

پایگاه‌های اطلاعاتی ایکس ام ال و کاربرد آن در کتابخانه‌های دیجیتال

مهدی علیپورحافظی^۱

چکیده

ایکس ام ال، زبان توصیفی مورد حمایت کنسرسیوم وب گسترده جهانی است که در نوامبر ۱۹۹۶ به جامعه کاربران معرفی شد. این زبان، شکل محدود شده اس جی ام ال است که با هدف انتشار، ذخیره و مبادله داده‌های ساخت یافته در اینترنت و اینترنت به وجود آمده است و به طراحان، امکان تعریف و استفاده از برجسب‌های خود را می‌دهد. امکان ذخیره اطلاعات با فرمت ایکس ام ال در پایگاه‌های اطلاعاتی، استفاده از آن برای مبادله داده‌ها بین نرم‌افزارهای نامتجانس و انتشار اطلاعات در وب در زمره مباحث مطرح در محافل اطلاعاتی در دهه حاضر می‌باشند. مقاله حاضر جهت روشن شدن زوایای مطالب فوق برای صاحب‌نظران، پژوهشگران و کتابداران علاقه‌مند به این حوزه تدوین شده است.

کلیدواژه‌ها

ایکس ام ال، پایگاه‌های اطلاعاتی، کتابخانه‌های دیجیتال

مقدمه

در شبکه‌ها و اینترنت کردند. در ادامه، به فکر ارائه اطلاعات تمام متن در شبکه جهان گستر و ایجاد کتابخانه‌های دیجیتالی افتادند. استفاده از رایانه در کتابخانه‌ها و استفاده از نرم‌افزارهای کتابخانه‌ای، این مراکز را متوجه ایجاد بانک‌های اطلاعاتی منابع اطلاعاتی خود کردند. در ابتدا، با توجه به نوپا بودن فن‌آوری‌های مربوط به پایگاه‌های اطلاعاتی، طبیعتاً پایگاه‌های اطلاعاتی مورد

کتابخانه‌ها به عنوان مجموعه‌هایی از منابع اطلاعاتی، نقش عمده‌ای در نگهداری و اشاعه دانش جوامع دارند. با گذشت زمان و ظهور فن‌آوری‌های رایانه‌ای، کتابخانه‌ها اولین مراکزی بودند که اقدام به استفاده از آن در راستای خودکارسازی فعالیت‌های خود کردند. در طول دوره گذار، کتابخانه‌ها اقدام به ارائه اطلاعات کتاب‌شناختی خود

۱. دانشجوی دکتری کتابداری و اطلاع‌رسانی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران meh.hafezi@gmail.com

استفاده در کتابخانه‌ها نیز بسیار ساده بودند. از دهه ۱۹۷۰ به بعد که فن‌آوری مربوط به مدیریت پایگاه‌های اطلاعاتی^۲ پا به عرصه وجود نهاد، به تدریج، پایگاه‌های اطلاعاتی پیچیده‌تر شده و قابلیت‌های بیشتر و بهتری یافتند. در ابتدا، پایگاه‌های اطلاعاتی به صورت ترتیبی، داده‌ها را در خود حفظ می‌کردند. در ادامه، فن‌آوری‌های مربوط به ذخیره داده‌ها به صورت رابطه‌ای، شیء‌گرا و ایکس ام ال مطرح شدند. کتابخانه‌ها به عنوان بستر مناسب جهت استفاده از این فن‌آوری‌ها در طول زمان و در دوره‌های خاص اقدام به استفاده از آنها کردند. ایکس ام ال در ابتدا به عنوان فرمتی برای مبادله داده‌ها مطرح شد و با توجه به سلسله‌مراتبی بودن آن بسیار مورد توجه کتابخانه‌ها و سازمان‌های دیگر قرار گرفت. اما امروزه با توجه به استفاده گسترده از ایکس ام ال و نیاز به ذخیره آن در پایگاه‌های اطلاعاتی، بحث مربوط به پایگاه‌های اطلاعاتی ایکس ام ال مطرح شده است. پایگاه‌های اطلاعاتی ایکس ام ال، نحوه استفاده از آن و کاربرد آن در ارتباط با پایگاه‌های اطلاعاتی دیگر، از جمله مباحثی است که در متن حاضر مورد بررسی قرار گرفته‌اند.

پایگاه داده و نسل‌های مختلف آن

پایگاه داده مرکب از چهار عنصر مهم داده، رابطه، محدودیت و ساختار^۳ است. «داده»، عبارت از موجودیت‌های منطقی

ذخیره شده به صورت دو دویی در رایانه است. «رابطه»، ارتباط بین داده‌های به هم مرتبط را نشان می‌دهد. «محدودیت»، گزاره‌ای است که حالات درست پایگاه داده را تعریف می‌کند و «ساختار»، سازمان‌دهی و روابط بین آنها را در پایگاه داده توصیف می‌کند (۱۸: ۵). ساختار، جنبه‌های فیزیکی ذخیره داده را برای استفاده از اجزای سیستم مدیریت پایگاه داده و امنیت برنامه‌های کاربردی توصیف می‌کند. «ساختار داخلی»، چگونگی و محل ذخیره داده‌ها را در رسانه فیزیکی ذخیره‌سازی مشخص می‌کند و «مدل ساختار مفهومی»، ساختار داده ذخیره شده را در مدل‌های مختلف (مثل رابطه‌ای^۴ یا شیء‌گرا^۵) توصیف می‌کند. «ساختار خارجی»، دیدگاه‌های کاربران (مدیر پایگاه، کتابدار و کاربر نهایی) را مشخص می‌کند. نهایت این که سیستم مدیریت پایگاه داده خدماتی را برای دسترسی به پایگاه داده در صورت ذخیره، صحت و دقت داده‌ها فراهم می‌کند. بنابراین، پایگاه داده به مجموعه داده‌هایی گفته می‌شود که روابط بین آنها مشخص و به شکل منطقی سازمان‌دهی شده و امکان دستیابی به اطلاعات موجود در آنها برای کاربران فراهم شده باشد.

پردازش داده‌ها از دهه ۱۹۵۰ تاکنون، فزاینده و نشیب بسیاری داشته است. در اوایل کار، کاربران به طور مستقیم با محیط فیزیکی یا سخت‌افزار رایانه تماس داشتند و داده‌ها را روی آنها ذخیره و بازیابی می‌کردند. با گذشت

2. Data Base Management Systems (DBMS)

3. Schema

4. Relational

5. Object-Oriented (OO)

زمان، نرم افزارهایی به نام شیوه دستیابی^۶ و نام‌های مشابه به وجود آمدند که رابط بین کاربر و رایانه بودند. کاربران، با استفاده از این نرم افزارها می‌توانستند داده‌ها را با فرم‌های متفاوتی روی رسانه‌های ذخیره نگهداری کنند. این نرم افزارها، مدیریت، ذخیره و بازیابی داده‌ها را برعهده داشتند. کاربر، شکل قرار گرفتن داده‌ها در کنار یکدیگر و نیازهای دستیابی به آنها را بیان می‌کرد و نرم‌افزار، آن را اجرا می‌کرد (۳: ۲).

هر چند وجود این نرم افزارها کار را آسان می‌نمود و پیشرفت قابل توجهی به حساب می‌آمد، با این همه، هنوز مشکلات زیادی خودنمایی می‌کرد. مثلاً برای هر بار استفاده باید داده‌های مربوطه را جداگانه آماده کرد و اشتراک داده‌ها در سطح قابل قبولی ممکن نبود. همچنین تأمین امنیت داده‌ها و حفاظت از آنها مشکل بود. این گونه مشکلات، باعث بروز پدیده‌ای به نام انقلاب بانک اطلاعاتی در اوایل دهه ۱۹۷۰ گردید. اصلی‌ترین تفاوت این روش با روش‌های قبلی، وجود نظام مدیریت بانک اطلاعات بود که هرگونه دستیابی به داده‌ها باید از طریق آن انجام شود.

مدل‌های بانک اطلاعاتی که مقبولیت عام یافته‌اند به سه دسته زیر تقسیم می‌شوند:

الف. مدل‌های قدیمی. مهم‌ترین این مدل‌ها، مدل سلسله‌مراتبی^۷ و مدل شبکه‌ای^۸ هستند. این دو مدل، امروزه مورد استفاده قرار نمی‌گیرند. هر دو مدل بر مبنای رکورد

پایه‌گذاری شده‌اند، یعنی داده‌ها در آنها به صورت رکوردهای مرتبط با یکدیگر سازمان‌دهی می‌شوند. در مدل سلسله‌مراتبی از درخت و در مدل شبکه‌ای از گراف برای سازمان‌دهی داده‌ها استفاده می‌شود. در ساختار سلسله‌مراتبی، هر بخش از اطلاعات زیرشاخه‌ای از بخش بالاتر است. بنابراین، اگر یک بانک اطلاعاتی شامل اطلاعات مربوط به کارکنان است، به زیرشاخه و بخش‌های مختلف، شامل کارکنان آن در سطوح مختلف، تقسیم می‌گردد. این روش، امکان بازیابی سریع اطلاعات با جزئیات مربوط به بخش‌های کاری را فراهم می‌کند. اگر جزئیات مربوطه مشخص نباشد، یک ساختار سلسله‌مراتبی می‌تواند جستجوی اطلاعات را مشکل کند (۵: ۱۹۱؛ ۳: ۷) ولی در ساختار شبکه‌ای بانک‌ها بر اساس ارتباط آشکار بین اجزای مرتبط با اطلاعات شکل می‌گیرند. مشهورترین الگوی شبکه‌ای به کداسیل^۹ معروف است. در یک الگوی شبکه‌ای ارتباط مستقیمی بین اجزای اطلاعات در سطوح مختلف برقرار می‌سازند. یک مدل شبکه‌ای به ارتباطات زیادی بین اجزای اطلاعات نیاز دارد و فضای زیادی را نیز اشغال می‌کند (۵: ۱۹۲؛ ۳: ۷؛ ۱۸: ۱۸۷).

ب. مدل سنتی (مدل رابطه‌ای). هرچند این مدل قدمتی چندین ساله دارد، اما هنوز هم در دنیای بانک اطلاعات حرف اول را می‌زند. موفقیت این مدل تا حدی است که امروزه اکثر کاربران، مفهوم بانک اطلاعات را در قالب

6. Access method

7. Hierarchical model

8. Network model

9. Conference on Database Systems Languages

(CODASYL)

این مدل می‌شناسند. در مدل رابطه‌ای، داده‌ها به صورت رکوردهای مرتبط در قالب سطرها و ستون‌ها سازمان‌دهی و بانک اطلاعات به صورت مجموعه‌ای از رابطه‌ها طراحی می‌شوند. مهم‌ترین عامل موفقیت آن، سادگی و پشتوانهٔ تئوریک بسیار قوی آن است (۴): ۳۳-۳۴؛ ۳۵؛ ۳۶۷؛ ۱۶؛ ۲۰-۲۱).

ج. مدل‌های جدید. مدل رابطه‌ای با همهٔ عظمتش برای کاربردهای تجاری و سنتی بانک اطلاعات ساخته شده و پاسخ‌گوی بسیاری از نیازهای کتابخانه‌های امروزی نمی‌باشد. مثلاً، بانک اطلاعات حاوی تصویر، صوت، و متن را به سادگی نمی‌توان در قالب رکورد و رابطه ریخت و در مدل رابطه‌ای سازمان داد. در دههٔ اخیر، کوشش‌های بسیار زیادی برای پاسخ‌گویی به این نیازها به عمل آمده و مدل‌های متنوعی پیشنهاد و پیاده‌سازی شده است. از مهم‌ترین این مدل‌ها می‌توان مدل شیء‌گرا، مدل تابعی^{۱۱}، مدل منطقی^{۱۲} و مدل استنتاجی^{۱۳} را نام برد. همچنین کوشش‌های موفقی در پیشبرد مدل رابطه‌ای در راستای نیازهای زمان صورت گرفته است (۱۶): ۲۱۳-۲۱۴؛ ۱۸؛ ۲۴۵).

علاوه بر مدل‌هایی که ذکر شد، از سال ۱۹۹۹ به بعد مدل دیگری بنام پایگاه اطلاعاتی ایکس ام ال ارائه شده است. علی‌رغم این‌که علت وجودی ایکس ام ال تسهیل مبادلهٔ

داده‌ها بین سیستم‌های اطلاعاتی و همچنین کاربران بود، با گذشت زمان کاربردهای جدیدی به آن افزوده شد که یکی از این قابلیت‌ها، پایگاه داده بود. این سیستم با توجه به نسل‌های مختلفی که ذکر آن رفت متعلق به نسل سیستم‌های قدیمی یعنی سلسله‌مراتبی می‌باشد ولی قابلیت‌ها و امکانات جدیدی به آن افزوده شده است که امکان ذخیره، بازیابی و مبادلهٔ داده‌ها را به صورت ساده و قابل فهم برای کاربران به ارمغان آورده است و در بخش بعدی به معرفی آن پرداخته می‌شود.

ایکس ام ال چیست و چگونه کار می‌کند؟

زبان‌های نشانه‌گذاری عمومی‌یافته^{۱۴} در دهه‌های ۶۰ و ۷۰ میلادی پا به عرصهٔ وجود نهادند. هدف این زبان‌ها، سازمان‌دهی اطلاعات اسناد براساس نشانه‌گذاری‌های انجام گرفته توسط نویسندگان بود. لذا تمامی این اسناد برای شروع سند نیاز به تعریفی داشتند تا به استفاده‌کننده از سند، خواه انسان، خواه ماشین، اطلاعات لازم برای شروع رمزگشایی نشانه‌ها را بدهند. با گذشت زمان و گسترش حوزهٔ نشر، نیاز به توسعهٔ این زبان‌ها و استانداردسازی آنها پیش آمد. از این رو، زبان نشانه‌گذاری عمومی‌یافتهٔ استاندارد (اس جی ام ال)^{۱۵} پا به عرصهٔ وجود نهاد. اس جی ام ال، زبان توصیفی^{۱۵}

10. Functional

11. Logical

12. Deductive

13. Generalized Markup Languages (GML)

14. Standard Generalized Markup Language (SGML)

۱۵. زبان توصیفی (metalanguage)، زبانی است که یک زبان دیگر را توصیف می‌کند. مثلاً فعل، اسم، صفت و قید کلماتی هستند که جزء دستور زبان محسوب می‌شوند و برای توصیف کلمات دیگر به کار می‌روند. مزیت یک زبان توصیفی در این است که می‌توان از آن برای توصیف زبان‌های مختلف استفاده کرد (۱: ۱۵).

است که چگونگی نشانه‌گذاری یک سند را نشان نمی‌دهد، اما به برنامه‌های کاربردی و محیط‌های مستقل اجازه کار می‌دهد. بنابراین، هدف اولیه آن، ارائه زبانی استاندارد برای نشانه‌گذاری ساختار منطقی مدارک مستقل از سخت‌افزار و نرم‌افزار بود، با این هدف که مبادله مدارک را در حلقه عملیات انتشار تسهیل کند. اسناد با استفاده از این زبان توصیفی قابل انتقال گشته و امکان اشتراک داده‌ها و بسته‌بندی چندباره^{۱۶} اطلاعات فراهم می‌گردد. با این همه بسیار پیچیده بوده و اجرا و نگهداری اسناد با استفاده از آن بسیار گران است. از طرف دیگر نرم‌افزارهای پشتیبان این استاندارد گران و پیچیده‌اند و به دلیل عدم پشتیبانی از روند کلی مرورگرها با محیط وب سازگار نیستند (۶: ۳۹).

لذا نیاز به زبانی بود تا عاری از این محدودیت‌ها و مشکلات باشد. به این منظور، زبان نشانه‌گذاری گسترش‌پذیر (ایکس‌ام‌ال)^{۱۷} به وجود آمد. با بررسی تعاریف به عمل آمده برای ایکس‌ام‌ال، با موارد متعددی مواجه می‌شویم که گاه در یک راستا و گاه در بخش‌هایی برخلاف یکدیگر هستند و گاه می‌توان آنها را مکمل یکدیگر دانست. برخی آن را زبان (۳۰؛ ۸؛ ۳۶؛ ۲۹؛ ۲۰؛ ۳۳؛ ۱۵)، سیستم (۳۱؛ ۱۹؛ ۲۳؛ ۴۶)، فرمت (۲۸؛ ۸؛ ۳۴؛ ۴۱)، استاندارد (۲۱؛ ۴۵؛ ۴۷) و برخی دیگر آن را یک زبان توصیفی (۲۱؛ ۱۴؛ ۲۴؛ ۱۵) معرفی می‌کنند و آن را مهم‌ترین ویژگی ایکس‌ام‌ال می‌دانند و اشاره دارند که

ایکس‌ام‌ال به‌عنوان یک زبان توصیفی اجازه می‌دهد تا برجسب‌های مورد نیاز خود را تعریف کنیم و آن را یک مزیت بزرگ نسبت به اچ‌تی‌ام‌ال^{۱۸} می‌دانند. اکثریت قریب به اتفاق تعاریف مورد بررسی آن را محصول کنسرسیوم وب گسترده جهانی می‌دانند، ولی زمان ارائه آن را نوامبر ۱۹۹۶ (۴۵؛ ۱؛ ۱۵) یا ۱۹۹۸ (۱۴؛ ۱۳؛ ۳؛ ۴۴؛ ۲؛ ۱۲؛ ۲) می‌دانند. در صورتی که کنسرسیوم وب گسترده جهانی در سایت وب خود، سال ۱۹۹۶ را زمان ارائه آن می‌داند (۴۲). اکثر تعاریف مورد بررسی آن را برگرفته از اس‌جی‌ام‌ال می‌دانند، ولی برخی از آنها تأکید دارند که علی‌رغم این که ایکس‌ام‌ال جهت تسهیل در استفاده از اس‌جی‌ام‌ال به وجود آمده، این دو از یک نسل نیستند (۱؛ ۱۵). این در صورتی است که کنسرسیوم وب گسترده جهانی آن را شکل محدود شده اس‌جی‌ام‌ال می‌داند (۴۲).

ایکس‌ام‌ال برای انتشار و مبادله داده‌های ساخت‌یافته در اینترنت و اینترنت مورد استفاده قرار گرفته و همچنین به طراحان امکان می‌دهد تا برجسب‌های خود را تعریف کنند. بنابراین، هدف اولیه آن را تسهیل اشتراک داده‌ها میان سیستم‌های مختلف به ویژه سیستم‌هایی که به اینترنت متصل هستند، می‌دانند. از این رو می‌توان گفت که ایکس‌ام‌ال برای تسهیل امکان مبادله داده‌ها بین برنامه‌های کاربردی در وب به وجود آمده است. کنسرسیوم وب گسترده جهانی، اهداف دهگانه زیر را برای طراحی ایکس‌ام‌ال بیان می‌کند:

16. Repackage

17. eXtensible Markup Language (XML)

18. Hyper Text Markup Language (HTML)

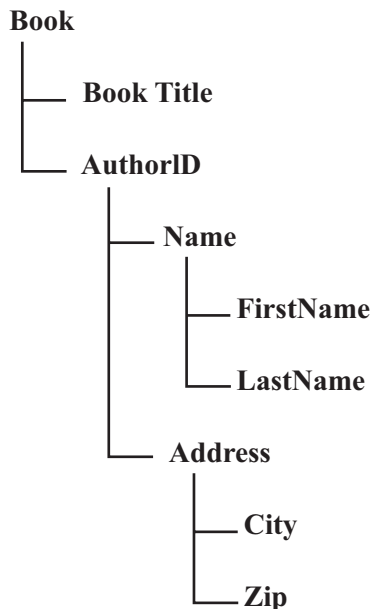
۱. ایکس ام ال باید مستقیماً از طریق اینترنت قابل استفاده باشد.
۲. ایکس ام ال باید انواع مختلف برنامه‌های کاربردی را پشتیبانی کند.
۳. ایکس ام ال باید با اس جی ام ال سازگار باشد.
۴. ایکس ام ال باید برای نوشتن برنامه‌هایی که اسناد ایکس ام ال را پردازش می‌کنند، آسان باشد.
۵. تعداد اجزای اختیاری در ایکس ام ال باید به حداقل برسد.
۶. اسناد ایکس ام ال باید واضح و قابل خواندن توسط انسان باشند.
۷. فایل ایکس ام ال باید به سرعت طراحی شود.
۸. فایل ایکس ام ال باید دقیق و ساده باشد.
۹. ایجاد اسناد ایکس ام ال باید آسان باشد.
۱۰. اختصار در نشانه‌گذاری ایکس ام ال از حداقل اهمیت برخوردار است (۴۲).
با توجه به موارد فوق می‌توان به سادگی، دقت، قابلیت سازگاری با اس جی ام ال و برنامه‌های کاربردی و قابل استفاده بودن از طریق اینترنت اشاره کرد که در تعاریف مورد بررسی نیز کمابیش به آنها اشاره شده است.
تنها در یکی از تعاریف مورد بررسی به ذخیره اطلاعات با فرمت ایکس ام ال اشاره شده است (۴۱). لذا با توجه به اهداف مطرح شده از طرف کنسرسیوم وب گسترده جهانی، که در آن اشاره‌ای به ذخیره اطلاعات

نشده است، نمی‌توان ذخیره اطلاعات را جزء اهداف اولیه ایجاد آن لحاظ کرد، ولی با بررسی‌های به عمل آمده در سایت وب کنسرسیوم به مدارکی در زمینه پایگاه‌های اطلاعاتی ایکس ام ال دسترسی حاصل شد که متعلق به اواخر دهه ۹۰ میلادی بودند. لذا امروزه در جهت استفاده از ایکس ام ال در راستای تولید پایگاه‌های اطلاعاتی نیز گام‌هایی برداشته شده است که از آن جمله می‌توان به پروژه‌های زوپ و کاستور در دانشگاه ویندسور اشاره کرد (۲: ۱۴۷-۱۶۲). علاوه بر پروژه‌های مذکور می‌توان به شرکت‌های نرم‌افزاری بزرگی چون سافت و رای جی^۹، سان، مایکروسافت و غیره اشاره کرد که سرمایه‌گذاری‌های بزرگی در این زمینه انجام داده‌اند.

ایکس ام ال برای حروف از استانداردهای یونی‌کد و ISO/IEC 10646، برای برجسب‌های شناسایی زبان از استاندارد IETF RFC 3066، برای کدهای نام زبان‌ها از استاندارد ISO 639 و برای کد نام کشورها از استاندارد ISO 3166 استفاده می‌کند.

در نهایت، با توجه به موارد اشاره شده می‌توان اظهار داشت که ایکس ام ال یک زبان توصیفی مورد حمایت کنسرسیوم وب گسترده جهانی است که در نوامبر ۱۹۹۶ به جامعه کاربران معرفی شد. این زبان، شکل محدود شده اس جی ام ال است که با هدف انتشار، ذخیره و مبادله داده‌های ساخت یافته در اینترنت و یا اینترنت به وجود آمده است و به طراحان امکان تعریف و استفاده از

بنابراین، ساختار سلسله‌مراتبی مثال فوق را می‌توان به شکل زیر نمایش داد:



تصویر ۲. ساختار درختی ایکس ام ال

با توجه به این که کاربر، خود می‌تواند برچسب‌ها را در ایکس ام ال تعریف کند، نیاز به ابزاری دارد تا بتواند این برچسب‌ها را تعریف کند. در این صورت است که برنامه‌های کاربردی می‌توانند این برچسب‌ها و ساختار سند را شناسایی کنند و مورد استفاده قرار دهند. دی تی دی^{۲۴} ابزاری است که این وظیفه را برعهده دارد. دی تی دی ساختار

برچسب‌های خود را می‌دهد. برخی آن را مانند چتری می‌دانند که برنامه‌های کاربردی دیگر نظیر ایکس اس ال^{۲۰}، ایکس لینک^{۲۱}، ایکس پث^{۲۲} و غیره را احاطه کرده است (۱۳:۳؛ ۱۷:۱۲؛ ۲۲:۱۷۲). به‌هرحال، ایکس ام ال دارای ویژگی‌هایی از قبیل خود مستندی^{۲۳}، توصیف داده‌ها در اسناد به‌صورت سلسله‌مراتبی، ایجاد برچسب‌ها توسط کاربر، توانایی انطباق با سایر برنامه‌های کاربردی و غیره است که آن را نسبت به سایر محصولات منحصربه‌فرد می‌سازد. همان‌طور که می‌دانیم ایکس ام ال قادر است محتوای اسناد و نیز خود اسناد را با ساختار سلسله‌مراتبی ذخیره و نمایش دهد.

```

<book>
  <booktitle> The Selfish Cene </booktitle>
  <authorid="dawkins">
    <name>
      <firstname> Richard </firstname>
      <lastname> Dawkins </lastname>
    </name>
    <address>
      <city> Timbuktu </city>
      <zip>99999 </zip>
    </address>
  </author>
</book>
  
```

تصویر ۱. ساختار سلسله‌مراتبی ایکس ام ال

20. XSL=Extensible Style Sheet Language

21. Xlink

22. Xpath

۲۳. یکی از مزایای مهم ایکس ام ال است که با یک نگاه می‌توان پی برد که چه کاری انجام می‌دهد (۱: ۱۹).

24. Document Type Definition (DTD)

یک سند ایکس ام ال را توصیف می‌کند و ساختار یک عنصر ایکس ام ال را با مشخص کردن نام عناصر فرعی و خصیصه‌های آنها مشخص می‌کند (۴۴: ۹؛ ۳۷: ۳۰۴).

```
<!ELEMENT book (booktitle, outhor)
<!ELEMENT article (title, author*, contactauthor)>
<!ELEMENT contactauthor EMPTY>
<!ATTLIST contactauthor author ID IDREF
IMPLIED>
<!ELEMENT monograph (title, author, editor)>
<!ELEMENT editor (monograph*)>
<!ATTLIST editor name CDATA # REQUIRED>
<!ELEMENT outhor (name, address)>
<!ATTLIST outhor id ID # REQUIRED>
<!ELEMENT name (firstname?, lastname)>
<!ELEMENT firstname (#PCDATA)>
<!ELEMENT lastname (#PCDATA)>
<!ELEMENT address ANY>
```

تصویر ۳. دی تی دی یک فایل ایکس ام ال

دی تی دی می‌تواند جزئی از یک فایل ایکس ام ال باشد. بدین معنی که دی تی دی مربوط به آن سند در ابتدای سند قرار می‌گیرد و برنامه‌های کاربردی برای نمایش یا مبادلهٔ سند، ابتدا دی تی دی مربوطه را ترجمه می‌کنند. در مراکز اطلاعاتی بزرگ مانند کتابخانه‌های دیجیتال که اسناد بسیار زیادی در آنها نگهداری و یا مبادله می‌شود، نیاز هست تا یک دی تی دی مشترک برای اسناد تعریف شود و همهٔ اسناد بتوانند از آن استفاده کنند. لذا برای این کار می‌توان دی تی دی را در یک فایل خارجی ذخیره کرد

و در مواقع مورد لزوم در فایل‌های ایکس ام ال آن را فراخوانی کرد.

حال که با ساختار فایل‌های ایکس ام ال آشنا شدیم، باید بدانیم که برای استفاده از آنها در محیط وب، نیاز به استفاده از برنامه‌ها و زبان‌های کاربردی دیگری است تا بتوان این اسناد را به شکل مطلوب در اختیار کاربر قرار داد. سال‌ها پیش وقتی که وب هنوز جوان بود و کاربران نیز توقع کمی داشتند، صفحات ایستا^{۲۵} همه را راضی می‌کرد. اما اکنون برای جلب (و حفظ) مخاطب، پویایی به یک عنصر حیاتی تبدیل شده است. لذا برای نمایش فایل‌های ایکس ام ال نیاز به کمک گرفتن از سایر برنامه‌ها و زبان‌ها مثل ایکس اس ال، اچ تی ام ال^{۲۶}، جاوا اسکریپت^{۲۷} و غیره است. ایکس اس ال می‌تواند با چند کد، فایل پایگاه داده ایکس ام ال را خوانده و سپس با استفاده از این اطلاعات صفحات اچ تی ام ال پویا ایجاد کرده و به رایانهٔ کاربر ارسال کند. بدین ترتیب، هر برنامه سه فایل خواهد داشت: یک فایل دادهٔ ایکس ام ال، یک فایل الگوی ایکس اس ال و یک فایل اچ تی ام ال و یا جاوا اسکریپت برای نمایش داده‌ها.



تصویر ۴. مجموعه عناصر مورد استفاده در کاربرد ایکس ام ال

پایگاه‌های اطلاعاتی ایکس ام ال

اسناد ایکس ام ال را می‌توان از دو دیدگاه مختلف و به هم مرتبط مورد بررسی قرار داد. از یک دیدگاه آنها را می‌توان به‌عنوان فایل‌هایی که شامل داده‌هایی هستند که در محیط وب بین سیستم‌های اطلاعاتی مختلف مبادله می‌شوند، بررسی کرد. از این دیدگاه می‌توان آنها را فایل‌های داده-مبنا^{۲۸} نامید. این اسناد دارای محتوایی با ساختار منظم و متجانس هستند. ابرداده‌های مورد استفاده در کتابخانه‌های دیجیتال می‌توانند در قالب فایل‌های ایکس ام ال ذخیره و بین سیستم‌های اطلاعاتی با یکدیگر و با کاربران مبادله شوند. از دیدگاه دیگر، آنها اسناد ایکس ام ال هستند که اشیاء مرتبط با برنامه‌های کاربردی می‌باشند. از این دیدگاه، آنها فایل‌های سند-مبنا^{۲۹} هستند. ساختار آنها نامنظم و داده‌ها نامتجانس هستند. به‌عنوان نمونه می‌توان به اشیاء دیجیتال مانند کتاب، پایان‌نامه، گزارش فنی و غیره که مورد استفاده در کتابخانه‌های دیجیتال هستند، اشاره کرد (۹: ۸۴؛ ۱۱: ۱؛ ۱۰: ۲).

دو دیدگاه اشاره شده در فوق (داده-مبنا و سند-مبنا) نشانگر دو نوع داده مورد استفاده در کتابخانه‌های دیجیتال هستند که می‌توانند از ایکس ام ال استفاده کنند. در کتابخانه‌های دیجیتال برای ذخیره اسناد ایکس ام ال سه روش وجود دارد:

• **ذخیره ایکس ام ال به‌عنوان یک فایل.** در این رویکرد، اسناد ایکس ام ال شبیه سایر فایل‌های متنی هستند. برای بازیابی اطلاعات

این فایل‌ها نیاز به تهیه نمایه‌هایی است که بتوان امکان دسترسی به محتوای آنها را مهیا ساخت. این روش زمانی می‌تواند عملیاتی باشد که با تعداد محدودی از فایل‌های ایکس ام ال، سروکار داریم.

• **استخراج داده.** در این رویکرد، اسناد ایکس ام ال تحلیل می‌شوند و اطلاعات موجود در آنها به پایگاه‌های اطلاعاتی منتقل می‌شوند. در این حالت فایل اصلی ایکس ام ال نگهداری نمی‌شود. زمانی که نیاز به اطلاعات مربوطه باشد، یک فایل ایکس ام ال جدید با استفاده از اطلاعات موجود در پایگاه اطلاعاتی دوباره تولید می‌شود (۲۶: ۱۸۶).

• **استفاده از پایگاه اطلاعاتی ایکس ام ال.** با این رویکرد، اسناد ایکس ام ال در پایگاه اطلاعاتی ذخیره می‌شوند که با ساختار اسناد ایکس ام ال منطبق هستند و قادر به اجرای پرسش‌ها و بازیابی داده‌ها در فرمت مذکور می‌باشند (۲۵: ۲).

محدودیت بزرگ رویکرد نخست این است که اطلاعات ساخت‌یافته موجود در اسناد ایکس ام ال نمی‌توانند برای بازیابی مورد استفاده قرار گیرند. بنابراین، نیاز به ایجاد نمایه‌هایی است تا بتوان آنها را بازیابی کرد، مانند نمایه عنوان یا نویسنده. اما اگر کاربری کلیدواژه‌های مربوطه را نداند یا با استفاده از کلیدواژه دیگری غیر از عنوان یا نویسنده دنبال مطلب باشد، با مشکل مواجه خواهد شد. در مجموع می‌توان گفت تنها چیزی که می‌تواند بازیابی شود، فایل اصلی است و بازیابی محتوای آن با مشکلات فراوانی

همراه است. از طرف دیگر، در مواردی که کاربر نیاز به طرح سؤالات پیچیده‌ای دارد، مانند کتاب‌های منتشر شده در موضوعی خاص پس از سال ۱۳۸۴، اسناد به صورت مجزا بازیابی خواهند شد و محتوای آنها باید جداگانه بررسی شوند.

انتقال داده‌های ایکس ام ال به پایگاه‌های اطلاعاتی رابطه‌ای، نیز دارای محدودیت‌های خاص خود است. از طرفی انتقال داده‌ها زمانی کارآمد خواهد بود که بتوان داده‌ها را در ردیف‌ها و ستون‌ها قرار داد. این روش برای اسناد توصیفی مانند گزارش‌ها و اسناد با ساختار پیچیده مانند کتاب‌ها پاسخ‌گو نیست، یا این‌که با مشکلات فراوانی همراه است. از طرف دیگر، در این روش سند اصلی نگهداری نمی‌شود. در صورتی که ممکن است به دلایلی به آن نیاز داشته باشیم. علاوه بر این، استخراج داده‌ها از یک سند و تقسیم آنها به ده‌ها قطعه کوچک و سپس دوباره‌سازی فایل اصلی با فرمت ایکس ام ال کار بیهوده‌ای است.

تأکید متن حاضر بر نوع سوم استفاده از پایگاه اطلاعاتی ایکس ام ال است. لذا در ادامه به این موضوع می‌پردازیم. در طراحی پایگاه‌های اطلاعاتی، در مدل‌های پیشین مانند پایگاه‌های اطلاعاتی رابطه‌ای، مراحل چهارگانه زیر رعایت می‌شود:

۱. گردآوری اطلاعات در ارتباط با نیاز اطلاعاتی کاربران و هدف مجموعه؛
۲. مدل‌سازی اشیاء، خصیصه‌ها و روابط، و به‌طور کلی تحلیل سیستم با استفاده از زبان‌هایی نظیر یو ام ال^{۳۰}؛

۳. ترجمه تحلیل انجام گرفته به مدل رابطه‌ای با استفاده از اصول نرمال‌سازی؛

۴. پالایش طراحی انجام گرفته در صورت ضرورت، جهت اطمینان از این که نیازمندی‌های اجرایی برنامه کاربردی مرتفع شده باشد (۲۵: ۶).

در عمل، چنین رویکردی به استثنای مرحله سوم می‌تواند برای ایجاد پایگاه اطلاعاتی ایکس ام ال به اجرا درآید (جایی که مدل به جای جداول رابطه‌ای به عناصر و خصیصه‌های ایکس ام ال ترجمه می‌شود). اما یک تفاوت بزرگ در اینجا وجود دارد و آن این که ذاتاً اسناد ایکس ام ال برای مبادله داده‌ها و نه برای ذخیره آنها طراحی شده‌اند. طراح پایگاه اطلاعاتی باید تصمیم بگیرد که آیا پایگاه اطلاعاتی ایکس ام ال برای نگهداری اسناد به شکلی که دریافت می‌شوند مورد استفاده قرار خواهد گرفت یا برای ساختاردهی مجدد محتوای اسناد مطابق با پرسش‌های ارائه شده از طرف کاربران. تصمیم طراح بستگی به نیازهای برنامه کاربردی و هدف سیستم دارد.

هدف هر سیستم اطلاعاتی، ذخیره، پردازش و بازیابی محتوای آن به آسان‌ترین و بهترین روش است. سیستم‌های پایگاه اطلاعاتی بهترین راه حل برای وصول به این امر می‌باشند. مهم‌ترین وظیفه سیستم‌های پایگاه اطلاعاتی، زبان پرس و جو^{۳۱} است. در پایگاه‌های اطلاعاتی ایکس ام ال نیز که با فایل‌های داده - مبنا و سند - مبنا سروکار دارند و به شکل سلسله‌مراتبی داده‌ها را

30. Unified Modeling Language (UML)

31. Query language

در خود ذخیره می‌کنند، نیاز به یک سیستم پرس‌وجوی مناسب است تا بتواند مرتبط‌ترین داده‌ها را به بهترین و سریع‌ترین وجه ممکن در اختیار کاربران قرار دهد. بدین منظور، ابتدا باید پرسش‌های کاربران را تجزیه و تحلیل کرد تا بهترین روش را برای بازیابی داده‌ها در رابطه با این پرسش‌ها طراحی کرد. اغلب پرسش‌های کاربران از فایل‌های ایکس ام ال یا به‌طور کلی از سیستم‌های اطلاعاتی را می‌توان به سه طبقه زیر، طبقه‌بندی کرد:

- بازیابی اسناد
- بازیابی حقایق
- بازیابی اطلاعات در رابطه با موضوعی خاص (۲۵:۳)

منظور از بازیابی اسناد، مکان‌یابی یک یا چند مدرک توسط کاربر است. مدارکی که بازیابی می‌شوند، همان منابعی هستند که در پایگاه اطلاعاتی ذخیره شده‌اند. اطلاعات مورد استفاده برای بازیابی اسناد ممکن است بسیار ساده، مانند کتاب ایکس ام ال در کتابخانه‌ها، یا پیچیده، مانند نقدهایی که درباره کتاب ایکس ام ال در کتابخانه‌ها نوشته شده است، باشند.

منظور از بازیابی حقایق، پرسش‌هایی است که واقعیت‌ها را از اسناد استخراج می‌کنند. اطلاعات مورد نیاز، ممکن است جملگی در یک مدرک یا در چندین مدرک باشند. به‌عنوان مثال، چه موقع آخرین آمار مربوط به صادرات فرش ایران ثبت شده است؟ منظور از بازیابی اطلاعات در رابطه با

موضوعی خاص، همان پرسش‌هایی هستند که کاربران در استفاده از موتورهای جستجوی وب طرح می‌کنند، به‌رحال، چیزی که در اینجا مهم است افزودن نشانه‌گذاری به اسناد توصیفی است، زیرا مواردی پیش خواهد آمد که تنها راه برای یافتن اسناد مرتبط، جستجوی متن خواهد بود. به‌عنوان مثال، بهترین روش برای یافتن اطلاعاتی در رابطه با صادرات فرش ایران جستجوی آن به صورت جستجوی تمام متن است. ویژگی این پرسش این است که هیچ پاسخ صریحی وجود ندارد و پاسخ آن بستگی به دقت موتور جستجویی دارد که می‌تواند اسنادی را بازیابی کند که با نیاز اطلاعاتی کاربر بیشتر مرتبط هستند.

این سه نوع پرسش، به وسیله انواع مختلف نرم‌افزارهای جستجوی پایگاه‌های اطلاعاتی مدنظر هستند. بازیابی اسناد می‌تواند به صورت ذخیره در فایل با نمایه‌سازی کلیدواژه‌ای ساده قابل اجرا باشد، به شرطی که کلیدواژه‌هایی که برای بازیابی مورد استفاده قرار می‌گیرند شناخته شده باشند. بازیابی حقایق، وظیفه اصلی پایگاه‌های اطلاعاتی رابطه‌ای است. حقایق، از منابع اصلی گزینش شده و به صورت جداگانه‌ای در پایگاه اطلاعاتی ذخیره می‌شوند. بنابراین، می‌توانند قابل جستجو و گردآوری باشند. بازیابی اطلاعات در رابطه با موضوعی خاص نیز وظیفه اصلی بسته‌های نرم‌افزاری بازیابی متن آزاد^{۳۲} و موتورهای جستجوی اینترنتی است (۲۵:۴). کنسرسیوم وب گسترده جهانی که مسئولیت توسعه ایکس ام ال را برعهده دارد، زبان

پرس و جویی را تحت عنوان ایکس کوئری^{۳۳} برای دستیابی به پایگاه‌های اطلاعاتی ایکس ام ال ایجاد کرده است. ایکس کوئری برای پاسخ‌گویی در سه دسته پرسش فوق طراحی شده است. سیستم پرس و جوی ایکس ام ال با استفاده از فن‌آوری‌های مختلف و مرتبط به هم فعالیت می‌کند که هر کدام از این فن‌آوری‌ها وظایف خاصی را برعهده دارند که در زیر به آنها اشاره می‌شود:

- **ایکس پث.** زبانی است که برنامه‌های کاربردی را قادر می‌سازد تا بخش‌های خاص یک سند را مکان‌یابی و دستکاری کنند.
- **دام (مدل شیء‌گرای سند)**^{۳۴}. برنامه‌ها و اسکریپت‌ها را قادر می‌سازد تا به محتوا و ساختار اسناد دسترسی پیدا کرده و آنها را روزآمد کنند.

- **ایکس لینک.** ارجاعات بین اسناد، گراف‌ها و یا سایر برنامه‌های کاربردی را فراهم می‌کند. لازم به ذکر است که قابلیت‌های پیوند فرامتن آن فراتر از اچ تی ام ال است، مانند قابلیت ایجاد ارتباط فرامتن به بیش از یک مرجع اطلاعاتی. مانند زمانی که می‌خواهید هم توصیف یک اصطلاح را نشان دهید و هم فایلی را در آن زمینه در اختیار کاربر قرار دهید (۲۶: ۱۸۴).

- **فضای نام**^{۳۵}. به سیستم‌ها اجازه می‌دهد تا تفاوت میان کلمات و معانی آنها را درک کنند. به عنوان مثال، کلمه «شیر» را در نظر بگیرید. اگر کسی بگوید «شیر خرید»، مطمئناً منظور او شیر آب یا شیر لبنی نیست، بلکه

آن شیری است که در جنگل زندگی می‌کند. بنابراین، فضای نام وسیله‌ای برای محدود کردن (صریح یا ضمنی) معنای یک نام است. این کار اغلب با استفاده از یک پیشوند که قبل از نام مورد نظر می‌آید، صورت می‌گیرد. مثلاً با گفتن «جنگل: شیر» یا «لبنیات: شیر»، دیگر کسی منظور شما را اشتباه نخواهد فهمید (۴۴: ۱۰-۱۳؛ ۳۲: ۴-۶؛ ۱۱: ۱۱-۱۲). البته لازم به ذکر است که استفاده از این مفهوم کاربرد محدودی دارد و همه جا نیاز به استفاده از آن نیست. به عنوان مثال، وقتی در متنی مربوط به جنگل این فضای نام مورد استفاده قرار می‌گیرد، دیگر نیازی به آوردن «جنگل: شیر» نیست و می‌توان اصطلاح «شیر» را به تنهایی به کار برد.

- **ایکس آپ دیت**^{۳۶}. قابلیت افزودن، جایگزینی، حذف و تغییر نام عملگرها را برای ایکس کوئری فراهم می‌کند (۲۷: ۲).

تبدیل فرمت پایگاه‌های اطلاعاتی رابطه‌ای و ایکس ام ال

زمانی که با فایل‌های ایکس ام ال سروکار داریم، همواره این مسئله ذهن ما را به خود مشغول می‌کند که چگونه می‌توان این فایل‌ها را در پایگاه‌های اطلاعاتی معمول (رابطه‌ای)، که در کتابخانه‌ها بیشتر مورد استفاده قرار می‌گیرند، ذخیره کنیم و یا این‌که چگونه اطلاعات موجود در پایگاه‌های اطلاعاتی خود را به فایل‌های ایکس ام ال جهت انتقال داده‌ها تبدیل کنیم. از آنجا که ساختار ذخیره

33. Xquery

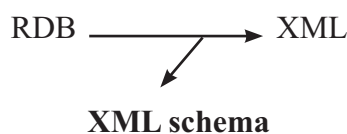
34. DOM=Document Object Model

35. Name space

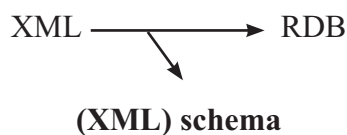
36. Xupdate

اطلاعات در پایگاه‌های اطلاعاتی رابطه‌ای و ایکس ام ال متفاوت هستند (پایگاه‌های اطلاعاتی رابطه‌ای داده‌ها را به صورت جداول و رابطه‌ها در خود ذخیره می‌کنند در حالی که در ایکس ام ال داده‌ها به صورت ساختار درختی و سلسله مراتبی ذخیره می‌شوند)، این مسئله یک معضل به نظر می‌رسد. برای وضوح هر چه بیشتر مسئله، آن را به سه جزء مستقل تقسیم کرده و مورد تجزیه و تحلیل قرار می‌دهیم:

- استخراج. استخراج داده‌ها از پایگاه اطلاعاتی رابطه‌ای و انتقال آن به فایل ایکس ام ال با استفاده از ساختار ایکس ام ال از پیش تعریف شده.



- تزریق. انتقال داده‌ها از فایل ایکس ام ال به پایگاه اطلاعاتی رابطه‌ای با استفاده از ساختار از پیش تعریف شده.



- پرس و جو. ترجمه امکان‌های اطلاعات مطابق با شرایط خاص مورد نظر در فرمت‌های تبدیل شده (ایکس ام ال یا پایگاه

اطلاعاتی رابطه‌ای) (۴۳: ۳؛ ۳۷: ۳۰۳). مراحل سه‌گانه فوق مشخص می‌کند که برای رسیدن به اهداف مورد نظر، به یک ساختار ایکس ام ال استاندارد در کنار پایگاه اطلاعاتی رابطه‌ای نیاز است تا بتوان با استفاده از آن اطلاعات را به پایگاه اطلاعاتی رابطه‌ای انتقال داد و یا از پایگاه اطلاعاتی خارج کرد. این امر به دلایل مختلف صورت می‌گیرد. در کتابخانه‌های دیجیتال، اطلاعات در فرمت‌های مختلف وجود دارند. بنابراین، نرم‌افزارهای کتابخانه دیجیتال باید قابلیت شناسایی و استفاده از آنها را داشته باشند. با توجه به قابلیت‌ها و توانایی‌های فایل‌های ایکس ام ال، می‌توان حضور گسترده آنها را در کتابخانه‌های دیجیتال به وضوح لمس کرد. از طرف دیگر، این کتابخانه‌ها ممکن است همچنان بخواهند از پایگاه‌های اطلاعاتی رابطه‌ای استفاده کنند. در این صورت نیاز به تبدیل فایل‌های ایکس ام ال به فرمت قابل هضم توسط پایگاه‌های اطلاعاتی ضرورت می‌یابد و نیز از طرف دیگر، نیاز به تبدیل فرمت دوباره اطلاعات موجود در پایگاه‌های اطلاعاتی به فایل ایکس ام ال جهت انتقال یا استفاده نیز ضرورت خواهد یافت.

برای تبدیل فایل‌های ایکس ام ال به پایگاه‌های اطلاعاتی، دو روش توسط واسط‌افزارها^{۳۷} مورد استفاده قرار می‌گیرد: مدل جدول-مبنا^{۳۸} و مدل شیء-رابطه^{۳۹}. مدل جدول-مبنا توسط بسیاری از واسط‌افزارهای موجود با هدف انتقال داده‌ها بین فایل ایکس ام ال و پایگاه اطلاعاتی رابطه‌ای

37. Middleware
38. Table-base

39. Object-relation

قرار می‌گیرد. این مدل، داده‌ها را در فایل ایکس ام ال مانند درختی از اشیاء، شبیه‌سازی می‌کند. در این مدل، موجودیت‌ها با ویژگی‌های خاص خود و محتوای موجودیت‌ها به‌عنوان کلاس شناخته می‌شوند. در استفاده از این مدل باید بتوانیم آن را با مدل شیء‌گرایی سند متمایز کنیم. مدل شیء‌گرایی سند، یک سند ایکس ام ال را مدل‌سازی می‌کند و برای تمامی اسناد ایکس ام ال یکسان است، در حالی که مدل مذکور، داده را در اسناد مدل‌سازی می‌کند و از سندی به سند دیگر در ایکس ام ال متفاوت است.

```

SalesOrder
 / | \
Customer Item Item
 | |
Part Part

```

در این روش، جداول رابطه‌ای شامل تمام اطلاعات مورد نیاز برای ذخیره اطلاعات موجود در فایل‌های ایکس ام ال و بالعکس خواهد بود. به این منظور، نرم‌افزارهای مدیریت پایگاه داده‌ها از مدت‌ها پیش به فکر ایجاد قابلیت ذخیره و بازیابی داده‌های ایکس ام ال افتاده‌اند. لذا امروزه شاهد ذخیره و بازیابی داده‌های ایکس ام ال در این نرم‌افزارها هستیم. با بررسی نرم‌افزار مدیریت پایگاه اطلاعاتی اس کیو ال^{۴۰}، اوراکل^{۴۱} و اکسس^{۴۲} می‌توانیم با این قابلیت آنها بیشتر

ایجاد شده‌اند، مورد استفاده قرار می‌گیرد. در این روش، اسناد ایکس ام ال مانند جداول، مدل‌سازی می‌شوند. ساختار فایل ایکس ام ال در این روش به شکل زیر باید باشد:

```

<database>
  <table>
    <row>
      <column1>...</column1>
      <column2>...</column2>
      ...
    </row>
    <row>
      ...
    </row>
    ...
  </table>
  <table>
    ...
  </table>
  ...
</database>

```

روش مذکور برای مبادله داده‌ها بین دو پایگاه اطلاعاتی رابطه‌ای مناسب است. طبیعی است که این روش برای هر فایل ایکس ام ال که با شکل فوق متناسب نباشد، نمی‌تواند مفید واقع شود (۷-۸:۱۱).

مدل شیء - رابطه در کلیه پایگاه‌های اطلاعاتی منطبق با ایکس ام ال و همچنین توسط بسیاری از واسط‌افزارها مورد استفاده

40. Structured Query Language (SQL)

41. Oracle

42. Access

آشنا شویم (برای کسب اطلاعات بیشتر در این زمینه به راهنمای این نرم‌افزارها مراجعه شود؛ نرم‌افزار مدیریت پایگاه‌های اطلاعاتی اکسس؛ اس کیو ال؛ اوراکل) (۷؛ ۳۸) متنها در استفاده از این نرم‌افزارها باید همواره متوجه تبدیل داده‌ها و خطاهای ایجاد شده و اصلاح آنها باشیم. چرا که در صورت عدم وجود دقت کافی در این زمینه ممکن است اطلاعات بسیاری را از دست بدهیم. البته، با توجه به مطالبی که در بخش پیشین به آنها پرداخته شد در مواردی که ما در کتابخانه‌های دیجیتال بیشتر با فایل‌های ایکس ام ال سرو کار داریم، نیاز به داشتن پایگاه‌های اطلاعاتی ایکس ام ال^{۴۳} است که خود قابلیت ذخیره و بازیابی داده‌های ایکس ام ال را داراست. بدین منظور از مجموعه نرم‌افزارها و زبان‌های مرتبط برای انجام فعالیت‌های مختلف استفاده می‌شود که هدف یکسانی را برعهده دارند و امور مربوط به ذخیره، پردازش، بازیابی، امنیت و مدیریت داده‌ها و غیره را به نحو مطلوب به ورطه اجرا در می‌آورند.

ضرورت استفاده از پایگاه‌های اطلاعاتی ایکس ام ال در کتابخانه‌های دیجیتال

کتابخانه‌های دیجیتال با انواع مختلف داده‌ها (صوت، متن، تصویر، گرافیک و غیره) با فرمت‌های مختلف (ورد، پی دی اف، اچ تی ام ال و غیره) سروکار دارند. بدین منظور به پایگاه‌های اطلاعاتی با قابلیت‌های بالا نیاز دارند. از آنجا که ایکس ام ال امروزه به‌خوبی توسعه یافته است و به‌عنوان یک ضرورت

در کتابخانه‌های دیجیتال مطرح شده است، نیاز به تأمل بیشتر دارد. ایکس ام ال به علت ساختار سلسله‌مراتبی که دارد در آینده‌ای نزدیک حجم بسیار زیادی از اطلاعات در کتابخانه‌های دیجیتال را به خود اختصاص خواهد داد. در این راستا است که بحث مربوط به پایگاه‌های اطلاعاتی ایکس ام ال مطرح می‌شود. اگر پایگاه اطلاعاتی، کاملاً متناسب با ایکس ام ال و ساختار سلسله‌مراتبی آن باشد، ما برای استفاده از آن نیاز به تبدیل فرمت ایکس ام ال به زبان پایگاه اطلاعاتی و بالعکس نخواهیم داشت. به منظور کسب آگاهی از ضرورت استفاده از این نوع پایگاه اطلاعاتی در کتابخانه‌های دیجیتال، ابتدا بهتر است با مزایای ایکس ام ال بیشتر آشنا شویم. چرا که کسب آگاهی در این زمینه به‌طور خودکار ما را به درک این ضرورت رهنمون خواهد شد. در اینجا به هفت مزیت عمده ایکس ام ال پرداخته می‌شود:

۱. ایکس ام ال یک زبان توصیفی است. ایکس ام ال می‌تواند هر نوع اطلاعاتی را تعریف و توصیف کند. گسترش‌پذیر است. مورد استفاده بسیاری از افراد، گروه‌ها و سازمان‌هاست. زبان‌های نشانه‌گذاری را برای اهداف خاص تعریف می‌کند و ساختارهایی را برای انواع مدل‌های داده‌ای ایجاد می‌کند. ایکس ام ال، همچنین برای منابع اطلاعاتی سند- مبنا و داده- مبنا می‌تواند مورد استفاده قرار گیرد.

۲. ایکس ام ال متن- مبنا و قابل فهم برای انسان و ماشین است. اسناد می‌توانند توسط

برنامه‌های کاربردی و انسان خوانده شوند. ایکس ام ال از فن‌آوری نشانه‌گذاری استفاده می‌کند. برچسب‌هایی مانند <author> یا <creator> می‌توانند برای نشانه‌گذاری در اسناد مورد استفاده قرار گیرند، بدون این‌که هیچ‌گونه محدودیتی داشته باشند. توصیف نوع سند یا ساختار ایکس ام ال می‌تواند برای هر سند مورد استفاده قرار گیرد و همچنین قابل فهم برای انسان و رایانه است، بدون این‌که نرم‌افزار کاربردی که آن را ایجاد کرده را در اختیار داشته باشیم. اسناد ایکس ام ال از استاندارد اسکی^{۴۴} و یونی‌کد^{۴۵} برای متون استفاده می‌کنند، بنابراین، می‌توانند برای تمامی زبان‌های زنده و مرده دنیا بدون مواجهه با مشکلی حاد، مورد استفاده قرار گیرند.

۳. ایکس ام ال برای اسناد ساخت یافته بهترین گزینه است. اسناد ایکس ام ال دارای ساختار سلسله‌مراتبی هستند. عناصر اسناد ایکس ام ال می‌توانند ساختار پیچیده اسناد را ایجاد کنند. در عمل، تبدیل اسناد ایکس ام ال پیچیده برای استفاده در پایگاه‌های اطلاعاتی رابطه‌ای امکان‌پذیر است. اما تبدیل عناصر اسناد ایکس ام ال به پایگاه‌های اطلاعاتی و از پایگاه‌های اطلاعاتی به ایکس ام ال فرایند زمان‌بری است. بنابراین، نیاز به نرمال‌سازی و ترمیم اسناد قبل از ذخیره یک سند ایکس ام ال می‌باشد.

۴. ایکس ام ال، فرمت نمایش نیست. ایکس ام ال، محتوای اسناد را از نمایش آنها

جدا می‌کند. برای نشان دادن نحوه نمایش اسناد ایکس ام ال از ایکس اس ال، اچ تی ام ال و غیره استفاده می‌شود.

۵. ایکس ام ال، چند زبانی است. ایکس ام ال بر مبنای یونی کد است. بنابراین قادر است تمام زبان‌ها نظیر ژاپنی، چینی، عربی و غیره را به درستی نشان دهد.

۶. ایکس ام ال، مبادلات را یکدست می‌کند. ایکس ام ال، محدودیت‌های شبکه‌های تجارت الکترونیکی جهانی را تقلیل می‌دهد و موجب می‌شود تا سازمان‌های کوچک مثل کتابخانه‌ها نیز بتوانند از قابلیت‌های آن به صورت بهینه‌ای برای مبادله الکترونیکی داده‌ها^{۴۶} استفاده کنند.

۷. ایکس ام ال، کاربرد گسترده‌ای دارد. ایکس ام ال مورد حمایت شرکت‌های بزرگ فن‌آوری اطلاعات مانند آی بی ام^{۴۷}، مایکروسافت، اوراکل، سان، سافت و رای جی، میکروسیستم و شرکت‌های دیگری است که سرمایه‌های زیادی را در ایکس ام ال خرج کرده‌اند. در نتیجه، آنها به‌طور فعال در فرایند استانداردسازی ایکس ام ال شرکت می‌کنند (۳۹؛ ۴۰).

با توجه به مزایای فوق، امروزه شاهد گسترده‌گی استفاده از ایکس ام ال هستیم. کتابخانه‌های دیجیتال به‌عنوان یکی از مراکز پراهمیت در جوامع امروزی از مراکزی هستند که به‌طور گسترده‌ای از ایکس ام ال در فعالیت‌های خود استفاده می‌کنند.

44. ASCII=American Standard Code for Information Interchange
45. Unicode

46. Electronic Data Interchange (EDI)
47. IBM=International Business Machines

نتیجه‌گیری

پایگاه‌های اطلاعاتی کارآمد، جزء ضروریات هر سیستم اطلاعاتی می‌باشند. کتابخانه‌های دیجیتال نیز به‌عنوان یکی از بزرگ‌ترین و پیچیده‌ترین سیستم‌های اطلاعاتی، دربرگیرنده پایگاه‌های اطلاعاتی برای ذخیره، پردازش، بازیابی، امنیت و مدیریت اطلاعات و منابع اطلاعاتی هستند. این کتابخانه‌ها با توجه به ماهیت فعالیتشان با انواع منابع اطلاعاتی (صوت، متن، تصویر، گرافیک و غیره)، با فرمت‌های مختلف (متن، پی‌دی‌اف، اچ‌تی‌ام‌ال، ایکس‌ام‌ال و غیره) سروکار دارند. امروزه، ایکس‌ام‌ال با توجه به ساختار سلسله‌مراتبی جایگاه خاصی را برای خود در مجموعه‌ها و فعالیت‌های کتابخانه‌های دیجیتال باز کرده است و پیش‌بینی می‌شود در آینده نزدیک نقش بسیار پررنگ‌تری را در این کتابخانه‌ها برعهده گیرد. ایکس‌ام‌ال، در ابتدا به‌عنوان فرمتی برای مبادله داده‌ها مطرح شد و نقش گسترده‌ای در فعالیت‌های انواع سیستم‌های اطلاعاتی نظیر سیستم‌های تجاری، بانکداری، کتابخانه‌ای و غیره برعهده گرفت. کنسرسیوم وب گسترده جهانی به‌عنوان بانی و حامی اصلی ایکس‌ام‌ال، در تلاش برای افزایش قابلیت و توسعه آن است. چنان‌که امروزه شاهد ظهور و گسترش پایگاه‌های اطلاعاتی ایکس‌ام‌ال هستیم، کتابخانه‌های دیجیتال نیز به‌عنوان مراکزی که استفاده گسترده‌ای از آن در فعالیت‌های خود دارند، به‌عنوان سازمان‌های پیشرو در این زمینه پا به عرصه

گذاشته‌اند و به تدریج به‌سمت استفاده از این پایگاه‌های اطلاعاتی در حرکت هستند. ایکس‌ام‌ال چتری است که زبان‌ها و برنامه‌های کاربردی دیگر نظیر ایکس‌کوئری، ایکس‌پث، ایکس‌آپ‌دیت، و غیره را در راستای اهداف خود تحت پوشش قرار می‌دهد؛ به همین علت است که می‌تواند به‌عنوان فعالیت‌های موردنیاز در جهت ذخیره، پردازش، بازیابی، مدیریت و غیره را که از یک بانک اطلاعاتی انتظار می‌رود، به اجرا درآورد. ایکس‌ام‌ال، هنوز جوان است و مرحله توسعه خود را طی می‌کند و در آینده نزدیک شاهد افزایش قابلیت‌های آن و گسترش میزان استفاده از آن در سازمان‌های مختلف خواهیم بود. از این‌رو، کتابخانه‌های دیجیتال نیز باید در جهت استفاده از توانایی‌های آن حرکت کنند تا بتوانند نقش و جایگاه شایسته خود را در جوامع اطلاعاتی حفظ نموده و گسترش دهند. البته، شایان ذکر است که ایکس‌ام‌ال به‌عنوان فرمتی برای مبادله امروزه در کتابخانه‌های دورگه^{۴۸}‌ای که از نرم‌افزارهای کتابخانه‌ای مختلف استفاده می‌کنند، نیز نقش بسیار مهمی را ایفا می‌کند. در حال حاضر این نرم‌افزارها به‌علت استفاده از فن‌آوری‌ها، معماری و فرایندهای اجرایی مختلفی که استفاده می‌کنند، قابلیت مبادله داده‌ها و استفاده از داده‌های یکدیگر را ندارند. در صورتی که با استفاده از ساختار ایکس‌ام‌ال و یا دیگر قابلیت‌های آن این امر به‌راحتی امکان‌پذیر است.

Databases.htm

11. Bourret, Ronald. "XML database products". 2006. [on-line]. Available: www.rpbouret.com/xml/XMLDatabaseProds.htm

12. Bray, T.; Paoli, J.; Sperberg-McQueen, C.M. "XML 1.0". 1998. [on-line]. Available: www.w3.org/TR/REC-xml

13. Brown, David; Byard, Terry; Hart, Cynthia. "The impact of XML on information content organization in libraries". 2001. [on-line]. Available: www.tennessee.edu/digitallibrary/xml/howisxmlgoingaffectlibraries.htm

14. Cornell University glossary. [on-line]. Available: www.cs.cornell.edu/wya/DigLib/MS1999/glossary.html. [10Sep.2006].

15. Data direct. [on-line]. Available: www.datadirect.com/developer/xquery/references/definitions/index.ssp. [10Sep.2006].

16. Delobel, Clade; Lecluse, Christophe; Richard, Philippe. *Databases: from relational to object-oriented systems*. London: International Thomson Publishing, 1995.

17. Desmarais, Norman. *The ABC's of xml: the librarian's guide to the eXtensible Markup Language*. Houston: Houston Technology Press, 2000.

18. Fortier, Paul J. *Database systems*

منابع

۱. اشباکر، چارلز. آموزش برنامه‌نویسی XML در ۲۴ ساعت. ترجمه علیرضا زارع‌پور. تهران: مؤسسه علمی و فرهنگی نص، ۱۳۸۰.

۲. تنانت، روی. ایکس ام ال در کتابخانه‌ها. ترجمه مهدی علیپورحافظی، یعقوب نوروزی و حمیدرضا رادفر. تهران: چاپار، ۱۳۸۴.

۳. حق جو، مصطفی. بانک اطلاعات علمی - کاربردی. تهران: دانشگاه علم و صنعت ایران، مرکز انتشارات، ۱۳۷۷.

۴. دیت، سی جی. سیستم‌های بانک اطلاعاتی. ترجمه عین‌الله جعفرنژاد قمی. بابل: علوم رایانه، ۱۳۷۹.

۵. رولی، جنیفر. مبانی سیستم‌های اطلاعاتی. ترجمه زهرا سیف‌کاشانی و نجیبه افغانی. تهران: سازمان مطالعه و تدوین علوم انسانی دانشگاه‌ها (سمت)، ۱۳۸۰.

۶. زارع‌زاده، فاطمه. «ایکس ام ال چیست و چه کاربردهایی در کتابخانه دارد؟». فصلنامه کتاب، دوره پانزدهم، ۲ (تابستان ۱۳۸۳): ۳۷-۴۴.

7. "About XML data and access". Microsoft Access Database management system (Xp version).

8. AEP. [on-line]. Available: www.aepedi.com/glossary.htm. [10Sep.2006].

9. Bertino, Elisa; Catania, Barbara. "Integrating XML and databases". *IEEE Internet Computing*, (Jul.Aug.2001): 84-88.

10. Bourret, Ronald. "XML and databases, tech. report, Technical Univ. Darmstadt". 2000. [on-line]. Available: www.rpbouret.com/xml/XMLAnd

XConstraint_WITS04.pdf. [10Sep.2006].

28. Maptrax Australia. [on-line]. Available: www.maptrax.com.au/standard_scompliance/glossaryoftermsatoh/. [10Sep.2006].

29. Motoroola dealer glossary. [on-line]. Available: www.moto-zone.com.au/motoglossary/motoglossary.asp. [10Sep.2006].

30. Naude, Frank. "Glossary". 2006. [on-line]. Available: www.orafaq.com/glossary/faqglosx.htm

31. Net project Glossary. [on-line]. Available: www.netproject.com/docs/migoss/v1.0/glossary.html. [10Sep.2006].

32. Obasanjo, Dare. "An exploration of xml in database management systems". [on-line]. Available: www.mygreatdataguide.info/xml-database. [10Sep.2006].

33. The Online iTV Dictionary. [on-line]. Available: www.itvdictionary.com/x.html. [10Sep.2006].

34. Poet glossary. [on-line]. Available: www.x-solutions.poet.com/eu/newsevents/glossar. [10Sep.2006].

35. Ryan, Nick; Smith, Dan. *Database systems engineering*. London: International Thomson Computer Press, 1995.

36. Seven design avenue. [on-line]. Available: www.7designavenue.com/glossary.htm. [10Sep.2006].

37. Shanmugasundaram, Jayavel... [et

handbook. New York: McGraw-Hill, 1997.

19. Glossary of Internet terms 4umi web. [on-line]. Available: <http://4umi.com/web/glossary.htm>. [10 Sep. 2006].

20. Graphic penguin. [on-line]. Available: www.graphicpenguin.com/services/term3.html. [10Sep.2006].

21. Guide to network resource tools. 2006. [on-line]. Available: www.acad.bg/beginner/gnrt/appendix/glossary.html

22. Gunn, Stan. "The web is only the beginning". *Texas Library Journal*, (Winter 1999): 172-173.

23. Internet terminology. [on-line]. Available: www.unitedyellowpages.com/internet/terminology.html

24. Introduction to Internet terminology. [on-line]. Available: <http://mason.gmu.edu/~montecin/netterms.htm>. [10Sep.2006].

25. Kay, Michael. *XML databases*. Germany: Software AG, 2003. [on-line]. Available: www.softwareag.com

26. Kim, Hyun-Hee; Choi, Chang-Seok. "XML: how it will be applied to digital library systems". *The Electronic Library*, Vol.18, No.3 (2000): 183-189.

27. Madiraju, Praveen; Sunderraman, Rajshekhar; Navathe, Shamkant B. "Semantic integrity constraint checking for multiple XML databases". [on-line]. Available: <http://tinman.cs.gsu.edu/~cscpmx/research/XML/WITS04/>

xml. [10Sep.2006].

43. Valikov, Aleksei; Kazakos, Wassili; Schmidt, Andreas. "Building updatable XML views on top of relational databases". [on-line]. Available: <http://citeseer.ist.psu.edu/cis?q=Andreas+Schmidt>. [10Sep.2006].

44. Waldt, Dale. "Using XML and databases". 2005. [on-line]. Available: www.gilbane.com

45. Webasyst. [on-line]. Available: www.webasyst.net/glossary.htm. [10Sep.2006].

46. Web men. [on-line]. Available: <http://web-men.net/glossary.htm>. [10Sep.2006].

47. Wifi wlan glossary. [on-line]. Available: www.indie.dk/buzz.htm. [10Sep.2006].

al]. Relational databases for querying xml documents: limitations and opportunities. Proceeding of the 25th VLDB conference, Edinburgh, Scotland, 1999.

38. "SQL server and XML support". SQL server package software.

39. Staken, Kimbro. "Introduction to native xml databases". 2001. [on-line]. Available: www.xml.com/pub/a/2001/10/31/nativexmlldb.html

40. Tamino XML benefits. [on-line]. Available: http://softwareag.com/Corporate/products/tamino/prod_info/7goodreasons.asp. [10Sep.2006].

41. The University of Liverpool. [on-line]. Available: www.liv.ac.uk/webteam/glossary. [10Sep.2006].

42. The World Wide Web Consortium. [on-line]. Available: www.w3.org/TR/REC-

تاریخ دریافت: ۱۳۸۵/۵/۲۹

