

نشانه‌ها: پیدایش

مروری بر ریشه‌ها؛ روند پیدایش و رواج XML

«احمد شریف پور»



برخی پیشرفت‌ها و فناوری‌ها در هنگام ارائه چنان شور و هیجانی می‌آفرینند که گویی آمده‌اند تا آینده دنیا را متحول کرده و نژاد بشر را به آینده‌ای بهتر رهنمون کنند. بخش زیادی از این تب و هیجان ساخته و پرداخته مشاوران تبلیغاتی و صاحبان شرکت‌ها است که در این کار هدفی جز افزایش درآمد شرکت متبوعشان را دنبال نمی‌کنند و ناگفته پیداست در این میان مواردی که موفق به عملی کردن وعده‌های خود می‌شوند، انگشت شمار خواهند بود. اما گاهی اکتشافات و اختراعاتی وجود دارند که بی‌هیچ هیاهویی ظهور می‌کنند و رفته رفته تا تبدیل شدن به شالوده فعالیت‌های روزمره ما انسان‌ها پیش می‌روند و ما بی‌آن‌که بخواهیم یا حتی از وجود چنین فناوری‌هایی آگاه باشیم، به قابلیت‌های آن‌ها وابسته می‌شویم. XML از همین گروه است. اگر نمی‌توانید با رابط زیبا و سریع gmail یا قالب‌های مرسوم اسناد بسته آفیس مایکروسافت خداحافظی کنید باید بدانید که شما نیز به قابلیت‌های فراوان XML تکیه و عادت کرده‌اید.

اما به راستی XML چیست؟ در تعریف مشخصات XML نسخه اول که توسط W3C منتشر شده، چنین آمده است: «زبان نشانه‌گذاری قابل توسعه یا XML مجموعه قوانینی است که برای کدگذاری اسناد در قالبی قابل درک توسط ماشین مورد استفاده قرار می‌گیرد.» این قوانین و دستورالعمل‌ها در قالب استانداردهای باز فراهم شده‌اند تا استفاده گسترده از آن‌ها در کاربردهای متنوع و به‌ویژه کاربردهای اینترنتی تضمین شود. در طراحی استاندارد XML تمرکز اصلی بر سادگی، عمومیت (قابلیت تعمیم) و قابلیت استفاده در اینترنت بوده است. به همین دلیل XML قالب داده‌ای متنی است که با پشتیبانی قوی از یونی‌کد، امکان استفاده را برای تمام زبان‌های دنیا فراهم می‌آورد. اگرچه در ابتدا تمرکز اصلی XML روی اسناد بود، اما در حال حاضر برای ایجاد انواع مختلف ساختارهای داده به‌خصوص در سرویس‌های وب استفاده می‌شود. تاکنون صدها زبان و رابط برنامه‌نویسی پیشرفته (API) برای کار با داده‌های XML طراحی شده‌اند که در قسمت انقلاب XML برخی از آن‌ها را بررسی خواهیم کرد.

XML

پیش از XML

در شماره ۳۹ مجله شبکه (کتاب سال ۱۳۸۲) مقاله‌ای با عنوان «XML، زبانی برای تمام کاربردها» منتشر شده است که امروز نیز با گذشت بیش از هشت سال از تاریخ انتشار، می‌تواند به عنوان مرجعی کامل در زمینه تاریخ و نحوه شکل‌گیری زبان XML مورد استفاده قرار بگیرد. اما در کنار توصیه به مطالعه آن مقاله، در این بخش تاریخ و روند پیدایش XML به صورت خلاصه آورده می‌شود.

در اواخر دهه ۶۰ میلادی، زمانی که تقریباً صنعت IT پا به بیست سالگی گذاشته بود، رشد قدرت پردازش سیستم‌ها و رواج بیش از پیش استفاده از کامپیوتر در کسب و کارها، کاربردهای کامپیوتر را از محاسبات عددی به زمینه‌های دیگری از جمله پردازش متن نیز گسترش داد. در چنین زمانی، برنامه‌نویسان برای ساده کردن فرآیند پردازش متون، علامت‌ها و نشانه‌هایی را در متن قرار می‌دادند تا هنگام استخراج اطلاعات در این اسناد با مشکلات کمتری مواجه شوند. زبان‌های نشانه‌گذاری در واقع نتیجه تعریف رسمی فرهنگ واژگان و قوانین استفاده از این نشانه‌ها برای جلوگیری از تداخل آن‌ها با متن هستند. زبان‌های نشانه‌گذاری اولیه کاملاً تک منظوره بودند. زبان‌هایی مختص انتقال اطلاعات مربوط به داده‌های آب‌وهوایی یا مختص قالب‌بندی متون برای چاپ و... توسعه می‌یافتند که اغلب اختصاصی بوده و تنها به همان یک کاربرد محدود می‌شدند. به عبارت دیگر، حتی زبانی که برای قالب‌بندی چاپ یک سند به کار می‌رفت برای نمایش همان سند در صفحه نمایشگر قابل استفاده نبود.

در سال ۱۹۶۷ ویلیام تونیکلیف برای نخستین بار به طور رسمی این ایده را مطرح کرد که اطلاعات محتوای سند باید از اطلاعات نشانه‌گذاری تفکیک شود. همزمان فردی به نام استنلی رایس برای توصیف ساختار یک کتاب از روش مشابهی استفاده کرد. حاصل همکاری این دو به همراه عده‌ای دیگر در قالب کمیته تولید کد انجمن ارتباطات گرافیکی (سرنام Graphical Communication Association Gencode)

نسل نوینی از زبان‌های نشانه‌گذاری را به وجود آورد. این نسل جدید زبان‌ها، دو دستاورد اصلی با خود به همراه داشت: نخست این که هم برای انسان و هم برای ماشین قابل فهم بودند و دیگر این که با حرکت به سمت زبان‌های نشانه‌گذاری توصیفی، دیگر به کاربرد خاصی محدود نمی‌شدند، سپس در سال ۱۹۶۹ چارلز گلدفارب که تازه به آی.بی.ام پیوسته بود، با الهام از کارهای رایس و تونیکلیف و به کمک دو تن از همکارانش به نام‌های موشر و لوری زبان نشانه‌گذاری جدیدی را به وجود آوردند که در آی.بی.ام آن را GML می‌نامیدند (این نام که در آن زمان و براساس حروف اول نام‌های گلدفارب، موشر و لوری انتخاب شده بود، اکنون برای زبان نشانه‌گذاری جغرافیایی یا Geography Markup Language به کار می‌رود). مؤسسه

ملی استاندارد آمریکا در سال ۱۹۸۰ و براساس GML، زبان SGML را تدوین کرد که ارائه رسمی آن تا سال ۱۹۸۶ به طول انجامید. زبان SGML یک فرازبان یا Meta Language است که از آن برای توصیف زبان‌های نشانه‌گذاری دیگر استفاده می‌شود. این زبان تا سال ۱۹۹۰ به طور معمول تنها توسط نهادهای نظامی و بخش‌های دولتی استفاده می‌شد. اما در سال ۱۹۹۰ و پس از گسترش آرپانت و رونق استفاده از استاندارد TCP/IP تیم برنرز لی با تکیه بر مبانی SGML فراگیرترین زبان نشانه‌گذاری دنیا، یعنی HTML را پایه‌گذاری کرد.

تولد XML

نخستین ناشران رسانه‌های دیجیتال در اواخر دهه ۱۹۸۰ میلادی و بسیار پیش‌تر از رواج اینترنت، به قدرت SGML در نمایش داینامیک اطلاعات پی برده بودند. در اواسط دهه ۱۹۹۰ برخی کاربران حرفه‌ای SGML، در کار با وب جهان گستر (World Wide Web) نیز به مهارت دست یافتند و به این نتیجه رسیدند که SGML راه‌حل‌هایی را برای برخی مشکلاتی که وب در آینده و در ادامه توسعه با آن روبه‌رو خواهد شد، ارائه کرده است. دن کانولی (Dan Connolly) در سال ۱۹۹۵ و هنگام پیوستن به کنسرسیوم وب یا W3C، زبان SGML را به فهرست فعالیت‌های این کنسرسیوم افزود. کار روی این پروژه به طور رسمی زمانی آغاز شد که جون بوساک (Jon Bosak) از مهندسان سان مایکروسیستمز، اساس نامه‌ای را برای ادامه کار روی این زبان طرح کرد و به جذب نیرو پرداخت. بوساک با جماعت کوچکی که هم در SGML و هم در وب مهارت داشتند، ارتباط داشت. در نهایت، XML به وسیله گروهی یازده نفره که توسط تیمی تقریباً ۱۵۰ نفره از طرفداران (Interest group) به معنی کسانی که دارای علائق مشترک در زمینه پیش‌برد یکی از علوم هستند) حمایت می‌شدند، تألیف شد. مذاکرات و مباحثات مرتبط با پروژه از طریق ایمیل بین اعضا انجام می‌شد و موارد مختلف به اتفاق آرا یا با مراجعه به رای اکثریت حل و فصل می‌شدند. در چهارم دسامبر ۱۹۹۷، سند تصمیمات طراحی و دلایل مرتبط با آن‌ها توسط مایکل اسپربرگ مک کوین (Michael Sperberg-McQueen) ارائه شد. جیمز کلارک (James Clark) که در واقع رهبر تکنیکی گروه بود، علاوه بر فرم نوشتاری المان خالی (empty-element) به فرم `<empty/>` نام XML را نیز برای این مولود تازه پیشنهاد کرد. از دیگر نام‌های پیشنهادی می‌توان به MAGMA (سرنام Minimal Architecture for Generalized Markup Applications)، SLIM (سرنام Structured Language for Internet Markup) و MGML (سرنام Minimal Generalized Markup Language) اشاره کرد. تدوین‌کنندگان اصلی مشخصات زبان در کنار جیمز کلارک، اغلب تیم بری (Tim Bray) و

گروه کاری XML هیچ‌گاه با یکدیگر ملاقات نکردند و طراحی این زبان به کمک ترکیبی از ایمیل و کنفرانس‌های راه‌دور هفتگی تکمیل شد.



رمز موفقیت این زبان سادگی آن بود. XML زبان تازه ای نبود. روشی برای ساخت زبان های تازه بود. زبانی که از ابتدا برای کار با اسناد ساخته شده بود، به تدریج جای خود را در کاربردهایی نظیر انتقال اینترنتی اطلاعات، پایگاه های داده و... نیز باز کرد.



اسپربرگ مک کویین بودند. در اواسط کار پروژه، بری قرارداد مشاوره ای را با نت اسکپ پذیرفت که باعث بروز مخالفت های شدیدی از سوی مایکروسافت شد و به طور موقت از بری خواسته شد که از مسئولیت خود کناره گیری کند. این امر مشاجره های شدیدی را در گروه توسعه دهنده این زبان برانگیخت که در نهایت با انتصاب یکی از مهندسان مایکروسافت به نام جین پاولی (Jean Paoli) به سمت تدوین کننده سوم این مناقشات خاتمه یافت.

گروه کاری XML هیچ گاه با یکدیگر ملاقات نکردند و طراحی این زبان به کمک ترکیبی از ایمیل و کنفرانس های راه دور هفتگی تکمیل شد. بیشتر تصمیمات اساسی و پایه مربوط به طراحی XML بین جولای تا نوامبر ۱۹۹۶ و در طی بیست هفته کاری بی وقفه و فشرده گرفته شدند و دست آخر در نوامبر ۱۹۹۶ نخستین نسخه قابل استفاده پیش نویس مشخصات XML منتشر شد. روند طراحی و توسعه این زبان در سال ۱۹۹۷ نیز ادامه یافت و در نهایت نتیجه تمام این تلاش ها در دهم فوریه ۱۹۹۸ به صورت نسخه XML 1.0 توسط W3C ارائه شد.

در حال حاضر، دو نسخه جاری از XML مورد استفاده قرار می گیرد. نسخه اول همان XML 1.0 است که در سال ۱۹۹۸ ارائه شد. از آن زمان تاکنون بازنگری های جزئی را پشت سر گذاشته، اما شماره نسخه آن تغییر نکرده است و آخرین ویرایش، یعنی ویرایش پنجم آن در ۲۶ نوامبر

۲۰۰۸ ارائه شده است. این نسخه هم اکنون هم به صورت گسترده مورد استفاده قرار می گیرد و زبان پیشنهادی برای کاربردهای معمول است.

نسخه دیگر XML 1.1 است که برای نخستین بار چهارم فوریه ۲۰۰۴ و درست همزمان با ویرایش سوم XML 1.0 ارائه شده و ویرایش دوم آن نیز در شانزدهم آگوست ۲۰۰۶ ارائه شده است. این نسخه حاوی امکاناتی است که انتظار می رود استفاده از XML را در برخی کاربردها ساده تر کند. این زبان به اندازه نسخه 1.0 مورد استفاده قرار نمی گیرد و تنها به کسانی پیشنهاد می شود که واقعا به امکانات خاص آن احتیاج دارند.

انقلاب XML

اما زبانی که توسط گروهی یازده نفره و بدون تعلق به غول های نرم افزار دنیا توسعه یافته بود، حرف های زیادی برای گفتن داشت. رمز موفقیت این زبان سادگی آن بود. XML زبان تازه ای نبود. روشی برای ساخت زبان های تازه بود. زبانی که از ابتدا برای کار با اسناد ساخته شده بود، به تدریج جای خود را در کاربردهایی نظیر انتقال اینترنتی اطلاعات، پایگاه های داده و... نیز باز کرد. اگر چه در کادر «گستره های کاربرد XML» می توانید زمینه های اصلی استفاده از XML را ببینید، اما بد نیست به نمونه های خاص کاربرد این زبان هم نگاهی داشته باشیم تا اهمیت و تأثیر این زبان بر زندگی روزمره ما آشکار تر شود.

گستره های استفاده از XML

توانمندی اصلی XML در سادگی آن نهفته است. XML به تنهایی هیچ کاری انجام نخواهد داد. سندها را زیباتر یا حتی خواناتر نخواهد کرد، بلکه تنها مجموعه مقرراتی را وضع می کند که به کمک آن می توان اسناد را نشانه گذاری کرد. خواه اسناد متنی ساده باشند یا اسناد مربوط به یک پایگاه داده یا حتی سایتی نظیر ویکی پدیا. در عین حال XML در سه زمینه بیش از سایر زمینه ها کاربرد دارد:

- توسعه وب: بسیاری از خدمات آنلاین و برنامه هایی که بر بستر اینترنت اجرا می شوند، از قالب XML برای تبادل اطلاعات بین سرورها و کلاینت ها استفاده می کنند. XHTML در واقع نسخه بهینه شده HTML 4.0 است و بسیاری از سایت های مبتنی بر HTML 4.0 در واقع از نوعی XHTML نامعتبر استفاده می کنند. مزیت دیگر XML این است که به کمک XSLT می توان اسناد مبتنی بر XML را به HTML تبدیل کرد. خوراک های Atom، RSS، ایزارک های نمایش اطلاعات آب و هوایی و بسیاری موارد مشابه، از دیگر کاربردهای XML در توسعه وب محسوب می شوند و به این ترتیب، وب به نوعی دربرگیرنده

گسترده ترین استفاده از فناوری XML است. راستی حرف X در انتهای کلمه AJAX را فراموش نکنید.

- کار با اسناد: این جا همان جایی است که XML کار خود را آغاز کرد. XML از ابتدا قالب بندی اسناد را هدف گرفته بود و به نظر می رسد که در رسیدن به هدف خود کاملاً موفق بوده است. بسیاری از شرکت ها برای تولید اسناد داخلی خود به XML روی آورده اند. مشهورترین پلتفرم در این زمینه DocBook است که زبان نشانه گذاری است که از ابتدا برای تولید متون تخصصی کامپیوتر طراحی شده، متون تهیه شده در این قالب به راحتی و بدون هیچ تغییری در قالب های epub، pdf، html و... قابل انتشار هستند.

مهم ترین مزیت XML این است که هم محتویات و هم نشانه ها، علاوه بر ماشین ها برای انسان ها نیز به سادگی قابل درک است. به این ترتیب، XML برای تمام انواع اسناد، از نشر روزمیزی خانگی گرفته تا محصولات تجاری بازار قابل استفاده است. حرف X در انتهای docx قالب جدید اسناد برنامه مایکروسافت شاهی بر این ماجرا است.

- توسعه پایگاه های داده: پایگاه های داده طبیعی ترین محل استفاده از XML است، زیرا فلسفه وجودی XML کار با داده ها است. برخلاف سیستم کار با اسناد که در آن کدهای XML هم برای ماشین و هم برای انسان قابل درک باشد. به عبارت ساده تر داده ها به شکلی نوشته می شوند که به ماشین ها امکان خواندن آن ها را بدهد و برای پایگاه داده قابل دسترس شان کند، پایگاه های داده به دو شکل از XML استفاده می کنند. برخی قابلیت کار با XML را دارند و برخی از پایه بر مبنای XML بنا شده اند. اوراکل غول پایگاه های داده امکان استفاده از XML را توسط رابط XQJ API فراهم می کند.

اما استفاده از XML به همین سه کاربرد محدود نمی شود. شرکت های زیادی بنا به دلایل متعدد، از نیاز به انتقال مداوم داده در اینترنت تا پردازش داده های متنی و... به استفاده از XML روی آورده اند. زمینه های کاری این شرکت ها از مهندسی هوافضا و بانکداری و تبلیغات گرفته تا آموزش و صنعت نفت متغیر است. (www.stcwc.org/galley/0107/xml.tm)

اسناد

XML تولید اسناد، ویرایش و به اشتراک گذاری آن‌ها و انتقال آن‌ها بین برنامه‌ها و پلتفرم‌های مختلف را به شدت دگرگون کرده است. در دنیای نرم‌افزار، دو بسته برنامه‌های اداری به نسبت پرکاربردتر از سایرین هستند. در رده نخست آفیس مایکروسافت قرار دارد که از نسخه ۲۰۰۷ به بعد به استفاده از پیاده‌سازی خاصی از XML در فرمت فایل تمام نرم‌افزارهای خود روی آورده است. فایل‌های ایجاد شده توسط نسخه‌های ۲۰۰۷ و ۲۰۱۰ آفیس مایکروسافت گرچه همانند نسخه‌های قبلی فایل‌هایی دودویی به نظر می‌رسند، اما در واقع از تعدادی فایل با فرمت XML و تصاویر و فایل‌های استایل‌گذاری و کنترل متناظر با آن‌ها تشکیل شده‌اند که در قالب یک فایل zip فشرده شده است. بسته دوم یا این آفیس نیز تقریباً از بدو تولد با XML سازگار بوده و فایل‌های خود را با همین فرمت مبتنی بر XML که اکنون دیگر OpenXML Formats نامیده می‌شود، ذخیره می‌کرده است (به سادگی و با باز کردن یک فایل آفیس مایکروسافت یا این آفیس در یک برنامه مدیریت فایل‌های فشرده نظیر WinRAR یا 7-Zip می‌توانید این موضوع را کنترل کنید). آزاد و غیرانحصاری بودن این فرمت، امکان تبادل ساده اطلاعات با سایر نرم‌افزارها به واسطه استاندارد باز XML، پایداری و کارایی بالا و امنیت اطلاعات به واسطه امکان کنترل و بررسی تمام محتویات و کدهای سند و حجم کم فایل حاصل به واسطه استفاده از فرمت متنی و فشرده‌سازی zip، از مزایای متعدد این شیوه ذخیره‌سازی اسناد هستند.

Atom و RSS

هر کاربر کم‌وبیش حرفه‌ای وب حتماً با مربع نارنجی رنگی که نشانه خوراک‌های وب یا Web Feeds است، آشنایی دارد. به‌طور عملی کسانی که به‌طور مداوم با اینترنت و محتوای اینترنتی سروکار دارند، نمی‌توانند بدون استفاده از خوراک‌ها فعالیت‌های معمول روزانه خود را به انجام برسانند. خوراک‌ها امکان اطلاع از محتویات و تازه‌های سایت‌ها یا وبلاگ‌هایی را که به‌طور مداوم در حال تغییر هستند، فراهم می‌آورد.

این خوراک‌ها توسط نرم‌افزارهای کوچکی که به اصطلاح خوراک‌خوان یا Feed Reader نامیده می‌شوند، پردازش شده و سرتیتیر یا محتوای خبرها و تغییرات و تازه‌ها را از خوراک‌ها استخراج می‌کنند. خوراک‌های وب در دو قالب اصلی RSS (سرنام Really Simple Syndication) و Atom (قالبی جدیدتر که برای برطرف مشکلات RSS توسط بن تروت ایجاد شد) ارائه می‌شوند. هر دو این قالب‌ها از زبان XML برای ایجاد ساختار داده‌ای ساده با قابلیت انتشار از طریق اینترنت استفاده می‌کنند. مبتنی بودن XML به متن، باز بودن استاندارد این زبان، سادگی انتقال از طریق اینترنت و رایگان بودن آن، امکان توسعه خوراک‌خوان‌های متعدد و فراگیر شدن استفاده از خوراک‌ها را فراهم آورده است.

AJAX

پس از اسناد، XML محتوای تعاملی وب را بیش از هر چیزی دیگری تحت تأثیر قرار داده است. این کار از طریق AJAX (سرنام Asynchronous JavaScript and XML) صورت پذیرفته است. AJAX در واقع مجموعه‌ای از متدهای طراحی وب در سمت کاربر است که امکان دریافت اطلاعات توسط برنامه‌های وب بدون تأثیر روی محتوای مشاهده شده یا ماهیت صفحه وب فعلی را فراهم می‌آورد. نتیجه این سیستم، برنامه‌های وبی است که رفتاری بسیار شبیه برنامه‌های دستکتاب داشته و امکان تعامل سریع و روان با کاربر را فراهم می‌کنند. رابط‌های کاربری Gmail و YahooMail از جمله نمونه‌های بارز AJAX هستند. تمام این تبادلات داده‌ای و این تعامل ساده مدیون ویژگی‌های منحصر به فرد XML است که امکان انتقال ساده متن و سایر عناصر را از طریق اینترنت فراهم آورده است.

GUI

تعریف عناصر شکل دهنده ظاهر یک نرم‌افزار، تعیین موقعیت، شکل و روابط بین آن‌ها به عرصه دیگری برای خودنمایی XML تبدیل شده است. با توصیف ضمنی عناصر و تعریف روابط بین آن‌ها، امکان طراحی ظاهری واحد و یکنواخت برای نرم‌افزار در پلتفرم‌ها و کاربردهای مختلف فراهم خواهد شد. از سوی دیگر، با جدا شدن کدهای عملیاتی برنامه از پوسته ظاهری آن‌ها، امکان تغییر ظاهر برنامه در حین اجرا یا تغییر عملکرد برنامه بدون تغییر ظواهر آن فراهم می‌شود. WPF (سرنام Windows Presentation Foundation) که شرکت مایکروسافت از نسخه سوم NET Framework شروع به ارائه آن کرده است، بر پایه XAML یکی از مشتقات XML بنا شده است. XUL (سرنام Xml User interface Language) که توسط بنیاد موزیلا توسعه یافته و برای تولید رابط بصری محصولاتی نظیر فایرفاکس مورد استفاده قرار می‌گیرد، نیز از دیگر رابط‌های کاربری مبتنی بر XML است.

اما کار XML هنوز به پایان نرسیده است. این زبان کم‌وبیش نوپا، رفته رفته در حال گسترش است. قالب‌های گرافیکی جدیدی بر اساس آن شکل می‌گیرد. نحوه نمایش معادلات و تصاویر را در وب بهبود می‌بخشد. در زمینه صوت و موسیقی به کار گرفته می‌شود و اداره پایگاه‌های داده را بر عهده می‌گیرد. این فناوری نه تنها خود انقلابی عظیم در زیربنای عرصه IT ایجاد کرده است، بلکه منشأ بسیاری از زبان‌ها و تحولاتی نیز شده است که هر یک به تنهایی دستاوردی بزرگ در عرصه‌ای خاص محسوب می‌شوند. تنها کافی است به آدرس http://en.wikipedia.org/wiki/List_of_XML_markup_languages مراجعه کنید تا با مشاهده فهرست زبان‌های مبتنی بر XML، یقین پیدا کنید که در آینده بیشتر و بیشتر به این زبان زیبا وابسته خواهیم شد، حتی اگر برخلاف بسیاری از همتایانش، بدون جار و جنجال و هیاهو گام بردارد.



تعریف عناصر شکل

دهنده ظاهر یک

نرم‌افزار، تعیین

موقعیت، شکل و

روابط بین آن‌ها

به عرصه دیگری

برای خودنمایی

XML تبدیل شده

است. با توصیف

ضمنی عناصر و

تعریف روابط

بین آن‌ها، امکان

طراحی ظاهری

واحد و یکنواخت

برای نرم‌افزار

در پلتفرم‌ها و

کاربردهای مختلف

فراهم خواهد شد.

