



اس.جی.ام.ال^۱ و تحول فهرست‌نویسی ماشین خوان

نیلوفر برهمند^۲

چکیده

در عصر شبکه‌ها، سازماندهی و ارائه اطلاعات در همه جنبه‌ها دچار تغییرات زیادی شده‌اند. این تغییرات نه تنها روش‌ها و فناوری‌های مورد استفاده را در بردارد، بلکه استانداردهای مورد استفاده برای ارائه و جستجوی اطلاعات را نیز شامل می‌شود. استاندارد مارک در زمینه‌های زیادی نتوانسته است با محیط اطلاعاتی جدید همسو شود. اس.جی.ام.ال به دلیل انعطاف‌پذیری زیاد می‌تواند از بسیاری جهات نواقص مارک را جبران کند و به مثابه جایگزین و یا مکملی برای مارک به کار رود.

مقاله حاضر مارک را به طور خلاصه معرفی می‌کند و نواقص آن را در محیط جدید اطلاعاتی برمی‌شمرد. همچنین به معرفی اس.جی.ام.ال و ویژگی‌های آن می‌پردازد و کاربرد مارک در اس.جی.ام.ال را بررسی می‌کند.

کلیدواژه‌ها

نشانه‌گذاری، پیشینه‌کتابشناختی، مارک، اس.جی.ام.ال

مقدمه

همواره سازماندهی و ارائه اطلاعات از مهم‌ترین وظایف کتابخانه‌ها بوده است. در عصر شبکه‌ها، سازماندهی و ارائه اطلاعات در همه جنبه‌ها دچار تغییرات زیادی شده‌اند. این تغییرات نه تنها روش‌ها و فناوری‌های مورد استفاده برای ایجاد فهرست‌ها و نمایه‌های تسهیل‌کننده دسترسی به اطلاعات را دربرمی‌گیرند، بلکه استانداردهایی که در موفقیت و اثربخشی ابزارهای جستجو و بازیابی اطلاعات نقشی

اساسی دارند شامل می‌شوند. امروزه دیگر فهرست کتابخانه‌ها فقط ابزاری برای دسترسی به مجموعه و کمک به استفاده‌کنندگان کتابخانه‌ای خاص نیستند، بلکه گره‌هایی از یک شبکه هستند که استفاده‌کنندگان در هر جای جهان که باشند می‌توانند با استفاده از رایانه‌هایشان، اتصال به اینترنت، و پیوستن به فضای جهانی اطلاعات، به آنها دست یابند (۱۲: ۸۳-۸۷).

ورود به فضای جهانی اطلاعات در گرو استانداردهایی است

که ارتباط انواع مختلف داده‌ها را در نظام‌های نامتجانس میسر می‌سازند. «تاریخچه کتابداری مدرن با ایجاد، توسعه و پذیرش استانداردها اجین بوده است» (۳:۴).

در این میان شاید بتوان گفت: «اختراع پیشینه‌های مارک و استانداردهای مرتبط یکی از مهم‌ترین رویدادهای تاریخ کتابداری بوده است» (۴).

اختراع پیشینه‌های مارک، تولید پایگاه‌های اطلاعاتی چندملیتی را برای محققان و دانشمندان ممکن ساخت، کتابخانه‌ها را قادر ساخت که در فعالیت‌های خود از امکانات نظام‌های خودکار بهره بگیرند، استقرار پروتکل‌های استاندارد را در ارتقاء بخشیدن به تسهیم و پردازش اطلاعات ضروری ساخت، و کتابخانه‌ها، آرشیوها و سایر مراکز اطلاعاتی را در صف مقدم انقلاب اطلاعات الکترونیکی قرار داد. با وجود این در بیشتر زمینه‌ها، کتابخانه‌ها دیگر در صف مقدم این انقلاب اطلاعاتی قرار ندارند. محیط اطلاعات الکترونیکی در بیرون از کتابخانه‌ها به نحوی که بیشتر کتابداران با آن آشنا هستند ولی از درک ابعاد آن ناتوانند در حال تحول است. وظیفه پیش‌روی کتابداران به مثابه نگهدارندگان و عرضه‌کنندگان اطلاعات محققانه، تلاش همه‌جانبه برای استفاده از ابتکاراتی است که به‌طور روز افزون توسعه می‌یابند. کتابداران برای تطبیق پایگاه‌ها و نظام‌های گسترده‌شان با محیط جدید یعنی اینترنت، وب و نظام‌های جایگزین آنها اجباری حرفه‌ای دارند. متأسفانه در بیشتر جهات داده‌های مارک در نظام‌هایی که برای پشتیبانی از خودکارسازی عملیات سنتی کتابخانه‌ای طراحی شده‌اند محبوس شده‌اند. راه‌گریز داده‌های مارک استفاده از پروتکل زد ۳۹/۵۰ - پروتکل جستجو و بازیابی بین سیستمی - است که هنوز و شاید در آینده نیز مانند قیف باریکی باشد به دنیای بزرگ‌تر اطلاعاتی (۴).

برای رفع نواقص موجود در مارک و پیوستن به محیط جدید اطلاعات الکترونیکی و استفاده بهینه از آن تلاش‌های زیادی شده است. در این میان استاندارد بین‌المللی برای زبان نشانه‌گذاری اس.جی.ام.ال، تحسولی اساسی در فعالیت‌های توصیفی و دسترسی ایجاد کرده است. اس.جی.ام.ال ابزاری است که رویارویی با تغییرات روزافزون در محیط اطلاعات الکترونیکی را تسهیل کرده و از جهات متعددی می‌تواند نواقص مارک را جبران کند.

«استفاده از ویرایش‌های "مارک - اس.جی.ام.ال" کتابخانه‌ها را قادر می‌سازد تا در حالی که از قالب مارک و

نظام‌های مبتنی بر آن استفاده می‌کنند، داده‌های خود را به محیط جدید اطلاعات الکترونیکی وارد کنند بدون اینکه مجبور باشند با مشکلاتی که قالب مارک و نظام‌های مبتنی بر آن هرگز برای مقابله با آنها طراحی نشده‌اند روبه‌رو شوند» (۴).

مارک چیست و چرا مهم است؟

طبق استاندارد بین‌المللی ایزو ۲۷۰۹، مارک سرواژه‌ای برای کلمات Machine Readable Cataloging به معنای فهرست‌نویسی ماشین‌خوان است. ماشین‌خوان به معنای این است که نوع خاصی از ماشین (رایانه) می‌تواند داده‌های پیشینه‌های فهرست‌نویسی را بخواند و تفسیر کند (۶).

پیشینه فهرست‌نویسی، پیشینه کتابشناختی یا اطلاعاتی است که به‌طور سنتی بر روی کارت فهرست‌نویسی قرار می‌گیرد. پیشینه‌ها شامل (۱) توصیف مدرک، (۲) سرشناسه و شناسه‌های افزوده، (۳) سرعنوان موضوعی، و (۴) شماره رده‌بندی و یا شماره بازیابی است (پیشینه‌های مارک اغلب اطلاعات اضافی دیگر را نیز شامل می‌شوند) (۶).

اطلاعات کارت فهرست‌نویسی را نمی‌توان به‌سادگی در رایانه حروفچینی و فهرست خودکار تولید کرد. رایانه به وسیله‌ای برای تفسیر اطلاعات موجود در کارت فهرست‌نویسی نیاز دارد. این وسیله کدگذاری نام دارد و مارک یکی از استانداردهای کدگذاری پیشینه‌هاست. با استفاده از مارک می‌توان داده‌های موجود در یک فهرست‌برگه را به‌گونه‌ای آماده ساخت که رایانه آنها را بخواند و بنا بر دستور بر روی هر داده به‌گونه‌ای مناسب عمل کند. «پیشینه مارک راهنمایی به داده‌های فهرست‌نویسی است و شامل علائم یا نشانه‌هایی است که قبل از هر قطعه اطلاعات کتابشناختی قرار می‌گیرد» (۶)

کدگذاری پیشینه‌ها سه مزیت عمده دارد:

۱. کدگذاری پیشینه‌ها این امکان را فراهم می‌کند که هر بخش از پیشینه‌ها به‌طور مجزا پردازش شود. به این ترتیب می‌توان برنامه‌های رایانه‌ای را به‌نحوی نوشت که هر منطقه بدان صورت که افراد می‌خواهند در جای مشخص قرار گرفته و نمایش داده شود.
۲. کدگذاری، دسترسی به پیشینه‌ها را امکان‌پذیر می‌سازد. بر این اساس می‌توان برنامه‌ای برای جستجو در مناطق مختلف نوشت.

۳. با کدگذاری می‌توان بسیاری از زبان‌ها و خط‌ها را برای نمایش و جستجو در بایگانی واحد ادغام کرد. مراکزی که اطلاعات را سازماندهی می‌کنند همواره با زبان‌ها و خط‌های مختلف سروکار دارند. زبان‌هایی که به خط رومی هستند همواره قابلیت درهم‌کرد دارند، اگرچه الفبایی کردنشان معمولاً دشوار است. زبان‌های با خطوط دیگر همواره باید به خط رومی آوانویسی شوند تا بتوانند با دیگر زبان‌ها الفبایی شوند. اگر نمایش خطوط غیررومی به روش پیوسته فراهم باشد، کدگذاری اجازه می‌دهد که مناطق پیشینه‌ها بدون دانش انسان‌هایی که سازماندهی می‌کنند، تعیین هویت شوند. کدگذاری پیشینه‌ها از طریق برچسب‌ها، نمرات، حروف و کلمات انجام می‌شود. برای مثال نام پدیدآور در کدگذاری مارک با نشانه ۱۰۰، در اچ.تی.ام. ال با آغازه «author» و پایانه «author» کدگذاری می‌شود (۲:۶۳).

به‌طور منطقی هر پیشینه کتابشناختی به حوزه‌های مختلف تقسیم می‌شود. مثلاً حوزه نویسنده، حوزه اطلاعات و حوزه عنوان. این حوزه‌ها به حوزه‌های فرعی دیگری نیز تقسیم می‌شوند. از آن‌جا که اسامی متنی پیشینه‌ها برای همراهی با مارک بسیار طولانی خواهد بود، هر حوزه با کدهای سه رقمی همراهی می‌شود. این نکته را نباید از نظر دور داشت که اسامی حوزه‌ها که در فهرست‌های پیوسته نمایش داده می‌شوند، توسط نرم‌افزار سیستم و نه پیشینه مارک ارائه می‌شوند (۶).

مثال زیر نحوه کدگذاری و نمایش پیشینه‌ها در مارک را نشان می‌دهد:

100 taga Personal name entry (author)

100 # Prising Robert M.100\$

اگر پیشینه‌های کتابشناختی درست کدگذاری و در فایل‌های رایانه‌ای ذخیره شوند می‌توان برنامه‌های رایانه‌ای را به نحوی نوشت که اطلاعات را به درستی برای چاپ برگه‌های فهرست‌نویسی یا نمایش آنها بر روی نمایشگر نقطه‌گذاری و قالب‌بندی کنند. همچنین می‌توان برنامه‌هایی را برای جستجو و بازیابی انواع معینی از اطلاعات موجود در مناطق خاص نوشت و فهرستی از نتایج جستجوها را نشان داد. هر کتابخانه می‌تواند روش سازماندهی خاص خود را برای اطلاعات کتابشناختی ابداع کند، ولی با این کار کتابخانه مهجور شده و ناگزیر به انجام کارهای اضافه می‌شود. استفاده از استاندارد مارک از دوباره کاری جلوگیری می‌کند و

کتابخانه‌ها را قادر می‌سازد که داده‌های فهرست‌نویسی مطمئن را فراهم کنند. اگر کتابخانه‌ای بخواهد نظام خاص خود را بدون استفاده از پیشینه‌های مارک بسازد، استاندارد را که هدف اولیه‌اش گسترش تبادل اطلاعات است از دست خواهد داد (۶).

استفاده از استاندارد مارک کتابخانه‌ها را قادر می‌سازد که از نظام‌های خودکار کتابخانه‌ای تجاری موجود برای مدیریت فعالیت‌های کتابخانه‌ای استفاده کنند. بسیاری از نظام‌های کتابخانه‌ای برای کار با قالب مارک طراحی شده‌اند. این نظام‌ها توسط کارگزاران یا فروشندگان حفاظت می‌شوند و توسعه می‌یابند. کتابخانه‌ها می‌توانند با بهره‌گیری از مارک از آخرین دستاوردهای فناوری استفاده کنند (۶).

با توسعه شبکه‌ها و رشد فزاینده مواد اطلاعاتی الکترونیکی، نواقص مارک در مواجهه با محیط جدید اطلاعات الکترونیکی مشخص شد. استفاده‌کنندگان و توسعه‌دهندگان مارک، اس.جی.ام. ال را جایگزین و یا مکملی برای مارک معرفی کردند و تلاش‌های زیادی برای توسعه ویرایش‌های مارک-اس.جی.ام. ال انجام دادند: «در حقیقت دلیل ایجاد و توسعه ویرایش‌های مارک - اس.جی.ام. ال این بود که استفاده‌کنندگان در بسیاری از موقعیت‌ها اس.جی.ام. ال را دارای ساختار مناسب‌تری نسبت به مارک می‌یافتند» (۱۰).

پیوستن به شبکه‌ها و نظام‌های مبتنی بر وب، فضای بیشتری را برای تجربه در فعالیت‌های فهرست‌نویسی ایجاد می‌کند و استفاده از موتورهای بازیابی جدید و ایجاد نظام‌های فراداده پیچیده را فراهم می‌سازد. توسعه ویرایش‌های مارک-اس.جی.ام. ال در نهایت می‌تواند پایگاه‌های اطلاعاتی مارک را همچون قسمتی از کل ساختار اطلاعاتی وب به هم متصل کند (۹:۲۳۵).

مسائل مارک

با وجود تلاش‌هایی که کتابداران انجام داده‌اند، پیشینه مارک همچنان به صورت ویرایش الکترونیکی فهرست برگه باقی مانده است. این مسئله محدودیت‌هایی را به همراه دارد. وقتی مارک در اواسط دهه ۱۹۶۰ طراحی شد، کارکرد اولیه‌اش فراهم کردن امکان توزیع الکترونیکی پیشینه‌های فهرست‌نویسی در پشتیبانی از تولید کارت‌ها بود. فهرست برگه پیشینه‌ای یکپارچه برای شیء فیزیکی است که شامل انواع گوناگون داده‌های کتابشناختی می‌شود. برنامه‌هایی که

اطلاعات کامل کتابشناختی را همراه با اطلاعات غیرکتابشناختی ارائه می‌دهند، به تجزیه و تحلیل‌هایی که توسط ماشین یا انسان انجام می‌شود کمک می‌کنند. در پیشینه مارک برای بیشتر این اطلاعات مکانی در نظر گرفته نشده است. اگر هم لازم باشد این اطلاعات گنجانده شود باید در فیلد یادداشت‌های ساخت نیافته^۳ (xx5) قرار گیرد. این فیلد ساختاری برای بازیابی ماشین و تجزیه و تحلیل اطلاعات ارائه نمی‌دهد (۵).

مارک نمی‌تواند اطلاعات ساختاریافته سلسله مراتبی را در خود بگنجانند. در وهله اول مارک برای ذخیره کردن توصیف‌ها و اطلاعات دسترسی که برای اقلام کتابشناختی مجزا به کار می‌رود طراحی شد. توصیف و فراهم کردن دسترسی به مجموعه‌های پیچیده با سطوح مختلف تحلیل، بار اضافی است بر ساختار مارک. می‌توان سطح ثانویه توصیف را در ساختار مارک قرار داد ولی نوع اطلاعاتی که در سطوح مختلف تحلیل به کار می‌رود بسیار محدود است. یکی از راه‌های ممکن برای این مشکل به کار بردن پیشینه‌های چندگانه است که به صورت سلسله‌مراتبی به هم مرتبط هستند و به هم پیوند دارند. استفاده از پیشینه‌های چندگانه از لحاظ کنترل، مشکلات میان‌سیستمی ایجاد می‌کند که هرگز در طراحی مارک برای این مشکلات راه حلی پیش‌بینی نشده است.

رویکرد یک‌لایه‌ای به فهرست‌ها در حال از بین رفتن است. چون میلیون‌ها اقلام اطلاعاتی از طریق شبکه‌ها دسترسی پذیر شده‌اند و پیوندهایی که در شبکه، بین کتابشناسی‌های پیوسته و پیشینه‌های کتابشناختی ایجاد شده است قدرت فهرست‌ها را فراتر از تصور کتابداران برده است. در محیط الکترونیکی برای دانستن اینکه یک اثر خاص در کتابخانه وجود دارد و یا اینکه چه آثاری در کتابخانه موجود است، در هر فهرست رویکردی سلسله‌مراتبی مورد نیاز است تا استفاده‌کنندگان بتوانند به راحتی از یک سطح به سطح دیگر حرکت کنند. مثل اثر، ویرایش‌های مختلف اثر، قسمت‌های مختلف اثر، متن کامل اثر، انواع آثار مرتبط و ... از آنجا که رویکرد یک‌لایه‌ای در فهرست‌های آینده مؤثر نخواهد بود این فهرست‌ها باید چندلایه‌ای و انعطاف‌پذیر باشند (۵).

مارک نمی‌تواند نیاز نظام‌های کتابخانه‌ای جدید را برآورده کند. البته این به معنای قابل استفاده نبودن مارک نیست. چون میلیون‌ها پیشینه کتابشناختی در قالب مارک کدگذاری شده‌اند. هزینه، زمان و منابع بسیاری برای ساخت پیشینه‌ها

صرف شده است، ولی تحولی که در دنیای اطلاعات رخ داده است نشان می‌دهد که مارک دیگر تنها قالب برای کدگذاری پیشینه‌ها در نظام‌های کتابخانه‌ای نخواهد بود.

مارکی که در حال حاضر برای کدگذاری پیشینه‌ها استفاده می‌شود به قدر کافی برای کتابخانه‌ها انعطاف‌پذیر به نظر نمی‌رسد تا آنها بتوانند از فناوری‌های بازیابی اطلاعات به‌ویژه شبکه جهانی وب که مرتباً در حال رشد است استفاده کنند. اس.جی.ام.ال برای رفع این محدودیت‌ها راه‌حلی را ارائه می‌دهد. اس.جی.ام.ال بسیار انعطاف‌پذیر است و همچون استاندارد مبادله برای هر نوع داده ساختاریافته می‌تواند برای مبادله اطلاعات کتابشناختی نیز به کار رود.

اس.جی.ام.ال و رابطه آن با مارک

اس.جی.ام.ال در ۱۹۷۰ به مثابه زبان نشانه‌گذاری عمومی^۴ شکل گرفت و در ۱۹۸۶ به استاندارد بین‌المللی ایزو ۸۸۷۹ تبدیل شد. این استاندارد به‌نحو گسترده‌ای در امریکا، اروپای غربی و ژاپن پذیرفته شده است و کاربردهای تجاری، صنعتی و دانشگاهی فراوانی دارد. سازمان‌های بسیاری چون سازمان جهانی استاندارد، شبکه تحقیقاتی آلمان، وزارت انرژی امریکا و سازمان‌هایی از این دست از این استاندارد استفاده می‌کنند (۱:۴۵).

اس.جی.ام.ال از جامعه ناشران متولد شد. هدف اولیه ایجاد 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 (۸).

اس.جی.ام.ال استاندارد برای توصیف متون الکترونیکی نشانه‌گذاری شده است و در واقع یک فرازبان است. به این معنا که دیگر زبان‌های نشانه‌گذاری را توصیف می‌کند. «به عبارت دیگر اس.جی.ام.ال مجموعه‌ای از قواعد و چارچوب‌ها برای توصیف زبان‌های نشانه‌گذاری خاص است» (۵).

از دیدگاه تاریخی، واژه نشانه‌گذاری به گزارمان‌ها یا نشانه‌های درون یک متن اطلاق می‌شد که ماشین‌نویس یا حروفچین را در چگونگی صفحه‌آرایی یا چاپ یک متن کمک می‌کردند. برای مثال خط‌های مودج‌دار زیر حروف برای حروف بزرگ، نشانه‌های مخصوص برای متنی که باید حذف شوند یا با خط‌های خاصی چاپ شوند و نظایر آنها. در روند خودکار شدن قالب‌بندی و چاپ، واژه نشانه‌گذاری، انواع

3.Unstructured Notes Fiee 1

4.GML= Generalized Markup Language

رمزهای نشانه‌گذاری را که به متون الکترونیکی متصل می‌شد تا قالب‌بندی، چاپ و سایر پردازش‌ها را هدایت کند دربرگرفت. «با عمومیت دادن مفهوم فوق می‌توان نشانه‌گذاری را هر نوع وسیله‌ای که تفسیر یک مدرک را واضح می‌سازد تعریف کرد» (۵).

دو نوع نشانه‌گذاری وجود دارد، نشانه‌گذاری رویه‌ای^۵ و نشانه‌گذاری توصیفی^۶. نشانه‌گذاری رویه‌ای عبارت از دستورالعمل‌هایی است که به دستگاه حروفچین داده می‌شود و در آن چگونگی صفحه‌آرایی متن، سبک و اندازه حروف مورد استفاده، عناصر خاص حروفچینی و مسائلی از این قبیل توضیح داده می‌شود. نشانه‌گذاری رویه‌ای به محتوای مدرک کاری ندارد و تمرکزش بر ظاهر فیزیکی مدرک است. بیشتر نرم‌افزارهای کاربردی واژه‌پرداز از نشانه‌گذاری رویه‌ای استفاده می‌کنند که فقط محدود به همان نظام است. به این معنا که نظام کدهای مخصوص به خود را دارد که فرایندها را در آن نظام یا در نرم‌افزارهای مربوط انجام می‌دهد. به‌طور کلی این نوع نشانه‌گذاری با یک نتیجه واحد در ذهن طراحی شده است، برای مثال صفحه‌های چاپی با یک سبک خاص. اگر لازم باشد محتوای یک مدرک با سبک و قالب دیگری دوباره استفاده شود، باید کدهای نشانه‌گذاری اولیه پاک و کدهای جدید برای قالب جدید اضافه شوند. بسیاری از کتابدارانی که پیش از یک نظام کتابخانه‌ای پیوسته را به کار برده‌اند می‌توانند تأیید کنند که وقتی نرم‌افزار نظام تغییر می‌کند ترجمه داده‌ها از یک نظام به نظام دیگر بسیار هزینه‌بر و وقت‌گیر خواهد بود. مبادله مدارک مبتنی بر نشانه‌گذاری رویه‌ای فقط زمانی که فرستنده و گیرنده از یک نظام استفاده می‌کنند امکان‌پذیر است (۵).

نشانه‌گذاری توصیفی، ساختار و یا محتوای یک مدرک را به جای ظاهر فیزیکی‌اش بر روی صفحه یا مانیتور توصیف می‌کند. بنابراین محتوای مدرک از سبک ارائه‌اش جدا خواهد بود. عناصر درون یک مدرک مثل فصل، بیت، پانویس و کتابشناسی با استفاده از کدهایی که مشخص می‌کنند که آن عنصر چیست (نه اینکه ظاهرش چه باشد) طبقه‌بندی می‌شوند. محتوای مدارکی که به‌صورت توصیفی نشانه‌گذاری می‌شوند می‌توان به سبک‌های مختلف ارائه کرد و با اهداف مختلف استفاده نمود (۵).

یک زبان نشانه‌گذاری باید نشانه‌گذاری مجاز، نشانه‌گذاری‌های مورد نیاز، چگونگی تشخیص نشانه‌گذاری

از متن، و معنای نشانه‌گذاری را مشخص کند. اس. جی. ام. ال سه مورد اولیه را تحت پوشش قرار می‌دهد. علاوه بر این اس. جی. ام. ال سه ویژگی بارز دارد که آن را از سایر زبان‌های نشانه‌گذاری متمایز می‌کند.

این سه ویژگی عبارتند از:
- تأکید آن بر نشانه‌گذاری توصیفی به جای نشانه‌گذاری رویه‌ای.

- استقلال آن از نظام‌های ارائه متون.

- مفهوم نوع مدرک (۱۳).

نظام نشانه‌گذاری توصیفی از کدهای نشانه‌گذاری استفاده می‌کند. این کدها اسامی را برای دسته‌بندی کردن قسمت‌های طبقه‌بندی شده یک مدرک ارائه می‌دهند. کدهای نشانه‌گذاری مثل <para> یا <end list> به سادگی بخشی از مدرک را شناسایی می‌کنند و مثلاً تأیید می‌کنند که «مورد بعدی پاراگراف است» و یا «این انتهای فهرستی است که اخیراً شروع شده است». برخلاف نشانه‌گذاری توصیفی، نشانه‌گذاری رویه‌ای پردازش‌های خاصی را که باید در بخش‌های ویژه یک مدرک انجام شود تعریف می‌کند. مثلاً «یک خط را جا بگذار و به حاشیه جدید سمت چپ برو». در اس. جی. ام. ال دستورالعمل‌های مورد نیاز برای پردازش یک مدرک برای اهداف خاص مثلاً قالب‌بندی مدرک، از نشانه‌گذاری توصیفی درون مدرک متفاوت است. عموماً این دستورالعمل‌ها از رویه‌ها و برنامه‌های خارج از مدرک گردآوری می‌شوند.

با استفاده از نشانه‌گذاری توصیفی، می‌توان مدارک را با قطعه‌های نرم‌افزاری فراوان و متفاوتی پردازش کرد. هر کدام از این قطعه‌ها می‌توانند دستورالعمل‌های پردازش متفاوتی را در بخش‌هایی که مناسب به نظر می‌رسند به کار گیرند. مثلاً یک برنامه تحلیل محتوا ممکن است از همه پانویس‌های یک متن صرف‌نظر کند، در حالی که برنامه قالب‌بندی ممکن است همه آنها را برای چاپ در انتهای فصل استخراج و جمع‌آوری کند. دستورالعمل‌های مختلف پردازش می‌توانند به قسمت‌های معین از یک فایل مربوط شوند. برای مثال یک برنامه ممکن است اسامی اشخاص و مکان‌ها را از یک مدرک برای تولید یک نمایه یا پایگاه اطلاعاتی استخراج کند درحالی‌که برنامه دیگر که در همان متن عمل می‌کند ممکن است اسامی اشخاص و مکان‌ها را با حروف متفاوت از حروف متن اصلی چاپ کند (۱۳).

استقلال داده‌ها

خاصی از مدرک را تعریف می‌کند، یعنی به تعریف موارد زیر می‌پردازد:

- همه عناصری که ممکن است بخشی از نوع خاصی از مدرک باشند؛

- اسامی عناصر و تکرارپذیری آنها؛

- عناصر با چه نظمی باید قرار بگیرند؛

- محتوای عناصر به‌طور کلی و نه به‌طور دقیق؛

- چه انواعی از نشانه‌گذاری می‌توانند حذف شوند؛

- صفات نشانه‌ها و ارزش‌های از پیش تعیین شده آنها؛

- اسامی شناسه‌های مجاز (۲: ۸۳).

با استفاده از اس.جی.ام.ال می‌توان برای انواع مدارک

تعریف نوع مدرک کرد. برای مثال ا.ج.تی.ام.ال^{۱۰}

تعریف نوع مدرک برای ساختار صفحات وب و مارک دی.تی.دی تعریف نوع مدرک برای ساختار مارک است.

اس.جی.ام.ال می‌تواند در همه مدارک مثل کتاب، مقاله

مجله، اسناد فنی، واژه‌نامه و البته پیشینه‌های کتابشناختی

به کار رود. موجودیت‌های خارجی مثل متن، تصویر و ویدئو

را می‌توان در مکان‌های خاصی از مدرک تثبیت کرد. این

مسئله ایجاد چندرسانه‌ای‌ها را ممکن می‌سازد. با استفاده از

زبان اس.جی.ام.ال می‌توان پیوندهایی درون یک مدرک یا

میان مدارک مختلف در سطوح مختلف سلسله‌مراتبی ایجاد

کرد و کاوش‌های فرامنتی انجام داد (۸).

هدف اصلی طراحی اس.جی.ام.ال اطمینان از این مسئله بود که مدارک کدگذاری شده بتوانند از یک محیط سخت‌افزاری و نرم‌افزاری به محیط دیگر بدون از دست رفتن اطلاعات انتقال یابند. نشانه‌گذاری توصیفی و مفهوم نوع مدرک در حدی بسیار مختصر به این هدف جامه عمل می‌پوشاند. استقلال داده‌ها، این هدف را حتی در سطح نویسه‌های^۷ تشکیل دهنده مدارک برآورده می‌کند. اس.جی.ام.ال روشی عمومی را برای جایگزینی رشته نویسه‌ها ارائه می‌دهد. این روش، ساده و مستقل از ماشین است و به رشته‌ای از نویسه‌های یک مدرک می‌گوید که وقتی مدرک پردازش می‌شود با رشته‌ای دیگر جایگزین شوند. یکی از کاربردهای آشکار این روش اطمینان از انسجام نظام نامگذاری است و کاربرد بسیار مهم‌تر آن، مقابله با ناتوانی نظام‌های رایانه‌ای متفاوت برای فهمیدن مجموعه نویسه‌های یکدیگر و یا هر نظام دیگر است که همه نویسه‌های گرافیکی موردنیاز را برای یک کاربرد خاص تهیه می‌کند. این کار از طریق ارائه طراحی‌های توصیفی برای نویسه‌های غیرقابل انتقال انجام می‌شود (۱۳).

«استقلال داده‌ها، دوام اطلاعات و استفاده دوباره از آنها را برای مدت زمان طولانی که عامل بسیار مهمی در کتابخانه‌هاست تضمین می‌کند» (۸).

تعریف نوع مدرک^۸

مارک - دی.تی.دی

واژه مارک-دی.تی.دی به استفاده از اس.جی.ام.ال در

مارک اشاره دارد. هدف اولیه ایجاد مارک دی.تی.دی تولید

دی.تی.دی‌های استاندارد بود که تبدیل داده‌های مارک به

اس.جی.ام.ال و برعکس را بدون از دست رفتن داده‌ها میسر

سازد (۷).

مارک و اس.جی.ام.ال مفاهیم قابل مقایسه بسیاری دارند.

با این وجود از کلمات بسیار متفاوتی برای توصیف قسمت‌های

مختلف خود استفاده می‌کنند. اس.جی.ام.ال ساختار عمومی

داده‌هاست که به صورت رسمی ایزو ۸۸۷۹ نامیده می‌شود.

ساختار استاندارد قابل مقایسه با آن ایزو ۲۷۰۹، قالب مبادله

اطلاعات یا "اینکس"^{۱۱} نامیده می‌شود. مارک در حقیقت

مجموعه‌ای از قالب‌هاست که دارای ساختار اینکس باشند.

هیچ کدام از این دو قالب نمی‌توانند به کار گرفته شوند، مگر

اینکه برچسب‌ها و قوانین به کار رفته در آنها با استانداردهای

مفهوم نوع مدرک و تعریف نوع مدرک با اس.جی.ام.ال مطرح شد. شبیه همه چیزهایی که رایانه‌ها پردازش می‌کنند، مدارک هم انواع مختلفی دارند. برای مثال تعریف یک گزارش می‌تواند متشکل از عنوان، احتمالاً نویسنده، چکیده و متعاقباً یک یا دو پاراگراف باشد. طبق این تعریف، مثلاً هر چیز که فاقد عنوان باشد گزارش نخواهد بود. بنابراین اگر مدارک انواع مختلفی دارند، یک برنامه مخصوص در نقش تجزیه‌کننده^۹ می‌تواند برای پردازش مدرکی که نوع خاصی دارد استفاده شود. این برنامه همه مؤلفه‌های مورد نیاز در آن نوع مدرک و نظم صحیحشان را بررسی می‌کند. به‌طور واضح‌تر مدارک مختلف از این نوع می‌توانند به یک روش واحد پردازش شوند (۱۳).

تعریف نوع مدرک یا دی.تی.دی با پیروی از قواعد خاص نشانه‌گذاری در اس.جی.ام.ال ساخته می‌شود و ساختار نوع

ساختار ایزو مطابقت داشته باشد. مجموعه برچسب‌های به کار رفته در ساختار اینکس قالب^{۱۲} نامیده می‌شود در حالی که همین مجموعه برچسب‌ها در ساختار اس.جی.ام.ال تعریف نوع مدرک نامیده می‌شود. بنابراین قالب و تعریف نوع مدرک مفاهیم قابل مقایسه‌اند (۱۱).

دو استاندارد فوق از رویکردهای بسیار متفاوتی برای نشانه‌گذاری استفاده می‌کنند. رویکرد مارک به نشانه‌گذاری، رویکردی واژه‌نامه‌ای است. ساختار مارک به نشانه‌گذاری کامل مؤلفه‌ها اما با طول‌های متغیر ولی نه خیلی طولانی تمرکز دارد. این رویکرد، نمایه‌سازی‌های پیچیده و طبقه‌بندی داده‌ها را تسهیل می‌کند. علاوه بر این دسترسی مؤثر به داده‌ها را به منظور آمایش مقذور می‌سازد و استفاده از ترکیبی از زیر مجموعه‌های متفاوت و همچنین نمایه‌سازی قسمت‌های انتخاب شده را امکان‌پذیر می‌سازد (۱۱).

ساختار داده‌های اس.جی.ام.ال با هدف کدگذاری مدارک با هر طولی و معمولاً بسیار طولانی، طراحی شده است. محتوای این مدارک باید نظم خطی خود را حفظ کنند. علاوه بر این خصوصیات سلسله‌مراتبی داده‌ها نیز باید حفظ شود. عدم دسترسی فهرستی به داده‌ها - مسئله‌ای که در مارک بر آن تأکید زیادی می‌شود - برای بخش‌های^{۱۳} داده‌ای طولانی و بسیار متغیر مناسب است. برچسب‌گذاری در اس.جی.ام.ال بیشتر بر روی ویژگی‌های ساختاری و ارائه‌ای تأکید دارد تا ویژگی‌های بازاریابی و طبقه‌بندی کردن^{۱۴} (۱۱).

همان‌طور که در بالا هم به این مسئله اشاره شد، مارک و اس.جی.ام.ال هر دو استانداردهای رسمی هستند که ساختار داده‌ها را تعیین می‌کنند و روش‌های استاندارد را برای کدگذاری پیشینه‌ها به کار می‌برند. بنابراین می‌توان از راه‌های بسیاری میان آن دو ارتباط برقرار کرد به شرط اینکه مارک - دی.تی.دی‌های استاندارد تهیه شوند.

در سال‌های گذشته اقدامات زیادی برای ایجاد امکان دسترسی به پیشینه‌های مارک از طریق اس.جی.ام.ال انجام شده است. از آن میان می‌توان به تلاش‌های میشل والپ^{۱۵} اشاره کرد. وی در سال ۱۹۹۰ یک مارک دی.تی.دی را به صورت تجربی نوشت. آزمایش‌های او نشان داد که می‌توان پیشینه‌های مارک را به روش‌های زیادی در اس.جی.ام.ال به کار برد. جروم^{۱۶} استاد مدرسه «مدیریت اطلاعات و نظام‌های» دانشگاه برکلی کالیفرنیا، یک مارک - دی.تی.دی برای برنامه فهرست‌نویسی این دانشگاه با نام چسایر^{۱۷} نوشت.

مهم‌ترین معیار وی در ایجاد مارک - دی.تی.دی مزبور، قابلیت برگشت پیشینه‌ها به قالب مارک بود. در پاییز سال ۱۹۹۵ دفتر توسعه شبکه و استاندارد مارک کتابخانه کنگره مارک امریکا^{۱۸}، مکانی که یو.اس. مارک در آن نگهداری می‌شود و توسعه می‌یابد، برنامه‌ای را برای ایجاد و توسعه مارک - دی.تی.دی انجام داد. مارک - دی.تی.دی‌هایی که در این مرکز نوشته شده کامل‌ترین در نوع خود است. برای اجرای این برنامه، گروهی از متخصصان مارک و اس.جی.ام.ال دور هم گرد آمدند. تعدادی طرح اولیه نوشته شد و نواقص موجود مشخص گردید. در سال ۱۹۹۶ اولین ویرایش مارک - دی.تی.دی به نام آلفا^{۱۹} در دسترس عموم قرار گرفت. نسخه کامل این مارک - دی.تی.دی در وب‌سایت دفتر توسعه شبکه و استانداردهای کتابخانه کنگره موجود است. برنامه مزبور ایجاد نرم‌افزارهایی که تبدیل داده‌های مارک به اس.جی.ام.ال و برعکس را میسر می‌سازد نیز در بر می‌گرفت (۱۱).

اصول کلی که در این طراحی در نظر گرفته شد عبارت بودند از:

• مارک - دی.تی.دی‌ها نباید منحصر به هیچ کاربرد و یا ابزار مبتنی بر اس.جی.ام.ال و یا مارک باشند. برای اینکه بشود آنها را برای مبادله و ایجاد پیشینه‌های مارک در اس.جی.ام.ال و یا به عنوان دی.تی.دی برای وارد شدن در مدارک اس.جی.ام.ال به کار برد (۷).

• داده‌های مارک و اس.جی.ام.ال باید به هم برگشت‌پذیر باشند و طی تبدیل این دو ساختار به هم، محتوای هیچ کدام از این دو نباید از بین برود (۷).

• دی.تی.دی‌ها باید اختیاری بوده، تحمیلی نباشند. به این ترتیب که برای تشکیل دی.تی.دی‌ها، مجموعه داده‌های مارک الزاماً نباید به صورت مشابه در اس.جی.ام.ال وجود داشته باشند (۷).

• دی.تی.دی‌ها باید کاربرپسند باشند و بتوان بدون ابزارهای اس.جی.ام.ال به سادگی از آنها استفاده کرد (۱۱).

• داده‌های کتابشناختی مارک باید با فراداده‌های کتابشناختی که در بخش‌های مشخصی از مدارک تعریف شده تی.ای.آی^{۲۰} کدگذاری شده‌اند ارتباط داشته باشند. تی.ای.آی تلاشی بود که محققان برای تسهیل مبادله داده‌ها انجام دادند و به حرکتی برای توسعه اصول راهنمای

نشانه‌گذاری توصیف داده‌های پیچیده و تحلیل آنها تبدیل شد (۹: ۲۳۳). «تی.ای. آی برنامه مشترک بین‌المللی است که اصولی را برای آماده‌سازی و مبادله متون الکترونیکی برای پژوهش‌های علمی ارائه می‌دهد» (۵). در طول توسعه مدارک تعریف شده تی.ای. آی اهمیت فراداده‌های کتابشناختی و مارک مشخص شد. مؤلفه‌های تعریف شده در ساختار تی.ای. آی با داده‌های فهرست‌نویسی ارتباط دارند. پیوند میان مارک و تی.ای. آی می‌تواند چه در مارک و چه در اس.جی.ام.ال برقرار شود (۷).

پنج قالب مارک فقط با دو دی.تی.دی اداره می‌شوند. دی.تی.دی قالب کتابشناختی مارک شامل اطلاعات اضافی درباره قالب داده‌های موجودی^{۲۱} و اطلاعات محلی^{۲۲} است. این مسئله باعث می‌شود که بتوان سه قالب مارک را در یک چارچوب اس.جی.ام.ال ایجاد کرد. این تصمیم بنا بر تاریخچه و خصوصیات سه قالب مذکور گرفته شد. همیشه این امکان وجود دارد که مؤلفه‌های داده‌های کتابشناختی و مؤلفه‌های داده‌های موجودی را کنار هم آورد.

قالب موجودی به‌ندرت اسامی مشابه با قالب کتابشناختی را به مؤلفه‌های خود اختصاص می‌دهد. این مسئله در مورد داده‌های قالب اطلاعات محلی نیز صادق است. اضافه بر این، دی.تی.دی قالب داده‌های مستندات^{۲۳} شامل اطلاعات اضافی درباره قالب داده‌های رده‌بندی^{۲۴} خواهد بود. این مسئله ایجاد فقط یک دی.تی.دی را برای هر دو پیشینه مذکور میسر می‌سازد. مزیت این رویکرد این است که نیاز کمتری به توسعه اس.جی.ام.ال خواهد بود. بر این اساس وقتی دی.تی.دی‌ها در برنامه‌ها بارگذاری می‌شوند، استفاده‌کنندگان می‌توانند انواع متنوعی از پیشینه‌ها را در یک دی.تی.دی ایجاد کنند (۷).

برچسب‌گذاری باید به جای نام مؤلفه‌های داده‌ای، منعکس‌کننده برچسب‌های مارک باشد (۷).

مثلاً Field tag: mrcb245 title statement
- شناسگرهای حوزه‌های فرعی به عنوان مؤلفه‌های منحصر به فرد تعریف می‌شوند. مثلاً:

Subfield tag mrcb 245 - a - "title"--
- اختصاص مؤلفه‌های منحصر به فرد هر حوزه فرعی باعث می‌شود که ارتباط اسامی به آنها امکان‌پذیر باشد. این مسئله استفاده از ابزارهای اس.جی.ام.ال را برای ایجاد پیشینه‌ها و برگشت‌پذیری میان مارک و اس.جی.ام.ال تسهیل می‌کند

(۱۱).

- از آنجا که مؤلفه‌های ساختاری مارک هنگام تبدیل به اس.جی.ام.ال خود به خود محاسبه و تنظیم می‌شوند، مؤلفه‌های مزبور الزاماً نباید به پیشینه اس.جی.ام.ال منتقل شوند. برای مثال راهنمای پیشینه‌ها^{۲۵} ساختاری است که نباید به اس.جی.ام.ال برده شود. چون وقتی داده‌های مارک به اس.جی.ام.ال تبدیل می‌شوند مبدل خود به خود راهنما را تولید خواهد کرد (۱۱).

مثال زیر نمونه‌ای از مارک-دی.تی.دی را نشان می‌دهد (۱۱).

بخشی از قالب مارک، حوزه عنوان:

245 TITLE STATEMENT (NR)

Indicators

First Title added entry

0 No title added entry

1 Title added entry

Second non filing characters

0-9 Number of non filing characters present

Subfield codes

\$a Title (NR)

\$b Remainder of title (NR)

\$c Remainder of title page

transcription/statement

responsibility (NR)

بخشی از مارک - دی.تی.دی:

<ELEMENT mrcb 245 - - ((mrcb 245 a I mrcb 245 b I mrcb 245c I mrcb 245d

I mrcb 245e I mrcb 245 f I mrcb 245 g I mrcb 245 h I mrcb 245 i I mrcb 245 p I mrcb 245 s I mrcb 245) *)

<!ATTLIST mrcb 245

name CDATA #FIXED "Title atement"

obsolete CDATA #FIXED "no"

repeatable CDATA #FIXED "no"

نتیجه گیری

با ظهور اینترنت و توسعه شبکه‌ها، پیوستن به فضای جهانی اطلاعات به الزامی اجتناب‌ناپذیر برای کتابخانه‌ها تبدیل شده است. وقایع در محیط الکترونیکی به‌طور سریع و غیرقابل پیش‌بینی در حال تحولند. از یک طرف کتابداران در انتهای فونونی هستند که در قرن گذشته شکل گرفتند و از طرف دیگر در ابتدای تحولاتی هستند که در حال شکل‌گیری‌اند. کتابداران در عصر اطلاعات الکترونیک، برای رویارویی با این تحولات نیازمند به رویکردهای ابتکاری و بهره‌برداری از فناوری‌های جدید هستند. کتابداران برای اینکه بتوانند اطلاعات موجود در کتابخانه‌های خود را به محیط جدید اطلاعاتی منتقل کنند، باید فعالیت‌های فهرست‌نویسی خود را مطابق با محیط جدید شکل دهند. بدون ایجاد انعطاف در فهرست‌نویسی و کدگذاری مدارک، داده‌های کتابشناختی نمی‌توانند به شکل سیاهه‌ها و نمایه‌های اثربخش و کارآمد از مجموعه کتابخانه‌ها وارد محیط جدید و متغیر اطلاعات الکترونیکی شوند. کتابداران مدتی از مارک برای کدگذاری پیشینه‌های خود استفاده کرده‌اند اما تجربه ثابت کرده است که در بعضی جنبه‌ها مارک نتوانسته است با تحولات جدید همسو شود. وقتی کتابداران استفاده از مارک را آغاز کردند آنها فقط پیشینه‌های چاپی خود را برای ذخیره‌سازی و بازیابی اطلاعات به قالب الکترونیکی تبدیل می‌کردند. موادی که با این پیشینه‌ها ارائه می‌شدند فیزیکی و ایستا بودند. اما در محیط اطلاعاتی جدید، مواد اطلاعاتی مجاری و پویا هستند. «در محیط اطلاعاتی جدید نقش کتابخانه‌ها از دروازه‌بانان اطلاعات به کاشفان اطلاعات تبدیل شده است» (۱۲: ۸۶). برای اینکه اطلاعات موجود در کتابخانه‌ها به‌نحو مؤثری وارد محیط جدید اطلاعاتی شوند کتابداران باید بتوانند با استفاده از رویکردهای نوین نواقص مارک را جبران کنند. اس.جی.ام. ال یکی از روش‌هایی است که به دلیل انعطاف‌پذیری بسیار خود می‌تواند جایگزین یا مکملی برای استاندارد مارک باشد.

منابع

۱. تزری، محمدرضا. «نوشته جات ساخت یافته و زبان SGML». گزارش کامپیوتر، ۱ (فروردین و اردیبهشت ۱۳۷۳): ۴۴-۴۸.
۲. تیلور، آرلین جی. سازماندهی اطلاعات. ترجمه محمد حسین دیانی. مشهد: کتابخانه رایانه‌ای، ۱۳۸۱.

il (i10/i11/i1fill)

i2 (i20/i21/i22/i23/i24/i25/i26/i27/i28/i29/i2fill) #REQUIRED>

<!ELEMENT mrcb 245 a - (#PCDATA)>

<!ATTLIST mrcb 245b

name CDATA #FIXED "Title"

obsolete CDATA #FIXED "no"

repeatable CDATA #FIXED "no">

<! ELEMENT mrcb 245b - (# CDATA) >

<!ATTLIST mrcb245b

name CDATA #FIXED "Remainder of title"

obsolete CDATA #FIXED "no"

repeatable CDATA #FIXED "no">

<! ELEMENT mrcb 245c - (# PCDATA) >

<!ATTLIST mrcb245c

name CDATA #FIXED "Remainder of title page transcription..."

obsolete CDATA #FIXED "no"

repeatable CDATA #FIXED "no">

بخشی از پیشینه کدگذاری شده:

[245]10\$ aSGML: \$ban athor's guide
to the standard generalized morkup
language/ \$ cMartin Bryan

اس.جی.ام. ال

<mrcb 245 il =li2=o>mrcb 245-a>

SGML:<mrcb 245-b> anuthor's guide /
to the standard generalized
markup language/<mrcb245-c>
Martin Bryan

9. Lowry, Charls B; Chestnut, David R. "Managing technology: SGML and the digital libraries of tomorrow". *Journal of Academic Librarianship*, Vol 24, No. 3 (1998): 232-237.

10. "MARC-SGML and SGML-MARC conversion program userguide". [on-line]. Available: <http://www.loc.gov/marc/arcddt/usermanual.html>. [21 Mar. 2004].

11. McCallum, Sally H. "The future of communication formats: MARC data in SGML structure". [on-line]. Available: <http://XML.coverpages.org/McCallumMARC.html>. [21 Apr. 2004].

12. Qin, Jian. "Representation and organization of information in the web space: from MARC to XML". *Informing Science*, Vol. 3, No. 2 (2000): 83-87.

13. "TEI guidelines for electronic text encoding and interchange". [on-line]. Available: <http://etext.virginia.edu/bin/tei-tocs?div-DIVI&id=SG>. [21 Apr. 2004].

تاریخ دریافت: ۱۳۸۳/۲/۱۳

3. Bolin, Mary K. "Catalog design, catalog maintenance, catalog governance". *Library collection, acquisition and technical services*, Vol. 24, No. 1 (2000): 53-63.

4. Davis, Stephen Paul. "SGML-MARC: incorporating library cataloging into the TEI environment". [on-line]. Available: <http://XML.coverpages.org/davis.MARC.html>. [7 Apr. 2004].

5. Gaynor, Edward. "from marc to markup: SGML and online library system". [on-line]. Available: <http://XML.coverpages.org/GaynorMARC96.html>. [7 Apr. 2004].

6. Library of Congress. Network Development and Marc Standard office. "What is marc record and why is it important?". [on-line]. Available: <http://www.loc.gov/mark/html>. [21 Apr. 2006].

7. *Ibid.* "MARC DTDs background and developments". [on-line]. Available: <http://www.loc.gov/mark/marcdtd/marcdtdback.html>. [18 Apr. 2004].

8. Loupovoci, Catherine "Cataloging in SGML: from tagging to markup". [on-line]. Available: <http://www.ifla.org/IV/ifla68/httpc.html>. [7 Apr. 2004].

