39.796 :

```
<li class="transcript-clip">
<a href="javascript:void(0)" ng-click="launchPlayerWindow('http://pluralsight.com/training',
'author=scott-allen&name=mvc4-building-m1-intro&mode=live&clip=0&course=mvc4-
building');">Introduction</a><br>
```

```
<div>
<a href="javascript:void(0)" ng-click="launchPlayerWindow('http://pluralsight.com/training',
'author=scott-allen&name=mvc4-building-m1-intro&mode=live&clip=0&course=mvc4-
building&start=39.796');">and also have an understanding of the design goals of the MVC
framework.</a>
  <a href="javascript:void(0)" ng-click="launchPlayerWindow('http://pluralsight.com/training',
'author=scott-allen&name=mvc4-building-m1-intro&mode=live&clip=0&course=mvc4-
building&start=43.796');">So, let's get started.</a>
</div>
</li>
```

نویسنده: وحید نصیری
تاریخ: ۱۷:۲۱ ۱۳۹۲/۰۷/۲۷

- البته من عضو نیستم و به نظر جدیداً عنوان کردند «Sorry, transcripts are only available to subscribers».
- در کدهای فوق، فقط این چند سطر باید تغییر کنند:

```
//var dataS = childNode.Attributes.First(x => x.Name == "data-s");
var dataS = childNode.Attributes.First(x => x.Name == "ng-click");
var startTime = new Regex("(?s)start=(.+)").Matches(dataS.Value)
                                .OfType<Match>()
                                .First()
                                .Groups[1]
                                .Value;

itemsList.Add(new TranscriptItem
{
    StartTime = double.Parse(startTime),
    Text = HttpUtility.HtmlDecode(childNode.InnerText.Trim())
});
```

نویسنده: پویان
تاریخ: ۱۶:۲۴ ۱۳۹۳/۰۱/۲۴

با سلام
الان که حتماً باید در سایت plural sight عضو باشیم راهی نیست تا زیرنویس بگیریم ؟
ایا شما زیرنویس بعضی از فیلم‌ها را دارید ؟

نویسنده: وحید نصیری
تاریخ: ۱۶:۳۸ ۱۳۹۳/۰۱/۲۴

[فایل تورنت](#) پیوست شده حاوی مثال‌ها و زیرنویس‌های 52 دوره هست:
[srt_only.zip](#)