

مقدمه ای بر ایکس.ام.ال. (XML)

ترجمه لیلا خلیلی

دانشجوی کارشناسی ارشد کتابداری و اطلاع رسانی دانشگاه تهران

"ایکس.ام.ال." (XML: Extensible markup language) برای توصیف داده طراحی شده است و بر آنچه از نوع داده است تمرکز دارد. "اچ.تی.ام.ال." (HTML: Hypertext Markup Language) بر ای نمایش داده طراحی شده است و بر چگونگی نمایش داده تمرکز دارد.

"ایکس.ام.ال." چیست؟

"ایکس.ام.ال." مخفف Extensible markup language زبان نشانه گذاری گسترش پذیر است.

تعاریف عمده "ایکس.ام.ال." و "اچ.تی.ام.ال."

"ایکس.ام.ال." جانشینی برای "اچ.تی.ام.ال." نیست. "ایکس.ام.ال." برای توصیف داده و "اچ.تی.ام.ال." نمایش داده طراحی شده است
"ایکس.ام.ال." برای انجام دادن کاری طراحی نشده است
ممکن است درک آن سخت باشد اما "ایکس.ام.ال." کاری انجام نمی دهد. "ایکس.ام.ال." برای ساخت، ذخیره و ارسال اطلاعات ایجاد شده است. مثال زیر یک سند "ایکس.ام.ال." را نشان می دهد:

```
<note>  
<to>Tove</to>  
<from>Jani</from>  
<heading>Reminder</heading>  
<body>Don't forgout me this weekend</body>  
</note>
```

این یادداشت یک سرصفحه (header) و یک بدنه (body) دارد، همچنین یک گیرنده و یک فرستنده اطلاعات دارد اما هنوز این سند کاری انجام نمی دهد. در واقع اطلاعات ساده پیچیده شده در برچسب های "ایکس.ام.ال." است و فرد باید یک نرم افزاری برای ارسال، دریافت یا نمایش آن بنویسد.

"ایکس.ام.ال." مستقل و گسترش پذیر است

برچسب ها در "ایکس.ام.ال." از پیش تعریف شده نیست و شما باید برچسب های خود را ابداع کنید. برچسب هایی که در نشانه گذاری اسناد "اچ.تی.ام.ال." استفاده شده اند و ساختار اسناد "اچ.تی.ام.ال." از پیش تعریف شده اند. مثل <p>، <h1>

در توسعه آینده وب به احتمال خیلی زیادی از "ایکس.ام.ال." برای توصیف داده استفاده خواهد شد، در حالی که "اچ.تی.ام.ال." برای فرمت و نمایش داده ها استفاده خواهد شد.

سرعت توسعه استانداردهای "ایکس.ام.ال." و سرعت انطباق تهیه کنندگان نرم افزارها با این استانداردها تعجب آور است. نویسنده معتقد است که "ایکس.ام.ال." برای آینده وب به اندازه "اچ.تی.ام.ال." که پایه وب بوده است مهم خواهد بود. "ایکس.ام.ال." مهمترین ابزار برای کاربرد درست داده و انتقال داده خواهد بود.

"ایکس.ام.ال." چگونه استفاده خواهد شد

"ایکس.ام.ال." می تواند داده را از "اچ.تی.ام.ال." جدا کند و در فایل "ایکس.ام.ال." ذخیره نماید. در این شیوه شما می توانید از "اچ.تی.ام.ال." برای نمایش و تنظیمات صفحه استفاده کنید و مطمئن باشید که تغییر در داده ای زیربنایی نیازمند هیچ تغییری در "اچ.تی.ام.ال." شما نیست.

"ایکس.ام.ال." برای مبادله داده استفاده می شود

ب"ایکس.ام.ال."، داده می تواند میان سیستمهای ناسازگار منتقل شود. در دنیای واقعی، سیستمهای کامپیوتری و پایگاههای داده شامل داده هایی در فرمت های ناسازگار(متضاد) هستند و یکی از چالش های زمان بر برای توسعه دهندگان مبادله داده میان چنین سیستمهایی روی اینترنت است. تبدیل داده به "ایکس.ام.ال." به مقدار زیادی این پیچیدگی را کاهش می دهد و داده هایی را ایجاد می کند که بوسیله انواع برنامه های کاربردی متفاوت می تواند خوانده شود.

مبادله اطلاعات مالی در اینترنت

با "ایکس.ام.ال." اطلاعات مالی در اینترنت می تواند مبادله شود. "ایکس.ام.ال." زبان اصلی برای مبادله اطلاعات مالی شرکتها در اینترنت خواهد بود. تعدادی از برنامه های کاربردی جالب شرکت به شرکت در حال توسعه هستند.

"ایکس.ام.ال." برای اشتراک داده می تواند استفاده شود

با "ایکس.ام.ال." فایل های متنی ساده می تواند برای اشتراک داده استفاده شود. "ایکس.ام.ال." یک شیوه مستقل سخت افزاری و نرم افزاری را برای اشتراک داده فراهم می کند که با این کار ایجاد داده ای که برنامه های کاربردی متفاوت براحتمی بتوانند با آن کار کنند را آسان می کند. "ایکس.ام.ال." همچنین گسترش و ارتقاء سیستم به سیستم های عملیاتی جدید، سرورها، برنامه های کاربردی و مرورگرهای جدید تسهیل می نماید.

"ایکس.ام.ال." داده را بسیار پراستفاده می نماید

با "ایکس.ام.ال." داده شما برای کاربران زیادی قابل دسترس خواهد بود. چون "ایکس.ام.ال." از نرم افزار، سخت افزار و برنامه کاربردی مستقل است شما می توانید داده خود را همچنین برای مرورگرهای دیگری غیر از مرورگرهای استاندارد "اچ.تی.ام.ال." قابل دسترس نمایند. برنامه های کاربردی و مشتریان دیگر می توانند به فایل های "ایکس.ام.ال." شما به عنوان منابع داده دسترسی پیدا کنند مثل اینکه آنها به پایگاههای داده دسترسی پیدا می کنند. داده شما ممکن است برای تمام انواع بنگاههای ماشین خوان قابل دسترس باشد و امکان دسترسی افراد نابینا و دیگر افراد معلول به داده های شما را تسهیل می گردد.

"ایکس.ام.ال." برای ایجاد زبان جدید می تواند استفاده شود

"ایکس.ام.ال." مادر زبان (wireless Markup language) WMI است. زبان نشانه گذاری بدون سیم استفاده شده در نشانه گذاری برنامه های کاربردی اینترنت برای ابزارهای دستی مثل تلفن همراه در "ایکس.ام.ال." نوشته شده است. "ایکس.ام.ال." در آینده ممکن است به ما برنامه های کاربردی صفحه گسترده و پردازشگر ورد و پایگاه های داده که می توانند داده های یکدیگر را در فرمت متن ساده بدون هیچ برنامه کاربردی تبدیلی مابینشان بخوانند را ارائه دهد.

ترکیب "ایکس.ام.ال."

ترکیب قوانین "ایکس.ام.ال." بسیار ساده و روشن است. یادگیری و استفاده قوانین بسیار آسان است، برای همین ایجاد نرم افزاری که بتواند "ایکس.ام.ال." را بخواند و بکار برد بسیار آسان است. نمونه یک سند "ایکس.ام.ال.":

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

اولین خط در سند(اعلان "ایکس.ام.ال.") نسخه "ایکس.ام.ال." و کاراکترکد شده مورد استفاده در سند را توصیف می کند. خط بعدی عنصر ریشه از سند را توصیف می کند. چهار خط بعدی چهار عنصر فرزند ریشه را توصیف می کند و خط آخر انتهای عنصر ریشه را توصیف می کند.

همه عناصر(elements) "ایکس.ام.ال." باید یک برچسب (Tag) پایانی داشته باشند

در "ایکس.ام.ال." حذف برچسب پایانی غیرقانونی است. در "اچ.تی.ام.ال." تعدادی از عناصر مجبور نیستند که یک برچسب پایانی داشته باشند.

<p> این یک پاراگراف است

<p> این یک پاراگراف دیگر است

اما در "ایکس.ام.ال." همه عناصر باید یک برچسب پایانی داشته باشند مثل:

<p/> این يك پاراگراف است
<p/> این يك پاراگراف ديگر است

توجه: اعلان "ايكس.ام.ال." برچسب پاياني ندارد. این يك اشتباه نيست. اعلان بخشي از سند "ايكس.ام.ال." نيست . اعلان ، يك عنصر "ايكس.ام.ال." نيست و مي تواند برچسب پاياني را نداشته باشد. برچسب هاي "ايكس.ام.ال." نسبت به بزرگي و كوچكي حروف حساس هستند: برخلاف "اچ.تي.ام.ال." برچسب هاي "ايكس.ام.ال." نسبت به بزرگي و كوچكي حروف حساس هستند. با "ايكس.ام.ال." برچسب <letter> از برچسب <Letter> متفاوت است. بنابراین برچسب هاي آغازی و آخري بايد به يك حالت نوشته شود. مثال:

<Message>this is incorrect</message>

<message>this is correct</message>

همه عناصر "ايكس.ام.ال." بايد بطور مناسب قرار بگيرند با قرار گرفتن نامناسب برچسب ها، هيچ معني در "ايكس.ام.ال." نخواهند داشت. در "اچ.تي.ام.ال." بعضي از عناصر مي توانند بطور نامناسبي با يكدیگر قرار بگيرند مثل: <i>this is</i> <b and italic<i>

همه اسناد "ايكس.ام.ال." بايد يك عنصر ريشه (root) داشته باشند

تمامي اسناد "ايكس.ام.ال." بايد شامل يك جفت برچسب واحد براي تعريف عنصر ريشه باشند. همه عناصر بايد درون این عنصر ريشه باشند. همه عناصر مي توانند عناصر فرزند (child) داشته باشند. عناصر فرعي بايد بطور صحيحی درون والدين قرار بگيرند.

<root>

<child>

<Subchild>...</subchild>

</child>

</root>

ویژگی مقادير بايد همیشه داخل " " قرار بگيرند

در "ايكس.ام.ال." حذف " " اطراف مقادير توصيفي غير قانوني است. در دو سند "ايكس.ام.ال." زیر سند اول نادرست و سند دوم درست است:

<?xml version="1.0" encoding="ISO-8859-1"?>

<note date=12/11/2002>

<to>Tove</to>

<from>Jani</from>

</note>

اشكال سند اول این است که ویژگی (attribute) تاريخ داخل " " قرار نگرفته است.

<?xml version="1.0" encoding="ISO-8859-1"?>

<note date="12/11/2002">

<to>Tove</to>

<from>Jani</from>

</note>

در "ايكس.ام.ال." فضاي خالي حفظ مي شود

برخلاف "اچ.تي.ام.ال." در "ايكس.ام.ال." فضاي خالي در سند کوتاه نمي شود. جمله اي مثل:

Hello my name is Tove,

در "اچ.تي.ام.ال." بصورت زیر نمایش داده خواهد شد:

Hello my name is Tove,

توضیح (comments) در "ايكس.ام.ال."

ترکیب نوشتن توضیح در "ايكس.ام.ال." شبیه "اچ.تي.ام.ال." است.

<!-- This is a comment -->

"ايكس.ام.ال." يك متن ساده بعلاوه تعدادي برچسب داخل علامت <> مي باشد، نرم افزاري که متن ساده را کنترل مي

کند همچنین می تواند "ایکس.ام.ال." را هم کنترل نماید.

عناصر "ایکس.ام.ال."

عناصر "ایکس.ام.ال." گسترش پذیر و مرتبط هستند. این عناصر قواعد نامگذاری ساده ای دارند. اسناد "ایکس.ام.ال." برای انتقال اطلاعات بیشتر می تواند گسترش داده شود. به مثال زیر توجه کنید:

```
<note>
<to>Tove</to>
<from>Jani</from>
<body>Don't forget me this weekend!</body>
</note>
```

با این تصور که ما یک برنامه کاربردی ایجاد کرده ایم که عناصر <body>, <from>, and <to> را از اسناد "ایکس.ام.ال." برای تولید این خروجی بیرون می کشد:

MESSAGE

To: Tove

From: Jani

Don't forget me this weekend

تصور کنید که مولف سند "ایکس.ام.ال." مقداری اطلاعات اضافی به سند اضافه می نماید.

```
<note>
<date>2002-08-01</date>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

برنامه کاربردی باید هنوز قادر به یافتن عناصر <body>, <from>, and <to> در اسناد "ایکس.ام.ال." باشد و همان خروجی را تولید نماید.

عناصر "ایکس.ام.ال." بصورت والدین و فرزندان با هم مرتبط هستند. برای درک "ایکس.ام.ال." شما باید بدانید چگونه ارتباطات مابین عناصر "ایکس.ام.ال." نامگذاری می شود و چگونه مندرجات عناصر توصیف می شود. فرض کنید این توصیف یک کتاب است:

My First XML

Introduction to XML

What is HTML

What is XML

XML Syntax

عناصر باید برچسب پایانی داشته باشند و بطور مناسبی جاگذاری شوند. فرض کنید این سند "ایکس.ام.ال." توصیف یک کتاب است:

```
<book>
<title>My First XML</title>
<prod id="33-657" media="paper"></prod>
<chapter>Introduction to XML
<para>What is HTML</para>
<para>What is XML</para>
</chapter>
<chapter>XML Syntax
<para>Elements must have a closing tag</para>
<para>Elements must be properly nested</para>
</chapter>
```

</book>

book عنصر ریشه است. Title, prod and chapter عناصر فرزند book است. Book والدین Title , prod and chapter است. Title, prod and chapter عناصر خواهر هستند چون آنها والدین یکسانی دارند. عناصر دارای مندرجات (content) هستند: عناصر می توانند مندرجات متفاوتی داشته باشند. یک عنصر "ایکس.ام.ال." از عنصر برچسب آغازی تا عنصر برچسب پایانی را شامل می شود. یک عنصر می تواند عصر مندرجات و مندرجات مرکب، مندرجات ساده یا مندرجات خالی داشته باشد. یک عنصر همچنین می تواند ویژگی (صفت مشخصه) داشته باشد. در مثال بالا: book عنصر مندرجات دارد، برای اینکه آن شامل عناصر دیگر است. capter مندرجات مرکب دارد، برای اینکه هم متن و هم عناصر دیگر دارد. para مندرجات ساده دارد (مندرجات متنی) برای اینکه فقط شامل متن است. prod مندرجات خالی دارد برای اینکه هیچ اطلاعاتی را منتقل نمی کند. در مثال بالا فقط عنصر prod صفت مشخصه دارد. صفت مشخصه که ID نامیده می شود، دارای مقدار "۶۵۷-۳۳" است. صفت مشخصه ای که media نامیده می شود، دارای مقدار "paper" است.

نامگذاری عناصر

عناصر "ایکس.ام.ال." باید این قواعد نامگذاری را دنبال کنند: نامها باید شامل حروف، شماره ها و دیگر کاراکترها باشند. نامها نباید با شماره و کاراکترهای نقطه گذاری شروع شوند. نامها نباید با حروف "ایکس.ام.ال." شروع شوند. نامها نمی توانند فضای خالی داشته باشند.

ویژگی های "ایکس.ام.ال."

عناصر "ایکس.ام.ال." دقیقاً مثل "اچ.تی.ام.ال." می توانند ویژگی هایی را در برچسب آغازی داشته باشند. ویژگی ها برای فراهم کردن اطلاعات اضافی درباره عناصر استفاده می شوند. در مثال "اچ.تی.ام.ال." زیر ویژگی SRC اطلاعات اضافی را درباره عنصر IMG فراهم می کند.

```
<IMG SRC="computer.gif."
```

ویژگی ها (صفات مشخصه) اغلب اطلاعاتی فراهم می کنند که قسمتی از داده نیست. در مثال زیر file type غیرمرتبط با داده است، اما برای نرم افزار مهم است که می خواهد این عنصر را بکار ببرد.

```
<file type="gif">computer.gif</file>
```

شیوه کوتیشن مارك (Quote) بصورت " " یا ' '

مقادیر ویژگی ها (صفات مشخصه) باید همیشه داخل کوتیشن مارك قرار گیرند، هر دو کوتیشن مفرد (' ') و جفت (" ") می تواند استفاده شود. برای جنسیت شخص، برچسب شخص بصورت زیر می تواند نوشته شود:

```
<person sex="female">
```

یا بصورت:

```
<person sex='female'>
```

توجه:

اگر مقدار صفت مشخصه، خود دارای کوتیشن جفت باشد، ضروری است که کوتیشن مفرد بکار برده شود و بالعکس. مثال:

```
<gangster name='George "Shotgun" Ziegler'>
```

```
<gangster name="George 'Shotgun' Ziegler">
```

داده می تواند در عناصر فرزند یا ویژگی ها (عناصر مشخصه) ذخیره شود. مثال:

```
<person sex="female">
```

```
<firstname>Anna</firstname>
```

```
<lastname>Smith</lastname>
```

```
</person>
```

```
<person>
```

```
<sex>female</sex>
```

```
<firstname>Anna</firstname>
```

```
<lastname>Smith</lastname>
```

```
</person>
```

در مثال اول جنسیت يك ويژگي و در مثال دوم يك عنصر فرزند است. هر دو مثال اطلاعات يكساني را فراهم مي کنند. قواعدي براي اينكه چه وقت از ويژگي ها و چه وقت از عناصر فرزند بايد استفاده شود ، وجود ندارد. سه سند "ايكس.ام.ال." زير دقيقا شامل اطلاعات يكساني است:
يك ويژگي date در مثال اول استفاده شده است:

```
<note date="12/11/2002">
```

```
<to>Tove</to>
```

```
<from>Jani</from>
```

```
<heading>Reminder</heading>
```

```
<body>Don't forget me this weekend!</body>
```

```
</note>
```

عنصر date در مثال دوم استفاده شده است:

```
<note>
```

```
<date>12/11/2002</date>
```

```
<to>Tove</to>
```

```
<from>Jani</from>
```

```
<heading>Reminder</heading>
```

```
<body>Don't forget me this weekend!</body>
```

```
</note>
```

يك عنصر date گسترش يافته در مثال سوم استفاده شده است:

```
<note>
```

```
<date>
```

```
<day>12</day>
```

```
<month>11</month>
```

```
<year>2002</year>
```

```
</date>
```

```
<to>Tove</to>
```

```
<from>Jani</from>
```

```
<heading>Reminder</heading>
```

```
<body>Don't forget me this weekend!</body>
```

```
</note>
```

بعضي از مشکلات استفاده از ويژگي ها(صفات مشخصه) عبارتند از:
ويژگي ها نمي توانند شامل چندين مقدار باشند(عناصر فرزند مي توانند)
ويژگي ها به آساني قابل گسترش نيستند(براي تغييرات آينده)
ويژگي ها نمي توانند ساختارها را توصيف کنند(عناصر فرزند مي توانند)
از ويژگي ها فقط براي فراهم نمودن اطلاعاتي كه با داده مرتبط نيست استفاده كنيد.

اعتبار "ايكس.ام.ال."

يك سند "ايكس.ام.ال." با تركيب صحيح، سندي است كه مطابق با قواعد تركيب "ايكس.ام.ال." باشد . سند "ايكس.ام.ال." خوب فرمت شده ، تركيب درست "ايكس.ام.ال." را دارد.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<note>
```

```
<to>Tove</to>
```

```
<from>Jani</from>
```

```
<heading>Reminder</heading>
```

```
<body>Don't forget me this weekend!</body>
```

```
</note>
```

اسناد معتبر "ایکس.ام.ال."

همچنین يك سند معتبر "ایکس.ام.ال." با "دي.دي.تي." (DTD:Document type definition) مطابقت دارد. يك سند معتبر "ایکس.ام.ال." يك سند "ایکس.ام.ال." خوب فرمت شده است که همچنین با قواعد توصیف نوع سند (DTD) مطابقت دارد.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE note SYSTEM "InternalNote.dtd">
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

يك DTD عناصر قانوني يك سند "ایکس.ام.ال." را توصیف مي کند. هدف از يك DTD توصیف قوانین ساختمان يك سند "ایکس.ام.ال." است.

منبع

http://www.w3schools.com/xml/xml_what_is.asp

سایر منابع

هیات مولفان و ویراستاران انتشارات میکروسافت. (۱۳۷۹). فرهنگ تشریحی اصطلاحات کامپیوتری میکروسافت. مترجم فرهاد قلی زاده نوری. تهران: آذریاد.