

بررسی تأثیر بستر نحوی بر نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای توسط موتورهای کاوش وب: مقایسه پیشینه‌های فراداده‌ای مارک ۲۱ و هسته دابلین

فهیمة باب‌الحوائجی^۱
سید مهدی طاهری^۲
زهرا آقاعابدی^۳

مطالعات دانش‌شناسی
سال اول، شماره ۳، تابستان ۹۴

تاریخ دریافت: ۹۳/۰۹/۲۱

تاریخ پذیرش: ۹۴/۰۴/۰۲

چکیده

هدف: پژوهش حاضر با هدف تبیین وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دابلین و مارک ۲۱ در بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر (ایکس. ام. آل.) مبتنی بر رویکرد پیوند توسط موتورهای کاوش وب انجام شد. **روش:** روش پژوهش، رویکرد تحلیلی-مقایسه‌ای بود. جامعه پژوهش را یک‌صد پیشینه فراداده‌ای مبتنی بر ایکس. ام. آل. در قالب دو گروه تشکیل می‌دادند. گروه نخست، دارای عناصر فراداده‌ای هسته دابلین و گروه دیگر، دارای عناصر فراداده‌ای مارک ۲۱ بودند که بر روی وب‌سایت <http://www.aghaabedi.com> انتشار یافتند. وب‌سایت یادشده به صورت مستقیم به موتورهای کاوش انتخابی (گوگل و یاهو) معرفی شدند. ابزار گردآوری داده‌ها سیاهه واری بود. **یافته‌ها:** یافته‌ها حاکی از آن است که تمامی ارزش‌های عناصر پیشینه‌های فراداده‌ای هسته دابلین و مارک ۲۱ در بستر ایکس. ام. آل. توسط موتورهای کاوش گوگل و یاهو نمایه شدند و امکان بازیابی هر یک از پیشینه‌ها بر اساس ارزش عناصر آن‌ها در محیط موتورهای کاوش فراهم است. بر اساس نتایج آزمون آماری تفاوت معناداری میان نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دابلین و مارک ۲۱ مبتنی بر ایکس. ام. آل. توسط موتورهای کاوش گوگل و یاهو مشاهده نشد. **نتیجه‌گیری:** بستر نحوی پیشینه‌های فراداده‌ای مورد مطالعه موجب افزایش دسترس‌پذیری آن‌ها از طریق موتورهای کاوش وب گردید. این امکان برای

۱. دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران

f.babalhvaeji@gmail.com

۲. استادیار گروه علم اطلاعات و دانش‌شناسی دانشگاه علامه طباطبائی taherismster@gmail.com

۳. کارشناس ارشد علوم کتابداری و اطلاع‌رسانی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران

eaghaabedi@gmail.com

بستر نحوی پیشین (اچ.تی.ام.آل.) پیشینه‌های فراداده‌ای فراهم نبود. همچنین ساختار داده‌ای پیشینه‌ها بر اساس ایکس.ام.آل. تأثیری بر روی رتبه‌بندی آن‌ها نداشته، و از این جنبه هیچ‌یک از طرح‌های هسته دوبلین و مارک ۲۱ بر دیگری برتری نشان نداد.

واژگان کلیدی: زبان نشانه‌گذاری گسترش‌پذیر، مارک ۲۱، موتورهای کاوش وب، نمایه‌سازی و رتبه‌بندی، هسته دوبلین.

مقدمه

فراداده، داده‌هایی درباره اشیای محتوایی است که امکان توصیف، کشف، دسترسی، بازنمون، و مدیریت آن‌ها را فراهم می‌آورد. کارآمدی این داده‌ها در محیط‌های اطلاعاتی شبکه‌ای به شیوه‌های گوناگونی مورد ارزیابی قرار گرفته، و به‌عنوان داده‌های ساختارمند، مبنای تحقق نسل جدید وب (وب معنایی) به شمار می‌آید (برنرزی-لی^۱، ۲۰۰۶). برای پاسخ به تحولات و نیاز به پردازش و سازمان‌دهی اشیای محتوایی وب، از زمان پیدایش این شبکه تاکنون، استانداردهای فراداده‌ای گوناگونی توسعه یافته‌اند، و استانداردهای پیشین نیز با هدف سازگاری با این محیط تمهیداتی را اندیشیده‌اند. در این میان، برخی از استانداردهای فراداده‌ای مانند طرح فراداده‌ای هسته دوبلین و قالب فراداده‌ای مارک ۲۱ (و مشتقات آن) از جمله پرکاربردترین‌ها به شمار می‌آیند. به کارگیری فراداده به دو روش امکان‌پذیر است: نخست، جاسازی (درج) پیشینه فراداده‌ای در شیء محتوایی، و دیگر، پیوند پیشینه فراداده‌ای با شیء محتوایی. همچنین نام‌های برجسب عناصر برخی از استانداردها مبتنی بر زبان طبیعی است مانند هسته دوبلین، و برخی با زبان طبیعی همخوانی ندارند مانند قالب فراداده‌ای مارک (فتاحی و طاهری، ۱۳۸۸).

از سوی دیگر، موتورهای کاوش محبوب‌ترین ابزار جستجوی محتوای وب نزد کاربران هستند. به‌نحوی که حدود ۹۰ درصد از زمانی که کاربران صرف جستجوی محتوای وب می‌کنند، در محیط موتورهای کاوش وب است (بیفت^۲ و همکاران، ۲۰۰۵). کاربردپذیری بالای موتورهای کاوش (سهولت در درک و یادگیری قابلیت‌های آن‌ها) که به دلیل محیط رابط کاربر بسیار ساده فراهم شده است، در برداشتن حجم گسترده‌ای از

1. Burners-Lee

2. Bifet

محتوای وب در پایگاه داده‌ای، ارائه فهرستی فراوان از پیشینه‌ها در برابر یک درخواست، و روزآمدسازی مداوم محتوا از مهم‌ترین دلایل این محبوبیت به شمار می‌آیند. ارائه خدماتی متنوع افزون بر خدمات جستجو و بازیابی اطلاعات، دلیل دیگری بر افزایش گرایش به بهره‌گیری از آن‌ها از سوی کاربران است. این قابلیت‌ها و نیز علاقه روزافزون کاربران وب به موتورهای کاوش وب و رفع نیازهای اطلاعاتی خود در بستر این موتورها باعث گردیده، اشخاص حقیقی و حقوقی تولیدکننده محتوا و ارائه‌دهنده خدمات پیوسته، به دسترس‌پذیری محتوای خود از طریق موتورهای کاوش تمایلی چشمگیر نشان دهند. بافت‌های اطلاعاتی که از فراداده استفاده می‌کنند، از این قاعده مستثنا نیستند.

بدین منظور اقدامات متعددی از سوی طراحان استانداردهای فراداده‌ای و توسعه‌دهندگان موتورهای کاوش برای افزایش میانگین‌پذیری این دو ابزار مؤثر در دسترسی به محتوای وب صورت گرفته است. گروه نخست سعی نموده‌اند شرایط اشاره‌شده از سوی توسعه‌دهندگان موتورهای کاوش برای افزایش نمایه‌پذیری و کسب رتبه مناسب در نتایج رعایت نمایند، و گروه دیگر نیز در راستای شناسایی و اولویت قرار دادن استخراج فراداده‌های اشیای محتوایی کوشش نموده‌اند. اقبال نرم‌افزارهای خزنده-نمایه‌ساز (روبات) موتورهای کاوش به عناصر و ارزش‌های عناصر فراداده‌ای پیش و بیش از دیگر محتوای یک شیء محتوایی، و انتخاب بستر نحوی جدید (زبان نشانه‌گذاری گسترش‌پذیر) از سوی طراحان استانداردهای فراداده‌ای از مهم‌ترین اقدامات محسوب می‌شوند.

بر پایه آنچه پیش‌تر بیان گردید، مسئله‌ای که اکنون مطرح می‌گردد آن است که آیا تغییر بستر نحوی به‌طور عملی باعث افزایش نمایه‌پذیری و بهبود رتبه پیشینه‌های فراداده‌ای توسط موتورهای کاوش وب خواهد شد؟ موتورهای کاوش وب چگونه پیشینه‌های فراداده‌ای را نمایه می‌کنند و در نتایج جستجو ارائه می‌دهند؟ آیا ساختار و نوع نام‌های برچسب عناصر (از لحاظ وابستگی به زبان طبیعی) پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر که در برخی از استانداردها همانند هسته دوبلین و مارک ۲۱ متفاوت است، بر روی شیوه نمایه‌سازی و جایگاه (رتبه) آن‌ها در میان نتایج جستجو

تأثیر گذار است؟ کدام یک از استانداردهای فراداده‌ای هسته دوبلین و مارک ۲۱، به عنوان پرکاربردترین فراداده‌ها، برای سازمان‌دهی اشیای محتوایی از لحاظ دسترس پذیری در محیط موتورهای کاوش وب مناسب تر هستند؟ پژوهش حاضر در راستای پاسخگویی به این مسئله طرح ریزی شده است.

هدف و پرسش‌های پژوهش

هدف اصلی پژوهش حاضر، تبیین وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر نحوی زبان نشانه‌گذاری گسترش پذیر (ایکس. ام. آل). مبتنی بر رویکرد پیوند توسط موتورهای کاوش وب بود. برای تحقق این هدف، پرسش‌های ذیل طراحی شدند:

۱- وضعیت نمایه‌سازی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر نحوی ایکس. ام. آل. مبتنی بر رویکرد پیوند توسط موتورهای کاوش وب (گوگل و یاهو) چگونه است؟

۲- وضعیت رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در قالب ایکس. ام. آل. مبتنی بر رویکرد پیوند توسط موتورهای کاوش وب (گوگل و یاهو) چگونه است؟

۳- کدام یک از طرح‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر نحوی ایکس. ام. آل مبتنی بر رویکرد پیوند برای تولید پیشینه‌های فراداده‌ای دسترس پذیر در محیط موتورهای کاوش وب (گوگل و یاهو) مناسب تر است؟
همچنین این پژوهش دارای یک فرضیه بود:

میان وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر نحوی ایکس. ام. آل مبتنی بر رویکرد پیوند توسط موتورهای کاوش وب تفاوت معناداری وجود دارد.

مروری بر پژوهش‌های پیشین

میانکنش پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب با هدف بهبود دسترسی به اشیای محتوایی و استفاده از بستر معناشناختی داده‌های ساختارمند (فراداده‌ها) در بازیابی و

پدیدار نمایی اشیای محتوایی همواره مورد توجه پژوهشگران و متخصصان علوم رایانه و اطلاعات بوده است. در پژوهش‌های انجام شده در این حوزه، قابلیت‌های ساختارمندی داده‌ها که در پیشینه‌ها و عناصر فراداده‌ای تبلور می‌یابد، و رفتار نرم‌افزارهای نمایه‌ساز موتورهای کاوش نسبت به این پیشینه‌ها و عناصر و چگونگی نمایش آن‌ها در نتایج جستجو، و ارائه پیشنهادها و راهکارهایی برای سازگاری بیشتر مورد مطالعه قرار گرفته است. طراحان خدمات وبی نیز از یافته‌های پژوهش‌های انجام شده در راستای افزایش تعامل دو نظام یادشده، و به پیروی از آن، کسب رضایت کاربران نهایی بهره جسته‌اند.

با بررسی این پژوهش‌ها می‌توان آن‌ها را به دو گروه کلی تقسیم نمود. مبنای این گروه‌بندی بستر نحوی پیشینه‌های فراداده‌ای و واکنش موتورهای کاوش وب به ساختار نحوی و معنایی بستر مورد اشاره است. زیرا مهم‌ترین نقطه بارز و تفاوتی که در پیاده‌سازی فراداده‌ها و واکنش موتورهای کاوش به چشم می‌خورد، به بستر نحوی وابسته است. گروه نخست، پژوهش‌هایی بودند که به شیوه‌های نمایه‌سازی و پیدا نمایی موتورهای کاوش در رویارویی با پیشینه‌ها و عناصر فراداده‌ای پیاده‌سازی شده در قالب زبان نشانه‌گذاری فرا متن (اچ.تی.ام.آل.) اختصاص داشته‌اند، و گروه دیگر، پیشینه‌های فراداده‌های مبتنی بر زبان نشانه‌گذاری گسترش پذیر (ایکس.ام.آل.) و نمایه پذیری و پیدا نمایی آن‌ها در نتایج جستجوی موتورهای کاوش را مدنظر قرار داده‌اند.

گروه نخست، پژوهش‌هایی که به مطالعه واکنش موتورهای کاوش وب به عناصر و پیشینه‌های فراداده‌ای مبتنی بر اچ.تی.ام.آل. پرداختند. در این پژوهش‌ها سه گرایش مشاهده می‌شود. گرایش اول به تأثیر بر چسب‌ها و فرابرجسب‌های زبان نشانه‌گذاری فرا متن بر بهبود بازیابی و رتبه‌بندی اشیای محتوایی توسط موتورهای کاوش ارتباط دارد. عمده یافته‌های این پژوهش‌ها بر اثربخشی برچسب‌ها و فرابرجسب‌های اچ.تی.ام.آل. تأکید دارند (ترنر و برکیل^۱، ۱۹۹۸؛ کوئودو-توررو^۲، ۲۰۰۴؛ و ژانگ^۳ و دیمیتروف^۴، ۲۰۰۵ الف). بااین وجود، محدود بودن این برچسب‌ها و فرابرجسب‌ها، ناتوانی آن‌ها در توصیف اشیای

1. Turner & Brackbill
2. Quevedo-Torrero
3. Zhang
4. Dimitroff

محتوایی وب را نشان می دهد. گرایش دوم، پژوهش هایی بودند که با هدف پژوهش های قبلی یعنی بهبود قابلیت بازیابی و رتبه بندی اشیای محتوایی انجام شدند اما جامعه پژوهش آن ها را به جای برچسب ها و فرابرجسب ها اچ.تی.ام.آل، عناصر مبتنی بر طرح های فراداده ای استاندارد (مانند هسته دوبلین) تشکیل می داد. بر پایه نتایج این پژوهش ها، موتورهای کاوش وب به عناصر جدید واکنش مثبت نشان نمی دهند، و این عناصر تأثیری بر روی بازیافت پذیری و رتبه بندی اشیای محتوایی مرتبط ندارند (سوکویتن، ۲۰۰۰؛ صفری، ۲۰۰۵).

استفاده از رویکرد ترکیبی، به دلیل کاستی های رویکردهای پژوهش های دو گرایش پیشین، مبنای طراحی پژوهش های گرایش سوم بود. بررسی میزان تأثیر برچسب ها و فرابرجسب های اچ.تی.ام.آل و عناصر مبتنی بر طرح های فراداده ای استاندارد به طور همزمان، و مقایسه تأثیر هر یک به صورت مقایسه ای در این پژوهش ها انجام گرفت. برچسب ها و فرابرجسب های اچ.تی.ام.آل در مجموع عملکرد بهتری را در افزایش دسترسی پذیری اشیای محتوایی نسبت به عناصر مبتنی بر طرح های فراداده ای استاندارد نشان دادند (هنشاو و والاسکاس، ۲۰۰۱؛ ژانگ و دیمیتروف، ۲۰۰۴؛ ژانگ و دیمیتروف، ۲۰۰۵؛ محامد، ۲۰۰۶؛ شریف، ۱۳۸۶). تمرکز صرف بر روی عناصر فراداده ای هسته دوبلین ضعف اصلی پژوهش های گرایش دوم و سوم بود (حریری و همکاران، ۱۳۹۲).

افزون بر این گرایش ها، پژوهش هایی دیگری را می توان در این گروه جای داد که کارایی خرده های غنی^۵ در بستر نحوی نسخه جدید (۵) اچ.تی.ام.آل در بهبود قابلیت بازیابی و پیدا نمایی اشیای محتوایی را مورد بررسی قرار دادند. پیشینه های فراداده ای تولید شده بر اساس استانداردهای داده های خرد، قالب های خرد، و چهارچوب توصیف منبع (خصایص) به دلیل سازگاری بالا با موتورهای کاوش وب به صورت ساختارمند و با

1. Sokvitne
2. Safari
3. Henshaw & Valauskas
4. Mohamed
5. Rich Snippets

استفاده از بستر معناشناختی پیشینه‌ها نمایه‌سازی شده، و در به‌صورت ساختارمند نیز نمایش داده می‌شوند. عناصر، خصایص، و ارزش‌های عناصر این استانداردهای باز، با استانداردهای فراداده‌ای بافت میراث فرهنگی (کتابخانه، موزه، و آرشیو) متفاوت هستند اما کارکردهای مشابهی را پشتیبانی می‌کنند. با این وجود، کارآمدی آن‌ها در توصیف اشیای محتوایی وب نسبت به استانداردهای فراداده‌ای بافت میراث فرهنگی در سطح پایین‌تری قرار می‌گیرد (طاهری و همکاران، ۲۰۱۳).

گروه دیگر پژوهش‌ها بر واکنش موتورهای کاوش وب به پیشینه‌های فراداده‌ای پیاده‌سازی شده در قالب زبان نشانه‌گذاری گسترش‌پذیر (ایکس. ام. آل). متمرکز شدند. دلیل این تمرکز، تغییر بستر نحوی استانداردهای فراداده‌ای از قالب اچ. تی. ام. آل. به ایکس. ام. آل. بود. کاستی‌های بستر نحوی پیشین باعث این تغییر گردید. بنابراین میزان سازگاری بستر نحوی جدید با موتورهای کاوش وب باید مورد مطالعه قرار می‌گرفت. آیا بستر نحوی جدید باعث افزایش نمایه‌پذیری و پیدا‌نمایی پیشینه‌های فراداده‌ای توسط موتورهای کاوش خواهد شد. استفاده از رویکرد جاسازی فراداده در شیء محتوایی (طاهری، حریری، و فتاحی، ۱۳۸۸) و واکنش موتورهای کاوش به فراداده‌های محیط‌های اطلاعاتی پویا (طباطبائی امیری و همکاران، ۱۳۹۱) مبنای اجرای این پژوهش‌ها بود. یافته‌های این پژوهش‌ها، تأثیر مثبت بستر نحوی ایکس. ام. آل. بر نمایه‌پذیری و پیدا‌نمایی پیشینه‌های فراداده‌ای و به پیروی از آن افزایش دسترس‌پذیری به اشیای محتوایی را نشان می‌دهد.

مرور پیشینه‌های پژوهش بیانگر آن است که تغییر بستر نحوی از زبان نشانه‌گذاری فرا متن به زبان نشانه‌گذاری گسترش‌پذیر موجب بهبود میانکنش‌پذیری موتورهای کاوش وب و نظام‌های فراداده‌ای بوده است. به عبارت دیگر، استفاده از فناوری ایکس. ام. آل. به‌عنوان بستر پیاده‌سازی پیشینه‌های فراداده‌ای، افزون بر رفع کاستی‌ها و مشکلات پیشین اچ. تی. ام. آل.، دسترس‌پذیری اشیای محتوایی از طریق ابزارهای محبوب وب را افزایش می‌دهد. رویکردی که پژوهش‌های گروه دوم در استفاده از فراداده مورد توجه قرار داده بودند، رویکرد جاسازی (درج) فراداده در شیء محتوایی مرتبط بود. در حالی که بیشتر پایگاه‌های اطلاعاتی از رویکرد پیوند فراداده با شیء محتوایی بهره می‌برند. بنابراین جای

خالی پژوهشی که به رویکرد پیوند فراداده پردازد، و رفتار نرم افزارهای نمایه ساز موتورهای کاوش به پیشینه های مبتنی بر این رویکرد را مورد مطالعه قرار دهد، محسوس است.

روش شناسی پژوهش

پژوهش حاضر از نظر نوع، پژوهشی کاربردی است. چراکه بر حل مسئله سازمان دهی اشیای محتوایی وب و دسترس پذیری آن ها از طریق ابزارهای کاوش عمومی وب در شرایط واقعی و عملی تأکید دارد، و از رویکرد تحلیلی-مقایسه ای^۱ برای اجرا استفاده می کند. جامعه پژوهش را یک صد پیشینه فراداده ای در قالب دو گروه ۵۰ پیشینه ای در بستر زبان نشانه گذاری گسترش پذیر تشکیل می دادند. پنجاه پیشینه آن بر اساس قالب فراداده ای مارک ۲۱ و پنجاه پیشینه دیگر بر اساس طرح فراداده ای هسته دوبلین تولید شده، و بر روی وبسایت <http://www.aghaabedi.com> منتشر شدند. پیشینه های یاد شده از وبسایت <http://www.archive.org> و از مجموعه منابع کتابخانه دیجیتالی کالیفرنیا انتخاب گردیدند. روش گزینش به این صورت بود که با استفاده از امکان مرور موضوعی در مجموعه منابع کتابخانه دیجیتالی کالیفرنیا، یکی از موضوعاتی که بیش از ۵۰ شیء محتوایی دارا بود انتخاب، و با استفاده از امکان اچ. تی. پی. اس.^۲ که برای هر شیء فراهم شده، فراداده های مبتنی بر هسته دوبلین و مارک ۲۱ آن منتقل گردیدند. هر دو گروه پیشینه ها از لحاظ فنی با یکدیگر شرایط یکسانی داشتند.

```
<?xml version="1.0"?><dc
xmlns:dc="http://purl.org/dc/elements/1.1/"><dc:title>The anatomy of
knowledge: an essay in objective logic </dc:title><dc:creator> Hooper, Charles
E. </dc:creator><dc:type>text</dc:type><dc:publisher>London:
Watts,</dc:publisher><dc:date>1906.</dc:date><dc:language>eng</dc:langua
ge><dc:description>Contents: Introduction--pt.1. The meanings of reality and
truth--pt.II. The distinctive grounds of the
sciences.</dc:description><dc:subject>Classification
of
sciences.</dc:subject><dc:subject>Logic.</dc:subject><dc:subject>Knowledge,
Theory f.</dc:subject></dc>
```

نمونه ۱. نمونه ای از یک پیشینه فراداده ای مبتنی بر هسته دوبلین در بستر نحوی ایکس. ام. آل.

1. Comparative-analytical Approach
2. HTTPS


```
<?xml version="1.0" encoding="UTF-8"?><record
xmlns="http://www.loc.gov/MARC21/slim">
<leader>00826nam^a22002291a^4500</leader><controlfield tag="001">
GLAD235110-B</controlfield><controlfield
tag="005">20060401233913.0</controlfield><controlfield
tag="008">850425s1906 enka 000 0 eng d</controlfield><datafield tag="035"
ind1=" " ind2=" " ><subfield
code="a">ocm12210623</subfield></datafield><datafield tag="040" ind1=" "
ind2=" " ><subfield
code="a">AAM</subfield><subfield
code="c">AAM</subfield>
<subfield
code="d">CUY</subfield></datafield><datafield tag="050" ind1=" " ind2=" "
"><subfield code="a">BD241</subfield><subfield code="b">.H7</subfield>
</datafield><datafield tag="100" ind1="1" ind2=" " ><subfield
code="a">Hooper, Charles E.</subfield></datafield><datafield tag="245"
ind1="1" ind2="4"> <subfield code="a">The anatomy of
knowledge:</subfield><subfield code="b">an essay in objective logic
/</subfield><subfield code="c">by Charles E.
Hooper.</subfield></datafield><datafield tag="260" ind1=" " ind2=" "
"><subfield
code="a">London:</subfield><subfield
code="b">Watts,</subfield><subfield
code="c">1906.</subfield></datafield><datafield tag="300" ind1=" " ind2=" "
"> <subfield code="a">226 p.</subfield><subfield code="b">ill.</subfield>
<subfield code="c">21 cm.</subfield></datafield><datafield tag="500" ind1=" "
ind2=" " ><subfield code="a">t.p.: issued for the Rationalist Press
Association, Limited.</subfield></datafield><datafield tag="505" ind1="0"
ind2=" " > <subfield code="a">Contents: Introduction--pt.1. The meanings of
reality and truth--pt.II. The distinctive grounds of the
sciences.</subfield></datafield> <datafield tag="650" ind1=" " ind2="0">
<subfield code="a">Classification of sciences.</subfield></datafield><datafield
tag="650" ind1=" " ind2="0"> <subfield
code="a">Logic.</subfield></datafield><datafield tag="650" ind1=" "
ind2="0"><subfield code="a">Knowledge, Theory of.</subfield></datafield>
<datafield tag="852" ind1=" " ind2=" " ><subfield
code="a">GLAD</subfield> <subfield code="b">NRLF</subfield><subfield
code="h">Q177</subfield> <subfield code="i">.H56</subfield><subfield
code="p">$B 279 041</subfield> </datafield><datafield tag="901" ind1=" "
ind2=" " ><subfield
code="a">GLAD</subfield><subfield
code="b">235110</subfield></datafield></record>
```

نمونه ۲. نمونه‌ای از یک پیشینه فراداده‌ای مبتنی بر مارک ۲۱ در بستر نحوی ایکس. ام.آل.

موتورهای کاوش گوگل و یاهو نیز به‌عنوان بخشی از جامعه پژوهش انتخاب شدند. ارزیابی و گزینش موتورهای کاوش از طریق بررسی سایت‌هایی مانند سرچ انجین واچ^۱ و

سرچ انجین گاید^۱ انجام شد. سپس وبسایتی که پیشینه‌ها بر روی آن منتشر شده، به صورت مستقیم و با استفاده از امکان XML sitemap، از طریق ابزار Webmaster tools به موتورهای کاوش گوگل و یاهو معرفی شدند. برای گردآوری داده‌ها از روش مشاهده طراحی شده استفاده شد. نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای دوبلین و مارک ۲۱ توسط موتورهای کاوش وب مقوله‌های مورد مشاهده بودند. ابزار گردآوری داده‌ها سیاهه واری طراحی شده توسط پژوهشگر بود که متناسب با پرسش‌های پژوهش طراحی شد. برای تعیین ارزش‌های سیاهه واری از راهبرد جستجوی زیر استفاده شد:

“keyword” or “keyphrase” site:aghaabedi.com

تجزیه و تحلیل داده‌های گردآوری شده بر اساس آمار توصیفی و استنباطی صورت گرفت. برای تجزیه و تحلیل داده‌های مربوط به پرسش‌های پژوهش، از جدول (آمار توصیفی) و سنجش فرضیه پژوهش از آزمون یو مان-ویتنی در قالب نرم‌افزار اس.پی.اس.اس. نسخه ۲۳ استفاده شد.

یافته‌های پژوهش

پژوهش حاضر برای نیل به اهداف خود، سه پرسش و یک فرضیه طراحی نمود. برای پاسخ به پرسش‌های اول و دوم، جداول ۱ و ۲ طراحی شدند. پاسخ پرسش سوم از برآیند پاسخ پرسش‌های اول و دوم به دست می‌آید. نتیجه آزمون یو مان-ویتنی برای سنجش تنها فرضیه پژوهش استفاده شد. داده‌های گردآوری شده برای پاسخ به پرسش‌های پژوهش تصادفی بودند. وبسایت مورد مطالعه به مدت یک سال در محیط وب قابل دسترس بود، و فرایند گردآوری داده‌ها برای اطمینان سه ماه به طول انجامید.

پرسش اول پژوهش: وضعیت نمایه‌سازی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر نحوی ایکس.ام.آل. مبتنی بر رویکرد پیوند توسط موتورهای کاوش وب (گوگل و یاهو) چگونه است؟

بررسی تأثیر بستر نحوی بر نمایه‌سازی و رتبه‌بندی پیشینه‌های ...

جدول ۱. وضعیت نمایه‌سازی ارزش‌های عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر ایکس.

ام.آل. توسط موتورهای کاوش گوگل و یاهو

طرح فراداده‌ای (در گوگل و یاهو)	وب‌سایت	تعداد پیشینه‌ها	درصد نمایه‌سازی
مارک	aghaabedi.com	۵۰	٪۱۰۰
هسته دوبلین	aghaabedi.com	۵۰	٪۱۰۰

همان‌طور که در جدول شماره ۱ مشاهده می‌شود، ارزش‌های عناصر تمامی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ مورد مطالعه که در قالب زبان نشانه‌گذاری گسترش پذیر پیاده‌سازی شده، و بر مبنای رویکرد پیوند فراداده با شیء محتوایی طراحی گردیده بودند، توسط هر دو موتور کاوش گوگل و یاهو نمایه‌سازی شدند، و پیشینه‌های یادشده بر اساس محتوای هر یک از عناصر در محیط موتورهای کاوش قابل بازیابی هستند.

پرسش دوم پژوهش: وضعیت رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در قالب ایکس. ام.آل. مبتنی بر رویکرد پیوند توسط موتورهای کاوش وب (گوگل و یاهو) چگونه است؟

جدول ۲. وضعیت رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر ایکس. ام.آل.

نسبت به یکدیگر در موتورهای کاوش گوگل و یاهو

طرح فراداده‌ای (در گوگل و یاهو)	وب‌سایت	تعداد پیشینه‌ها	تعداد پیشینه‌هایی که بالا قرار گرفته‌اند
مارک	aghaabedi.com	۵۰	۲۵
هسته دوبلین	aghaabedi.com	۵۰	۲۵

داده‌های جدول شماره ۲ بیانگر آن است که تعداد پیشینه‌های هسته دوبلین که در نتایج جستجو پیش از پیشینه‌های مارک ۲۱ قرار گرفته‌اند، یعنی رتبه بالاتری داشته‌اند ۲۵ پیشینه، و تعداد پیشینه‌های مارک ۲۱ که وضعیتی مشابه، ۲۵ پیشینه بوده‌اند. به عبارت دیگر، رتبه پیشینه‌های فراداده‌ای مورد مطالعه در فهرست نتایج موتورهای کاوش گوگل و یاهو یکسان بوده، و موتورهای کاوش گوگل و یاهو اولویتی برای این پیشینه‌ها قائل نشده‌اند، و بدون توجه به ویژگی‌های خاص هر یک، آن‌ها را در نتایج جستجو پدیدار نموده‌اند.

پرسش سوم پژوهش: کدام یک از طرح‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر نحوی ایکس. ام. آل مبتنی بر رویکرد پیوند برای تولید پیشینه‌های فراداده‌ای دسترس پذیر در محیط موتورهای کاوش وب (گوگل و یاهو) مناسب تر است؟

چنان که قبلاً گفته شد، پاسخ پرسش سوم پژوهش، برآیند پاسخ پرسش‌های اول و دوم است. جداول شماره ۱ و ۲ نشان داد، تمامی عناصر پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر ایکس. ام. آل. که بر پایه رویکرد پیوند طراحی شده بودند، توسط نرم‌افزارهای نمایه‌ساز موتورهای کاوش وب نمایه شده، و تعداد پیشینه‌های هر استاندارد که رتبه بالاتری نسبت به پیشینه‌های استاندارد دیگر به دست آورده بود، یکسان بودند. بدین معنا که هیچ یک از دو استاندارد رتبه بهتری در نتایج به دست نیآورده بود. بنابراین تحلیل داده‌ها و عدم تفاوت میان نتایج به دست آمده حاکی از آن است که استانداردهای هسته دوبلین و مارک ۲۱ از لحاظ دسترس پذیری در محیط موتورهای کاوش وب بر یکدیگر برتری ندارند، و هر دو استاندارد از جنبه دسترس پذیری مناسب هستند.

آزمون فرضیه پژوهش: این پژوهش افزون بر پرسش‌ها، یک فرضیه داشت. برای آزمون فرضیه از آمار استنباطی، و به منظور تعیین تفاوت میان وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش وب، همانند دیگر پژوهش‌های مشابه، از آزمون یو-مان-ویتنی استفاده شد. دلیل به کارگیری این آزمون، ویژگی رتبه‌ای بودن داده‌های مربوط به نتایج موتورهای کاوش است.

فرضیه پژوهش: میان وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ در بستر نحوی ایکس. ام. آل مبتنی بر رویکرد پیوند توسط موتورهای کاوش وب تفاوت معناداری وجود دارد.

در آزمون اجرا شده، $\alpha = 5\%$ است، و بر مبنای ρ به دست آمده که برابر با یک است ($\rho=1$)، آزمون تفاوت معنی داری را میان وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش وب نشان نمی‌دهد.

Test Statistics

	F
Mann-Whitney U	1250.000
Wilcoxon W	2525.000
Z	.000
Asymp. Sig. (2-tailed)	1.000

a. Grouping Variable: Metadata

b. SE= Google & Yahoo

در هر دو موتور کاوش گوگل و یاهو، ۲۵ پیشینه فراداده‌ای هسته دابلین و ۲۵ پیشینه مارک ۲۱ در بستر نحوی ایکس.ام.آل. که بر اساس رویکرد پیوند طراحی شده بودند، نسبت به یکدیگر رتبه بهتری کسب نموده بودند. با توجه به تعداد یکسان پیشینه‌های با رتبه بالاتر، تفاوت معناداری میان وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دابلین و مارک ۲۱ توسط موتورهای کاوش وب وجود ندارد.

بحث و نتیجه‌گیری

همان‌طور که داده‌های گردآوری شده در فرایند پژوهش نشان داد، تغییر بستر نحوی پیشینه‌های فراداده‌ای و پیاده‌سازی آن‌ها در قالب زبان نشانه‌گذاری گسترش‌پذیر (ایکس.ام.آل.) موجب افزایش دسترس‌پذیری به اشیای محتوایی در محیط محبوب‌ترین ابزارهای کاوش وب می‌گردد. نمایه‌پذیر شدن ارزش‌های تمامی عناصر فراداده‌ای توسط نرم‌افزارهای نمایه‌ساز موتورهای کاوش، بیانگر بهبود میانگین پذیرش پذیرش نظام‌های فراداده‌ای و ابزارهای کاوش عمومی است. به عبارت دیگر، موتورهای کاوش وب نقش یک دروازه اطلاعاتی را برای کتابخانه‌ها و مراکز اطلاعاتی ایفا خواهند نمود، و کاربران را به سوی این مجموعه‌های ارزشمند هدایت می‌کنند. نمایه‌شدن ارزش‌های عناصر پیشینه‌های فراداده‌ای توسط پرکاربردترین موتورهای کاوش وب، نویددهنده سازگاری بالای زبان نشانه‌گذاری گسترش‌پذیر با این موتورهاست. نکته مهم دیگر آن‌که رویات‌های موتورهای کاوش تفاوتی میان پیشینه‌های فراداده‌ای مبتنی بر ایکس.ام.آل. با ساختار مسطح یا درختی ساده (مانند هسته دابلین) و با ساختار درختی خانوادگی (مانند قالب مارک) قائل نمی‌شوند، و هر دو نوع را به یک‌شکل نمایه‌سازی می‌کنند. همچنین خط‌مشی و شیوه نمایه‌سازی

موتورهای کاوش نسبت به طرح‌های فراداده‌ای با نام‌ها برچسب مبتنی بر زبان طبیعی و غیر زبان طبیعی یکسان است.

از سوی دیگر، یافته‌های پژوهش حاکی از آن است که تفاوت ساختاری پیشینه‌های فراداده‌ای و همخوانی و عدم همخوانی نام‌های برچسب آن‌ها با زبان طبیعی تأثیری بر روی رتبه پیشینه‌ها در فهرست نتایج جستجوی موتورهای کاوش ندارد، و موتورهای کاوش به پیشینه‌های هیچ‌یک از استانداردهای فراداده‌ای اهمیت و جایگاه بهتری اختصاص ندادند. چنین به نظر می‌رسد که موتورهای کاوش برای اشیای مبتنی بر ایکس.ام.آل. به دلیل گسترش پذیر بودن این قالب و امکان تعریف هر نام برچسب دلخواه توسط توسعه‌دهنده آن، الگوریتم‌های متمایزکننده‌ای طراحی نکرده‌اند. بر پایه داده‌های گردآوری شده، ۲۵ بار پیشینه‌های فراداده‌ای هسته دابلین و ۲۵ پیشینه‌های فراداده‌ای مارک ۲۱ در نتایج بالاتر قرار گرفتند.

در مقایسه با یافته‌های پژوهش‌های پیشینی که به کارآمدی عناصر و پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری فرا متن در بازیابی و دسترس‌پذیری اشیای محتوایی پرداخته بودند، پژوهش حاضر نشان داد، پیاده‌سازی پیشینه‌های فراداده‌ای قالب ایکس.ام.آل. بهبود نمایه‌پذیری پیشینه‌های یادشده و به پیروی از آن، افزایش دسترسی و بازیافت پذیری اشیای محتوایی را در بر خواهد داشت. نقطه مثبت یافته‌های پژوهش در ارتباط با یافته‌های پژوهش‌هایی که برچسب‌ها و فراب‌چسب‌های اچ.تی.ام.آل. را مورد مطالعه قرار داده بودند (ترنر و برکیل، ۱۹۹۸؛ کوئودو-توررو، ۲۰۰۴؛ و ژانگ و دیمیتروف، ۲۰۰۵ الف)، تنوع، غنا، ساختارمندی، و بستر معناساختی قوی‌تر پیشینه‌های مبتنی بر ایکس.ام.آل. بود. مزیت دیگر یافته‌ها نسبت به پژوهش‌های متمرکز بر روی پیشینه‌های فراداده‌ای طراحی شده بر اساس استانداردهای فراداده‌ای و در قالب زبان نشانه‌گذاری فرا متن این بود که برخلاف واکنش موتورهای کاوش به عناصر این گروه، واکنش در مورد پیشینه‌های فراداده‌ای عضو جامعه پژوهش حاضر مثبت بود (سوکویتن، ۲۰۰۰؛ صفری، ۲۰۰۵).

و در خصوص پژوهش‌های مربوط به رویکرد ترکیبی، یعنی مقایسه نمایه‌پذیری برچسب‌ها و فراب‌چسب‌های اچ.تی.ام.آل. و پیشینه‌های مبتنی بر استانداردهای فراداده‌ای

که اثربخشی بیشتر برچسب‌ها و فرابرچسب‌ها را نشان می‌دادند (هنشاو و والاسکاس، ۲۰۰۱؛ ژانگ و دیمیتروف، ۲۰۰۴؛ ژانگ و دیمیتروف، ۲۰۰۵؛ محمد، ۲۰۰۶؛ شریف، ۱۳۸۶)، نتایج این پژوهش، بر قابلیت‌های بیشتر پیشینه‌های فراداده‌ای در قالب ایکس. ام. آل. بر برچسب‌ها و فرابرچسب‌های اچ. تی. ام. آل. تأکید می‌نمود.

اما نتایج پژوهش با نتایج پژوهش‌های طاهری، حریری، و فتاحی (۱۳۸۸) و طباطبائی امیری و همکاران (۱۳۹۱) مشابه بود. بدین معنی که همان نتایج پژوهش طاهری، حریری، و فتاحی (۱۳۸۸) که بررسی پیشینه‌های فراداده‌ای با رویکرد به کارگیری جاسازی (درج) را انجام دادند، و با نتایج پژوهش طباطبائی امیری و همکاران (۱۳۹۱) که پیشینه‌های موجود در محیط‌های اطلاعاتی پویا را بررسی کرده بود، زبان نشانه‌گذاری گسترش‌پذیر بهبود وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌ها را در پی داشت. همچنین در هر سه پژوهش، روبات‌های موتورهای کاوش هنگام نمایه‌سازی عناصر پیشینه‌ها رویکرد حذف برچسب را برگزیدند.

بنابراین تغییر بستر نحوی پیشینه‌های فراداده‌ای به ایکس. ام. آل. از سوی طراحان استانداردهای فراداده‌ای اقدامی مناسب است، و به تقویت میانکنش‌پذیری موتورهای کاوش و نظام‌های فراداده‌ای منجر خواهد شد. بدین صورت کاربران هنگام جستجو در محیط موتورهای کاوش افزون بر دسترسی به منابع گوناگون وب، به محتوای کتابخانه‌ها و مراکز اطلاعاتی که از غنا و کیفیت بیشتر برخوردار است، دست یابند. با وجود دستاوردهای مثبت پژوهش حاضر، هنوز امکان افزایش میانکنش‌پذیری موتورهای کاوش و نظام‌های فراداده‌ای وجود دارد تا بر مزایای دسترسی به محتوای کتابخانه‌ها و مراکز اطلاعاتی از طریق موتورهای مذکور بیافزاید. در همین راستا، از آنجا که نرم‌افزارهای خزنده-نمایه‌ساز موتورهای کاوش هنگام نمایه‌سازی عناصر فراداده‌ای از رویکرد حذف برچسب استفاده می‌کنند، گمان می‌رود اجرای پژوهشی بر روی بررسی امکان نمایه‌سازی و پیدا‌نمایی نام‌های برچسب عناصر فراداده‌ای مفید باشد.

منابع

- حریری، نجلا؛ طاهری، سید مهدی؛ فتاحی، سید رحمت‌الله؛ نوشین فرد، فاطمه (۱۳۹۲). میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب: چالش‌ها و رویکردهای جاری. *پژوهشنامه کتابداری و اطلاع‌رسانی*، ۳ (۲)، ۲۳۳-۲۵۲.
- شریف، عاطفه (۱۳۸۶). بررسی میزان اثربخشی عناصر ابر داده‌ای بر رتبه‌بندی صفحات وب توسط موتورهای کاوش. *فصلنامه کتابداری و اطلاع‌رسانی*، ۱۰ (۲)، ۲۴۱-۲۵۸.
- طاهری، مهدی؛ حریری، نجلا؛ فتاحی، رحمت‌الله (۱۳۸۸). بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی. *فصلنامه کتابداری و اطلاع‌رسانی*، ۱۲ (۴)، ۱۴۱-۱۶۲.
- طباطبایی امیری، فائزه السادات؛ طاهری، سید مهدی؛ فرج پهلوی، حسین؛ عصاره، فریده؛ معرف زاده، عبدالحمید (۱۳۹۱). موتورهای کاوش وب، و نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای در محیط‌های پویا. *فصلنامه مدیریت و پردازش اطلاعات*، ۲۷، شماره ۴، ۹۰۷-۹۲۰.
- فتاحی، رحمت‌الله؛ طاهری، سید مهدی (۱۳۸۸). *فهرست‌نویسی رایانه‌ای: مفاهیم، شیوه‌ها، و کاربرد نرم‌افزارهای رایانه‌ای در سازمان‌دهی اطلاعات*. تهران: کتابدار.
- Berners-Lee, T (2006). *Linked Data*. Retrieved 05 Feb. 2015 from <http://www.w3.org/DesignIssues/LinkedData.html>.
- Bifet, A., Castillo, C., Chirita, P. A., & Weber, I. (2005, April). An Analysis of Factors Used in Search Engine Ranking. In *AIRWeb* (pp. 48-57). Retrieved 14 June 2015 from <http://airweb.cse.lehigh.edu/2005/bifet.pdf>
- Henshaw, R., & Valauskas, E. J. (2001). Metadata as a catalyst: experiments with metadata and search engines in the Internet journal, First monday. *Libri*, 51(2), 86-101.
- Mohamed, K. A. (2006). The impact of metadata in web resources discovering. *Online Information Review*, 30(2), 155-167.
- Quevedo-Torrero, J. U. (2004). *Improving web retrieval by mining the HTML tags for keywords and exploring the hyperlink structures of web pages*. Unpublished doctoral Dissertation, University of Houston.

- Retrieved March 25, 2009 from
<http://dl.acm.org/citation.cfm?id=1087297>
- Safari, M. (2005). Search Engines and Resource Discovery on the Web: Is Dublin Core an Impact Factor?. *Webology*, 2(2), 80-94.
- Sokvitne, L. (2000). An evaluation of the effectiveness of current Dublin Core metadata for retrieval. In *VALA Conference*. Retrieved March 29, 2009 from www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF.
- Taheri, S.M., Hariri, N., & Fattahi, S.R. (2014). Using data island method for creating metadata records with indexability and visibility of tag names in web search engines. *Library Hi Tech*, 32(1), 83-97.
- Turner, T. P., & Brackbill, L. (2011). Rising to the top: evaluating the use of the HTML meta tag to improve retrieval of World Wide Web documents through Internet search engines. *Library resources & technical services*, 42(4), 258-271.
- Zhang, J., & Dimitroff, A. (2004). Internet search engines' response to metadata Dublin Core implementation. *Journal of Information Science*, 30(4), 310-320.
- Zhang, J., & Dimitroff, A. (2005a). The impact of webpage content characteristics on webpage visibility in search engine results (Part I). *Information Processing & Management*, 41(3), 665-690
- Zhang, J., & Dimitroff, A. (2005b). The impact of metadata implementation on webpage visibility in search engine results (Part II). *Information processing & management*, 41(3), 691-715.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی