

ایجاد فهرستگان اسناد آرشیوی با استفاده از پروتکل OAI-PMH

♦ دکتر محمد توکلی زاده راوری^۱

سید حسین میرجلیلی^۲

تاریخ دریافت مقاله: ۱۳۸۷/۴/۲

تاریخ پذیرش: ۱۳۸۸/۵/۲۰

چکیده

این نوشتار، با معرفی پروتکل پیشنهادی PMH، بر آن است تا نشان دهد که چگونه می‌توان از این پروتکل برای ایجاد فهرستگان اسناد آرشیوی بهره گرفت. هدف نهائی این پروتکل، تسهیل در انتقال اطلاعات به صورت فراداده در محیط شبکه است. PMH، با بهره‌گیری از عناصر فرادادهٔ دوبلین کور، قالب (فرمت) XML (زبان نشانه‌گذاری گسترش‌پذیر) و پروتکل HTTP این امر را میسر می‌سازد. بدین منظور، مقاله به تعریف فراداده و معرفی عناصر فراداده‌ای دوبلین کور می‌پردازد و ساختار پیشنهادی رکورد فراداده را براساس الگوی پیشنهادی در قالب XML نشان می‌دهد. همچنین، توضیحاتی دربارهٔ شش فرمان قراردادی برای درخواست فراداده از طریق پروتکل HTTP ارائه می‌کند، و نشان می‌دهد که چگونه دارندگان اسناد، می‌توانند نقش داده‌فرآور و متولیان امور اسناد، نقش خدمت‌فرآور را ایفا کنند. در پایان، به مزایای اصلی و جنبی استفاده از این پروتکل اشاره شده است.

کلید واژه‌ها:

فهرستگان‌ها / اسناد الکترونیکی / انتقال / پروتکل OAI-PMH / زبان‌های نشانه‌گذاری / فراداده‌ها / پروتکل دوبلین کور / زبان نشانه‌گذاری گسترش‌پذیر / پروتکل فرا متن

۱ استادیار گروه کتابداری و اطلاع‌رسانی دانشگاه یزد.

۲ مربی گروه کتابداری و اطلاع‌رسانی دانشگاه یزد. hmirljalili@yahoo.com

مقدمه

عموماً آرشیوها متشکل از اسنادی همچون نامه‌ها، سوابق، عکس‌ها، فایل‌های رایانه‌ای، بریده جراید، سوابق مالی، و نظایر این‌هاست. (موسی‌چمنی، ۱۳۸۷، ص ۵۲) از دیدگاه کتابداری و اطلاع‌رسانی، سند در دو معنای خاص و عام به کار می‌رود. در معنای خاص، متنی را گویند که برای ثبت رویداد خاصی، در یک یا چندین نسخه محدود تولید شده است. این‌گونه اسناد را، معمولاً در آرشیوها و بایگانی‌ها نگه می‌دارند و در شمار انتشارات عمومی قلمداد نمی‌شوند. در مفهوم عام، سند به هر نوع نوشته خطی، چاپی، عکس ... اطلاق می‌شود که بتوان از محتوای آن اطلاعاتی به دست آورد. (اکرمی، سوری، ۱۳۸۷، ص ۷۸)

باتوجه به حیطة کاری هر نهاد، ممکن است سند با مصداقی خاص تعریف شود. از نظر آرشیو ملی، سند به «کلیه اوراق، مراسلات، دفاتر، پرونده‌ها، عکس‌ها نقشه‌ها، کلیشه‌ها، نمودارها، فیلم‌ها، نوارهای ضبط صوت و سایر اسنادی که در دستگاه دولت تهیه شده و یا به دستگاه دولت رسیده است و...» (دیانی، ۱۳۷۷، ص ۴۰) اطلاق می‌شود. در این تعریف، ماهیت دولتی بودن است که مصداق پیدا می‌کند. بحث‌های نظری فراوانی نیز وجود دارد که به ماهیت سند اشاره دارد و به این نکته می‌پردازد که چه موادی را می‌توان به عنوان سند یا مدرک لحاظ کرد.

برخی از صاحب‌نظران مانند رانگاناتان، بیش‌تر بر شکل و ماهیت فیزیکی مدرک و سند توجه دارند، و برخی مانند لیوی، به ماهیت کارکردی سند اشاره دارند و سند یا مدرک را آن چیزی می‌دانند که به جای انسان سخن می‌گوید و عملاً وظیفه رساندن سخن انسان را برعهده دارد. (Levy, 2005, p 25) مفهوم این سخن، آن است که سند برای هدف و مخاطب مشخصی تدوین شده است. نقطه مشترک همه تعاریف، این است که سند آن چیزی است که ارزش حفظ و نگهداری دارد و از دسته مواد زودگذر^۱ محسوب نمی‌شود. این نوع از مواد، طبعاً

ارزش سازماندهی و در دسترس قرار گرفتن نیز دارند، و برای این کار، باید از اصول و فنون کتابداری و دبیزش بهره برد.

گردآوردن همه اسناد موردنظر در یک مکان امکان‌پذیر نیست. زیرا اسناد در مکان‌های مختلف پراکنده‌اند و در مراکز گوناگون آرشیو می‌شوند؛ شمار آن‌ها محدود به یک یا چند نسخه است و امکان جستجوی آن‌ها در منابع کتابشناختی معمول بسیار کم است. از این رو، این نوع از اسناد عموماً در رده ادبیات خاکستری^۲ قرار می‌گیرند که گاهی ممکن است از اهمیت فراوانی برخوردار باشند. برای حل این موانع، یکی از بهترین روش‌ها، گردآوری اطلاعات مربوط به این اسناد و نشان دادن محل نگهداری آن‌هاست. در بین ابزارهای بازیابی اطلاعات، فهرستگان‌ها چنین نقشی را برای مکانیابی مواد کتابخانه‌ای در بین گروهی از کتابخانه‌ها برعهده دارند.

در این راستا، پروتکل‌هایی برای تبادل داده‌های منابع اطلاعاتی ارائه شده است که در طول یا عرض هم قرار می‌گیرند، و هر کدام از آن‌ها مسئله‌ای خاص را مدنظر قرار می‌دهد. معمولاً این نوع از استانداردها و پروتکل‌ها، از ترکیب دو یا چند استاندارد مرتبط حاصل می‌شوند، و هر کدام از آن‌ها دارای مزایا و معایبی است. PMH^۳، نام پروتکلی در این زمینه است که یکی از پروژه‌های OAI^۴ به شمار می‌رود. پیدایش OAI در اصل ریشه در رشد صعودی اطلاعات علمی در شبکه جهانی وب دارد. این مسئله باعث شده است تا تلاش‌های فراوانی برای سازماندهی اطلاعات صورت گیرد. یکی از این تلاش‌ها، نشست در ساتتافی^۵ در اکتبر ۱۹۹۹ برای یافتن سازوکاری برای حل مسائل مربوط به مواد نوشتاری الکترونیکی بود. اعضای این نشست، جملگی بر این اعتقاد بودند که اگر منابع اطلاعاتی موجود در وب یکپارچه شوند، فرصت اشاعه به موقع تولیدات علمی حوزه آموزش و پژوهش فراهم خواهد آمد. (Sompel & Lagoze, 2000)

گرچه هدف و تمرکز این پروتکل بر تبادل خودکار اطلاعات الکترونیکی حوزه دانشوری است، اما براساس

Initiative :

یافتن معادلی که دربردارنده تمامی مفاهیم آن باشد، بسیار دشوار است. مفهوم آن را می‌توان با توجه به این نکته استخراج کرد که OAI برای هماهنگ و یکپارچه‌سازی منابع الکترونیکی چارچوبی (استانداردی) ساده و «دست‌وپانگیر»^{۱۱} فراهم می‌آورد تا فرایند تبادل خودکار اطلاعات الکترونیکی در حوزه دانشوری میسر شود. (Logoze & Sompel, 2001)

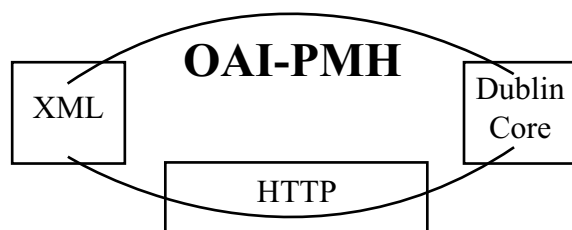
PMH، یکی از دو طرحی است که OAI ارائه کرده است. این طرح، سه پروتکل و استاندارد اصلی دیگر را به خدمت می‌گیرد تا از آن طریق بتوان اطلاعات را در وب تحت یک قالب درآورد و به طور خودکار منتقل کرد. این سه استاندارد، عبارت است از:

۱- Dublin Core ؛

۲- XML ؛

۳- HTTP .

تصویر ۱: بهره‌گیری پروتکل OAI-PMH از Dublin Core، XML و HTTP



در فرایندی که OAI برای درخواست و دریافت اطلاعات در نظر گرفته است، به طور طبیعی دو گروه جداگانه شرکت دارند:

۱. خدمات‌دهندگان اطلاعات؛

۲. تولیدکنندگان اطلاعات.

دسته اول، مراکزی است که اطلاعات را از مکان‌های مختلف گردآوری می‌کنند و در سامانه‌ای واحد ذخیره می‌کنند تا یک فهرستگان ایجاد کنند. به این گروه اصطلاحاً خدمت‌فرآور^{۱۲} گفته می‌شود. دسته دیگر،

ساختار آن، می‌توان برای متمرکز ساختن داده‌های مربوط به اسناد با مصادیق گوناگون (مثل دولتی یا تاریخی و غیره) نیز بهره برد. با استفاده از این پروتکل، می‌توان اطلاعات اسنادی را - که در جاهای گوناگون قرار دارند - شناسائی، گردآوری، ذخیره و مکانیابی کرد. در عمل، به کارگیری این پروتکل امکان به اشتراک‌گذاری منابع را فراهم می‌سازد و هدف اصلی فهرستگان‌ها نیز همین است. با به اشتراک‌گذاری مؤثر منابع، می‌توان امکان جستجوی منابع چاپی و الکترونیکی مورد علاقه کاربران را فراهم آورد. (Yu & Chen & Chang, 2005)

OAI-PMH

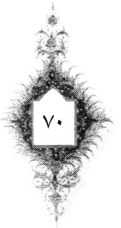
OAI، نهادی است که دو پروتکل را با نام‌های PMH و ORE برای انتقال و تبادل اطلاعات در محیط وب ایجاد کرده است. واژه‌های تشکیل‌دهنده آن، مفاهیم زیر را دربردارد:

Open^۶:

این واژه، دلالت بر باز بودن آرشیوهای الکترونیکی از لحاظ معماری آن‌ها دارد و نشانگر رایگان بودن استفاده از منابع الکترونیکی نیست. هدف این سامانه، یافتن سازوکاری است که بتوان آن را هم به منابع تجاری و هم غیرتجاری موجود در وب تعمیم داد. نگاه معمارگونه به واژه باز، در اصطلاح «مخزن باز» نیز قابل مشاهده است. هنگامی که ما از باز بودن مخزن کتابخانه‌ای صحبت می‌کنیم، منظورمان نحوه معماری آن برای بهره‌گیری است که همگان می‌توانند بدون واسطه وارد مخزن شوند و کتاب‌های موردنظر خود را از فکسه‌ها بردارند.

Archives^۸:

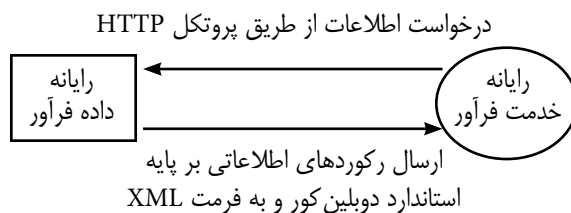
در این‌جا مقصود از آرشیو، گنجینه داده‌های^۹ منابع الکترونیکی است که در شبکه‌های بزرگ اطلاعاتی (مانند اینترنت) به صورت ساختاریافته^{۱۰} قرار دارند. نمونه بارز منابع الکترونیکی ساختاریافته، پایگاه‌های کتابشناختی و منابع تمام متن است که در آن‌ها عناصر فراداده به کار رفته است.



مراکزی است که تولیدکننده یا صاحب اطلاعات هستند و اصطلاحاً داده‌فرآور^{۱۳} خوانده می‌شوند.

به طور مختصر، می‌توان گفت که هر خدمت‌فرآور، از داده‌فرآوران مختلف به طور خودکار از طریق پروتکل HTTP درخواست اطلاعات می‌کند و داده‌فرآوران نیز، به طور خودکار در پاسخ به درخواست آن‌ها، رکوردهای اطلاعاتی را از گنجینه داده‌های خود بیرون می‌آورند و بر پایه استاندارد فراداده دابلین کور نمایه‌سازی می‌کنند و در قالب XML برای درخواست‌دهنده ارسال می‌کنند. پس از آن خدمت‌فرآور متن XML را پویش و داده‌ها را از آن، استخراج می‌کند و در فهرستگان قرار می‌دهد تا برای کاربران اطلاعات قابل جستجو شود. تصویر زیر، این فرایند را تا دریافت اطلاعات درخواستی نشان می‌دهد:

تصویر ۲: فرایند کار طبق پروتکل OAI-PMH



استاندارد فراداده دابلین کور

برای این که بتوان اطلاعات را به طور خودکار و گسترده از جاهای مختلف در شبکه وب گردآوری کرد، باید همه این اطلاعات به صورت یکسان و در قالبی از پیش تعیین شده تعریف شود تا رایانه بتواند پس از دریافت آن‌ها از مراکز مختلف، بدون عامل انسانی اجزای مختلف اطلاعاتی به دست آمده را پویش و به منظور بازیابی ذخیره کند. عناصر فراداده دابلین کور، دقیقاً برای رفع چنین نیازی به وجود آمده است، و اساس آن بر فراداده استوار است. از این رو، لازم است مفهوم فراداده تشریح شود تا بتوان کارکرد این استاندارد را بهتر درک کرد.

به‌رغم اجماع و همگرایی بین‌المللی در مفهوم فراداده، تعریف آن از جهاتی چالش‌برانگیز است. زیرا از ابتدا این

اصطلاح بدون تعریف دقیقی مورد استفاده قرار گرفته است. شاید بتوان این چالش را نتیجه نبود تعریف و برداشت یکسان از اصطلاحاتی چون داده، اطلاع، و دانش خواند. از آن جا که هر کس اصطلاحات یاد شده را از زاویه‌ای خاص مورد توجه قرار داده، ارائه مفهومی یکدست برای فراداده دشوار است. در ادبیات مربوط به منابع الکترونیکی، عموماً واژه داده، معادل مفهیمی قرار گرفته است که ما ممکن است به آن‌ها اطلاعات یا حتا دانش بگوئیم.

فراداده را معمولاً با عباراتی مانند «داده‌هائی درباره داده دیگر» (Gardner, 1998, pp 60-52) یا «اطلاعاتی درباره اطلاع دیگر» (Henshaw & Valouskas, 2001, pp 86-101) تعریف کرده‌اند. بنابراین هر چیزی که با اطلاعات یا داده‌های دیگر توصیف شود، خود داده شمرده می‌شود و اطلاعات یا داده‌های توصیف‌کننده آن داده، فراداده خوانده می‌شوند. به عنوان نمونه، هر سند آرشیوی خود یک داده محسوب می‌شود و اجزائی چون پدیدآور، تاریخ نگارش، تاریخ ثبت، مخاطب، نوع رسانه و طول یا اندازه سند، داده‌هائی است که به توضیح درباره سند (داده) می‌پردازند که اصطلاحاً به این‌گونه داده‌ها، فراداده می‌گویند. بنابراین، فراداده موضوعی تازه نیست. کتابداران عهد باستان، از آن زمان که به سازماندهی منابع اقدام می‌کردند، با آن، اما با نامی دیگر سر و کار داشتند؛ ولی اطلاق اصطلاح فراداده محصول مباحث مربوط به ادبیات سازماندهی و ساختاردهی مواد الکترونیکی در دهه‌های اخیر است.

فراداده، خود از دو جزء تشکیل شده است: عنصر فراداده و مقدار^{۱۴} آن؛ به عنوان مثال، در هر پایگاه داده، مانند فهرست اسناد سازمان، مجموع داده‌های توصیف‌کننده سند، رکورد را تشکیل می‌دهد که در ادبیات مواد الکترونیکی، اصطلاحاً به آن رکورد فراداده می‌گویند. هریک از اجزای^{۱۵} این رکوردها، خود دارای نام یا برجستگی نظیر «پدیدآور»، «تاریخ نگارش» و غیره

OAI، تنها پانزده عنصر سادهٔ دوبلین کور را توصیه می‌کند. داده‌فرآوران، این اختیار را دارند تا بنابر نیاز خود یک یا چند عنصر را استفاده کنند. به عبارت دیگر، به کارگیری هر کدام از این عناصر اختیاری است و هر فراهم‌آورنده، ممکن است تنها از چند عنصر، برحسب نیاز بهره بگیرد. وارد، در مطالعهٔ خود نتیجه گرفت که به طور متوسط در هر رکورد OAI در عمل هشت عنصر دوبلین کور به کار می‌رود و عناصر توضیحی استفادهٔ چندانی ندارند. (Ward, 2003, pp 315,317)

علاوه بر اختیاری بودن کاربرد هر کدام از عناصر ساده و پالایشگرها، امکان به کارگیری یک عنصر در هر رکورد بیش از یک بار نیز ممکن است. مثلاً به هر سند معمولاً بیش از یک موضوع داده می‌شود. لذا عنصر سادهٔ موضوع را می‌توان در هر رکورد فراداده چند بار به کار برد.

مشکلی که ممکن است هنگام تبدیل منابع فارسی به وجود آید، این است که معمولاً ترجمهٔ برخی از اجزای اطلاعاتی این نوع منابع به زبان دیگری (معمولاً انگلیسی) نیز داده شده است. عنوان یا چکیده، دو جزء اطلاعاتی است که نویسندگان معادل انگلیسی آن‌ها را نیز ارائه می‌کنند. در این نوع رکوردها - که زبان اصلی فارسی است - می‌توان از طریق عنصر زبان، فارسی بودن منبع را نشان داد و اجزائی را که به زبان انگلیسی یا زبان دیگری آمده است، یک به یک مشخص ساخت و به سامانه‌های خودکار رایانه‌ای این امکان را داد که به صورت خودکار این موارد را بازشناسند.

رکوردهای «XML»

پس از دریافت اطلاعات، باید آن‌ها را با برنامه‌های رایانه‌ای از پیش نوشته پویش کرد، تا بتوان داده‌ها را استخراج و قابل ذخیره و بازیابی کرد. اطلاعات را می‌توان به شکل‌های مختلف ارسال کرد؛ مثلاً می‌توان رکورد را براساس پروتکل دوبلین کور به صورت متنی ارسال کرد، و در هر خط نام عنصر و پس از درج علامت متمایزکننده، و سپس مقدار آن را آورد، مثلاً:

است؛ هر جزء، علاوه بر نام، دارای یک مقدار است. نام یا برچسب هر جزء اطلاعاتی اصطلاحاً «عنصر» و اطلاعات مربوط به آن، «مقدار» خوانده می‌شود. به عنوان نمونه، در عبارت «پدیدآور: امین‌السلطان» که جزئی از رکورد فراداده است، عبارت «پدیدآور»، عنصری از آن رکورد و «امین‌السلطان» مقدار آن است.

دوبلین کور، اساساً به عناصر فراداده (نه مقادیر آن‌ها) توجه دارد و آن‌ها را به دو سطح «ساده»^{۱۶} و «توضیحی»^{۱۷} تقسیم کرده است. سطح ساده، پانزده عنصر را دربرمی‌گیرد که عبارت است از: عنوان، پدیدآور، موضوع، توصیف، ناشر، همکار، تاریخ، نوع، شکل، شناسه، مأخذ، زبان، ارتباط، پوشش، و حق استفاده.

سطح توضیحی، سه عنصر دیگر را نیز شامل می‌شود که عبارت است از: «مخاطب»^{۱۸}، «مالک اثر»^{۱۹} و «مالک حقوقی»^{۲۰}. در کنار عناصر ساده، گروهی از «پالایشگرها»^{۲۱} نیز وجود دارد که به آن‌ها «توضیحگر»^{۲۲} گفته می‌شود. در عمل، پالایشگرها نیز به دو گروه «پالایشگر عناصر» و «طرح‌های نشانه‌گذاری»^{۲۳} تقسیم می‌شوند. پالایشگرها، معنا و مفهوم عنصر فراداده را خاص‌تر و دقیق‌تر می‌کند. مثلاً یکی از عناصر پانزدهگانهٔ ساده، عنصر تاریخ است. این عنصر را می‌توان به مفاهیم و معانی خاص‌تری مانند تاریخ خلق، تاریخ اعتبار، تاریخ دسترس بودن، تاریخ صدور، و تاریخ تغییرات، پالایش کرد. به زبانی دیگر، پالایشگرها نسبت به عناصر، خاص‌تر و دقیق‌تر هستند.

رده‌بندی‌های کنگره و دیوئی، دو نمونه از طرح‌های نشانه‌گذاری است که طبق این استاندارد، از آن‌ها برای پالایش عنصر موضوع (یکی از عناصر پانزدهگانهٔ ساده) استفاده می‌شود. این نوع پالایشگرها، به شناخت و تفسیر مقدار عنصر کمک می‌کنند. به عنوان مثال، ردهٔ دیوئی هر اثر را، بر پایهٔ عناصر دوبلین کور، می‌توان به عنوان پالایشگر موضوع به کار برد. حال اگر در رکوردی توضیح داده شود که مقدار این عنصر همان شمارهٔ دیوئی است، کاربر به سادگی می‌تواند بفهمد که منظور از موضوع در این جا رده‌بندی اثر است.

جدول ۱: عناصر پانزدهگانه فراداده دوبلین کور و پالایشگرهای آن‌ها همراه با توضیحات

عناصر متاداده ساده دوبلین کور		پالایشگرهای عناصر متاداده ساده دوبلین کور	
Element	توضیح عنصر	طرح‌های نشانه‌گذاری	Refinement(s)
Title	نامی که پدیدآور یا ناشر به اثر می‌دهد.		Alternative
Creator	شخص/اشخاص یا سازمانی / سازمان‌های مسئول در برابر محتوای اثر از لحاظ فکری. خالق اثر		-
Subject	موضوع اثر، همچنین کلیدواژه‌ها عبارات یا رده‌بندی که موضوع یا محتوای اثر را توصیف می‌کند.	LCSH MeSH DDC LCC UDC	-
Description	توصیف محتوای اثر به صورت متن، شامل چکیده مربوط به اثر مکتوب یا توصیف محتوای اثر دیداری یا شنیداری		Table of Contents
			Abstract
Publisher	مسئول دسترس‌گذار اثر به شکل موجود، مانند ناشر، گروه آموزشی در دانشگاه یا سازمان		-
Contributor	شخص یا اشخاص، سازمان یا سازمان‌هایی که از لحاظ فکری در خلق اثر شرکت داشته‌اند، اما نقش درجه دوم دارند.		-
Date	تاریخی که اثر به شکل موجود در دسترس قرار گرفته است.		Created
			Valid
			Available
			Issued
			Modified
Type	نوع اثر، مانند home page، داستان، شعر، مقاله، گزارش فنی، وازه‌نامه و ...		-
Format	شکل ارائه اثر، مانند text/HTML، ASCCI، PDF، کاست و ...	IMT	Extent Medium
Identifier	یک رشته یا عدد مورد استفاده برای متمایز ساختن اثر از سایر آثار. نمونه‌ای از شناسه‌گرها در محیط شبکه نشانی اینترنتی اثر (URL) است.	URI	-
Source	مأخذ و منشأ اثر که از آن طریق می‌توان به اصل اثر دست یافت.	URI	-
Language	زبان یا زبان‌های اثر	ISO 639-2 RFC 1766	-





URI	Relation			
ارتباط اثر با آثار دیگر	Is Version Of			
بیان می‌کند که اثر حاضر نسخه، ویراست یا برگرفته از چه اثر یا آثار دیگری است.	Has Version			
بیان می‌کند که چه آثار دیگری نسخه، ویراست یا برگرفته از اثر حاضر هستند.	Is Replaced By			
بیان می‌کند که اثر حاضر دارای چه ضمایم، نسخه‌های مشابه یا مرتبط با چه آثار قبلی است.	Replaces			
بیان می‌کند که اثر حاضر ضمیمه، نسخه مشابه یا اثر مقدم بر چه اثر یا آثاری است.	Is Required By			
بیان می‌کند که اثر حاضر، مورد نیاز چه اثر یا آثار دیگری نیاز دارد.	Requires			
بیان می‌کند که اثر حاضر، بخشی فیزیکی یا منطقی از یک مأخذ دیگر است.	Is Part Of			
بیان می‌کند که اثر حاضر، از لحاظ فیزیکی یا منطقی برای کار به چه اثر یا آثار دیگری نیاز دارد.	Has Part			
بیان می‌کند که اثر حاضر، را چه آثار دیگری ارجاع داده، یا استناد کرده یا مورد اشاره قرار داده‌اند.	Is Referenced By			
بیان می‌کند که اثر حاضر، به چه آثار دیگری ارجاع داده یا استناد کرده یا چه آثاری را مورد اشاره قرار داده است.	References			
بیان می‌کند که نسخه حاضر، مشابه نسخه دیگری با شکل دیگر است.	IS Format Of			
بیان می‌کند که اثر دیگری مشابه نسخه حاضر است.	Has Format			
پوشش مکانی اثر	Spatial	DCMI ISO 3166 DCMI Box TGN	پوشش زمانی و مکانی اثر	Coverage
پوشش زمانی اثر	Temporal	DCM Period W3C-DFT		
-			یک پیوند مانند یک نشانی اینترنتی (در صورت وجود) که به مسائل مربوط به حق مؤلف و مواردی از این دست اشاره دارد.	Rights

پوشش اطلاعاتی که به صورت الکترونیکی ارسال می‌شود شکل بالا یا شکل‌های مشابه چندان پذیرفته نیست. از میان همه شکل‌های ممکن، قالب XML مقبولیت زیادی یافته است. زبان‌های XML و HTML متعلق به دسته‌ای از زبان‌های رایانه‌ای است که به «زبان نشانه‌گذاری»^{۳۴} موسوم‌اند. اگر غالب فایل‌ها، بجز فایل‌های متنی به صورت متن باز شود، چیزی بجز شناسه (کد)ها و علائم

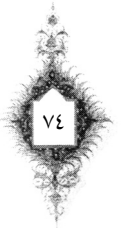
Creator : امین السلطان

Type : نام

Source : سازمان اسناد ملی ایران؛ شماره ۲۸-۱۹

Language : فارسی.

در مثال بالا creator, type, source و language چهار عنصر ساده دابلین کور هستند و آنچه بعد از دو نقطه «:» آمده، مقدار آن‌هاست. اما برای ایجاد امکان



نامفهوم مشاهده نخواهد شد.

باز کردن فایل‌های XML و HTML در نرم‌افزارهای متنی، محتویات آن را همراه با اضافاتی درون علائم «>>» (کوچک‌تر) و «<<» (بزرگ‌تر) نشان می‌دهد. از این رو، اطلاعاتی که در قالب این زبان‌ها ذخیره می‌شود، همانند یک متن معمولی مستقیماً درخور خواندن برای انسان است و با پویش متن آن‌ها از طریق برنامه‌های رایانه‌ای، می‌توان به راحتی و به صورت خودکار عناصر و مقادیر اصلی آن‌ها را استخراج و به یک پایگاه داده وارد کرد.

طبق پروتکل PMH، هر داده‌فرآور باید در پاسخ به درخواست خدمت فرآور، رکوردهای فراداده را به شکل XML و در قالب الگوی پیشنهادی ارسال کند. طبق این الگو، هر رکورد فراداده دارای سه بخش است:

۱. **سرصفحه:**^{۲۵} این بخش، خود ممکن است دارای سه قسمت باشد:

الف- شناسه‌گری،^{۲۶} برای متمایز کردن هر رکورد از رکوردهای دیگر (مانند OAI_EXMPL00354):

ب- مهر تاریخ،^{۲۷} شامل تاریخ ایجاد، حذف، یا آخرین تغییرات در رکورد فراداده؛

پ- گستره،^{۲۸} یا محدوده موضوعی آن (مثل قاجار، صفویان و غیره). هر رکورد ممکن است دارای صفر گستره و بیش‌تر باشد.

اطلاعات موجود در این بخش، به خدمت فرآوران کمک می‌کند تا رکوردهای موردنظر خود را از رکوردهای دیگر متمایز سازند.

۲. **فراداده‌ها:** این بخش، فراداده مربوط به هر رکورد را مشخص می‌کند و دارای دو قسمت «اعلانات» و «معرفی فراداده» است. در این‌جا - همان‌گونه که قبلاً آمد - وجود حداقل یک عنصر ساده دابلین کور ضروری است. در کنار آن، امکان به کارگیری سایر عناصر فراداده‌ای (مثل عناصر فراداده کتابخانه کنگره، مدلاین و غیره) نیز وجود دارد.

۳. **درباره:**^{۲۹} این بخش، اختیاری و حاوی مطالبی درباره رکورد فراداده است. ساختار داخلی آن در پروتکل،

تعریف نشده و برعهده داده‌فرآوران است. بنابراین، آن‌ها می‌توانند برای ساختار معنایی آن تصمیم بگیرند. به طور معمول این بخش، اطلاعاتی درباره حق استفاده از فراداده و «مالکیت» آن به دست می‌دهد.

نمونه زیر، یک رکورد فراداده از یک سند فرضی است که داده‌فرآوری فرضی با نام National Document Organization براساس الگوی بالا ارسال کرده است.

```
<?xml version="1.0" encoding="UTF-8"
?>
- <OAI-PH xmlns="http://www.
openarchives.org/OAI/2.0/" xmlns:
xsi="http://www.w3.org/2001/
XMLSchema-instance" xsi:
schemaLocation="http://www.
openarchives.org/OAI/2.0/
http://www.
openarchives.org/OAI/2.0/OAI-PMH.xsd">
<responseDate>2003-03-
15T16:16:51+01:00</responseDate>
<request verb="GetRecord"
metadataPrefix="oai_dc" identifier="oai:
National-Documents-Organization:000007">
>http://edoc.nlai.ir/OAI-2.0</request>
- <GetRecord>
- <record>
- <header>
<identifier>oai:NDO:000007</identifier>
<datestamp>1997-07-18</datestamp>
<setSpec>تاریخ</setSpec>
</header>
- <metadata>
- <oai_dc:dc xmlns:oai_dc="http://
www.openarchives.org/OAI/2.0/oai_
```


پروتکل HTTP

HTTP ، پروتکلی است که غالباً کاربران اینترنت با آن کار می‌کنند. همین که نشانی‌ای مانند `http://example.com` را در بخش نشانی مرورگری مانند Internet Explorer وارد می‌کنیم و صفحه یا وبگاه موردنظر خود را روی صفحه نمایش خود می‌آوریم، از این پروتکل استفاده کرده‌ایم. نشانی‌ای که در مرورگر وارد می‌شود، اصطلاحاً URL نام دارد. OAI نشانی گنجینه داده‌های داده‌فرآور را URL پایه می‌نامند. برای درخواست اطلاعات، خدمت فرآور پس از درج URL پایه، با افزودن علامت «؟» (نشانه پرسش) به آخر آن، درخواست‌هایش را باید با یکی از دو روش Get یا Post برای داده فرآور بفرستد. در این پروتکل، شش نمونه درخواست یا دستور تعریف شده است:

۱- `GetRecord` : درخواست ارسال رکورد اطلاعاتی خاص.

۲- `Identify` : درخواست اطلاعاتی درباره محل نگهداری اطلاعات (گنجینه داده‌ها). در پاسخ به این دستور، می‌توان اطلاعاتی مانند نام گنجینه، URL پایه، شماره نسخه PMH (هم‌اکنون تنها دو نسخه ۱ و ۲ موجود است)، و قدیمی‌ترین رکورد موجود در هر گنجینه را دریافت کرد.

۳- `ListRecords` : درخواست بازبازی رکوردهای موردنظر از هر گنجینه داده. می‌توان این درخواست را به تاریخ و گستره موضوعی محدود کرد.

۴- `ListIdentifiers` : همانند `ListRecords` عمل می‌کند، با این تفاوت که به جای کل رکورد فراداده، فقط بخش آغازگر رکوردها را بازبازی می‌کند.

۵- `ListMetadataFormat` : این دستور، می‌پرسد که خدمت‌فرآور، فراداده را با چه شکل‌هایی ارائه می‌کند. مثلاً دوبلین کور، کتابخانه کنگره، و مدلاین هر کدام به شکلی خاص فراداده‌ها را ساختار بندی و عرضه می‌کنند. نهادهای دیگر، ممکن است الگوی آن‌ها را به کار بگیرند.

۶- `ListSets` : درخواست گستره موضوعی رکوردها

```
dc/" xmlns:dc="http://purl.org/dc/
elements/1.1/" xmlns:xsi="http://www.
w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.
openarchives.org/OAI/2.0/oai_dc/ http://
www.openarchives.org/OAI/2.0/oai_
dc.xsd">
<identifier>oai:NDO:000007</identifier>
<dc:type>نامه</dc:type>
<dcterms:medium>کاغذ ۲۱×۲۶ سانتیمتری
رنگ سفید با مهر برجسته قرمز و طلایی شیر و خورشید
</dc:format>
<dc:date>[بی تا]</dc:date>
<dc:creator>امین السلطان</dc:creator>
<dc:description>این سند، نامه‌ای است از
امین السلطان به ناصرالدین شاه درباب قضیه رژی،
امین السلطان پس از نقل گفتگوی خود با رشتی، کاردار
سفارت روس در تهران، نکاتی از مواضع میرزا حسن
آشتیانی، رهبر روحانی دارالخلافه، در مورد قضیه رژی
برای شاه بازگو کرده است.
</dc:description>
<dc:source>سازمان اسناد ملی ایران</dc:source>
<dc:source>۱۹-شماره ۲۸</dc:source>
<dc:language>فارسی</dc:language>
<dc:subject>ناصرالدین شاه</dc:subject>
<dc:subject>قاجاریه</dc:subject>
<dc:subject>امین السلطان</dc:subject>
</oai_dc:dc>
</metadata>
</record>
</GetRecord>
</OAI-PH>
```

در هر گنجینه.

مثال زیر، نمونه‌ای از به‌کارگیری دستورهای بالاست:
<http://www.example.com/oai-repository?Verb=GetRecord&set=Yazd&identifier=oai:example.com:5000>

درخواست بالا در خط آغازین، URL پایه گنجینه فرضی مربوط به داده‌فرآوری را به نام example آورده است. پس از درج آن نشانی، علامت پرسش می‌آید و آن‌گاه از داده‌فرآور خواسته می‌شود تا به او رکوردهائی بدهد (GetRecord) که گستره موضوعی آن یزد (set=Yazd) و شناسه آن ai:examples.com:5000 است.

در example پاسخ، رکوردهائی به شکل فراداده ارسال می‌کند. شکل ارسال داده‌ها همانند نمونه‌ای است که در بخش XML توضیح داده شد. از آن‌جا که اطلاعات دریافتی به زبان XML است، آن مرکز به راحتی می‌تواند این صفحات را ذخیره و فیلدهای اطلاعاتی رکوردهای دریافتی را از طریق یک برنامه ساده رایانه‌ای، پویس و در پایگاه داده‌های خود ذخیره و اشاعه کند.

پیاده‌سازی PMH برای فهرستگان اسناد آرشیوی

مطالب بالا، این پرسش را پیش می‌آورد که چه شرایطی نیاز است تا نهادی بتواند در این سامانه به صورت داده‌فرآور یا خدمت‌فرآور عمل کند و فراداده‌های اسناد خود را به اشتراک بگذارد یا فراداده‌ها را از گنجینه اسناد دیگران برداشت کند تا فهرستگان به وجود آید. در پاسخ باید گفت که هیچ محدودیتی برای این کار وجود ندارد. هر مرکز آرشیوی و هر خدمات‌دهنده اطلاعات اسناد آرشیوی، می‌تواند با فراهم آوردن زمینه‌های فنی لازم، بدون هیچ مانعی، به ترتیب به عنوان داده‌فرآور و خدمت‌فرآور ایفای نقش کند.

بنیاد OAI، معتقد است که OAI-PMH به گونه‌ای طراحی شده که به راحتی قابل پیاده‌سازی است. برای این

کار، لازم است خدمتگر (سرور) تحت وب، درخواست‌ها را از خدمت‌فرآور دریافت و تجزیه و تحلیل کند. این کار، برای کسی که تجربه کار با خدمتگرهای تحت وب و CGI^{۳۰} اسکریپت را دارد، کم‌تر از یک روز طول خواهد کشید.^{۳۱}

در این فرایند، طبیعتاً کار اصلی برعهده داده‌فرآور است تا بتواند به راحتی درخواست‌های ارسالی مطابق این پروتکل را تجزیه و تحلیل و فراداده‌ها را از گنجینه داده‌های خود استخراج و به شکل XML ارسال کند. اگر داده‌ها قبلاً به‌خوبی سازماندهی و فراداده‌ها نیز بر پایه دابلین کور نمایه‌شده و سازوکاری برای استخراج و جستجوی خودکار فراداده‌ها از پایگاه یا گنجینه فراداده‌ها وجود داشته باشد، پیاده‌سازی پروتکل ساده خواهد بود.

شرط مهم دیگر برای ایفای نقش داده‌فرآور، افزودن اجزائی اضافی - بجز آنچه معمول است - به رکوردهای کتابشناختی است. این اجزای اضافی، ممکن است شامل مواردی نظیر مهر، تاریخ و شناسه‌گر باشند. همچنین پیگیری و اعلان رکوردهای حذف شده، یکی دیگر از نیازمندی‌هاست که باید داده‌فرآور انجام دهد. (Shleman, 2002, p 414)

خدمت‌فرآوران نیز، می‌توانند درخواست‌های خود را از طریق مرورگر اینترنتی ارسال کنند، و این به سادگی با زبان‌های رایانه‌ای موجود طراحی‌پذیر است. برنامه‌ها باید این امکان را داشته باشند که دستورهای لازم را، طبق فرمان‌های ششگانه بالا، به صورت URI ارسال کنند. حتا بدون داشتن این‌گونه برنامه‌ها، تنها با تایپ URL پایه و فرمان‌های یاد شده در محل نشانی مرورگری اینترنتی مانند Internet Explorer، اطلاعات لازم قابل درخواست و دریافت است. سپس، لازم است که برنامه پویسگر متن، پس از دریافت فراداده‌ها، داده‌هائی را که داده‌فرآور به صورت XML ارسال کرده است، استخراج و به پایگاه موردنظر خود بیفزاید،^{۳۲} تا فهرستگان موردنظر را ایجاد کند یا توسعه دهد.

بحث و نتیجه‌گیری

گردآوری اسناد آرشیوی، باتوجه به نوع و ویژگی‌های متفاوت آن‌ها در یک محل دشوار است. این‌گونه اسناد، دارای ارزش تاریخی، فرهنگی، اجتماعی، سیاسی و اقتصادی است که به راحتی نهادها و افراد متولی آن‌ها حاضر نیستند آن‌ها را در اختیار دیگران قرار دهند. به علت ارزش و اهمیت این اسناد، مالکان و متولیان آن ترجیح می‌دهند که آن‌ها را نزد خود حفظ و یا محدودیت‌های خاصی برای دسترس‌پذیری آن‌ها وضع کنند. در این صورت، ایجاد فهرست مشترک از این‌گونه اسناد، می‌تواند به مرئی کردن آن‌ها کمک کند، بدون این که نیازی به انتقال سند از محل اصلی خود باشد.

مرئی شدن این اطلاعات، ممکن است از جهاتی برای مالکان و متولیان آن‌ها سودمند باشد. به عنوان مثال، دسترس‌پذیر کردن اطلاعات کتابشناختی اسناد، ممکن است بهره‌ اقتصادی برای سازمان‌ها و نهادهای متولی اسناد داشته باشد. مشکل اصلی در تهیه چنین فهرستی، فرایند درخواست و دریافت اطلاعات مربوط به آن اسناد است. فناوری اطلاعات، این امکان را به وجود آورده است تا گردآوری اطلاعات بدون واسط انسانی به سرعت و دقت صورت گیرد. پروتکل PMH، با بهره‌گیری از پروتکل‌های برگرفته از فناوری اطلاعات مانند HTTP و XML، این امر را میسر می‌سازد.

بهره‌گیری از عناصر فراداده پیشنهادی دوبلین کور برای نمایه‌سازی این اسناد، بستری فراهم می‌کند تا تمامی مراکزی که اطلاعات اسناد آرشیوی خود را به اشتراک می‌گذارند، رویه‌ای یکدست و جهانی برای ارائه اجزای اطلاعاتی لازم در پیش گیرند. این امر، سبب می‌شود تا بستری مناسب برای تبادل اطلاعات اسناد، فراتر از سطح محلی یا ملی، در سطح جهانی فراهم آید. محدود کردن اجزای اطلاعاتی به پانزده عنصر ساده و سه عنصر توضیحی، ممکن است محدودیت‌هایی در ارائه جزئیات هر سند پیش آورد که این موضوع را می‌توان از معایب دوبلین کور دانست. در برابر، عناصر فراداده دوبلین

کور، شهرت جهانی دارند و هر سامانه، به راحتی می‌تواند به صورت خودکار آن‌ها را تجزیه و تحلیل و ذخیره کند. نکته دیگر این که، بخش عمده‌ای از اطلاعات ارزشمند موجود در شبکه جهانی وب (مانند اسناد آرشیوی)، در پایگاه داده‌هایی ذخیره شده که بسیاری از آن‌ها از دید پیمایشگرها^{۳۳} پنهان مانده است. موتورهای جستجوی سنتی نیز نمی‌توانند به درون آن‌ها بخزند. یکی از راه‌های مؤثر برای ایجاد تعامل با این پایگاه‌ها، استفاده از فراجستجوگرهاست که به جستجوگران کمک می‌کند تا از طریق یک عبارت جستجو، همزمان در چندین پایگاه جستجو کنند. (گارسیا و دیگران، ۲۰۰۶)

هر فراجستجوگر، سه کار مهم انجام می‌دهد: پس از دریافت عبارت جستجو، بهترین پایگاه را برای جستجو برمی‌گزیند (انتخاب پایگاه)، عبارت جستجو را به صورتی برمی‌گرداند که مناسب آن پایگاه باشد (برگردان عبارت جستجو)، و سرانجام اطلاعات را از پایگاه‌های مختلف بازبایی و ادغام (درهم) می‌کند و نتایج را به جستجوگر نشان می‌دهد. از لحاظ عملی، ساخت فراجستجوگر، دشوار و چالش‌برانگیز است؛ بویژه روشی فراگیر برای همخوانی میان پایگاه‌های مختلف وجود ندارد. (Ipeivotis, Bavry & Gravano, 2002)

پروتکل Z39.50، یکی از طرح‌هایی است که در این راستا ایجاد شده است. اما ثابت شده که بسیار مشکل است تا بتوان خدمات جستجو را با کیفیت بالا از طریق این استاندارد ایجاد کرد. علت آن قائم بودن هر کدام از خدمتگرهای اطلاعاتی به استانداردی خاص است. (Yu, 2006, pp 288-293)

دیگر پروتکل‌هایی که تا به امروز در این حوزه ارائه شده است، هر کدام کاستی‌های خود را دارد. مثلاً مشکل ایزو ۲۷۰۹، این است که محدود به سه شناسه (کد کنترلی (نوع فیلد، فیلد فرعی، و نشانه‌گر) است و نمی‌تواند مارک‌های گوناگون (یو اس مارک، ایران مارک و غیره) و کدهای کاراکتری (یونی کد، کد اسکی و غیره) را تشخیص دهد. (Yu, 2006, pp 288-293)





20. right holder.
21. Refinements.
22. Qualifiers.
23. encoding schemes.
24. markup languages.
25. Header.
26. Identifier.
27. Datestamp.
28. Set.
29. About.
30. Common Gateway Interface.
31. <http://www.openarchives.org/documents/FAQ.html#How%20long%20will%20it%20take%20me%20to%20implement%20the%20protocol>.
32. Crawlers.
33. Meta-searchers.

منابع:

۱. اکرمی، صفیه؛ سوری، علی. «آرشیو ملی یا اسناد ملی»، گنجینه اسناد، ۶۹ (بهار ۱۳۸۷)، ص ۷۸.
۲. دیانی، محمدحسین. مقدمه‌ای بر آرشیو. مشهد: دانشگاه فردوسی مشهد، ۱۳۷۷، ص ۴۰.
۳. موسی‌چمنی، اعظم. «آرشیوهای دانشگاهی با تأکید بر آرشیو دانشگاه هاروارد [HUA]»، گنجینه اسناد، ۷۱ (پائیز ۱۳۸۷)، ص ۵۲.
- levy, David M. "Documents and libraries: A sociotechnical perspective", In Ann Peterson Bishop, Nancy A. van House, and Barbara Battenfield, ed., Digital

افزون بر این کارکرد، رعایت قواعد پروتکل PMH به طور غیرمستقیم تأثیرات دیگری نیز داشته است. یکی از این تأثیرات، امکان نمایه شدن منابع در موتورهای جستجو است. طبق یک پژوهش، ۶۵٪ منابع عرضه شده برپایه این پروتکل در یاهو، ۴۴٪ در گوگل، و ۷٪ در ام. اس. ان. نمایه شده‌اند. (M, Cown.etal, 2006, pp 66-73)

بنابراین پروتکل مورد بحث علاوه بر مزایای یاد شده، امکان پویش فراداده‌های اسناد آرشیوی را برای موتورهای جستجو نیز فراهم می‌آورد.

◆ پی‌نوشت‌ها

1. Emephera.
2. gray literature.
3. Protocol for Metadata Harvesting.
4. open archive initiative.
5. Santa Fee.
6. Object Reuse and Exchange.
7. Open Archives Initiative (FAQ)", from: <http://www.openarchives.org/documents/FAQ.html> , retrieved 02.04.1387.
8. Ibid.
9. Repository.
10. Structured.
11. low-barrier.
12. service provider.
13. data provider.
14. Value.
15. Field.
16. Simple.
17. Qualified.
18. Audience.
19. Provenance.

Joint Conference on Digital Libraries (JCDL 2003), Institute of Electrical and Electronics Engineers, Inc., Los Alamitos, CA. (2003): 315-317.

Suleman, H. "Introduction to the Open Archives Initiative Protocol for Metadata Harvesting", Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries. (2002): 414.

Garcia, E. A. „Experiences in using the OAI-PMH through the construction of the OAHermes Metasearch engine“, International Conference on Dublin Core and Metadata Applications, 3-6 October. (2006).

Ipeirotis, P. G.; Barry, T.; Gravano, L. "Extending SDARTS: extracting metadata from Web databases and interfacing with the Open Archives Initiative", Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries, Portland, OR, USA, 13-17 July. (2002): 162-70, from: <http://doi.acm.org/0.1145/544220.544254> , retrieved 02.04.1387.

Yu, S. C. „Study on digital archives standard for library automation system“ Proceedings of the Asia-Pacific Conference on Library & Information Education & Practice. (2006): 288-293.

McCown, F et al. "Search engine coverage of the OAI-PMH corpus". Internet Computing, IEEE. Vol. 10, No. 2(2006): 66-73.

Garcia, E. A. „Experiences in using the

Library Use: Socila Practice in Design and Evaluation. India: Ane Book ,2005, P.25-42.

Van de Sompel, H; Lagoze, c. "The Santa Fe Convention of the Open Archives Initiative", D-Lib Magazine. Vol. 6, No 2(2000). From: <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html> , retrieved 02.04.1387.

Yu ,shien-chiang ;Chen,Hsueh-hua ; Chang , Huai-wen. "Building and Open Archive Union Catalog for Digital Archives", The Electronic Library, Vol. 23, No. 4 (2005): 410-418.

Lagoze, C.; Van de Sompel, H. "The Open Archives Initiative: Building a Low-Barrier Interoperability Framework", (2001). From: <http://www.openarchives.org/documents/oai.pdf>, retrieved 02.04.1387.

Gardner, S. R. "Building the Data Warehouse", Communications of the ACM, vol. 41, No. 9 (1998): 52 – 60.

Henshaw, Robin; Valauskas, Edward J. "Metadata as a Catalyst: Experiments with Metadata and Search Engines in the Internet Journal, First Monday", Libri, Vol. 51, No. 2 (2001): 86–101.

Ward, J. "A Quantitative Analysis of Unqualified Dublin Core Metadata Element Set Usage within Data Providers Registered with the Open Archives Initiative", Proceedings: 2003

OAI-PMH through the construction of the OAHermes Metasearch engine“, International Conference on Dublin Core and Metadata Applications, 3-6 October. (2006).

