

## پایگاه‌های داده مبتنی بر ایکس.ام.ال

حسین پورسلطانی

این مقاله بر اساس کار ارزشمندی از رونالد بارت تهیه شده است.<sup>1</sup>

### توضیح در باب شیوه نگارشی

ادبیات رایانه و فن‌آوری اطلاعات آکنده از واژگان تخصصی است، آنچنان که تقید به حفظ خط و زبان فارسی را برای نگارندگان دشوار می‌سازد (البته اگر چنین قیدی نزد آنان متصور باشد). متأسفانه به دلیل فقدان مجموعه واژگان فارسی مناسب اغلب از معادل‌های خارجی آنها استفاده می‌شود. این امر در کوتاه‌مدت منجر به سرگشتگی خوانندگان و در درازمدت به ضعف محتوای فارسی در این حوزه خواهد انجامید. از آنجا که واژگان معانی یکسانی را در ذهن خوانندگان متبادر نمی‌سازند و وحدت رویه در نگارش آنها وجود ندارد بنابراین جستجو و بازیابی محتوا بر اساس کلیدواژه‌ها بسیار دشوار خواهد شد.

در متنی که پیش رو دارید دو اصل مدنظر قرار گرفته‌اند: اول) انتقال مطلب به ساده‌ترین شکل ممکن و دوم) پای‌بندی به زبان پارسی و خط که چهره مکتوب آن است. در پاره‌ای از موارد اصل دوم به نفع اصل اول نادیده گرفته شده است. این موضوع به خصوص در به‌کارگیری معادل‌های فارسی واژگان تخصصی مشهود است. معادل‌سازی لغات بر اساس کیفیت معنایی و رواج به‌کارگیری آنها انجام شده است. به این معنا که اگر برای یک واژه تخصصی معادل فارسی رایج و صحیحی وجود داشته باشد به‌کار گرفته می‌شود، در غیر این صورت از اصل واژه خارجی استفاده می‌شود. واژگان خارجی به همان شیوه‌ای که تلفظ می‌شوند نگاشته شده‌اند و اغلب برای تأکید بر این موضوع از علائم سجاوندی یا حرکات حروف استفاده شده است. به عنوان مثال برای کلمه Database به جای استفاده از عبارت بانک اطلاعاتی از عبارت پایگاه داده استفاده شده است. زیرا اگرچه بانک اطلاعاتی اصطلاحی رایج است اما صحیح نیست (زیرا نه بانک است و نه اطلاعات در آن ذخیره می‌شود!). پایگاه داده اصطلاحی است صحیح که شاید به اندازه واژه بانک اطلاعاتی متداول نباشد اما واژه مهجوری نیز نیست. در مورد اختصارات به طور کلی از معادل خارجی آنها استفاده شده است. برای مثال به جای اختصار XML، عبارت ایکس.ام.ال به کار گرفته شده است.

در نگارش متن تلاش شده است تا حداکثر انطباق با دستور خط فارسی، مصوب فرهنگستان زبان و ادب فارسی حاصل شود. در متن کلمات تخصصی (چه اصل و چه معادل) به صورت پانویس آمده‌اند. هر جا که از این کلمات در متن استفاده شده برای جداسازی آنها از سایر عبارات از قلم کج استفاده شده است.

<sup>1</sup> See: XML and Databases by Ronald Bourret (Last Update September 2005).

## داده در برابر سند: اسناد داده‌محور<sup>۲</sup> و اسناد سندمحور<sup>۳</sup>

اسناد داده‌محور اسنادی هستند که از قالب ایکس.ام.ال برای جایجائی داده استفاده می‌کنند. این اسناد اغلب برای استفاده توسط ماشین طراحی و تهیه شده‌اند و قالب ایکس.ام.ال تنها یک واسطه است. اسناد داده‌محور اغلب دارای ساختاری معین و ساختاریافته<sup>۴</sup> بوده، اجزاء تشکیل دهنده کوچکی داشته و معمولاً فاقد داده‌های متنی تلفیقی‌اند. به طور کلی ترتیب قرار گرفتن عناصر<sup>۵</sup> و مشخصه‌ها<sup>۶</sup> در این گونه از اسناد بی‌معنا است، مگر برای اعتبارسنجی<sup>۷</sup> سند. این اسناد گاهی برای تبادل داده بین پایگاه‌های داده مختلف مورد استفاده قرار می‌گیرند.

در مقابل اسناد داده‌محور، اسناد سندمحور قرار دارند که معمولاً برای استفاده انسانی طراحی و تهیه می‌شوند. صفحات وب نمونه بارزی از این گونه اسناد هستند. این گونه اسناد اغلب نیمه‌ساختاریافته<sup>۸</sup> و یا ساختاریافته<sup>۹</sup> اند، اجزاء سازنده بزرگتری نسبت به اسناد داده‌محور دارا بوده و مملو از محتوای متنی تلفیقی‌اند. ترتیب قرار گرفتن متون، عناصر و مشخصه‌ها معنادار بوده و مؤثر است. تفکیک اجزاء این گونه اسناد و نگهداری آنها در پایگاه‌های داده اغلب دشوار است.

### ان.ایکس.دی<sup>۱۰</sup>

ان.ایکس.دی‌ها، پایگاه‌های داده‌ای هستند که به طور ویژه برای نگهداری اسناد ایکس.ام.ال مورد استفاده قرار می‌گیرند و همانند سایر پایگاه‌های داده، از قابلیت‌هایی مانند تراکنش<sup>۱۱</sup>ها، زبان‌های پرس و جو<sup>۱۲</sup>، دسترسی هم‌زمان و امنیت پشتیبانی می‌کنند. تفاوت عمده آنها با دیگر پایگاه‌های داده، استفاده از مدل درختی ایکس.ام.ال به جای مدل رابطه‌ای<sup>۱۳</sup> در ذخیره و بازیابی است.

ان.ایکس.دی‌ها اغلب برای نگهداری اسناد سندمحور مورد استفاده قرار می‌گیرند. مهم‌ترین دلیل این کاربرد پشتیبانی این پایگاه‌های داده از زبان‌های پرس و جوی مبتنی بر ایکس.ام.ال است که قابلیت انعطاف فراوانی را در استفاده از این گونه اسناد به کاربران می‌دهند. دلیل دیگر این است که این پایگاه‌های داده اسناد را با تمام جزئیات نگهداری می‌کنند، جزئیاتی که در تبدیل این گونه اسناد به اسناد داده‌محور اغلب از دست می‌روند.

<sup>2</sup> Data Centric

<sup>3</sup> Document Centric

<sup>4</sup> Structured

<sup>5</sup> Elements

<sup>6</sup> Attributes

<sup>7</sup> Validation

<sup>8</sup> Semi-Structured

<sup>9</sup> Unstructured

<sup>10</sup> NXD (Native XML Database)

<sup>11</sup> Transaction

<sup>12</sup> Query

<sup>13</sup> Relational

ان.ایکس.دی‌ها علاوه بر این برای یک‌پارچه‌سازی<sup>۱۴</sup> داده‌ها نیز مورد استفاده قرار می‌گیرند. روش متداول برای یک‌پارچه‌سازی داده‌ها استفاده از یک الگو<sup>۱۵</sup>ی واحد رابطه‌ای است که در برابر الگوها و انواع خاص داده، کاستی‌هایی دارد. در مقابل قالب ایکس.ام.ال از انعطاف کافی برای نگهداری و یک‌پارچه‌سازی انواع و الگوهای مختلف داده برخوردار است، به خصوص که ایکس.ام.ال به طور کلی در مقابل تغییرات در الگوی اسناد حساس و آسیب‌پذیر نیست. کاربرد دیگر ان.ایکس.دی‌ها در ذخیره داده‌های نیمه‌ساخت‌یافته است، شرایطی که الگوی مشخصی برای داده‌ها وجود ندارد. همان گونه که پیش‌تر شرح داده شد، ان.ایکس.دی‌ها به الگوی داده‌ها حساس نبوده و از این رو توانایی نگهداری این گونه داده‌ها را دارا هستند. این ویژگی ان.ایکس.دی‌ها (عدم تقید به الگوی داده) باعث می‌شود تا حتی در صورت تغییر یا تکمیل سازمان داده‌ها، اسناد موجود اجباری به انطباق با الگوی جدید نداشته باشند.

### تعریف ان.ایکس.دی

عبارت ان.ایکس.دی برای اولین بار در برنامه‌های بازاریابی تامینو<sup>۱۶</sup> (یک پایگاه داده مبتنی بر ایکس.ام.ال محصول سافتور ای.جی<sup>۱۷</sup>) استفاده شد. به دلیل موفقیت این محصول این عبارت وارد ادبیات فنی شده و در مورد محصولات مشابه نیز مورد استفاده قرار گرفت، اما به طور دقیق تعریف نشد.

در ادامه تعریفی از گروه کاری ایکس.ام.ال:دی.بی<sup>۱۸</sup> ارائه شده است:

... یک مدل منطقی برای اسناد ایکس.ام.ال تعریف می‌کند که بر اساس آن ذخیره و بازیابی می‌شوند. این مدل باید حداقل شامل عناصر، مشخصه‌ها، جایگاه متون و ترتیب قرار گرفتن آنها در سند باشد. نمونه‌هایی از این مدل‌ها عبارتند از مدل داده‌ای ایکس.پس<sup>۱۹</sup>، ایکس.ام.ال اینفوسِت<sup>۲۰</sup> و مدل‌هایی که توسط دی.ا.ام<sup>۲۱</sup> و اس.ای.ایکس<sup>۲۲</sup> ارائه می‌شوند. این موارد اجزاء سازنده یک سند ایکس.ام.ال هستند، همان گونه که سطرها، ستون‌ها و رابطه آنها با یکدیگر سازنده جداول در پایگاه‌های داده رابطه‌ای هستند. این که اسناد چگونه ذخیره و بازیابی می‌شوند فاقد اهمیت است.

بخش اول این تعریف مشابه تعریفی است که از سایر پایگاه‌های داده ارائه می‌شود، با این تفاوت که بر مدلی ویژه (که مبتنی بر ساختار ایکس.ام.ال است) تأکید می‌نماید. البته در عمل ان.ایکس.دی داده‌های بیشتری نسبت به آن چه که در مدل در نظر گرفته شده ذخیره می‌نماید.

<sup>14</sup> Integration

<sup>15</sup> Schema

<sup>16</sup> Tamino

<sup>17</sup> Software AG

<sup>18</sup> XML:DB

<sup>19</sup> XPath (XML Path)

<sup>20</sup> XML Infoset

<sup>21</sup> DOM (Document Object Model)

<sup>22</sup> SAX (Simple API for XML)

بخش دوم این تعریف مشخص می‌کند که این.ایکس.دیها اسناد ایکس.ام.ال را ذخیره و بازیابی می‌کنند. در واقع این.ایکس.دیها می‌توانند این مسئولیت را به تکه‌های<sup>۲۳</sup> سند ایکس.ام.ال نیز تعمیم دهند بدون این که تغییری در ماهیت عمل کرد یا تعریف آنها حاصل شود.

و بالاخره بخش سوم این تعریف بیان می‌دارد، لایه فیزیکی پایگاه داده اهمیتی ندارد. این مسأله کاملاً مشابه این وضعیت است که ماهیت رابطه‌ای بودن یک پایگاه داده ارتباطی به رابطه‌ای بودن محل ذخیره‌سازی داده‌ها ندارد.

## معماری‌های متداول برای این.ایکس.دی

معماری‌های متداول برای این.ایکس.دیها را می‌توان به دو گروه عمده تقسیم کرد: مبتنی بر متن<sup>۲۴</sup> و مبتنی بر مدل<sup>۲۵</sup>.

### معماری مبتنی بر متن

این معماری سند ایکس.ام.ال را به صورت متنی ذخیره می‌نماید. محل ذخیره‌سازی این متن می‌تواند در پایگاه داده یا بر روی حافظه سخت<sup>۲۶</sup> باشد. شاخص<sup>۲۷</sup>ها در این معماری برای جستجوی سریع و آسان بسیار متداول هستند و این فرصت را فراهم می‌آورند که بر اساس پرس و جوی مشخص اسناد سازگار به سرعت بازیابی شوند. این ویژگی به این گونه پایگاه‌های داده در شرایطی که تمام اسناد بازیابی می‌شوند، سرعت بالایی می‌دهد. این موضوع به این دلیل است که پایگاه داده می‌تواند با یک بار بازیابی یک شاخص تمام متن سند یا اسنادی که با پرس و جو سازگار هستند را بازیابی کند. اگر سند ایکس.ام.ال در پایگاه داده رابطه‌ای یا مبتنی بر مدل ذخیره شود، از آن جایی که هر بخش از سند در قسمتی مجزاً نگهداری می‌شود (که شاخص‌های مرتبط با خود را دارند)، برای بازیابی تمام متن سند باید چندین شاخص از نظر انطباق با پرس و جو مورد بررسی قرار گرفته و نتایج از این قسمت‌ها گردآوری و سرهم‌بندی شود تا تمام متن سند به دست آید. از این رو در بازیابی تمام متن اسناد معماری مبتنی بر متن از دیگر معماری‌های ذخیره‌سازی سریع‌تر عمل می‌نماید.

معماری مبتنی بر متن از جهاتی مشابه پایگاه‌های داده سلسله مراتبی<sup>۲۸</sup> هستند. هر دو در بازیابی داده بر اساس یک سلسله مراتب از پیش تعیین شده بهتر از پایگاه‌های داده رابطه‌ای عمل می‌نمایند و همچنین هر دو از کارایی پائینی در بازیابی داده در قالبی غیر از قالب پیشین برخوردارند. به بیان دیگر تجدید آرایش داده‌ها در این پایگاه‌های داده با مشکلاتی مواجه است.

<sup>23</sup> Fragment

<sup>24</sup> Text-Based

<sup>25</sup> Model-Based

<sup>26</sup> Hard Disc

<sup>27</sup> Index

<sup>28</sup> Hierarchal

## معماری مبتنی بر مدل

این معماری به جای ذخیره‌سازی سند ایکس.ام.ال به صورت متنی آن را با استفاده از یک مدل شیء گرا<sup>۲۹</sup> داخلی شرح داده و این مدل را ذخیره می‌نماید. این که این مدل چگونه ذخیره می‌شود به پایگاه داده بستگی دارد. برخی از پایگاه‌های داده مبتنی بر این معماری مدل دی.ام.را در پایگاه‌های داده رابطه‌ای ذخیره می‌نمایند.<sup>۳۰</sup>

ان.ایکس.دی‌هایی که از یک قالب ذخیره‌سازی اختصاصی استفاده می‌کنند در شرایطی که قصد بازیابی داده‌ها را با حفظ ساختار دارند، از کارآئی مشابهی با معماری مبتنی بر متن برخوردارند. این موضوع به این دلیل است که اشاره‌گرهایی را به صورت فیزیکی و در ساختار ذخیره‌سازی بین اجزاء ایجاد می‌نمایند. بدیهی است که بازیابی سند به صورت متنی در معماری مبتنی بر متن سریع‌تر و بازیابی مدل داده‌ای در معماری مبتنی بر مدل سریع‌تر است.

همانند معماری مبتنی بر متن، مشکلات مشابهی در کارآئی عملیات بازیابی داده‌ها در قالبی غیر از آنچه که مطابق آن ذخیره شده‌اند وجود دارد. این که در این خصوص نسبت به معماری مبتنی بر متن از مزایایی برخوردارند مشخص نیست.

## قابلیت‌های ان.ایکس.دی

در این بخش برخی از قابلیت‌هایی که در اغلب ان.ایکس.دیها یافت می‌شود، معرفی شده‌اند.

### گردایه اسناد<sup>۳۱</sup>

تمامی ان.ایکس.دیها به نحوی مفهوم گردایه اسناد را در خود گنجانده‌اند. گردایه اسناد نقشی مشابه جداول در پایگاه‌های داده رابطه‌ای ایفا می‌نماید و نسبت آن به یک سند معادل نسبت یک سطر به جدول است. در واقع اسناد ایکس.ام.ال مشابه در یک گردایه اسناد گنجانده می‌شوند. البته این تقسیم‌بندی یک تقسیم‌بندی منطقی است و هیچ محدودیت استواری بر روی ساختار اسناد اعمال نمی‌شود از این رو صرف قرار گرفتن اسناد در یک گردایه اسناد تضمینی بر هم‌ریختی آنها و مشابهت ساختاری نیست. برخی از ان.ایکس.دیها از گردایه‌های اسناد تودرتو پشتیبانی می‌نمایند و در نتیجه می‌توان اسناد ایکس.ام.ال را در یک ساختار درختی سازماندهی کرد.

## زبان‌های پرس و جو

تمامی ان.ایکس.دیها از یک یا چند زبان پرس و جو پشتیبانی می‌نمایند. متداول‌ترین این زبان‌ها ایکس.پس<sup>۳۲</sup> (با اضافاتی برای امکان پرس و جو بر روی چندین سند) و ایکس.کوئری<sup>۳۳</sup> هستند. برخی نیز از زبان‌های پرس و جو

<sup>29</sup> Object-Oriented

<sup>30</sup> See: *XML-L Mailing List* by Mark Birbeck.

<sup>31</sup> Document Collection

<sup>32</sup> XPath

اختصاصی برای اهداف خاص<sup>۳۴</sup> (مانند جستجوی تمام متن<sup>۳۴</sup>) استفاده می‌کنند. انتظار می‌رود در آینده اغلب این.ایکس.دیها از زبان ایکس.کوئری که زبانی ساخت‌یافته و کامل برای پرس و جو در اسناد ایکس.ام.ال است پشتیبانی نمایند.

### به‌روزرسانی<sup>۳۵</sup> و حذف<sup>۳۶</sup>

این.ایکس.دیها از راه‌بردهای متفاوتی برای به‌روزرسانی و حذف اسناد ایکس.ام.ال استفاده می‌نمایند. به طور کلی این راه‌بردها را می‌توان به سه دسته تقسیم نمود که عبارتند از: جایگزینی سند موجود با سند جدید یا حذف کامل آن، اعمال تغییرات به‌هنگام از طریق یک مدل داده به‌هنگام (مثل پی.دی.ام.آ<sup>۳۷</sup>) و استفاده از زبان‌های مارک‌آپ<sup>۳۸</sup> ویژه برای تشریح تغییرات. دو مورد از این زبان‌ها که تا حدی استاندارد شده‌اند عبارتند از:

- ایکس.آپ‌دیت<sup>۳۹</sup> که توسط گروه کاری ایکس.ام.ال:دی.بی<sup>۴۰</sup> ارائه شده یک زبان مبتنی بر ایکس.ام.ال است. این زبان از ایکس.بِس برای تعیین عناصر و مشخصه‌هایی که باید تغییر کنند استفاده نموده و مشخص می‌کند که این تغییرات باید چگونه اعمال شوند (با اضافه یا حذف عناصر یا مشخصه‌هایی قبل یا بعد از آنها). بسیاری از این.ایکس.دیها در حال حاضر از این زبان پشتیبانی می‌نمایند.
- مجموعه‌ای از اصلاحات و اضافات توسط گروه کاری ایکس.کوئری پیشنهاد شده است. این تغییرات در تعدادی از این.ایکس.دیها پیاده‌سازی شده و به نظر می‌رسد به عنوان مبنای تغییرات آتی ایکس.کوئری برای امکان حذف و به‌روزرسانی مورد استفاده قرار گیرد. انتظار می‌رود در آینده اغلب این.ایکس.دیها از این ویرایش اصلاح شده ایکس.کوئری برای این مقصود استفاده نمایند.

### تراکنش‌ها، قفل‌گذاری<sup>۴۱</sup> و هم‌زمانی<sup>۴۲</sup>

بسیاری از این.ایکس.دیها از تراکنش‌ها پشتیبانی می‌نمایند. مفهوم تراکنش‌ها کاملاً مشابه آن چیزی است که در مورد پایگاه‌های داده رابطه‌ای مطرح است. قفل‌گذاری اما تنها بر روی کل یک سند ایکس.ام.ال امکان‌پذیر است و امکان قفل‌گذاری بر روی اجزاء سازنده سند میسر نیست. از این رو سطح هم‌زمانی دسترسی کاربران به اسناد پائین است.

---

33 XQuery  
34 Full-Text Search  
35 Update  
36 Delete  
37 PDOM (Persistent DOM)  
38 Markup [Language]  
39 XUpdate  
40 XML:DB  
41 Locking  
42 Concurrency

مشکل اساسی در راه قفل‌گذاری بر روی اجزاء، پیاده‌سازی آن است. قفل‌گذاری یک جزء، قفل‌گذاری جزء والد آن را به دنبال خواهد داشت و این روند تا آنجا ادامه می‌یابد که به جزء ریشه‌ای برسد و در نهایت منجر به قفل‌گذاری بر روی تمام سند شود. برای حل این مسأله مطالعاتی انجام شده و راه‌حلهایی نیز پیشنهاد شده‌اند<sup>۴۳</sup> که انتظار می‌رود در آینده در اغلب این.ایکس.دیها پیاده‌سازی شوند.

## واسط‌های برنامه‌نویسی<sup>۴۴</sup>

تمامی این.ایکس.دیها واسط‌های برنامه‌نویسی‌ای را ارائه می‌نمایند. این واسط‌ها امکاناتی را برای اتصال به پایگاه داده، اجرای پرس و جوها و دستورات و دستیابی به نتایج و داده‌ها فراهم می‌آورند. نتایج اغلب به صورت متن سند ایکس.ام.ال یا مدل دی.ا.ام یا اس.ای.ایکس آن بازگردانده می‌شوند. اغلب این.ایکس.دیها از واسط‌های برنامه‌نویسی اختصاصی خود استفاده می‌کنند اما دو واسط برنامه‌نویسی متداول‌تر از سایرین است، این دو عبارتند از:

- واسط برنامه‌نویسی ایکس.ام.ال:دی.بی که توسط گروه کاری ایکس.ام.ال:دی.بی ارائه شده. این واسط برنامه‌نویسی به صورت مستقل از زبان طراحی شده (پیاده‌سازی این واسط برنامه‌نویسی به زبان جاوا توسط این گروه کاری ارائه شده است) و از ایکس.پس به عنوان زبان پرس و جو استفاده می‌نماید و در حال توسعه برای استفاده از ایکس.کوئری است. اغلب این.ایکس.دیها از این واسط برنامه‌نویسی پشتیبانی می‌نمایند.
- جی.اس.آر.آر<sup>۴۵</sup> یا ایکس.کیو.جی<sup>۴۶</sup>: که توسط جی.سی.پی<sup>۴۷</sup> ارائه شده و یک واسط برنامه‌نویسی مبتنی بر جی.دی.بی.سی<sup>۴۸</sup> است که از ایکس.کوئری به عنوان زبان پرس و جو استفاده می‌نماید. کار بر روی این واسط برنامه‌نویسی توسط دو شرکت آی.بی.ام<sup>۴۹</sup> و اوراکل<sup>۵۰</sup> هدایت می‌شود و انتظار می‌رود اولین پیاده‌سازی آن در محصولات این دو شرکت انجام شود.

برخی از این.ایکس.دیها از واسط‌هایی مبتنی بر وب سرویس<sup>۵۱</sup> و اس.ا.ی.پی<sup>۵۲</sup> را برای اجرای دستورات و دستیابی به داده‌ها از طریق پروتکل<sup>۵۳</sup> اچ.تی.تی.پی<sup>۵۴</sup> برخوردار هستند.

<sup>43</sup> See: *Stijn Dekeyser, et al.*

<sup>44</sup> APIs (Application Programming Interfaces)

<sup>45</sup> JSR (Java Specification Recommendation) 225

<sup>46</sup> XQJ (XQuery API for Java)

<sup>47</sup> JCP (Java Community Process)

<sup>48</sup> JDBC (Java Database Connectivity)

<sup>49</sup> IBM

<sup>50</sup> Oracle

<sup>51</sup> Web Service

<sup>52</sup> SOAP (Simple Object Access Protocol)

<sup>53</sup> Protocol

<sup>54</sup> HTTP

## گردش حلقوی<sup>۵۵</sup>

یکی از مهمترین قابلیت‌های این.ایکس.دی‌ها گردش حلقوی اسناد ایکس.ام.ال است. به این معنی که یک سند ایکس.ام.ال به همان صورت که ذخیره می‌شود، بازیابی خواهد شد. این قابلیت در مورد اسناد سندمحور بسیار مهم است زیرا تمامی اجزاء سند ایکس.ام.ال (اعم از متون، عناصر، مشخصه‌ها، یادداشت‌ها<sup>۵۶</sup> و دستورالعمل‌های پردازشی<sup>۵۷</sup>) نگهداری می‌شوند.

تمامی این.ایکس.دیها دارای قابلیت گردش حلقوی در حد کوچکترین اجزاء اسناد ایکس.ام.ال هستند. این که گردش حلقوی در چه سطحی اجراء می‌شود به معماری این.ایکس.دی بستگی دارد. در معماری مبتنی بر متن گردش حلقوی دقیق و کامل اعمال می‌شود، حال آن که در معماری مبتنی بر مدل سطح گردش حلقوی به مدلی که برای ذخیره و بازیابی مورد استفاده قرار می‌گیرد بستگی دارد. در این نوع معماری بسیار محتمل است که اسناد ایکس.ام.ال پس از ذخیره‌سازی در حالت کانونی<sup>۵۸</sup> بازیابی شوند.

## داده دور<sup>۵۹</sup>

برخی از این.ایکس.دیها می‌توانند در اسناد نگهداری شده از داده‌هایی را که در محل جداگانه‌ای ذخیره شده‌اند (معمولاً پایگاه‌های داده رابطه‌ای) استفاده نمایند. این کار از طریق واسط‌های برنامه‌نویسی ویژه (مانند جی.دی.بی.سی) انجام می‌شود و یک تبدیل بین داده‌های رابطه‌ای و اسناد ایکس.ام.ال به وجود می‌آید. این ارتباط می‌تواند زنده باشد، به این معنی که با تغییر سند ایکس.ام.ال داده‌های متناظر با آن در پایگاه داده رابطه‌ای دست‌خوش تغییرات شود.

## شاخص‌ها

تمامی این.ایکس.دیها از شاخص‌ها برای افزایش سرعت پرس و جو بهره می‌گیرند. این.ایکس.دیها از سه نوع مختلف از شاخص‌ها استفاده می‌نمایند. شاخص‌های مقداری<sup>۶۰</sup> که مقادیر متون و مشخصه‌ها را نگهداری نموده و برای پرس و جوهایی که بر اساس مقادیر اجزاء نوشته شده‌اند مورد استفاده قرار می‌گیرند. شاخص‌های ساختاری<sup>۶۱</sup> که موقعیت عناصر و مشخصه‌ها را ثبت می‌کنند و برای پرس و جوهایی که بر اساس موقعیت اجزاء نوشته شده‌اند مورد استفاده قرار می‌گیرند. شاخص‌های مقداری و شاخص‌های ساختاری برای تجزیه و تحلیل پرس و جوهای پیچیده‌تر اغلب با یکدیگر

<sup>55</sup> Round-Tripping

<sup>56</sup> Comments

<sup>57</sup> Process Instructions

<sup>58</sup> Canonical Form

<sup>59</sup> Remote Data

<sup>60</sup> Value Indexes

<sup>61</sup> Structural Indexes

تلفیق می‌شوند. در نهایت شاخص‌های تمام‌متن<sup>۶۲</sup> هستند که برای جستجوی تمام‌متن بر روی تمام مقادیر مورد استفاده قرار گرفته و یا با تلفیق با شاخص‌های ساختاری به کار گرفته می‌شوند. تمام این‌یکس‌دیها از شاخص‌های مقداری و شاخص‌های ساختاری پشتیبانی می‌کنند و فقط برخی از آنها از شاخص‌های تمام‌متن پشتیبانی می‌نمایند.

### موجودیت‌های خارجی<sup>۶۳</sup>

روش‌های مختلفی برای برقراری ارتباط با موجودیت‌های خارجی در پایگاه‌های داده وجود دارد. دو استاندارد مختلف برای ارتباط با موجودیت‌های خارجی در اسناد ایکس‌ام‌ال وجود دارد. استاندارد ایکس‌لینک<sup>۶۴</sup> برای برقراری ارتباط بین اجزاء یک سند ایکس‌ام‌ال با یک منبع خارجی (مثل یک تصویر یا چندرسانه‌ای<sup>۶۵</sup>) مورد استفاده قرار می‌گیرد و استاندارد ایکس‌پوینتر<sup>۶۶</sup> که برای ارتباط بین اجزاء یک سند ایکس‌ام‌ال با اجزاء یک سند ایکس‌ام‌ال دیگر مورد استفاده قرار می‌گیرد.

### به‌هنجارسازی<sup>۶۷</sup>، جامعیت ارجاعی<sup>۶۸</sup> و مقیاس‌پذیری<sup>۶۹</sup>

این بخش به مباحث مشترک بین این‌یکس‌دیها با پایگاه‌های داده رابطه‌ای پرداخته است.

### به‌هنجارسازی

به‌هنجارسازی عبارت است از فرآیند طراحی یک الگوی پایگاه داده که هر داده در آن فقط در یک جا ذخیره می‌شود. به‌هنجارسازی از مزایای فراوانی مانند کاهش فضای ذخیره‌سازی و از میان برداشتن احتمال ناپایداری در نگهداری داده‌ها برخوردار است. این مفهوم در پایگاه‌های داده رابطه‌ای یک مفهوم اساسی است که در این‌یکس‌دیها نیز به بازنگری دارد.

پیش از آن که بحث به‌هنجارسازی به صورت دقیق مورد بررسی قرار گیرد، ذکر این نکته مفید است که نابه‌نجاری<sup>۷۰</sup> اسناد سندمحور به طور کلی یک ایراد محسوب نمی‌شود. همانند پایگاه‌های داده رابطه‌ای در این‌یکس‌دیها نیز هیچ الزامی به رعایت به‌هنجاری<sup>۷۱</sup> وجود ندارد.

<sup>62</sup> Full-Text Indexes

<sup>63</sup> External Entity Storage

<sup>64</sup> XLink

<sup>65</sup> Multi-Media

<sup>66</sup> XPointer

<sup>67</sup> Normalization

<sup>68</sup> Referential Integrity

<sup>69</sup> Scalability

شیوه به‌هنجارسازی در این‌ایکس‌دیها به مقدار زیادی مشابه پایگاه‌های داده رابطه‌ای است. کافی است الگوی اسناد خود را به گونه‌ای طراحی کنید که هیچ داده‌ای تکرار نشود یا تکرارها به حداقل برسد. اسناد ایکس‌ام‌ال یک تفاوت اساسی با داده‌های رابطه‌ای دارند که به‌هنجارسازی آنها را ساده‌تر و انعطاف‌پذیرتر می‌نمایند. این تفاوت آن است که در اسناد ایکس‌ام‌ال می‌توان از اجزاء چندمقداری<sup>۷۳</sup> استفاده نمود.

در مسائل پیچیده‌تر اغلب طراحی الگویی با مشخصات مذکور ناممکن است. از این رو لازم است از روش‌های کامل‌تری برای به‌هنجارسازی اسناد استفاده کرد. یکی از این روش‌ها بهره‌گیری از استانداردهای ایکس‌لینک و ایکس‌پوینتر برای تفکیک اجزاء و موجودیت‌ها و برقراری ارتباط بین آنها است. البته این موضوع به این بستگی دارد که این‌ایکس‌دیها از استانداردهای مذکور پشتیبانی نمایند.

در عمل اغلب داده‌ها در مسائل واقعی، به منظور بهبود کارایی دارای مقداری نابه‌هنجاری هستند. نابه‌هنجاری داده‌ها آنچنان که به نظر می‌رسد نامطلوب نیست. شیوه و سطح به‌هنجاری‌سازی اسناد ایکس‌ام‌ال در این‌ایکس‌دیها به کاربرد و نوع (داده‌محور یا سندمحور) آنها بستگی دارد.

### جامعیت ارجاعی

جامعیت ارجاعی ارتباط نزدیکی با به‌هنجارسازی دارد. جامعیت ارجاعی بر حفظ اعتبار اشاره‌گرهای داده‌های مرتبط به یکدیگر به منظور حفظ پایداری پایگاه داده، دلالت دارد. در پایگاه‌های داده رابطه‌ای جامعیت ارجاعی به این معنی است که کلیدهای خارجی<sup>۷۳</sup> به کلیدهای اصلی<sup>۷۴</sup> معتبر اشاره کنند. در این‌ایکس‌دیها نیز جامعیت ارجاعی مفهومی مشابه دارد، به این معنا که ارجاعات در اسناد ایکس‌ام‌ال به اجزاء، اسناد یا تگ‌های اسناد معتبر باشد، از این رو جامعیت ارجاعی در اسناد ایکس‌ام‌ال کمی پیچیده‌تر از پایگاه‌های داده رابطه‌ای است.

اشاره‌گرها در اسناد ایکس‌ام‌ال اقسام گوناگونی دارند، برخی بر اساس انواع داده‌های استاندارد<sup>۷۵</sup> در الگوی اسناد تعریف می‌شوند، برخی دیگر با استفاده از ایکس‌لینک برقرار می‌شوند و در نهایت تعدادی نیز از روش‌های اختصاصی در مدل داده استفاده می‌کنند. دو مورد اول روش‌هایی مبتنی بر زبان و روش آخر روشی مبتنی بر پایگاه داده است. روش‌های مبتنی بر زبان در این‌ایکس‌دیها متداول‌تر هستند.

جامعیت ارجاعی در این‌ایکس‌دیها را می‌توان در دو گروه کلی بررسی نمود، یک پارچگی اشاره‌گرهای درونی (اشاره‌گرهای درون سند) و یک پارچگی اشاره‌گرهای موجودیت‌های خارجی (اشاره‌گرهای بین اسناد). جامعیت ارجاعی

<sup>70</sup> Abnormality

<sup>71</sup> Normality

<sup>72</sup> Multi-Valued

<sup>73</sup> Foreign Keys

<sup>74</sup> Primary Keys

<sup>75</sup> See: ID/IDREF & key/keyref

اشاره‌گرهای داخلی‌ای که از روش‌های اختصاصی و غیراستاندارد استفاده می‌کنند قابل بررسی و اعتبارسنجی نیست زیرا ره‌گیری و درک آنها توسط این.ایکس.دی میسر نیست، در حالی که با استفاده از شیوه‌های استاندارد از طریق اعتبارسنجی روابط و حفظ آنها امکان اعمال این قاعده به طور نسبی فراهم می‌شود. علت این نسبی بودن آن است که اغلب این.ایکس.دیها این اعتبارسنجی را تنها هنگام افزودن اسناد انجام می‌دهند. اگر عملیات به‌روزرسانی در سطح سند انجام شود، برای اعتبارسنجی کافی است، یک‌پارچگی اشاره‌گرهای درونی بررسی شوند. اما اگر به‌روزرسانی در سطح اجزاء انجام شود علاوه بر اعتبارسنجی پیشین باید کار بیشتری برای بررسی تمامی تغییرات انجام شود. تعداد اندکی از این.ایکس.دیها از این قابلیت برخوردارند.

جامعیت ارجاعی موجودیت‌های خارجی اغلب پشتیبانی نمی‌گردد. اگر موجودیت‌های خارجی در محلی خارج از پایگاه داده ذخیره شوند ضرورتی به بررسی و اعتبارسنجی آنها وجود ندارد. بدیهی است که پایگاه داده توانایی نظارت بر این موجودیت‌ها را نداشته و از این رو الزام چنین رابطه‌ای بی‌معنی است.

### مقیاس پذیری

این.ایکس.دیها همانند پایگاه‌های داده رابطه‌ای و سلسله‌مراتبی از شاخص‌ها برای یافتن داده‌ها استفاده می‌نمایند. این بدان معنا است که سرعت پیدا کردن داده‌ها بیش از آن که تابعی از حجم اسناد باشد، تحت تأثیر حجم شاخص‌ها قرار دارد. این پایگاه‌های داده می‌توانند با استفاده از همان فن‌آوری و روش‌هایی که در پایگاه‌های داده رابطه‌ای یا سلسله‌مراتبی مورد استفاده قرار می‌گیرند، شاخص‌هایی از اسناد را تهیه نموده و در ازای یک پرس و جوی مشخص با همان سرعت سند یا اسناد مرتبط را مشخص نماید. بنابراین از این منظر این.ایکس.دیها همانند پایگاه‌های داده رابطه‌ای مقیاس‌پذیرند.

این.ایکس.دیها همانند پایگاه‌های داده سلسله‌مراتبی و بر خلاف پایگاه‌های داده رابطه‌ای پیوندهای فیزیکی بین داده‌های مرتبط برقرار می‌سازند. مشخصاً در معماری مبتنی بر متن داده‌های مرتبط به صورت فیزیکی در کنار یکدیگر و در یک گروه نگهداری می‌شوند و در معماری مبتنی بر مدل، که از یک مدل اختصاصی برای نگهداری داده‌ها استفاده می‌کنند نیز پیوندهای فیزیکی بین داده‌های مرتبط برقرار می‌گردد. از آنجا که پیوندهای فیزیکی سریع‌تر از پیوندهای منطقی عمل می‌کنند، این.ایکس.دیها همانند پایگاه‌های داده سلسله‌مراتبی اسناد ایکس.ام.ال را با سرعتی بیش از پایگاه‌های داده رابطه‌ای بازیابی نمایند. از این جهت این.ایکس.دیها همانند پایگاه‌های داده سلسله‌مراتبی مقیاس‌پذیرند.

متأسفانه این مقیاس‌پذیری محدود است. همانند پایگاه‌های داده سلسله‌مراتبی ذخیره‌سازی فیزیکی در این.ایکس.دیها تنها به یک ساختار خاص محدود می‌شود و بازیابی داده‌ها تنها در محدوده این ساختار مشخص بسیار سریع است و در ساختارهایی غیر از آن از چنین مزیتی برخوردار نیست.

برای فائق آمدن بر این مشکل، این.ایکس.دیها استفاده زیادی از شاخص‌ها کرده و اغلب تمامی عناصر و مشخصه‌ها را در شاخص‌ها نگهداری می‌نمایند. اگرچه این کار زمان جستجو و بازیابی را بهبود می‌بخشد اما زمان به‌روزرسانی را به منظور نگهداری شاخص‌ها افزایش می‌دهد. این مسأله شاید در شرایطی که اسناد فقط (یا در بدترین حالت اغلب) جهت خواندن مورد استفاده قرار می‌گیرند بروز نکند، اما در محیط‌های پرتراکنش بسیار حادّ خواهد بود. در نهایت این که این.ایکس.دیها در جستجوی داده‌های بدون شاخص بسیار ضعیف‌تر از پایگاه‌های داده‌ی رابطه‌ای عمل می‌نمایند. این.ایکس.دیها و پایگاه‌های داده‌ی رابطه‌ای هر دو باید با پیمودن یک فرآیند خطّی به دنبال داده‌ها بگردند و از آنجا که داده‌ها در این.ایکس.دیها کمتر به‌هم‌نجار شده‌اند این فرآیند طولانی‌تر خواهد بود.

### پی‌نوشت‌ها و مراجع

برای جزئیات بیشتر همچنین رجوع کنید به:

متن اصلی مقاله‌ی رونالد بارت؛

<http://www.rpbouret.com/xml/XMLAndDatabases.htm>

<http://www.rpbouret.com/xml/UseCases.htm>

فهرستی از محصولات مرتبط نیز که توسط رونالد بارت تهیه شده است؛

<http://www.rpbouret.com/xml/XMLDatabaseProds.htm>

در مورد ایکس.ام.ال و فن‌آوری‌های وابسته به نشانی‌های زیر رجوع کنید؛

<http://www.w3.org/XML/>

<http://www.w3schools.com/xml/>

<http://www.w3schools.com/dtd/>

<http://www.w3.org/XML/Schema>

<http://www.xml.com/pub/a/2000/11/29/schemas/part1.html>

<http://www.w3.org/TR/xpath>

<http://www.w3schools.com/xpath/>

<http://www.w3.org/XML/Query>

<http://www.w3schools.com/xquery/>

<http://www.xml.com/pub/a/2002/10/16/xquery.html>

<http://www.w3.org/XML/Linking>

<http://www.xml.com/pub/a/2000/09/xlink/index.html>

<http://www.w3schools.com/xlink/>

نشانی‌های زیر اطلاعات مفیدی در خصوص مدل‌های مورد استفاده در ایکس.ام.ال ارائه می‌نمایند؛

<http://www.w3.org/DOM/>

<http://www.w3schools.com/dom/>

<http://www.saxproject.org/>

<http://www.xml.com/pub/p/626>

وب‌گاه رسمی گروه کاری ایکس.ام.ال:دی.بی، در این وب‌گاه می‌توانید اطلاعات مفیدی در خصوص ایکس.آپ‌دیت و واسط برنامه‌نویسی بیابید؛

<http://xmldb-org.sourceforge.net/>

وب‌گاه رسمی جی.سی.پی، در این وب‌گاه می‌توانید در مورد مأموریت، روش انجام کار و کارهای در دست اجرای این گروه کاری اطلاعاتی به دست بیاورید؛

<http://www.jcp.org/en/home/index>

جزئیات کامل واسط برنامه‌نویسی جی.اس.آر ۲۲۵؛

<http://www.jcp.org/en/jsr/detail?id=225>

اطلاعات مفیدی در مورد معماری ایکس.ام.ال-ال.

<https://listserv.heanet.ie/xml-1.html>