

موتورهای کاوش وب و نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای دسترس‌پذیر در محیط‌های اطلاعاتی پویا

فائزه‌السادات طباطبایی امیری* | کارشناس ارشد علوم کتابداری و اطلاع‌رسانی
سیدمهدی طاهری^۱ | عضو هیئت علمی،

پژوهشگاه علوم و فرهنگ اسلامی

استاد،

عبدالحسین فرج‌پهلوی^۲

گروه کتابداری و اطلاع‌رسانی دانشگاه شهید چمران اهواز

استاد،

فریده عصاره^۳

گروه کتابداری و اطلاع‌رسانی دانشگاه شهید چمران اهواز

دانشیار،

عبدالحمید معرفزاده^۴

گروه کتابداری و اطلاع‌رسانی دانشگاه شهید چمران اهواز

دریافت: ۱۳۸۹/۱۲/۱۹ | پذیرش: ۱۳۹۰/۰۷/۰۳

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۸۲۲۳-۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS، LISA و ISC
http://jipm.irandoc.ac.ir
دوره ۲۷ | شماره ۴ | صص ۹۰۷-۹۲۰
تابستان ۱۳۹۱

نوع مقاله: پژوهشی

*faeze.tabatabai@gmail.com
1. taherismster@gmail.com
2. farajpahlou@gmail.com
3. osareh.f@gmail.com
4. amoorefzadeh@hotmail.com
5. content objects

چکیده: پژوهش حاضر با هدف تبیین نمایه‌سازی و رتبه‌بندی اشیای محتوایی^۵ حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا توسط موتورهای کاوش گوگل و یاهو و مقایسه آنها با یکدیگر با رویکرد تحلیلی-مقایسه‌ای انجام شد. جامعه پژوهش را ۱۰۰ شیء محتوایی (کتاب الکترونیکی) مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در قالب دو گروه تشکیل می‌داد. گروه اول، حاوی عناصر فراداده‌ای هسته دوبلین و گروه دوم، حاوی عناصر فراداده‌ای مارک ۲۱ که از اواخر مرداد ۱۳۸۸ تا خرداد ۱۳۸۹ روی وبسایت <http://www.marcdemi.ir> انتشار یافته بودند. ابزار گردآوری داده‌ها سیاه و ارسی بود. وبسایت مورد مطالعه براساس پیشرفته‌ترین روش‌ها به موتورهای کاوش گوگل و یاهو معرفی شد. یافته‌های پژوهش نشان می‌دهند که موتور کاوش گوگل توانست عناصر فراداده‌ای مربوط به دو طرح فراداده‌ای هسته دوبلین و مارک ۲۱ درج‌شده در اشیای محتوایی مورد مطالعه را به‌طور کامل نمایه‌سازی کند. ولی، موتور کاوش یاهو تا پایان مهلت انتشار (حدود ۱۰ ماه) واکنش مثبتی نشان نداد و فاقد توانایی در نمایه‌سازی و به‌تبع آن رتبه‌بندی اشیای محتوایی مورد مطالعه بود. از دیدگاه مقایسه‌ای نیز تفاوتی میان نمایه‌سازی و رتبه‌بندی عناصر فراداده‌ای مربوط به دو استاندارد مورد مطالعه توسط موتور کاوش گوگل مشاهده نگردید. در مجموع، یافته‌ها حاکی از آن بود که هیچ‌یک از دو طرح فراداده‌ای هسته دوبلین و مارک ۲۱ مبتنی بر XML منتشرشده در محیط‌های اطلاعاتی پیوسته پویا از لحاظ دسترس‌پذیری توسط موتورهای کاوش نسبت به یکدیگر برتری ندارند.

کلیدواژه‌ها: طرح فراداده‌ای هسته دوبلین، قالب فراداده‌ای مارک ۲۱، نمایه‌سازی عناصر فراداده‌ای، محیط‌های اطلاعاتی پیوسته پویا، موتورهای کاوش وب، رتبه‌بندی نتایج کاوش، زبان نشانه‌گذاری گسترش‌پذیر (XML)

۱. سخن آغازین

ویژگی‌های منحصر به فرد شبکه جهانی وب و دسترس‌پذیری بالای اطلاعاتی که بر روی آن منتشر شده، موجب رشد چشمگیر حجم اطلاعات بر روی آن گردیده است. همانند دیگر محیط‌های اطلاعاتی، سازماندهی اطلاعات محیط وب نیز با هدف تسهیل و تسریع دسترسی به آن ضروری است. توسعه فنون و نظام‌های سازماندهی اطلاعات و دانش همواره متناسب با رشد منابع اطلاعاتی و رسانه‌های دانش بوده است. این توسعه در قالب دو رویکرد تبلور یافته است: الف) سازگاری نظام‌ها و ابزارهای سنتی با محیط‌ها و رسانه‌های جدید (از جمله اصطلاحنامه‌ها و قالب فراداده‌ای مارک ۲۱)، و ب) طراحی و تولید نظام‌ها و ابزارهای جدید (مانند گسترش طرح فراداده‌ای هسته دوبلین) (طاهری ۱۳۸۷؛ Qin 2000). بهبود دسترسی و بهینه‌سازی فرایند مدیریت اطلاعات و منابع دانش، افزایش توجه به گسترش محیط‌های اطلاعاتی پیوسته پویا^۱ را موجب گردیده است. روزآمدی و حجم اشیای محتوایی^۲ مربوط به حوزه‌های مختلف دانش بشری از مهمترین ویژگی‌های محیط‌های پویا نسبت به محیط‌های ایستا^۳ است. از دهه ۱۹۹۰ به این سو، متخصصان حوزه‌های مدیریت اطلاعات، مدیریت فناوری اطلاعات، علوم رایانه، و برخی از حوزه‌های مرتبط دیگر، تلاش‌های فراوانی را برای اعمال کنترل و سازماندهی اشیای محتوایی موجود در شبکه جهانی وب (پیشوا و مجیدی ۱۳۸۶)، به‌ویژه اشیای محتوایی ذخیره‌شده در محیط‌های پویا انجام داده‌اند. از سوی دیگر، پژوهش‌ها نشان می‌دهند کاربران شبکه وب، بیش از ۸۸ درصد از زمانی را که صرف جستجو و بازیابی اشیای محتوایی مورد نیاز خود می‌کنند، به موتورهای کاوش اختصاص داده‌اند (Bifet and Castillo 2005). این محبوبیت موتورهای کاوش باعث گردیده است، افزایش پیمایش‌پذیری^۴ و نمایه‌پذیری^۵ پیشینه‌های فراداده‌ای برای موتورهای کاوش یکی از اهداف گسترش‌دهندگان طرح‌های فراداده‌ای باشد

1. dynamic online information environment

محیط‌هایی که دارای شکلی ثابت و محتوایی متغیر هستند. صفحات خروجی‌ای که از این محیط‌ها در برابر درخواست کاربران ایجاد می‌شود و محتوای آنها از پایگاه‌های داده‌ای است و پارامترهایی خاص (؟، !، = ...) در نشانی اینترنتی آنها وجود دارد (پاول ۲۰۰۲).

۲. موجودیتی که داده‌ها یا اطلاعات را دربرمی‌گیرد، یک شیء محتوایی خود می‌تواند از اشیای محتوایی دیگری ساخته شود. برای مثال یک مجله، یک شیء محتوایی است که خود از مجموعه‌ای از مقالات که هر یک از آنها نیز شیء محتوایی هستند، ساخته شده است. متن، تصاویر، و اشکالی که در یک مقاله هستند نیز می‌توانند شیء محتوایی باشند، حتی فراداده‌های مربوط به یک شیء محتوایی نیز خود، اشیای محتوایی هستند (National Information Standards Organization (NKOS) 2005).

3. static environment

4. crawlability

5. indexability

(طاهری و دیگران ۱۳۸۸). فراداده‌ها با دو رویکرد با اشیای محتوایی مرتبط می‌شوند: نخست، پیوند^۱ با اشیای محتوایی و دیگر، درج یا جاسازی^۲ درون آنها (فتاحی و طاهری ۱۳۸۸) که در این میان، رویکرد دوم از کارایی بالاتری برخوردار است (Thornely 2000).

همان‌طور که پیشتر اشاره شد، حجم گسترده اشیای محتوایی روزآمد وب در محیط‌های پویا ذخیره شده‌اند. این مسأله طراحی راهکارهایی برای بازیابی اشیای محتوایی دسترس‌پذیر در محیط‌های پویا را از سوی ابزارهای کاوش وب ضروری می‌سازد. این راهکارها زمانی مؤثر خواهند بود که زمینه لازم برای پیمایش و نمایه‌سازی محتوای محیط‌های پیوسته پویا توسط نرم‌افزارهای نمایه‌ساز ابزارهای کاوش وب، از سوی فراهم‌کنندگان محیط‌های پیوسته پویا فراهم شود. آنچه بیشترین اهمیت را در این زمینه‌سازی دارد، آماده‌سازی فراداده‌ها - به‌عنوان هسته کشف و بازیابی داده‌ها - است. محبوبیت موتورهای کاوش از سوی کاربران و اهمیت طرح‌های فراداده‌ای در سازماندهی اشیای محتوایی وب، گسترش تلاش‌هایی را از سوی توسعه‌دهندگان موتورهای کاوش وب و طرح‌های فراداده‌ای با هدف افزایش میانکنش‌پذیری^۳ میان آن دو در پی داشته است. به‌عنوان مثال، انتخاب بستر نحوی XML به‌عنوان مبنای پیاده‌سازی و ذخیره‌سازی از سوی نظام‌های فراداده‌ای و نمایه‌سازی اشیای (مدارک) مبتنی بر XML و امکان نمایه‌سازی محتوای محیط‌های پیوسته پویا توسط روبات‌های موتورهای کاوش از مصادیق این تلاش‌هاست (Taheri and Hariri, Forthcoming). بدیهی است افزایش و بهبود میانکنش‌پذیری میان موتورهای کاوش وب و نظام‌های فراداده‌ای، باعث افزایش دسترس‌پذیری اشیای محتوایی منتشرشده بر روی وب می‌گردد.

براساس اقدامات صورت گرفته از سوی توسعه‌دهندگان موتورهای کاوش و طراحان نظام‌های فراداده‌ای که در بالا اشاره شد، این پرسش‌ها مطرح می‌شود که آیا اشیای محتوایی حاوی عناصر فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) توسط موتورهای کاوش وب قابل نمایه شدن هستند؟ با توجه به ساختار متفاوت پیشینه‌های (رکوردهای) فراداده‌ای مبتنی بر طرح‌های فراداده‌ای گوناگون (با ساختار مسطح یا ساختار سلسله مراتبی) و با استفاده از برجسب‌های عناصر زبان-پایه و غیر زبان-پایه، کدام یک رتبه بالاتری را در نتایج بازیابی نسبت به یکدیگر دارند؟ پژوهش حاضر در راستای پاسخگویی به این پرسش‌ها طرح‌ریزی شده است. بدین منظور، مهمترین و پرکاربردترین طرح‌های فراداده‌ای، قالب

1. Attaching

2. Embedding

۳. منظور از میانکنش‌پذیری، تمهیدات و اقداماتی است که نظام‌های فراداده‌ای و موتورهای کاوش برای قابلیت کار متقابل با یکدیگر انجام داده‌اند.

فرا داده‌ای مارک ۲۱ بر مبنای رویکرد اول (نظام‌های سنتی سازگار شده با محیط جدید) با ساختاری سلسله مراتبی و عناصری غیر زبان-پایه و طرح فراداده‌ای هسته دابلین بر مبنای رویکرد دوم (نظام‌های جدید گسترش یافته متناسب با محیط جدید) با ساختاری مسطح و عناصری زبان-پایه انتخاب شدند. همچنین، پیشینه‌های فراداده‌ای مبتنی بر آن دو طرح پس از انتشار بر روی وب‌سایتی مستقل در معرض موتورهای کاوش گوگل و یاهو قرار گرفتند.

اهمیت پژوهش حاضر نیز از دو بعد قابل بررسی است: نخست اینکه جزء معدود پژوهش‌هایی است که به بررسی نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) می‌پردازد. دیگر اینکه پیشینه‌های فراداده‌ای ذخیره شده در محیط‌های پویا را مد نظر قرار می‌دهد. از بعد دوم، اولین پژوهش محسوب می‌شود و بیانگر ضرورت پژوهش نیز است، چرا که بیشترین حجم اطلاعات و روزآمدترین آن از محیط‌های پیوسته پویا قابل دسترس هستند.

۲. هدف و پرسش‌های پژوهش

هدف اصلی این پژوهش تبیین نمایه‌سازی و رتبه‌بندی اشیای محتوایی^۱ حاوی منتخبی از مجموعه عناصر فراداده‌ای هسته دابلین^۲ و مارک ۲۱ مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا^۳ توسط موتورهای کاوش عمومی و مقایسه آنها با یکدیگر است. به منظور نیل به این هدف، هفت پرسش طراحی شده است:

- نمایه‌سازی اشیای محتوایی حاوی بخشی از^۴ عناصر فراداده‌ای هسته دابلین^۵ مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا توسط موتورهای کاوش عمومی (گوگل و یاهو) چگونه است؟
- نمایه‌سازی اشیای محتوایی حاوی بخشی از عناصر فراداده‌ای مارک ۲۱^۶ مبتنی بر زبان

۱. در این پژوهش، منظور از اشیای محتوایی، ۱۰۰ کتاب الکترونیکی در قالب ایکس.ام.ال (XML) است که برخی عناصر فراداده‌ای هسته دابلین و مارک ۲۱ در آنها درج شده‌اند.

2. Dublin Core Metadata Elements Set (DCMES)

۳. در پژوهش حاضر، صفحات پویایی هستند که نشانی آنها از طریق وب سایت www.marcdcmi.ir در اختیار موتورهای کاوش جهت بازیابی اطلاعات قرار می‌گیرد.

۴. بدیهی است تمامی عناصر هسته دابلین و مارک ۲۱ برای توصیف یک شیء خاص قابل کاربرد نیستند.

۵. عناصر عنوان، پدیدآورنده، موضوع، توصیف، ناشر، همکار، تاریخ، نوع منبع، و زبان

۶. عناصر پدیدآورنده، عنوان، تکرار نام پدیدآور، محل نشر، ناشر، تاریخ نشر، موضوع، یادداشت، ویرایش، و فروست

- نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا توسط موتورهای کاوش عمومی (گوگل و یاهو) چگونه است؟
- تفاوت نمایه‌سازی اشیای محتوایی حاوی بخشی از عناصر فراداده‌ای هسته دابلین و مارک ۲۱ مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا توسط موتورهای کاوش عمومی (گوگل و یاهو) چیست؟
 - تفاوت رتبه‌بندی اشیای محتوایی حاوی بخشی از عناصر فراداده‌ای هسته دابلین و مارک ۲۱ مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا توسط موتورهای کاوش عمومی (گوگل و یاهو) چیست؟
 - واکنش موتورهای کاوش عمومی (گوگل و یاهو) نسبت به اشیای محتوایی محیط‌های اطلاعاتی پیوسته پویا و مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) که حاوی پیشینه‌های فراداده‌ای با ساختارهای مسطح (هسته دابلین) و سلسله‌مراتبی (مارک) هستند، چگونه است؟
 - واکنش موتورهای کاوش عمومی (گوگل و یاهو) نسبت به پیشینه‌های فراداده‌ای با برجسب‌های عناصر زبان-پایه (هسته دابلین) و با برجسب‌های عناصر غیر زبان-پایه (مارک ۲۱) چگونه است؟
 - کدام‌یک از طرح‌های فراداده‌ای هسته دابلین و مارک ۲۱ برای سازماندهی اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا از لحاظ دسترس‌پذیری (نمایه‌سازی و رتبه‌بندی) توسط موتورهای کاوش عمومی (گوگل و یاهو) مناسب‌تر است؟

۳. مرور تحلیلی پژوهش‌های مرتبط

چنانکه در مقدمه اشاره شد، بهبود میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش، افزایش دسترس‌پذیری اشیای محتوایی را در پی خواهد داشت. بدین منظور، اقداماتی از سوی توسعه‌دهندگان هر دو نظام صورت گرفت (Taheri and Hariri, forthcoming). تاکنون پژوهش‌هایی چند با هدف ارزیابی این تلاش‌ها انجام شده است. با بررسی تحلیلی این پژوهش‌ها، می‌توان آنها را به سه گروه کلی تقسیم نمود. نخست، پژوهش‌هایی که به بررسی اثربخشی استفاده از فرابرجسب‌ها و تقویت محتوای برجسب‌های زبان نشانه‌گذاری فرامتن

(HTML) بر رتبه‌بندی و بازیابی‌پذیری^۱ اشیای محتوایی وب از طریق موتورهای کاوش پرداخته‌اند. دیگر، پژوهش‌هایی که تأثیر استفاده از طرح‌های فراداده‌ای استاندارد (مانند هسته دوبلین) را در افزایش دسترس‌پذیری اشیای محتوایی وب مد نظر قرار داده‌اند. دو پژوهش دیگر، پژوهش‌هایی که واکنش موتورهای کاوش به پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) را - به‌عنوان بستر نحوی جدید پیاده‌سازی فراداده‌ها - بررسی نموده‌اند.

در پژوهش‌های گروه اول، بررسی میزان نمایه‌پذیری فرابرجسب‌ها و برجسب‌های زبان نشانه‌گذاری فرامتن (HTML) نشان داد، این برجسب‌ها و فرابرجسب‌ها برای نرم‌افزارهای نمایه‌سازی موتورهای کاوش شناخته شده هستند و ارزش‌های^۲ عناصر آنها به‌طور کامل توسط موتورهای کاوش نمایه‌سازی می‌شود. بنابراین، بر بازیابی‌پذیری اشیای محتوایی وب تأثیر مثبت دارند، همچنین باعث بهبود رتبه شی خواهند شد. اما، با وجود این تأثیرات مثبت، محدود بودن تعداد فرابرجسب‌ها و برجسب‌ها و نقاط دسترسی محدود، عدم امکان پیاده‌سازی طرح‌های فراداده‌ای استاندارد را فراهم می‌سازد. پژوهش‌های دیگری نیز به این مهم پرداخته‌اند. برخی از آنها، مانند ترنر و برکییل (Turner and Brackbill 1998)، کونودو- توررو (Quevedo-Torrero 2004)، و ژانگ و دیمیتروف (Zhang and Dimitroff 2005a)، فقط به فرابرجسب‌ها و برجسب‌های زبان نشانه‌گذاری فرامتن (HTML) و برخی دیگر، مانند هنشاو و والاسکاس (Henshaw and Valauskas 2001)، ژانگ و دیمیتروف (Zhang and Dimitroff 2004)، ژانگ و دیمیتروف (Zhang and Dimitroff 2005b)، موهامد (Mohamed 2006) علاوه بر فرابرجسب‌ها و برجسب‌های HTML، به عناصر مربوط به طرح‌های فراداده‌ای استاندارد (مانند هسته دوبلین) - گاهی با رویکرد مقایسه‌ای - اختصاص داشتند.

یافته‌های حاصل از گروه دوم پژوهش‌ها که شامل پژوهش‌های سوکوویتن (Sokvitne 2000)، هنشاو و والاسکاس (Henshaw and Valauskas 2001)، ژانگ و دیمیتروف (Zhang and Dimitroff 2004)، ژانگ و دیمیتروف (Zhang and Dimitroff 2005b)، صفری (Safari 2005)، و شریف (۱۳۸۶) می‌شود، حاکی از آن است که استفاده از عناصر یا پیشینه‌های تولیدشده براساس طرح‌های فراداده‌ای استاندارد مانند هسته دوبلین تأثیر چندانی بر بازیافت‌پذیری و رتبه‌بندی اشیای محتوایی وب ندارند. حتی برخی از این پژوهش‌ها استفاده از عناصر هسته دوبلین را فاقد تأثیر ارزیابی کردند. مقایسه یافته‌های پژوهش‌های گروه‌های اول و دوم بیانگر

1. retrievability

2. value

آن است که تعامل روبات‌های موتورهای کاوش با فرابرجسب‌های خود "زبان نشانه‌گذاری فرامتن (HTML)" بیش از عناصر هسته دوبلین - باوجود ارزش‌های افزوده این طرح نسبت به فرابرجسب‌های HTML - است. اکنون این پرسش مطرح می‌شود که برای حل این مشکل، طراحان و گسترش‌دهندگان نظام‌های فراداده‌ای استاندارد چه اقداماتی باید انجام می‌دادند؟

انتخاب زبان نشانه‌گذاری گسترش‌پذیر (XML) به‌عنوان بستر نحوی پیاده‌سازی پیشینه‌های فراداده‌ای راه‌حل مناسبی برای این مشکل بود. پژوهش‌های طاهری (۱۳۸۷الف) و آقاعابدی (۱۳۸۹) - به‌عنوان گروه سوم - افزایش میانگین پذیرش موتورهای کاوش با نظام‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) را تأیید نمودند. آنها به‌ترتیب براساس رویکرد درج (جاسازی) پیشینه‌های فراداده‌ای در اشیای محتوایی و رویکرد پیوند آنها با اشیای محتوایی، نمایه‌سازی، و رتبه‌بندی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ را توسط موتورهای کاوش مورد بررسی قرار دادند. یافته‌های هر دو پژوهش بر واکنش مثبت موتورهای کاوش نسبت به عناصر فراداده‌ای و نمایه‌سازی کامل ارزش‌های آن عناصر دلالت داشتند. نوآوری‌های دو پژوهش، بررسی پیشینه‌های فراداده‌ای مبتنی بر XML و قالب فراداده‌ای مارک ۲۱ برای نخستین بار بود. اشیای محتوایی و پیشینه‌های فراداده‌ای مورد مطالعه در آن دو پژوهش در محیط ایستا منتشر شده بودند.

بررسی تحلیلی پژوهش‌های انجام‌شده در حوزه تعامل فراداده و موتورهای کاوش بیانگر آن است که استفاده از زبان نشانه‌گذاری فرامتن (HTML) به‌عنوان بستر نحوی پیاده‌سازی پیشینه‌های فراداده‌ای مبتنی بر طرح‌های استاندارد، موجب کاهش نمایه‌پذیری آنها توسط موتورهای کاوش می‌شود. اگر چه ارزش‌های فرابرجسب‌ها و برجسب‌های خود زبان نشانه‌گذاری فرامتن (HTML) به‌راحتی توسط موتورهای کاوش نمایه‌سازی می‌شوند، محدودیت‌های متعدد این زبان، از ارزش آن برای سازماندهی اشیای محتوایی وب کاسته است. تغییر بستر نحوی از زبان نشانه‌گذاری فرامتن (HTML) به زبان نشانه‌گذاری گسترش‌پذیر (XML) توسط نظام‌های فراداده‌ای، موجب بهبود و افزایش بازیابی‌پذیری اشیای محتوایی شد. پژوهش‌های مرتبط با بستر نحوی جدید بر روی محیط‌های ایستا انجام شد. با توجه به افزایش حجم اطلاعات منتشرشده در محیط‌های پویا، جای خالی پژوهش‌هایی که به مطالعه واکنش موتورهای کاوش به پیشینه‌های فراداده‌ای مبتنی بر XML در محیط جدید پردازد، احساس می‌شد.

۴. روش‌شناسی

پژوهش حاضر از نظر نوع پژوهش، پژوهشی کاربردی^۱ است. از آنجایی که پژوهش‌های کاربردی با هدف رشد و بهبود وضعیت موجود و آزمون مفاهیم نظری در موقعیت‌های واقعی صورت می‌پذیرند (رشیدی ۱۳۷۷)، این پژوهش نیز که به آزمون تعامل و میانگنش‌پذیری اشیای محتوایی حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ در محیط‌های اطلاعاتی پیوسته پویا با موتورهای کاوش عمومی می‌پردازد، پژوهشی کاربردی محسوب می‌گردد. برای انجام این پژوهش، از رویکرد تحلیلی-مقایسه‌ای^۲ استفاده شده است. جامعه پژوهش را ۱۰۰ شیء محتوایی (کتاب الکترونیکی) مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) که از یک رده موضوعی خاص "Knowledge, theory of" مربوط به مجموعه منابع کتابخانه دیجیتالی کالیفرنیا^۳ ذخیره‌شده در وبسایت "آرشیو وب" (<http://www.archive.org>) استخراج شده بودند، تشکیل می‌داد. این اشیای محتوایی در قالب دو گروه تقسیم‌بندی شدند. در هر یک از اشیای محتوایی گروه اول، پیشینه‌های فراداده‌ای هسته دوبلین و در هر یک از اشیای محتوایی گروه دوم، پیشینه‌های فراداده‌ای مارک ۲۱ به‌عنوان برجسب فراداده^۴ بعد از عنصر ریشه^۵ درج (جاسازی) شدند. هر دو گروه بر روی وبسایت www.maredecmi.ir (ایجادشده توسط پژوهشگران) از اواخر مرداد ماه ۱۳۸۸ تا خرداد ۱۳۸۹ منتشر شدند و براساس پیشرفته‌ترین روش‌های معرفی‌شده از سوی موتورهای کاوش، به موتورهای کاوش گوگل (از طریق Webmaster Tools و XML Sitemap) و یاهو (از طریق Suggest a site، Yahoo! Search URL، Status Review Form، ROR & Text Sitemap) - به‌عنوان پرکاربردترین و بهترین موتورهای کاوش (Taheri and Hariri, Forthcoming; Alexa 2008; Campex. 2005) - معرفی شدند. گردآوری داده‌ها در فروردین ۱۳۸۹ صورت گرفت. روش گردآوری داده‌ها، مشاهده طراحی‌شده^۶ و ابزار آن سیاهه واریسی بود که براساس پرسش‌های پژوهش توسط پژوهشگران طراحی شده است. با بهره‌مندی از امکانات نرم‌افزار SPSS و مطابق با پرسش‌های پژوهش، جدول‌هایی تهیه و داده‌های استخراج‌شده از سیاهه واریسی در این جدول‌ها درج شدند. برای پاسخ به پرسش‌های اول تا چهارم از این راهبرد جستجو استفاده شد: "keyphrase" site:maredecmi.ir. این راهبرد در کادر جستجوی پایه^۷ موتورهای کاوش انتخابی وارد می‌شد. داده‌های به‌دست‌آمده از سیاهه واریسی در جدول‌های طراحی‌شده توسط نرم‌افزار SPSS درج گردیدند. برای نشان دادن وضعیت نمایه‌سازی در پرسش‌های اول و دوم، در صورت انجام

1. applied research
2. metadata tag
3. basic search

4. comparative-analytical approach
5. root element

6. california digital library
7. structured observation

نمایه‌سازی از علامت (+) و در صورت عدم نمایه‌سازی از علامت (-) استفاده شد. با تخصیص امتیاز ۰ و ۱ به هر یک از مؤلفه‌های سیاهه واریسی، جمع امتیازات تعلق گرفته به مؤلفه‌های به‌کارگرفته‌شده در پژوهش محاسبه گردید و به این روش، پاسخ پرسش‌های سوم و چهارم به‌دست آمد. براساس پاسخ پرسش سوم، پرسش‌های پنجم و ششم پاسخ داده شد و در مجموع، با توجه به پاسخ کل پرسش‌ها، پاسخ پرسش هفتم فراهم گردید. سپس، با استفاده از روش‌های آمار توصیفی، داده‌های گردآوری‌شده مورد تجزیه و تحلیل قرار گرفتند.

۵. یافته‌های پژوهش

جدول‌های شماره ۱ و ۲ حاوی پاسخ مربوط به پرسش اول و دوم پژوهش است. داده‌های جدول ۱ بیانگر آن است که مؤتور کاوش گوگل، محتوای تمامی ۹ عنصر پیشینه‌های مبتنی بر طرح فراداده‌ای هسته دوبلین و ۱۰ عنصر پیشینه‌های مارک ۲۱ مورد مطالعه را به‌طور کامل (۱۰۰٪) استخراج و نمایه‌سازی کرده است. بنابراین، اشیای محتوایی عضو جامعه پژوهش که در محیط اطلاعاتی پیوسته پویا پیاده‌سازی شده‌اند، از طریق ارزش‌های (مقادیر) این عناصر قابل بازیابی هستند، در حالی که مؤتور کاوش یاهو در مدت انتشار وب‌سایت حاوی اشیای محتوایی مورد مطالعه قادر به نمایه‌سازی هیچ‌کدام از پیشینه‌ها نبوده است.

جدول ۱. نمایه‌سازی عناصر پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ درج‌شده در اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا توسط مؤتور کاوش گوگل

طرح فراداده‌ای (گوگل)	وب سایت	تعداد عناصر مورد مطالعه	تعداد اشیای محتوایی	درصد نمایه‌سازی
هسته دوبلین	www.marcdemi.ir	۹	۵۰	۱۰۰٪
مارک ۲۱		۱۰	۵۰	۱۰۰٪

جدول ۲. نمایه‌سازی عناصر پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ درج‌شده در اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا توسط مؤتور کاوش یاهو

طرح فراداده‌ای (یاهو)	وب سایت	تعداد عناصر مورد مطالعه	تعداد اشیای محتوایی	درصد نمایه‌سازی
هسته دوبلین	www.marcdemi.ir	۹	۵۰	۰٪
مارک ۲۱		۱۰	۵۰	۰٪

بر اساس جدول‌های ۱ و ۲ می‌توان به پرسش‌های سوم و پنجم و ششم نیز پاسخ داد. داده‌های گردآوری شده نشان می‌دهند، موتورهای کاوش گوگل و یاهو نسبت به نمایه‌سازی پیشینه‌های فراداده‌ای درج شده در اشیای محتوایی مورد مطالعه کنش متفاوتی نشان داده‌اند. گوگل مقادیر تمامی عناصر فراداده‌ای هسته دابلین و مارک ۲۱ را نمایه‌سازی کرده است، اما یاهو توانایی چنین فرایندی را نداشته است (پاسخ پرسش سوم).

نمایه‌سازی تمامی عناصر مربوط به پیشینه‌های فراداده‌ای ایجادشده بر پایه طرح‌های فراداده‌ای هسته دابلین و مارک ۲۱ به صورت مشابه و یکسان، بیانگر عدم تفاوت برخورد نرم‌افزار نمایه‌ساز^۱ گوگل با پیشینه‌های فراداده‌ای دارای ساختار سلسله مراتبی (مارک ۲۱) و مسطح (هسته دابلین) و دارای عناصر زبان-پایه (مبتنی بر زبان طبیعی) و غیر زبان-پایه است. به عبارت دیگر، موتور کاوش گوگل الگوریتم خاصی برای نمایه‌سازی اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) با ساختار پیشینه‌ای و نوع عناصر مختلف ندارد. در مورد موتور کاوش یاهو نیز وضعیت مشخص است. این موتور هیچ پیشینه‌ای را نمایه‌سازی نکرده است، از این رو، کنش خاصی نداشته است.

جدول ۳ برای پاسخ به پرسش چهارم پژوهش طراحی شده است. یافته‌های مرتبط با این پرسش بر عدم تفاوت میان رتبه‌بندی اشیای محتوایی حاوی پیشینه‌های فراداده‌ای هسته دابلین و مارک ۲۱ دلالت دارد. از ۵۰ شیء محتوایی حاوی پیشینه‌های فراداده‌ای هسته دابلین، ۲۵ شیء نسبت به نمونه‌های مشابه خود که حاوی مارک ۲۱ هستند، بالاتر قرار گرفته‌اند و ۲۵ شیء پایین‌تر. این بدین معنی است که موتور کاوش گوگل خط‌مشی خاصی برای رتبه‌بندی اشیای محتوایی حاوی پیشینه‌های فراداده‌ای با ساختار سلسله مراتبی و مسطح و عناصر زبان-پایه و غیر زبان-پایه ندارد. در یاهو نیز به دلیل عدم نمایه‌سازی اشیای محتوایی تفاوتی وجود ندارد.

جدول ۳. نمایه‌سازی عناصر طرح‌های فراداده‌ای هسته دابلین و مارک ۲۱ درج شده در اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی پیوسته پویا توسط موتورهای کاوش گوگل و یاهو

جمع امتیاز پیشینه- هایی که بالا قرار گرفته‌اند.	موتورهای کاوش		وب سایت	طرح فراداده‌ای
	گوگل (امتیاز)	ياهو (امتیاز)		
۲۵	۰	۲۵	www.marcdcmi.ir	مارک ۲۱
۲۵	۰	۲۵		هسته دابلین

1. Spider-indexer

یافته‌های مربوط به پرسش‌های اول تا ششم تأمین‌کننده پاسخ پرسش هفتم هستند. براساس این یافته‌ها، هیچ کدام از طرح‌های فراداده‌ای هسته دویلین و مارک ۲۱ برای سازماندهی اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) منتشر شده در محیط‌های اطلاعاتی پیوسته پویا از لحاظ دسترس‌پذیری (نمایه‌سازی و رتبه‌بندی) توسط موتورهای کاوش عمومی (گوگل و یاهو) بر یکدیگر برتری ندارند.

۶. نتیجه‌گیری

پژوهش‌های طاهری (۱۳۸۷الف)، آقاعابدی (۱۳۸۹)، و یافته‌های پژوهش حاضر نشان می‌دهند، انتخاب بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر (XML) توسط نظام‌های فراداده‌ای موجب افزایش بازیابی‌پذیری آنها از طریق موتورهای کاوش می‌شود. یافته‌های هر سه پژوهش یادشده حاکی از آن است که موتور کاوش گوگل برای نمایه‌سازی اشیای محتوایی حاوی عناصر فراداده‌ای در محیط‌های ایستا و پویا و پیشینه‌های (رکوردهای) فراداده‌ای که به‌صورت مجزا منتشر شده بودند، عملکرد مثبت و یکسانی نشان داد و توانست محتوای^۱ (مقادیر یا ارزش‌های)^۲ عناصر فراداده‌ای مبتنی بر استانداردهای مارک ۲۱ و هسته دویلین را استخراج و قابل بازیابی نماید. نمایه‌سازی اشیای محتوایی حاوی عناصر فراداده‌ای توسط گوگل - به‌عنوان یافته ویژه پژوهش حاضر - نمایانگر عدم توجه روبات این موتور به پارامترهایی چون (؟)، (=)، و غیره در نشانی اینترنتی صفحات وب پویاست. همچنین، این موتور نسبت به رتبه‌بندی پیشینه‌های مبتنی بر دو طرح فراداده‌ای منتخب و نمایه‌سازی پیشینه‌های فراداده‌ای با ساختار سلسله‌مراتبی و مسطح و با برجسب‌های زبان-پایه و غیر زبان-پایه تفاوتی نشان نداد.

در مورد موتور کاوش یاهو، یافته‌های پژوهش‌های طاهری (۱۳۸۷الف) و آقاعابدی (۱۳۸۹) بر توانایی یاهو در نمایه‌سازی اشیای محتوایی حاوی عناصر فراداده‌ای مبتنی بر XML و پیشینه‌های فراداده‌ای منتشر شده بر مبنای رویکرد پیوند در محیط‌های ایستا تأکید نمودند. آنها همچنین اذعان داشتند موتور کاوش یاهو میان رتبه‌بندی پیشینه‌های فراداده‌ای مبتنی بر طرح فراداده‌ای هسته دویلین و قالب فراداده‌ای مارک ۲۱ و نیز نمایه‌سازی پیشینه‌های فراداده‌ای دارای ساختار سلسله‌مراتبی و مسطح و طراحی‌شده بر پایه برجسب‌های زبان-پایه و غیر زبان-پایه تفاوتی قائل نیست، در حالی که پژوهش حاضر عکس آن را اثبات نمود. یعنی یاهو را ناتوان در نمایه‌سازی (و به‌پیروی از آن رتبه‌بندی) اشیای محتوایی حاوی عناصر فراداده‌ای مبتنی بر XML منتشر شده در

1. Content

2. Value

محیط‌های پیوسته پویا دانست. همچنین، این پژوهش‌ها هیچ امتیاز ویژه و برتری برای پیشینه‌های مبتنی بر مارک ۲۱ و هسته دوبلین نسبت به یکدیگر نیافتند.

براساس آنچه پیشتر گفته شد، گرایش طرح‌های فراداده‌ای به پیاده‌سازی پیشینه‌های فراداده‌ای خود در بستر زبان نشانه‌گذاری گسترش‌پذیر (XML)، علاوه بر بهبود پردازش، تبادل و به‌طور کلی، مدیریت این پیشینه‌ها، موجب دسترس‌پذیری آنها از طریق محبوب‌ترین ابزارهای کاوش وب (موتورهای کاوش) گردیده و به‌نوعی نسبت به استفاده از زبان نشانه‌گذاری فرامتن (HTML) منجر به تولید ارزش افزوده شده است. پژوهش‌هایی که به مطالعه اثربخشی استفاده از پیشینه‌های فراداده‌ای هسته دوبلین مبتنی بر HTML در افزایش بازیابی‌پذیری و دسترس‌پذیری صفحات وب (اشیای محتوایی) پرداختند، این پیشینه‌ها را فاقد کارآمدی یا با میزان کارآمدی ضعیف ارزیابی کردند (Zhang and Dimitroff 2005b; Safari 2005; Zhang and Dimitroff 2004; شریف ۱۳۸۶).^۱ بر پایه یافته‌های پژوهش‌های انجام گرفته بر روی فراداده‌های مبتنی بر XML، این ضعف به دلیل پیاده‌سازی پیشینه‌های فراداده‌ای هسته دوبلین در بستر نحوی زبان نشانه‌گذاری فرامتن (HTML) است.

در مورد واکنش مثبت موتورهای کاوش به فرابرجسب‌ها و برجسب‌های HTML که در پژوهش‌های ترنر و برکییل (Turner and Brackbill 1998)، هنشاو و والاسکاس (Henshaw and Valauskas 2001)، کوئودو-توررو (Quevedo-Torrero 2004)، ژانگ و دیمیتروف (Zhang and Dimitroff 2004)، ژانگ و دیمیتروف (Zhang and Dimitroff 2005a)، ژانگ و دیمیتروف (Zhang and Dimitroff 2005b)؛ موهامد (Mohamed 2006) - مورد بررسی قرار گرفت نیز باید گفت، نخست اینکه تعداد این عناصر بسیار محدود است (چهار عنصر)، و نقاط دسترسی کمتری را نسبت به استفاده از طرح‌های فراداده‌ای استاندارد همچون مارک و هسته دوبلین در اختیار کاربران می‌گذارد. گذشته از این، زبان نشانه‌گذاری فرامتن (HTML) قالبی برای نمایش و نه انتقال و تبادل، و ذخیره‌سازی داده‌ها و فراداده‌هاست و امکان پیاده‌سازی آن نیز درون پایگاه‌های داده‌ای به‌عنوان قالب ذخیره‌سازی وجود ندارد. همچنین، استفاده از این عناصر برای توصیف اشیای محتوایی وب، به دلیل محدودیت فراوان (از لحاظ تعداد، ساختار، بستر معنی‌شناختی، و ... مناسب)، در بافت مدیریت اطلاعات و کتابخانه‌ای، و پایگاه‌های اطلاعاتی بزرگ و معتبر وب متداول نیست.

۱. عدم مطالعه بر روی قالب فراداده‌ای مارک در پژوهش‌های پیشین به دلیل عدم امکان پیاده‌سازی مارک در قالب زبان نشانه‌گذاری فرامتن (HTML) بود.

اما، با وجود مزایای انتخاب بستر نحوی XML برای پیشینه‌های فراداده‌ای، یافته‌های پژوهش‌های انجام‌شده بر روی پیشینه‌های فراداده‌ای مبتنی بر XML بیانگر آن است که موتورهای کاوش برای نمایه‌سازی اشیای محتوای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) رویکرد حذف برچسب^۱ را برگزیده‌اند. به عبارت دیگر، این موتورها هنگام نمایه‌سازی، برچسب‌ها یا نام‌های عناصر فراداده‌ای را در نظر نمی‌گیرند و یا اینکه در صورت نمایه‌سازی آنها - با احتمال قوی‌تر - از جستجوپذیر نمودن (یا پدیدار نمایی)^۲ آنها خودداری می‌کنند. بدیهی است انتخاب این رویکرد به دلیل قابلیت گسترش‌پذیری زبان XML است و با توجه به تجاری بودن موتورهای کاوش خردمندانه می‌نماید. از این رو، جای انجام پژوهش‌هایی که به بررسی تمهیدات و ترفندهایی برای نمایه‌پذیر شدن برچسب‌ها یا نام‌های عناصر، علاوه بر ارزش‌های عناصر پردازد، خالی نشان می‌دهد.

۷. منابع

- آقاعابدی، زهرا. ۱۳۸۹. بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران.
- پاول، توماس. ۲۰۰۲. راهنمای جامع طراحی و ساخت صفحات وب. ترجمه احمد شیدا. ۱۳۸۳. تهران: انتشارات سیمین‌دخت.
- پیشوا، پگاه، و موسی مجیدی. ۱۳۸۶. مطالعه میزان بکارگیری ابرداده‌های زبان نشانه‌گذاری فرامتن در مجله‌های دسترسی آزاد وب. فصلنامه علوم و فناوری اطلاعات ۲۲ (۳): ۱-۱۸.
- رشیدی، حجت‌الله. ۱۳۷۷. دستنامه روش تحقیق در مدیریت. تهران: گلپونه.
- شریف، عاطفه. ۱۳۸۶. بررسی میزان اثربخشی عناصر ابرداده‌ای بر رتبه‌بندی صفحات وب توسط موتورهای کاوش. فصلنامه کتابداری و اطلاع‌رسانی ۳۸ (۲): ۲۴۱-۲۵۸.
- طاهری، مهدی. ۱۳۸۷ الف. بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی. پایان‌نامه کارشناسی ارشد. کتابداری و اطلاع‌رسانی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران.
- طاهری، مهدی. ۱۳۸۷ ب. مقایسه کارآیی طرح فراداده‌ای هسته دوبلین و قالب فراداده‌ای مارک ۲۱ در سازماندهی منابع اطلاعاتی شبکه جهانی وب. فصلنامه کتابداری و اطلاع‌رسانی ۴۳ (۳): ۱۳۹-۱۵۸.
- طاهری، مهدی، نجلا حریری، و رحمت‌الله فتاحی. 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 ۴۸ (۴). ۱۴۱-۱۶۲.

1. tag name or label or element name

2. visible

فتاحی، رحمت‌الله، و مهدی طاهری. ۱۳۸۸. فهرست‌نویسی رایانه‌ای: مفاهیم، شیوه‌ها و ابزارهای فهرست‌نویسی در محیط رایانه‌ای. با همکاری فرشته ناقد احمدی. تهران: کتابدار.

- Alexa: The web information company. 2008. Global Top 500. http://www.alexa.com/site/ds/top_sites?ts_mode=global&lang=none (accessed 5 June 2011).
- Bifet, Albert, and Carlos Castillo. 2005. An analysis of factors used in search engine ranking. *AIRWeb* 8 (1): 48-57.
- Campex. 2005. Top search engines. <http://capmex.biz/resources/top-search-engines> (accessed 5 Jun. 2011).
- Henshaw, Robin, and Edward J. Valauskas. 2001. Metadata as a catalyst: Experiments with metadata and search engines in the internet journal, first monday. *Libri*. 51 (2): 86-101. www.librijournal.org/pdf/1999-3pp125-131.pdf (accessed 5 Nov. 2009).
- Mohamed, Khaled A. F. 2006. The impact of metadata in web resources discovering. *Online Information Review* 30 (2): 155-167.
- National Information Standards Organization (NKOS). 2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies. <http://www.niso.org/standards/resources/Z39-19-2005.pdf> (accessed 25 March 2009).
- Qin, J. 2000. Representation and organization of information in the web space: from MARC to XML. *Informing Science: International Journal of an Emerging Discipline* 3 (2): 83-87.
- Quevedo-Torrero, J. U. 2004. Improving web retrieval by mining the HTML tags for keywords and exploring the hyperlink structures of web pages, [Abstract] doctoral Dissertation. University of Houston. <http://www.lib.umi.com/dissertations/fullcit/3156028> (accessed 23 Oct. 2009).
- Safari, Mehdi. 2005. Search engines and resource discovery on the web. *Webology* 2 (2). www.webology.ir/2005/v2n2/a13.html (accessed 13 Nov. 2009).
- Sokvitne, Lloyd. 2000. An Evaluation of the effectiveness of current dublin core metadata for retrieval. www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF (accessed 13 Nov. 2009).
- Taheri, Sayyed Mahdi, and Nadjla Hariri. Forthcoming. A comparative study on the indexing and ranking of the content objects including the marcxml and dublin core's metadata elements by general search engines. *Electronic Library* 30 (4).
- Thornely, J. 2000. Metadata and the development of Dublin Core at State Library of Queensland and Education Queensland, Australia. *OCLC Systems & Services* 16 (3): 118-129.
- Turner, Thomas P., and Lise Brackbill. 1998. 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 and Technical Services* 42 (4): 258-271. <http://cat.inist.fr/?aModele=afficheN&cpsid=1748620> (accessed 25 sep. 2009).
- Zhang, Jin, and Alexandra Dimitroff. 2004. Internet search engine's response to metadata Dublin Core implementation. *Journal of Information Science* 30 (4): 310-320. <http://portal.acm.org/citation.cfm?id=1142111> (accessed 15 Nov. 2009).
- Zhang, Jin, and Alexandra Dimitroff. 2005a. The impact of Webpage content characteristics on webpage visibility in search engine result (Part I). *Information Processing & Management* 41 (3): 665-690. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VC8-4BHCBX4-1&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=33927751b92200b392f8c79b950dcd1 (accessed 15 Nov. 2009).
- Zhang, Jin, and Alexandra Dimitroff. 2005b. The impact of metadata implementation on Webpage visibility in search engine result (Part II). *Information Processing and Management* 41 (3): 691-715. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VC8-4BHCBX4-2&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=a853d410a866732d3f8ab5dd3217d412 (accessed 15 Nov. 2009).

Web Search Engines and Indexing and Ranking the Content Object Including Metadata Elements Available at the Dynamic Information Environments

Faeze Sadat Tabatabai Amiri*
MA in Library and Information science

Sayyed Mahdi Taheri¹
Faculty Member of Islamic Culture and Sciences Academy

Abdolhossein Farajpahlou²
Full Professor in Library and Information Science,
Chamran University

Farideh Osareh³
Full Professor in Library and Information Science,
Chamran University

Abdolhamid Moarrefzadeh⁴
Associate Professor in Library and Information Science,
Chamran University

Iranian Journal of
**Information
Processing &
Management**

Iranian Research Institute
For Science and Technology
ISSN 2251-8223
eISSN 2251-8231
Indexed in LISA, SCOPUS & ISC
Vol.27 | No.4 | pp: 907-920
summer 2012

Abstract: The purpose of this research was to make exam the indexing and ranking of XML content objects containing Dublin Core and MARC 21 metadata elements in dynamic online information environments by general search engines and comparing them together in a comparative-analytical approach. 100 XML content objects in two groups were analyzed: those with DCXML elements and those with MARCXML elements were published in website <http://www.marcdcmi.ir>. from late Mordad 1388 till Khordad 1389. Then the website was introduced to Google and Yahoo search engines. Google search engine was able to retrieve fully all the content objects during the study period through their Dublin Core and MARC 21 metadata elements; Yahoo search engine, however, did not respond at all. The indexing of metadata elements embedded in content objects in dynamic online information environments and different between indexing and ranking of them were examined. Findings showed all Dublin Core and MARC 21 metadata elements by Google search engine were indexed. And there was not observed difference between indexing and ranking DCXML and MARCXML metadata elements in dynamic online information environments by Google search engine.

Keywords: Dublin Core Metadata Initiative (DCMI), MARCXML, indexing metadata elements, dynamic online information environments, web search engine, ranking, Extensible Markup Language (XML)

*Corresponding author: faeze.tabatabai@gmail.com

1. taheerismster@gmail.com

2. farajpahlou@gmail.com

3. osareh.f@gmail.com

4. amoorefzadeh@hotmail.com