

نگاهی بر

ساختار فایل های

PDF

.....

گرد آورنده

محمد جواد عمرو

استاد

جناب آقای مهندس شکرانی



وزارت علوم، تحقیقات و فناوری
موسسه آموزش عالی خواجه

خدایا:

گرم تو روی آری چه باک که دیگران روی بگردانند
اگرم تو یاور باشی چه ترس که آفریدگانت یاوریم نکنند
وگر تو غمخوار گشتی چه غم که خلقت غم نخورند

خداوندگارا

مهربانا

معشوقا

گرم تو روی گردانی چه سود که دیگران روی آرند
اگرم تو یآوری نکنی چه حُسن که آفریدگانت یآوری کنند
وگر تو غم نخوری چه شادیست که خلقت غمخوار گردند

خداوندا : همنشینِی ، یآوری ، غمخواری

من خوب میدانم که تنها نیستم !!

مهربانا

اگر آنی ... همنشینیم، یاوریم ، غمخواریم نکنی

عمری تنهای تنهایم

معشوقا

من ، تنها با تو تنها نیستم

ای تنها ترین ، تنهایم مگذار...

تقدیم بر آستان مهربان مهر هشتم

تقدیم به :

پدر بزرگوارم و مادر مهربانم

و با سپاس از استاد گرامیم مهندس شکرانی

فهرست :

4	فصل اول : مقدمه
5	1.1 معرفی اسناد Pdf
5	1.2 تفاوت اسناد Pdf با اسناد دیگر
6	فصل دوم : نیم نگاهی بر Pdf
7	فصل سوم : اشیاء
7	3.1 بولین ها
7	3.2 اعداد
7	3.3 رشته ها و نوشته ها
8	3.3.1 نوشته ها
8	3.4 اسم ها
9	3.5 آرایه ها
9	3.6 واژه نامه
10	3.7 جریان ها و اعداد تهی
11	فصل چهارم : ساختار فایل
11	4.1 فایل های Pdf
12	4.2 هدر
12	4.3 بدنه
13	4.4 جدول ارجاع متقابل
16	4.5 تریلر
18	4.6 بروزآوری افزایشی
19	فصل پنجم : ساختار اسناد Pdf
20	5.1 مقدمه
20	5.2 کاتالوگ
21	5.3 درخت صفحات
23	5.4 درخت نقشه
24	5.4 عملیات ها
25	5.5 واژه نامه اطلاعات
	ضمائم : a : نمونه ای از فایل Pdf b : مراجع

1. مقدمه

این کتابچه بشرح در مورد اسناد با فرمت قابل حمل (سبک) PDF قالب اصلی ادوب اکروبات و خانواده تولیدات می پردازد.
(PDF: Portable Document Format)
هدف از این تولیدات این است که کاربران را برای معاوضه (تغییر) و نمایش اسناد الکترونیکی به آسانی و بطور قابل اعتماد قادر سازد ، جدا از محیطی که آن اسناد در آن ساخته شده اند.

پیشرفت های اخیر دنیای تجارت باعث شده است تا نیاز به تبادل اطلاعات روز به روز بیشتر شود . با استفاده از پست الکترونیکی ، کاربران کامپیوتر مجبور نیستند مدت زیادی منتظر تبادل اطلاعات خود بمانند و با استفاده از PDF ، اسناد و مدارک در همه کامپیوترهایکسان نشان داده میشوند و در همه چاپگرهایکسان چاپ میشوند.

(Pdf اثر گذاشته بر مدل های تصویری از زبان Postscript تا متن و گرافیک را با وضعیت وسایل مستقل و وضوح مستقل وصف کند).

بی شک PDF یکی از بهترین روش های توسعه کارایی نمایش مجازی است.
Pdf همچنین شامل اشیائی از قبیل حاشیه نویسی و پیوند های ما فوق متن (پیوندی که به سادگی از قسمتی از متن به قسمت دیگر میرود) و ... میباشد؛ و اینها جزئی از خود صفحات نیست بلکه برای نمایش محاوره ای (مجازی) مفید است .

به علت سرعت و کارایی بالای فرمت فایل Pdf و سازگاری آن با تمام سیستم های کامپیوتری ، این فرمت مورد پسند اکثر کاربران وب و شبکه های داخلی قرار گرفته است . فایل های Pdf حجم کمی دارند و به همین دلیل در سایت های وب و پیغام های الکترونیکی به کار می روند . قرارداد ها و اسناد حقوقی در فرمت Pdf و با پست

الکترونیکی ارسال میشوند زیرا با استفاده از ابزار امضاء دیجیتالی در Acrobat ، میتوان اینگونه اسناد را امضاء کرد.

1.1 معرفی اسناد PDF

در سال 1993 اولین نسخه Acrobat توسط شرکت Adobe منتشر شد و از آن زمان دنیای انتشار اسناد الکترونیکی متحول شد.

همین طور که تقاضای به اشتراک گذاشتن اسناد بیشتر می شد مسائل و مشکلات هم افزایش می یافت. بعضی از کاربران با سیستم های سازگار IBM و بعضی با سیستم های مکینتاش کار میکردند. همچنین هر سند فقط در برنامه ای باز میشد که سند را بوجود آورده بود. حتی زمانی که کامپیوتر و نرم افزار هم یکی بود، فونت ها متفاوت بودند. اسنادی که از طریق پست الکترونیکی ارسال میشدند ممکن بود در کامپیوتر مقصد باز نشوند یا با ظاهری متفاوت باز شوند.

دنیای تجارت به شکل واحدی از کاغذ دیجیتالی جهت تبادل اطلاعات نیاز داشت. Pdf مشکلات توزیع اسناد کاغذی را بر طرف کرد. با استفاده از Reader رایگان یا Adobe Acrobat هر کسی میتواند اسناد Pdf را دقیقاً یکسان مشاهده و چاپ کند. کاربرانی که با سیستم عامل لینوکس یا ویندوز کار میکنند فایل های Pdf را دقیقاً یکسان مشاهده میکنند. و حالا کمی شمارا با تفاوت Pdf با اسناد دیگر آشنا میکنم:

1.2 تفاوت اسناد PDF با اسناد دیگر

اسناد Pdf را میتوان از طریق اینترنت توزیع کرد. فایل های Pdf با فایل های Word, Excel و Page marker متفاوتند. فایل های Pdf به منظور ویرایش شدن ایجاد نمی شوند و نکته منفی فایل Pdf قابلیت های ویرایشی اندک است. معمولاً Pdf از فایل اصلی کوچکتر است. ظاهر سند اصلی در Pdf به خوبی حفظ میشود. خطوط و پارگراف ها و فونت ها تغییر نمی کنند. با استفاده از تنظیمات امنیتی میتوانید دسترسی بیا چاپ یا ویرایش فایل را محدود کنید. برای اینکه تغییرات انجام شده در هر فایل را به بقیه یادآوری کنید از ابزار Comments استفاده کنید.

ابزار Comments قابلیت هایی جهت بوجود آوردن ملاحظات تبادل آنها بین کپی یک سند Pdf جهت ویرایش و غلط گیری بینچند نفر توزیع شده باشد. در این صورت هر شخصی میتواند ملاحظاته را در فایل مربوط به خودش وارد و همچنین ملاحظات دیگران را نیز به آن بیفزاید.

2. نیم نگاهی بر PDF

PDF چیست؟؟ قبل از نگاه انداختن به جزئیات ساختار Pdf این مهم است که بدانیم Pdf چیست. PDF یک فرمت فایل استفاده شده برای بیان کردن اسناد در یک روش مستقل از برنامه های کاربردی ، سخت افزار و سیستم عامل است . برای ساختن آن ، یک فایل Pdf ، شامل یک سند Pdf و دیگر اطلاعات تعریف شده می شود.

یک فایل Pdf می تواند شامل یک یا چندین صفحه باشد. هر یک از صفحات در یک سند ، میتواند شامل ترکیبی از متن ، گرافیک و تصاویر در یک وسیله باشد و وضوح مستقل از نوع است. این بود شرح مختصر یک صفحه سند Pdf همچنین ممکن است محتوی اطلاعاتی که فقط برای نمایش الکترونیک باشد مثل لینک های مافوق متن ، صدا و حتی فیلم . در فزاش دادن یک سند ، یک فایل Pdf در بر دارد شرحی که مشخصه های استفاده شده در فایل Pdf و اطلاعات درباره ساختارها مهم در فایل را نشان میدهد (موقعیت یابی میکند)

برای فهمیدن یک Pdf مهم است بدانیم که چگونه یک سند Pdf درست میشود و استفاده میشود. بطوریکه با فراگیر شدن اسناد Pdf ، کاربرد های آن و خواندن فایل های Pdf نیز بیشتر شایع شده است.

یکی از اهداف این کتابچه این است که فایل با فرمتی قابل دسترس بطوریکه برنامه های اجرایی بتوانند از آن استفاده کنند بسازید.

در فصول بعدی بیشتر با ساختار و اشیائی که در این ساختار وجود دارند آشنا می شوید. همچنین در ادامه فصول نیز به طریقه ساخت امضاء الکترونیک برای اسناد و حفاظت از آنها آشنا میشوید.

3. اشیاء

PDF از هفت نوع اصلی پشتیبانی میکند :

شماره ها (Numbers) ، بولین ها (Boolean) ، رشته ها و نوشته ها ، اسم ها ، آرایه ها ، واژه نامه و جریان ، بعلاوه شیء Null (هیچ) برای کاربرد های متفاوت تدارک دیده شده است. اشیاء میتوانند برچسب زده شده باشند تا اینکه بوسیله اشیاء دیگر بتوانند به آنها اشاره نمود. یک شی برچسب زده شده یک مفعول غیر مستقیم (شی غیر مستقیم) نامیده میشود . در بخش های زیر انواع اشیاء و شی null را مورد وصف قرار میدهیم .

توجه : Pdf نسبت به حروف کوچک و بزرگ حساس میباشد و حروف کوچک و بزرگ متمایز از یکدیگرند!!

3.1 بولین ها (Booleans)

اشیاء بولین دارای کلمه کلیدی True و false است . و نشان دهنده مقدار صحیح (درست) یا غلط (نادرست) هستند.

3.2 اعداد (Numbers)

PDF از دو نوع عدد پشتیبانی میکند : اعداد صحیح / اعشاری . اعداد صحیح دارای مقادیر علامت دار یا بدون علامت میباشد و Real ها هم فقط در شکل اعشاری هستند. نکته : شکل توانی برای اعداد نظیر 1.0e3 پشتیبانی نشده است.

3.3 رشته ها و نوشته ها

یک رشته مجموعه ای بدون علامت از بایت های 8 بیتی و نوشته ها از ترکیب قرار گیری کاراکترها و بوسیله پرانتز از یکدیگر جدا شده اند. اگر طول یک رشته بسیار زیاد باشد بطوریکه نتوان آنرا در یک خط قرار داد میتوان آنرا در خطوط دیگر (خطوط متوال ادامه رشته) قرار داد . برای این کار فقط کافیست از علامت \ (Backslash) در انتهای یک خط شکافته شده استفاده نمود که نشان میدهد رشته روی خط زیری ادامه می یابد . در این قسمت مثال هایی از رشته ها آورده شده است :

(This is string number 1?)
(strangeonium spectroscopy)

(This string is split \

across \ three lines)

در طول یک رشته ممکن است کاراکتر '\ ' همراه با کاراکترهای دیگر بکار برده شده باشد که هر یک از ترکیبات '\ ' با آن کاراکترها مشخص کننده عملیاتی هستند. در پایین به چند نمونه از این ترکیبات اشاره شده است :

\n	تعویض خط
\v	تعویض سطر
\t	حرکت بصورت افقی tab
\b	Backspace
\\	Backslash
\(	پرانتز
\)	پرانتز
\ddd	کاراکترهای ddd (decimal)

3.3.1 نوشته ها (Text)

رشته های مطمئنی (مسلّمی) را در بر دارند که اطلاعات آن توسط انسان خواناست. این رشته ها شامل شرح متون ، شرح نشانه گذاری ها ، نام بندها ، اطلاعات اسناد و غیره میباشد.

3.4 اسم ها (Names)

یک اسم شبیه به رشته است و با ترتیبی از قرارگیری کاراکترها نوشته شده است . اسم باید با کاراکتر برش '/' شروع شود و با ترتیبی از کاراکترهای ASCII ادامه یابد. بعضی از کاراکترهای خاص مجا به استفاده در اسم ها نمی باشند که برخی از آنها کاراکترهای { , } , [,] , < , > , (,) , # , / , % و چند کد اسکی :

در این قسمت نمنه ای از چند اسم مجاز آورده شده است :

```
/Name1
/ASomewhatLongerName2
/A;Name_With-various***Characters?.
/1.2
/.notdef
```

کاراکتری به عنوان کاراکتر هیچ Null <00> میتواند در اسامی استفاده شود. برای ایکار باید از علامت # در بین نام استفاده کرد :

```
/Adobe#20Green
```

/PANTONE#205757#20CV
/paired#28#29parentheses
/TheKeyOfF#23Minor
/A#42

در ضمن میتوان بجای استفاده از خود کاراکتر از کد اسکی آن کاراکتر نیز استفاده نمود به عنوان مثال /A#42 راه دیگری برای نوشتن AB/ است.
** این خاصیت در رشته ها هم وجود دارد.
نکته: حداکثر شماره مربوط به کاراکترها 127 میباشد .
برای مثال اسم /A#20B به چهار کاراکتر (/ A فضا B) اشاره میکند ، نه شش کاراکتر.

3.5 آرایه ها

یک آرایه یک ترتیب از اشیاء Pdf است و ممکن است مخلوطی از انواع اشیاء را در بر داشته باشد.

ک آرایه با علامت کروشه چپ [شروع گشته و در ادامه وسیله یک ترکیبی از اشیاء ادامه یافته و با کروشه راست] نیز پایان می پذیرد.
این یک مثال از آرایه است :

[0 (Higgs) false 3.14 3 549 /SomeName]

3.6 واژه نامه(Dictionary)

یک واژه نامه جدولی رادر بر دارد که شامل جفت هایی از اشیاء است .
اولین عنصر هر جفت کلید نامیده میشود و عضو دومین ارزش نامیده شده است.
کلید باید یک اسم باشد و یک ارزش از هر نوع شی و نیز شامل یک واژه نامه میتواند باشد.
عموماً یک واژه نامه استفاده میشود (کاربرد دارد) تا اینکه با دیگر خصیصه های یک شیء پیچیده با مشخص کردن جفت 'کلید و ارزش' از یک خصیصه جمع شود و متصل شود (واژه نامه تو در تو) (این قسمت اشاره به استفاده واژه نامه در اشیائی نظیر صفحه و دیگر اشیاء مرکب دیگر دارد)
یک واژه نامه با استفاده از کروشه زاویه دار (<<) شروع و بایک ترکیب از جفت های کلید و ارزش ادامه یافته و در انتها نیز با کروشه زاویه دار راست (>>) پایان می پذیرد.
برای مثال :

>>..... ارزش کلید/ ارزش کلید / <<

<< /Type /Example /Key2 12 /Key3 (a string) >>

نمونه ای از واژه نامه های تو در تو (واژه نامه ای در داخل واژه نامه دیگر) :

```
<<
  /Type /AlsoAnExample
  /Subtype /Bad
  /Reason (unsure)
  /Version 0.01
  /MyInfo
    <<
      /Item1 0.4
      /Item2 true
      /LastItem (not!)
      /VeryLastItem (OK)
    >>
  >>
```

اشیاء واژه نامه ، قطعات (ساختار) اصلی یک سند Pdf هستند.

بخش های زیادی در مدرک Pdf مثل صفحات و فونت ها برای نمایش داده شدن از واژه نامه استفاده میکنند.

توسط مجمعی از انواع کلید یک واژه نامه مشخص میکند یک نوع شی را (شی مرکب) که در واژه نامه تعریف شده است .

و در برخی حالات کلید میتواند به یک زیر انواع (subtype) را مشخص کند که در مواردی چون تخصصی شدن یک نوع خاص استفاده میشود .

مثلاً در یک فونت ، ارزش کلید و چندین زیر عنوان شامل True type, Type3,MMType1,Type1 و ... است .

3.7 جریان ها (Streams) و اشیاء تهی (Null)

از دیگر اشیاء از مجموعه اشیاء بکار رفته می توات جریان ها و اشیاء تهی و اشیاء غیر مستقیم و ... نام برد

اشیاء تهی همانطور که از نامشان پیداست داری کلمه کلیدی Null میباشد و هیچ مقداری را نشان نمی دهند . (از موارد کاربردی آنها در واژه نامه هاست که در قسمت ارزش واژه نامه قرار گرفته و راه سریع تر و ساده تری برای واژه نامه است که در مجموع کلید را حذف میکند .(کلید حاوی مقدار تهی بی فایده است) .

و اشیاء غیر مستقیم نیز برای تعریف انواع جدید بکار میرود که اصطلاحاً به آنها مفعول بی واسطه میگویند.

در فصول بعدی با استفاده از اشیاء توصیف شده در این بخش به معرفی ساختار فایل و اسناد Pdf می پردازیم ...

4. ساختار فایل

در این فصل به شرح سازماندهی و ساختار یک فایل Pdf می پردازم .
 یک فایل Pdf آماده میکند ساختاری را که یک سند را (document) نمایش میدهد.
 این ساختار راهی را که به سرعت بازیابی کند هر بخش یک سند و یک سازو کار برای سازماندهی آن را آماده می کند.
 بدنه یک فایل Pdf شامل مراحل (مراتب) اشیاء Pdf که در شکل دادن Pdf استفاده می شوند.

4.1 فایل های Pdf

یک فایل Pdf استاندارد شامل چهار بخش است :
 سر صفحه (Header)
 بدنه (Body)
 جدول ارجاع متقابل (cross reference table)
 تریلر (Trailer)

- شکل زیر این ساختار را نمایش میدهد :

```
<PDF file> ::= <header>
                <body>
                <cross-reference table>
                <trailer>
```

یک فایل Pdf با داده های binary ممکن است دارای خطوط طویلی باشد .
 به هر حال ، برای سازگاری فایل Pdf با برنامه های کاربردی ، طول خطوط در یک فال Pdf بیشتر از 255 کاراکتر نخواهد بود.

اما یک استثنا در مورد محتویات رشته ای یک واژه نامه وجود دارد .

- شکل 5.1: ساختار یک فایل Pdf که تابحال به روز آوری نشده است :

header
Body
cross-reference table
trailer

4.2 هدر (سر صفحه) (Header)

اولین خط یک فایل Pdf که شماره نگارش مشخصه Pdf را مشخص میسازد تا به فایل بپیوندد مثلاً نگارش جاری 1.4 است. (با نگاهی به فایل کامل Pdf میتوان این مقدار را یافت).

به هر حال اگر یک فایل Pdf از یک یا چند ویرایش زود تر از 1.4 هم که باشد توسط برنامه کاربردی قابل اجرا میباشد . مثلاً اگر چندفایل مربوط به نگارش 1.3 و 1.2 و یا 1.1 باشند باز هم توسط برنامه کاربردی قابل خواندن هستند.

و هدر که در اولین خط یک فال Pdf واقع است نشان دهنده این نگارش است .

`<header> ::= <PDF version>`

هدر معرفی کننده یک سند Pdf است و توسط یک رشته " %PDF 1.3 " نمایان میشود.

و این همانند اعداد جادویی در یونیکس است .

4.3 بدنه

بدنه یک فایل Pdf تشکیل شده است از دنباله ای از اشیاء که غیر مستقیم یک سند (document) را نمایش میدهند.

ساختار بدنه سند شامل اطلاعاتی راجع به اشیاء بکار رفته در صفحات است.

و این یک مجموعه از اشیاء برای توصیف کردن درخت ساختار صفحه که در آن صفحات و عنصر ها (متن ، گرافیک و...) روی صفحات سازماندهی شده است.

اشیاء Pdf برگ های درخت Pdf می باشند ، که هر شی شامل سه مولفه است:

یک شماره / موقعیت شی در فایل Pdf (offset) / و محتویات (ظرفیت)

اشیاء در نوع های اصلی که در فصل قبلی (اشیاء) شرح آنها بیان شده است .
مولفه های نشان دادن یک مدرک (document) عبارتند از :
فونت ها ، صفحات ، تصاویر نمونه و

توضیحات (comments) می توانند در هر جا در بخش بدنه یک فایل ظاهر شود.
آنها با یک علامت % شروع میشوند و ممکن است در هر نقطه ای روی خط قرار گیرند. تمام
نوشته های ما بین % تا پایان خط به عنوان توضیحات در نظر گرفته میشود.
و قسمت های comment در مورد رشته ها یا جریان هایی که در طول یک بخش قرار دارد
توضیحاتی بیان می کند.

4.4 جدول ارجاع متقابل (Cross-reference table)

جدول ارجاع (جدول مختصات) اجازه میدهد یک تجزیه گر Pdf (مانند Adobe Acrobat) به
سرعت به اشیا دست بیابد .

اطلاعات جدول ارجاع متقابل شامل دستیابی تصادفی (دستیابی مستقیم) به مفعول های غیر
مستقیم در فایل ، که این امکان را فراهم میکند برای سازماندهی یک سی خاص ، فایل بطور کامل
خوانده نشود (عملکردی شبیه به شاخص ها)

برای هر مفعول غیر مستقیم در فایل ، جدول شام یک خط توصیف راجع به محل ورود شی در
فایل است. یعنی مختصاتی از اشیاء در جدول گردآوری شده است که برای دستیابی به شیء
خاصی در فایل ، با یافتن مختصات ، فقط به همان قسمت از فایل دستیابی داشته و پس از
خواندن آن قسمت به شی دست می یابد.

به همین منظور است که به جدول ارجاع متقابل ، جدول مختصات فایل نیز می گویند.
هر فایل Pdf شامل یک جدول ارجاع متقابل است که خود شامل یک یا چندین بخش می باشد.
اگر بروز آوری به فایل اضافه نشده باشد، جدول ارجاع شامل فقط یک بخش می باشد و بخش
اضافه شده به جدول ارجاع متقابل مربوط به زمان به روز آوری (Update) میباشد.

بخش ثابت (بخش اولیه) جدول مختصات در فایل Pdf به یک شکل ثابت است و برای دستیابی
تصادفی (مستقیم) به ورودی ها (entry) در جدول ارجاع اجازه میدهد.
این بخش با خطی شامل کلمه کلیدی xref شروع میشود .

و در ذیل این خط یک یا چندین زیر جدول مختصات است .

<cross-reference section> ::=
xref
<cross-reference subsection>+

توجه : اگر بروزآوری در فایل انجام نشده باشد جدول ارجاع فقط یک بخش را در بردارد.
 هر یک از ورودی های زیر بخش ها شامل حدود پیوسته از شماره های اشیاء می باشند.

ساختار زیر بخش های سیستم مختصات مفید برای بروزآوری افزایشی (که در قسمت های بعدی توضیح می دهیم) می باشد ؛ زیرا اجازه میدهد یک زیر بخش به سیستم مختصات اضافه شود تا اشیاء ورودی یا حذف شده در فایل Pdf به جدول ارجاع نیز اضافه شوند.
 هر زیر بخش سیستم مختصات شامل یک هدر که یا دو شماره شروع میشود :
 اول شماره شی در آن زیر بخش و شماره وارده (entry) در زیر بخش

در پایین نمونه ای بیان شده است :

<cross-reference subsection> ::=
<object number of first entry in subsection>
<number of entries in subsection>
<cross-reference entry>+

هر entry شامل 20 بایت می باشد و نیز شامل علامت end of line (eol) می باشد.

این دو شکل متفاوت برای وارده های جدول ارجاع هستند.

یکی برای اشیائی که در حال استفاده میباشند و دیگری برای اشیائی که حذف شده اند و آزاد میباشند:

<cross-reference entry> ::= <in-use entry> | <free entry>

برای یک شی در حال استفاده وارده شامل offset بایت (محاسبه بایت) (شماره بایت ها از شروع فایل به شروع شی) و شماره ایجاد شی (شماره نسل) و نیز کلمه کلیدی n می باشد.

<in-use entry> ::= <byte offset> <generation number> n <end-of-line>

بایت افسست (offset) شامل 10 رقم میباشد . در صورت لزوم میتوان باقیمانده آنرا با صفر های مقدم پر نمود.

شماره ایجاد شی توسط یک فاصله جدا شده است . شماره ایجاد شی یک پنج رقمیست و برای آن نیز می توان با صفر های مقدم لایه گذاشت .

بعد از شماره ایجاد (شماره نسل) یک فاصله و کلمه کلیدی n است ، در پایان هم کلمه کلیدی eol است (توجه کنید کلمه کلیدی n برای اشیاء در حال استفاده بکار میرود)

برای یک شی آزاد (delete/free) وارده شامل شماره شی آزاد بعدی و شماره نسل خودش و کلمه کلیدی f است.

<free entry> ::=
<object number of next free object>
<generation number> f <end-of-line>

وارده همان شکل و فرمتی را دارد که یک شی در حال استفاده دارد : یک ده رقمی برای شماره شی ، یک فضا ، یک پنج رقمی شماره ایجاد و یک کلمه کلیدی f و در پایان هم eol (end of line). اشیاء آزاد در فرم جدول ارجاع ، یک لیست پیوندیست که با entry برای هر شی آزاد ، اشاره به شی آزاد بعدی میکند.

همیشه اولین وارده (entry) در جدول (شیء شماره صفر) آزاد است و یک شماره نسل 65535 دارد و آن سر لیست پیوندی اشیاء آزاد میباشد.

وارده آزاد گذاشته در جدول ارجاع (سر لیست پیوندی) شیء شماره صفر ، برای دستیابی به شی آزاد بعدی بکار برده میشود .

موقعیکه یک مفعول غیر مستقیم حذف شده است ؛ entry جدول ارجاع آزاد علامتگذاری شده است . شماره نسل در وارده افزایش داده شده است تا شماره ایجادش توسط یک شی بکاربرده شده در زمانی دیگر استفاده شود.

هر زمانی که entry دوباره استفاده شده است ، شماره نسل (شماره ایجاد) افزایش داده شده است . بیشترین حد شمکاه ایجاد 65535 است . و این شماره به اولین وارده جدول ارجاع رسیده است.

• نمونه ای از جدول ارجاع (1):

نمایش یک سیستم مختصات شامل بخش اولیه با یک زیر بخش دارای 6 وارده که 4 تای آن در حال استفاده (1و2و4و5) و دو تای آن آزادند (3و0) .

```
xref
0 6
0000000003 65535 f
0000000017 00000 n
0000000081 00000 n
0000000000 00007 f
```

0000000331 00000 n

0000000409 00000 n

شیء شماره 3 حذف شده است و به شیء بعدی ایجاد شده با شماره شیئی 3، شماره ایجاد 7 داده خواهد شد.

• نمونه ای دیگر از جدول ارجاع (2):

نمایش یک سیستم مختصات با 4 زیر بخش شامل 5 وارده (درمجموع)
 زیر بخش اول که شامل وارده ای برای شیء شماره 0 که آزاد است
 زیر بخش دوم که دارای یک وارده برای شیء شماره 3 که در حل استفاده است .
 زیر بخش سومین شامل دو وارده با شماره اشیا 23 و 24 که هر دو در حال استفاده اند.
 شیء شماره 23، دوباره استفاده شده است بطوریکه دیده میشود که آن شماره ایجاد 2 دارد.
 و زیر بخش چهارمین یک وارده داشته ، شیء شماره 3 که در حال استفاده است .

xref

0 1

0000000000 65535 f

3 1

0000025325 00000 n

23 2

0000025518 00002 n

0000025635 00000 n

30 1

0000025777 00000 n

4.5 تریلر (ناظر trailer)

تریلر این امکان را فراهم می آورد که با درخواست خواندن یک فایل Pdf به سرعت به جدول ارجاع و اشیاء دسیابی پیدا شود .

برای این کار برنامه (منظور برنامه خواننده Pdf است) باید فایل Pdf را از پایانش بخواند .
 آخرین خط فایل Pdf [end of file] دارای علامت %eof است و دو خط قبل از آن نیز شامل کلمه کلیدی startxref که offset (فاصله) بایت از شروع فایل به شروع کلمه xref سیستم مختصات را در بر دارد .

و واژه نامه تریلر در دو خط قبل تر قرار دارد. (واژه نامه تریلر تشکیل شده از کلمه کلیدی trailer و در ادامه نیز مجموعه ای از ارزش و کلید که در گروه های زویه دار زوج شده اند)

- به نمای کلی یک تریلر توجه کنید :

```
<trailer> ::= trailer
<<
<trailer key-value pair>+
>>
startxref
<cross-reference table start address>
%%EOF
```

نکته : در **startxref** آدرس شروع سیستم مختصات در فایل ذخیره شده است یعنی میزان فاصله xref از ابتدای فایل

- و نمونه ای از یک تریلر برای یک فایل بروز آوری شده در پایین نشان داده شده است:

```
trailer
<<
/Size 22
/Root 2 0 R
/Info 1 0 R
>>
startxref
18799
%%EOF
```

و حالا به برخی از خصوصیات تریلر می پردازم و در باره آنها توضیحاتی میدهم:

Key	Type	Semantics
Size	integer	شماره مجموع وارده های جدول ارجاع فایل، شامل جدول اصلی و تمام به بروز آوری ها (لازم)
Prev	integer	(لازم فقط در حالتی که جدول ارجاع بیشتر از یک بخش داشته باشد)
		و شماره فاصله محل قبلی جدول ارجاع از ابتدای فایل
		– اگر جدول ارجاع بروز نشده باشد این کلید را در بر نخواهد داشت
Root	dictionary	شی کاتالوگ برای یک سند (در فصل بعد توضیح داده شده)
Info	dictionary	واژه نامه اطلاعات برای یک سند
ID	array	یک آرایه از دو رشته که هر دوی آنها یک شناسه اند شناسه اول زمانی را که فایل درست شده و شناسه دوم زمانی که

فایل تغییر کرده یا بروز رسانی شده است نشان میدهد (لازم)

زمانی که سند رمزی شده باشد لازم است و اطلاعاتی در مورد رمز شدن سند داراست dictionary Encrypt

4.6 بروز آوری افزایشی (Incremental update)

محتویات یک فایل Pdf بدون بازنویسی کردن فایل کامل ، به روزآوری می شود و تغییرات به پایان فایل اضافه می شود بطوریکه محتویت اصلی فایل دستنخورده باقی می ماند. موقعیکه یک فایل Pdf جدیدی به روزآوری شده است یا اشیاء عوض شده ، اضافه شده اند ، یک بخش سیستم مختصات اضافه شده اسن و یک تریلر جدید وارد شده است .

- در شکل زیر فایل با ساختار نتیجه را نشان می دهد:

```
<Updated PDF file> ::=
    <PDF file>
    {<update>}*
<update> ::= <body>
    <cross-reference section>
    <trailer>
```

سیستم مختصات موقعیکه یک ایل Pdf به روز آوری میشود یک بخش اضافه میشود. که شامل وارده های اشیائی بوده اند که تغییر یافته یا حذف شده باشند ، بعلاوه وارده ای برای شی 0 .

اشیاء حذف شده بی تغییر در فایل باقی میمانند اما علامت گذاشته شده اند بطوریکه در وارده (entry) های جدول ارجاع حذف شده اند .

شکل زیر نمایی از تاثیر یک به روزآوری افزایشی را نمایش می دهد :

Header
Original Body
Original cross-reference section
Original trailer
Body update 1
Cross-reference section 1
Updated trailer 1 Body update n
Cross-reference section n
Updated trailer n

5. ساختار اسناد Pdf

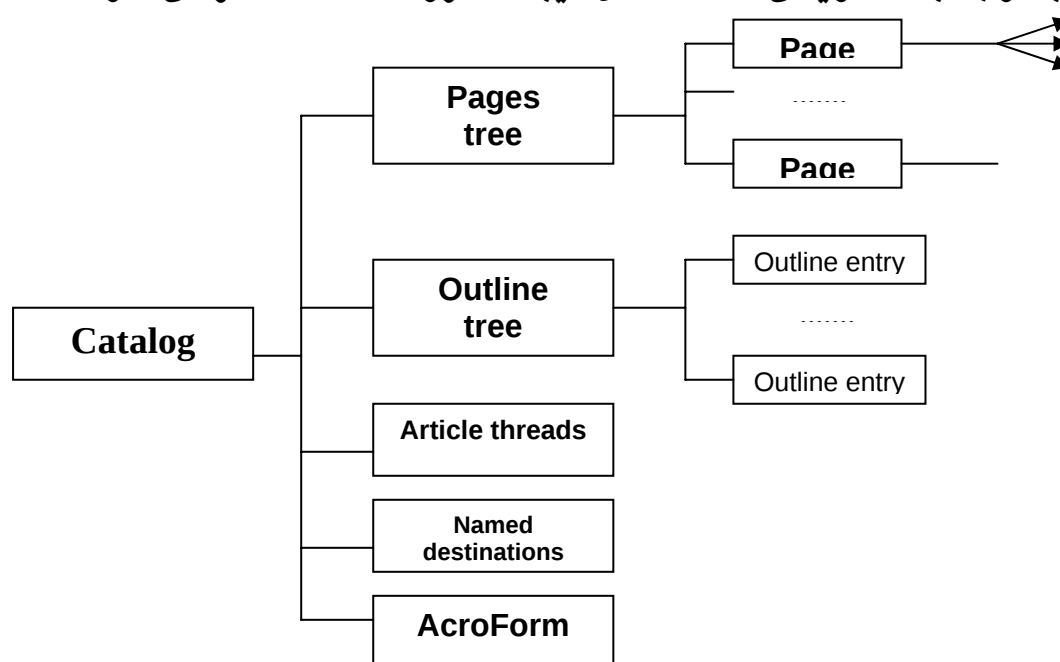
Pdf فراهم کرده است یک نمایش الکترونیکی از یک مدرک را که مجموعه‌ایست از صفحات که شامل متن، گرافیک و تصاویر به همراه دیگر اطلاعات مختصر مانند تصاویر ریز از صفحات، شرح متون، لینک‌های ما فوق متن و
 در فصول قبل اساس مقدماتی برای درک نمایش یک مدرک Pdf را توضیح دادم اما درباره خود نمایش توضیحی بیان نشده است.
 بدنه یک فایل Pdf از ترتیبی از اشیاء تشکیل شده است که مجموعاً یک مدرک Pdf را نشان میدهند. این فصل منحصراً روی محتویات بخش بدنه یک فایل Pdf تمرکز میکند و شامل توضیحات از اشیائی که یک مدرک Pdf در بردارد، میباشد.
 در این فصل فقط به اشیائی می پردازیم که بیشترین کاربرد را دارند و استفاده از آنها عموماًست.

1. مقدمه

یک مدرک Pdf توصیف میشود بعنوان ترکیبی از اشیاء که در بخش بدنه آنها را در بر دارد .
 شکل 6.1 ساختار یک مدرک Pdf را نشان میدهد.

اشیاء زیادی در این سلسله مراتب ، واژه نامه اند (اشیاء مرکب)
 والدین ، فرزند و نیز هم نسلان (برادران) که توسط زوج ، کلید و ارزش ، که بطور غیر
 مستقیم به والدین /فرزند و هم نسلان ارتباط دارند ، نشان داده شده اند.
 برای مثال شی کاتالوگ که ریشه درخت است شامل کلید صفحات میباشد و یک اشاره غیر
 مستقیم به شیئی که فرزند صفحه است دارد .

هر صفحه از مدرک با توجه به محتویاتی که شامل آن میباشد ، روی صفحه ظاهر می شود.



شکل 6.1

5.2 کاتالوگ

کاتالوگ یک واژه نامه است که گره ی ریشه ای مدرک است .
 و شامل یک اشاره به درخت صفحات در مدرک ،واشاره ای به خطوط اصلی نمایش دادن
 اشیاءسند ، و یک اشاره به رشته های بند مدرک ، و فهرست مقصد های نام دارد (شکل 6.1)
 • در پایین نمونه ای از شی کاتالوگ نشان داده شده است :

```

1 0 obj
<<
/Type /Catalog
/Pages 2 0 R
/Outlines 3 0 R
/PageMode /UseOutlines
>>
endobj
    
```

در ضمن کاتالوگ دارای صفات مختلفی است که چند تا از آنها عبارتند از : Type ، Dests ، Threads ، Pages ، Outlines و ... که به توضیح در باره تک تک آنها نمی پردازم .

5.3 درخت صفحات

صفحات یک مدرک در دسترس هستند توسط یک گره از درخت که به درخت صفحات مشهورند. این درخت ترتیب صفحات سند را مشخص می سازد. برای بهینه سازی نمایش در برنامه های کاربردی ، adobe Distiller و adobe PDF درخت ها را متعادل و موزون ساخته اند. ساختار درخت اجازه میدهد تا برنامه کاربردی برای یک مدرک شامل هزاران صفحه ، تنها حافظه محدودی را مصرف کند ، همچنین سرعت دستیابی به صفحات هم افزایش می یابد. بنابراین برنامه های نمایش دهنده Pdf بایستی این نوع درخت را بپذیرند . بشرط اینکه گره های درخت شامل کلید هایی باشد که در پایین ذکر شده اند . ساده ترین ساختار تشکیل شده از یک گره صفحه و آن مستقیماً به تمام اشیاء موجود در صفحه اشاره میکند. **** توجه :** ساختار درخت صفحات برای یک مدرک با محتوای مدرک بی ارتباط است . برای یک فایل Pdf به عنوان مثال برای یک کتاب ، تضمینی نیست که یک فصل توسط یک گره تنها در درخت صفحات نشان داده شده باشد.

Key	Type	Semantics
Type	name	نام نوع بکاررفته (بشتر اوقات <u>صفحه</u> نوع بکار رفته است) (لازم)
Kids	array	(لازم) لیستی از اشیاء غیر مستقیم به فرزندان بی واسطه گره صفحه
Count	integer	تصریح کننده شماره ی گره های ریشه (لازم)
Parent	dictionary	

برنامه هایی که فایل Pdf را مصرف میکنند (منظور نمایش دادن است و کاربردهای دیگر) یا تولید میکند نیاز به این ندارند تا ساختار موجود درخت صفحات را حفظ کنند .
ریشه و تمام گره های درخت صفحات ، واژه نامه هستند . که مقداری از محتویات آن در جدول زیر آورده شد است :

```
2 0 obj
<<
/Type /Pages
/Kids [4 0 R 10 0 R 24 0 R]
/Count 3
>>
endobj
```

5.3.1 به ارث بردن خواص (Inheritance of Attributes)

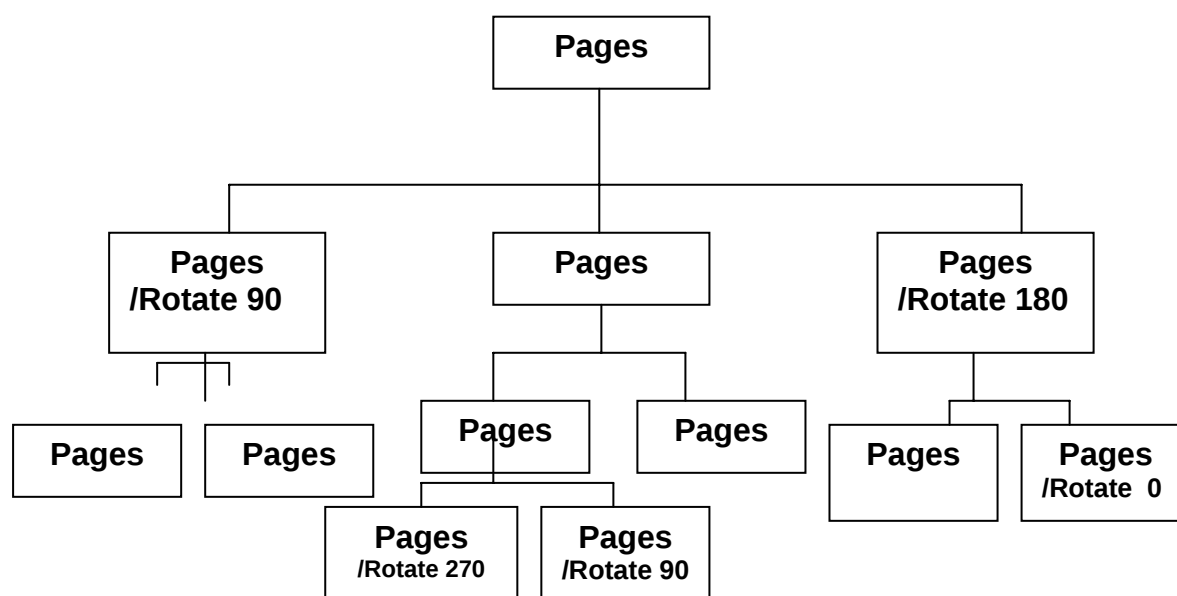
یک شی صفحه ممکن است در بردارنده خواص اضافی باشد و آنها ارزش هایی را برای فرزندان شیء صفحه فراهم می آورد .

چنانچه که ما می گوئیم آن مقادیر را به ارث برده اند.

برای مثال یک سند ممکن است یک mediabox را برای تمام صفحات به وسیله شی صفحه ریشه ، تعریف کند (شی صفحه در بخش بعدی توضیح داده شده)

یک صفحه منفرد در مدرک می توان mediabox را به طور اختصاص هم تعریف کند ، به عنوان مثال مشخص کردن یک mediabox در شی صفحه این خاصیت را برای آن صفحه بی اعتبار میکند (صفحات میتوانند با خواص اختصاصی مربوط به خود ، از خواص به ارث رسیده طبیعت نکنند)

وراثت خواص : خواص ذکر شده در باره کلید Rotate آورده شده



5.4 درخت نقشه (outline tree)

یک نقشه (خطوط اصلی outline) به یک کاربر اجازه میدهد تا به نمایشی از سند به وسیله نام ها دسترسی داشته باشد .

بطوریکه بایک لینک شرح ، قسمتی از خطوطی که کاربر درخواست کرده است (bookmark) فعال شود (نمایش داده شود). درخت نقشه شامل وارده هایی است که به آن خطوط اصلی میگویم. واردهای (entrys) درخت نقشه از ترکیبی از عناصرها تشکیل شده است . یک وارده ممکن است خود دارای یک یا چندید مرحله (بخش) باشد . که به آن زیروارده میگویندو آن زیر وارده ممکن است یزر وارده ، وارده های دیگر باشد و یا ممکن است زیر وارده ای از مجموعه خودش باشد.

یک وارده خطوط اصلی ممکن است باز یا بسته باشد .اگر آن وارده باز باشد ، فرزندانش هم قایل رویت هستند و موقعیکه خطوط اصلی نشان داده شده است و اگر آن وارده بسته باشد این طور نیست .

اگر یک مدرک (document) دارای درخت نقشه باشد ، در گاتالوگ به کلید (منظور کلید ریشه) خطوط اصلی دسترسی داده شده است (کلید در بخشی از شی کاتالوگ ذخیره میگردد)

مقدار این کلید ، شی خطوط اصلی است که ریشه درخت نقشه (درخت خطوط اصلی) است. محتویات واژه نامه خطوط اصلی را میتوان به کلید های : Count و First و Last نام برد .

- شیء خطوط اصلی با 6 وارده

```
21 0 obj
<<
/Count 6
/First 22 0 R
/Last 29 0 R
>>
endobj
```

واردی های خطوط اصلی سطح بالا ، در یک لیست پیوندی واقعند که با اشاره گر اول به سرسیت و با اشاره گر آخر به پایان لیست اشاره میکنند.

5.5 عملیات ها :

(شروع در Pdf 1.1) مشخص میکند که یک همی چه موقع انجام شده است ، مثلاً یک نمایش توسط لینک چه موقعی انجام شده و یا زمانی که وارده های خطوط اصلی فعال شده اند و یا حتی موقعیکه یک مدرک باز شده است.

چند Pdf زیر عنوان را برای عمل ها تعریف میکند:

GoTo ، GoToR ، Launch ، Url ، Sound ، Movie ، SetState ، Hide و ...

نکته : برخی از عناوین عمل ها در ویرایش 1.3 اضافه گشته است .

Action Type	Description
GoTo	تغییر یافتن شماره صفحه و رفتن به صفحه مبدا یا زوم کردن
GoToR(Go to Remote)	باز کردن فایل پی دی اف دیگری و یافتن شماره صفحه و رفتن به صفحه مبدا یا زوم کردن
Launch	باز کردن یک برنامه ، و معمولاً باز کردن یک فایل
Sound	Play a sound.
Movie	Play a movie.
Thread	Begin reading an article thread.
JavaScript	Execute a JavaScript script.
ImportData	Import field values from a file.

و ...

5.6 واژه نامه اطلاعات :

تریلر یک مدرک ممکن است یک اشاره به یک واژه نامه اطلاعات که آن اطلاعات آماده ساختن (فراهم آوردن) یک مدرک را در بر دارد ، داشته باشد .
این واژه نامه اختیاری ممکن است یک یا چندین کلید در بر داشته باشد ، که قادی هر یک از کلیدها ، رشته هستند
و در اغلب موارد رشته ها اغلب مکتون هستند یعنی رشته ها توسط افراد هم میتوانند خوانده شوند. که این رشته ها ممکن است نشان داده شوند و یا توسط کاربر فایل Pdf بروزآوری شود.

ضمیمه a : نمونه ای از یک فایل Pdf

در این قسمت نمونه ای از یک فایل Pdf کوچک آورده شده است .جدول زیر حاوی اشیاء بکاررفته در سند است .

<u>Object number</u>	<u>Object type</u>
1	Catalog
2	Outlines
3	Pages
4	Page
5	Contents
6	ProcSet array

```
%PDF-1.0
1 0 obj
<<
/Type /Catalog
/Pages 3 0 R
/Outlines 2 0 R
>>
endobj
2 0 obj
<<
/Type /Outlines
/Count 0
>>
endobj
3 0 obj
<<
/Type /Pages
/Count 1
/Kids [4 0 R]
```

```
>>
endobj
4 0 obj
<<
/Type /Page
/Parent 3 0 R
/Resources << /ProcSet 6 0 R >>
/MediaBox [0 0 612 792]
/Contents 5 0 R
>>
endobj
5 0 obj
<< /Length 35 >>
stream
endstream
endobj
6 0 obj
[/PDF]
endobj
xref
0 7
0000000000 65535 f
0000000009 00000 n
0000000074 00000 n
0000000120 00000 n
0000000179 00000 n
0000000300 00000 n
0000000384 00000 n
trailer
<<
/Size 7
/Root 1 0 R
>>
startxref
408
%%EOF
```

ضمیمه b : مراجع

b.a کتابها و کتب الکترونیک

1. Portable Document Format - Reference Manual / Version 1.3

Adobe Systems Incorporated / March 11, 1999

Authors—Tim Bienz, Richard Cohn, and Jim Meehan

2. Portable Document Format - Reference Manual / Version 1.2

Adobe Systems Incorporated / November 12, 1996

Authors—Tim Bienz, Richard Cohn, and Jim Meehan

3. خود آموز ادوب اکروبات / اسمیت- کوکس / مترجم : محمد رضا منصوری

1. <http://www.adobe.com/>
2. <http://www.wotsit.org/>
3. <http://www.Huckert.com>
4. <http://www.barnamenevis.org>
5. <http://www.Planetpdf.com>

با سپاس فراوان از جناب آقای مهندس اصغرپور
و خواهر عزیزم مهندس عمویی