

## میانگش‌پذیری نظام‌های فراداده‌ای مبتنی بر ایکس ام ال و موتورهای کاوش

● زهرا جعفری

**استاندارد فراداده‌ای  
هسته دوبلین،  
طرحی بین‌المللی و  
میان‌رشته‌ای است که  
مجموعه عناصری ساده  
و کارآمد برای توصیف  
طیف گسترده‌ای  
از منابع اطلاعاتی  
شبکه‌ای را ارائه  
می‌دهد**

سخنرانی ماهانه انجمن کتابداری و اطلاع‌رسانی ایران در شهریورماه به بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی اختصاص داشت.

سید مهدی طاهری، کارشناس ارشد کتابداری و اطلاع‌رسانی و عضو کمیته پژوهش انجمن کتابداری و اطلاع‌رسانی ایران، در آغاز این جلسه به تاثیر وب بر انتشار اطلاعات اشاره کرد و گفت: پیدایش وب به‌عنوان مهم‌ترین فناوری و رسانه اینترنت و با قابلیت‌های منحصر به‌فرد باعث تمایل سازمان‌ها، ناشران، مراکز اطلاع‌رسانی، و پدیدآورندگان به انتشار و ارائه اطلاعات و خدمات خود بر روی این رسانه شده است که عملاً منجر به انتشار حجم گسترده‌ای از اطلاعات بر روی آن شبکه گردیده است.

وی افزود: برای تسریع و تسهیل دسترسی به این حجم گسترده دو جنبش نرم‌افزاری در سطح بین‌المللی بوجود آمدند، نخست ایجاد و گسترش موتورهای کاوش، و دیگر، جنبش فراداده.

طاهری در تشریح جنبش گسترش موتورهای کاوش گفت: در جنبش نخست برخی اشخاص و سازمان‌های انتفاعی و غیر انتفاعی به تولید ابزارهای کاوشی چون واندکس<sup>۱</sup>، علیوب<sup>۲</sup>، و جامپ استیشن<sup>۳</sup>، تا اولین موتورهای کاوش پیمایش‌گر- پایه<sup>۴</sup> و تمام‌متن مانند وب کراولر<sup>۵</sup>، و موتورهای کاوش امروزی مثل گوگل، یاهو، و جدیدترین آنها کوئیل<sup>۶</sup> (با نرم‌افزارها و الگوریتم‌ها بسیار پیشرفته برای نمایه‌سازی و رتبه‌بندی) پرداختند.

وی افزود: در جنبش فراداده، نظام‌های سازماندهی دانش مدنظر قرار گرفت که همواره در طول تاریخ همگام با تحولات حوزه فناوری های اطلاعاتی گسترش یافته‌اند و شامل دو رویکرد سازگاری نظام های سنتی با محیط و رسانه های جدید، و ایجاد نظام هایی نوین برای حداکثر بهره‌وری از امکانات و قابلیت های محیط جدید بوده

است. طاهری درباره تفاوت این دو نظام خاطر نشان کرد: موتورهای کاوش علیرغم نقاط ضعف و کاستی هایی که نسبت به نظام های فراداده ای در دسترس پذیر نمودن اشیای محتوایی وب دارند، به دلیل سادگی استفاده، غنای پایگاه های داده ای، و عواملی دیگر، هشتاد و هشت درصد از زمانی که کاربران وب صرف جستجوی اطلاعات می کنند را به خود اختصاص داده اند.

وی ادامه داد: گرایش قابل ملاحظه کاربران به استفاده از موتورهای کاوش و قابلیت ها و توانایی های بالقوه نظام های فراداده ای، ضرورت میان کنش پذیری دستاوردهای دو جنبش را بدیهی و بر این اساس خط مشی هایی در موتورهای کاوش به منظور نمایه سازی برچسب ها و عناصر فراداده ای تدوین و تمهیداتی در نظام های فراداده ای برای افزایش بازیافت پذیری آنها توسط موتورهای کاوش در نظر گرفته شده است. انجام پژوهش های بیشتر در این زمینه از دیگر پیشنهادات این پژوهشگر بود.

طاهری در ادامه به مفهوم مارک اشاره کرد و گفت: مارک یک قالب است که امکان خواندن و پردازش اطلاعات فهرستنویسی هر اثر مشخص را به‌وسیله رایانه فراهم می کند و برای ثبت داده‌های فهرستنویسی و استفاده بعدی از آنها برای تبادل، جستجو، بازیابی، نمایش اطلاعات کتابشناختی، و یا چاپ پیشینه‌های کتابشناختی استفاده می‌شود.

وی همچنین درباره طرح فراداده‌ای هسته دوبلین گفت: استاندارد فراداده‌ای هسته دوبلین، طرحی بین‌المللی و میان‌رشته‌ای است که مجموعه عناصری ساده و کارآمد برای توصیف طیف گسترده‌ای از منابع اطلاعاتی شبکه‌ای را ارائه می‌دهد.

طاهری در ادامه درباره کیفیت نمایه‌سازی در موتورهای کاوش گفت: نمایه‌سازی موتورهای کاوش از

نوع نمایه‌سازی واژه‌ای است که در این تحقیق جامعیت نمایه‌سازی، و امکان بازیابی اشیای محتوایی نمایه‌سازی شده از طریق عبارات کلیدی - که بر اساس اطلاعات مربوط به محل و جایگاه قرار گرفتن واژه‌ها در شیء محتوایی انجام می‌گیرد - به عنوان کیفیت نمایه‌سازی در نظر گرفته شده است.

وی در ادامه به چگونگی انجام تحقیق اشاره کرد و گفت: در مرحله نخست ۱۰۰ شیء محتوایی (کتاب الکترونیکی) در قالب دو گروه ۵۰ پیشینه‌ای در بستر زبان نشانه‌گذاری گسترش‌پذیر<sup>۲</sup> قرار گرفت که هر پیشینه از گروه اول حاوی یک پیشینه فراداده‌ای هسته دوبلین (مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر) به‌عنوان برجسب فرا بعد از عنصر ریشه جاسازی (درج) و بر روی یک حوزه فرعی با آدرس: <http://www.dcmixml.is-lamicdoc.com> بر وب سایت مرکز اطلاعات و مدارک اسلامی انتشار یافت. همچنین هر یک از اشیای محتوایی گروه دوم حاوی پیشینه‌ای فراداده‌ای قالب مارک ۲۱ مبتنی بر ایکس ام ال روی حوزه فرعی دیگری در آدرس <http://www.marxml.islamicdoc.com> بر روی همان وب سایت منتشر شد. اشیای محتوایی از وب سایت <http://www.archive.org> و از مجموعه منابع کتابخانه دیجیتالی کالیفرنیا و از یک رده موضوعی (Knowledge theory of) انتخاب شده بودند.

وی ادامه داد: سپس حوزه‌های مورد مطالعه به گوگل و یاهو معرفی شدند. یافته‌ها نشان داد: تمامی عناصر فراداده‌ای هسته دوبلین و تمامی عناصر فراداده‌ای مارک ۲۱ مورد مطالعه، به‌طور کامل توسط نرم‌افزارهای نمایه‌سازی موتورهای کاوش گوگل و یاهو نمایه‌سازی شدند.

وی درباره نتایج این پژوهش گفت: نتایج این تحقیق گویای آن بود که برای موتورهای کاوش گوگل و یاهو، بین اشیای محتوایی با ساختار مسطح و سلسله مراتبی، و بین پیشینه‌های فراداده‌ای با برجسب‌های عناصر مبتنی بر زبان و بدون آن، تفاوتی وجود ندارد.

وی خاطر نشان کرد: گرایش طرح‌های فراداده‌ای، به استفاده از فناوری پیشرفته زبان نشانه‌گذاری گسترش‌پذیر باعث شده تا امکان جاسازی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در اشیای دیجیتالی در قالب‌های مختلف، از جمله قالب PDF با استفاده از RDF و امکان

نمایه‌سازی کامل آنها توسط موتورهای کاوش عمومی فراهم شود.

طاهری تسهیل تبدیل طرح‌های فراداده‌ای به یکدیگر و در نتیجه افزایش میان کنش‌پذیری نظام‌های فراداده‌ای و امکان ایجاد پروفایل‌های کاربردی<sup>۳</sup> به‌منظور تامین نیازهای خاص سازمان‌ها و محیط‌های اطلاعاتی مختلف در قالب متر<sup>۴</sup> را از دیگر مزایای استفاده از فناوری پیشرفته زبان نشانه‌گذاری گسترش‌پذیر خواند.

وی درباره سایر نتایج پژوهش خود گفت: براساس نتایج این تحقیق، اشیای محتوایی حاوی این عناصر، از طریق جستجوی کلیدواژه‌ای و عبارات کلیدی به آسانی قابل بازیابی هستند؛ همچنین تمامی عناصر فراداده‌ای برای نرم‌افزارهای نمایه‌سازی موتورهای کاوش گوگل و یاهو، ارزش نمایه‌شدن را دارند.

طاهری اضافه کرد: هیچ یک از دو طرح فراداده‌ای هسته دوبلین و مارک از لحاظ دسترس‌پذیری از سوی موتورهای کاوش عمومی، بر دیگری برتری نداشته، و هر دو طرح برای سازماندهی اشیای محتوایی وب مناسب هستند.

وی همچنین تدریس رسمی رویکردهای نوین نظام‌های فراداده‌ای در واحدهای درسی مربوط به سازماندهی اطلاعات در رشته کتابداری و اطلاع رسانی، برگزاری کارگاه‌های آموزشی در این زمینه، طراحی و پیاده‌سازی ابزارها و نرم‌افزارهای تولید و ایجاد پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در قالب زبان نشانه‌گذاری گسترش‌پذیر به صورت پیوسته و ناپیوسته، ایجاد نظام‌های میان کنش‌پذیر در شبکه‌های اطلاعاتی و انجام پژوهش‌های بیشتر در این زمینه را پیشنهاد کرد.

#### پی‌نوشت‌ها:

1. Wandex
2. Aliweb
3. Jumpstation
4. Crawler-based
5. Web crawler
6. Cuil
7. XML
8. Application profile
9. METS

هیچ یک از دو طرح فراداده‌ای هسته دوبلین و مارک از لحاظ دسترس‌پذیری از سوی موتورهای کاوش عمومی، بر دیگری برتری نداشته، و هر دو طرح برای سازماندهی اشیای محتوایی وب مناسب هستند