

واکنش موتورهای کاوش وب به پیشنهادها فرا داده‌ای مبتنی بر روش ترکیبی داده‌های خرد و روش داده‌های پیوندی

سید مهدی طاهری

دکتری علم اطلاعات و دانش‌شناسی،

استادیار گروه علم اطلاعات و دانش‌شناسی دانشگاه علامه طباطبائی

رضا نیکزاد بهله

کارشناسی ارشد علم اطلاعات و دانش‌شناسی

میترا صمیعی

دکتری علم اطلاعات و دانش‌شناسی

استادیار گروه علم اطلاعات و دانش‌شناسی دانشگاه علامه طباطبائی

پذیرش: ۹۶/۰۹/۲۹

دریافت: ۹۶/۰۹/۱۸

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS، LISA و ISC
<http://jlist.irandoc.ac.ir>
دوره XX | شماره X | صص XX-XX
۱۳XX X

نوع مقاله: مروری / پژوهشی

چکیده: پژوهش حاضر، با هدف تبیین واکنش موتورهای کاوش وب به پیشنهادها، داده‌های مبتنی بر روش ترکیبی داده‌های خرد و روش داده‌های پیوندی با رویکرد تجربی انجام شد. جامعه پژوهش را ۲۰۰ پیشنهاد داده‌ای در قالب دو گروه گواه و آزمون با تعداد پیشنهادی برابر تشکیل می‌دادند. پیشنهادها یاد شده به روش نمونه‌گیری تصادفی، از میان پیشنهادها، داده‌های دروازه اطلاعاتی کنسرسیوم محتوای ملی انتخاب شدند. در گروه آزمون، ۱۰۰ پیشنهاد داده‌ای انتخاب شده با رویکرد ترکیبی بر مبنای دو روش داده‌های خرد، به عنوان بهترین روش تولید خرده‌های غنی، و داده‌های پیوندی (قالب آردی‌اف/ایکس‌ام‌ال) ایجاد، و در گروه گواه، ۱۰۰ پیشنهاد بدون تغییر و با قالب اولیه آماده، و بر روی وب سایت <http://ww.rnikzad.ir> منتشر شدند. سپس وب سایت بیان شده به صورت مستقیم و بر اساس روش‌های توصیه شده، به موتورهای کاوش گوگل و بینگ معرفی شد. با استفاده از سياهه و ارسى، وضعیت نمایه‌پذیری و پیدانمایی پیشنهادها، داده‌های هر دو گروه در محیط موتورهای کاوش انتخابی مورد بررسی قرار گرفت. یافته‌های پژوهش حاکی از آن است، موتورهای کاوش گوگل و بینگ ارزش‌های عناصر پیشنهادها، داده‌های گروه‌های گواه و آزمون را نمایه‌سازی کردند، و پیشنهادها را با جستجوی ارزش‌های عناصر داده‌ای در نتایج

به این مقاله به شکل زیر استناد کنید:

دورن متن:

(طاهری، پژوهش نامه پردازش و مدیریت اطلاعات)

در فهرست منابع:

طاهری، پژوهش نامه پردازش و مدیریت اطلاعات. واکنش موتورهای کاوش وب به پیشنهادها، داده‌های مبتنی بر روش ترکیبی داده‌های خرد و روش داده‌های پیوندی.

پژوهشنامه پردازش و مدیریت اطلاعات.

(دسترسی در <http://Jipm.irandoc.ac.ir>)

روز/ماه/سال)

جستجو پدیدار می‌نمایند، اما نمایش آنها به شکل معنادار و همانند پیشینه‌های فراداده‌ای تولید شده فقط مبتنی بر روش‌های تولید خرده‌های غنی نبود. به عبارت دیگر، نرم‌افزارهای خزنده-نمایه‌ساز موتورهای کاوش وب توانایی شناسایی و تشخیص برچسب‌های (عناصر) توصیه شده براساس روش داده‌های خرد را که در قالب چارچوب توصیف منبع (آردی‌اف)، به عنوان چهارچوب و بستر نحوی روش داده‌های پیوندی، پیاده‌سازی شده‌اند، ندارند، و به پیروی از آن، این پیشینه‌ها را در نتایج جستجو به صورت معنادار نمایش نمی‌دهند. بنابراین موتورهای کاوش وب با وجود تلاش برای ایجاد وب معنایی، سازگاری کامل با استانداردهای وب معنایی ندارند.

کلیدواژه‌ها: پیشینه‌های فراداده‌ای؛ نمایه‌پذیری؛ پیدانمایی؛ روش داده‌های خرد؛ روش داده‌های پیوندی؛ موتورهای کاوش وب

* سیدمهدی طاهری Email: taherismster@gmail.com

۱. مقدمه

شبکه وب به عنوان مهمترین و بزرگترین بخش شبکه اینترنت، از زمان پیدایش تاکنون، سه نسل گوناگون را تجربه نموده است. در نسل نخست، قابلیت‌های فرارسانه‌ای و فراپیوند نمودن محتوا مد نظر قرار گرفت. نسل دیگر، به ایجاد بسترهای تعاملی بین انسان‌ها با استفاده از فناوری‌های وب و در محیط شبکه وب پرداخت. نسل دو دیگر، بر این نکته تأکید داشت که داده‌های دسترس‌پذیر کنونی در محیط وب به دلیل عدم پیوند با داده‌های دیگر، داده‌های خام به شمار می‌آیند، و برای معنادار شدن نیاز به پردازش (ساختارمند نمودن و ایجاد ارتباط میان آنها) دارند (طاهری و دیگران ۱۳۹۴، ۴۶-۴۸). به عبارت دیگر، داده‌های ساختارمندی که با دیگر داده‌های منتشر شده روی وب پیوند یابند، بازیابی دانش را در پی خواهند داشت، و افزون بر بهبود تعامل انسان و ماشین، موجب بهبود تعامل ماشین‌ها با یکدیگر می‌گردند. تیم برنرز-لی مخترع وب، نام نسل جدید را وب معنایی نهاد و روش داده‌های پیوندی^۱ را برای تحقق آن توصیه نمود (جمالی مهمویی ۱۳۸۲، ۵۱؛ طاهری، ۱۳۹۲؛ برنرز-لی^۲، ۲۰۰۶، ۵-۷).

^۱. برای کسب اطلاعات بیشتر در خصوص روش داده‌های پیوندی به این مقاله رجوع گردد:

طاهری، سیدمهدی. ۱۳۹۲. داده‌های پیوندی: مفاهیم و استانداردها، با تأکید بر جامعه میراث فرهنگی. فصلنامه اطلاع‌رسانی کنسرسیوم محتوای ملی (۱۲): ۴-۱۳.

^۲. Tim Berners-Lee

از سوی دیگر، برای جستجو و دسترسی به داده‌های منتشر شده بر روی

وب، ابزارها و نظام‌های گوناگونی طراحی شدند. در میان این ابزارها و نظام‌ها، موتورهای کاوش به دلیل قابلیت‌های منحصر به فرد، بیش از دیگران مورد اقبال کاربران واقع شدند. کاربران با استفاده از کلیدواژه‌ها یا عبارات کلیدی به جستجو در موتورهای کاوش می‌پردازند. این در حالی است که این کلیدواژه‌ها یا عبارات کلیدی نمی‌توانند به شیوه‌ای مناسب دربرگیرنده اطلاعات مورد نظر آنها باشد و کاربر با حجم وسیعی از اطلاعات به عنوان نتایج بازیابی روبرو می‌شود. زیرا موتورهای کاوش فقط امکان محدود نمودن محتوای درخواست بر اساس سه یا چهار برچسب را هنگام فرایند جستجو امکان‌پذیر می‌سازند. بنابراین عدم نمایش ساختارمند اطلاعات بازیابی شده، درک و تشخیص ارتباط آن‌ها با نیاز کاربر، و انتخاب گزینه‌های مناسب را دشوار می‌سازد. در صورتی که کاربران در اغلب موارد برای هر درخواستی به سراغ موتورهای کاوش می‌روند، این ویژگی، دسترس‌پذیر کردن اطلاعات از طریق موتورهای کاوش را برای فراهم‌کنندگان اطلاعات پیوسته ضروری می‌سازد (آزاد و شریف ۱۳۸۶، ۱۳۲؛ طاهری ۱۳۹۴). در چنین شرایطی، ابزارها و تمهیداتی مورد نیاز است که بتوانند درخواست کاربر را از نظر معنایی درک کرده، امکان جستجوی معنایی را فراهم نمایند. از این رو، ساختارمند نمودن داده‌ها و فراداده‌ها، گام مهمی به شمار می‌آید. با توجه به این که هدف هر موتور کاوش ارائه اطلاعات موجود در پایگاه داده‌ای خود به کاربران است، نمایه‌سازی و همچنین نمایش ساختارمند (معنادار) این اطلاعات توسط موتورهای کاوش برای کاربران، افزون بر بهبود ربط در فرایند بازیابی و افزایش رضایت کاربران، به تولید ارزش افزوده می‌انجامد. در راستای معنابخشی به محتوای قابل دسترس از طریق شبکه وب، تلاش دیگری از سوی موتورهای کاوش صورت گرفته است. این موتورها با طراحی و توصیه روش‌های تولید خرده‌های غنی، در جهت نمایه‌سازی و پدیدارنمایی معنادار داده‌های وب گام برداشتند. این روش نیز همانند روش داده‌های پیوندی، بر ساختارمند نمودن داده‌های ذخیره شده، و پیوند دادن آنها با داده‌های مرتبط تأکید می‌کند. پیاده‌سازی این روش و تولید پیشینه‌های فراداده‌ای مبتنی بر آن، سبب شناخت دقیق موتور کاوش از موجودیت توصیف شده در صفحه وبی، و نمایش معنادار اطلاعات در نتایج بازیابی می‌شد (وان در میر^۱ و دیگران، ۲۰۱۱، ۱؛ www.schema.org).

^۱ . Van der Meer

با وجود این که تاکنون اقدامات گوناگونی برای ایجاد قابلیت معنایی در محیط وب توسط فراهم‌کنندگان اطلاعات و خدمات پیوسته صورت گرفته است، اما در این میان، دو راهکار روش داده‌های پیوندی بر اساس ایده وب معنایی، و تولید خرده‌های غنی به عنوان لایه‌ی معنایی نسخه جدید زبان نشانه‌گذاری فرامتن (اچ‌تی‌ام‌ال) که از سوی موتورهای کاوش توصیه شده است، بیش از دیگران مورد توجه قرار گرفته است. بدیهی است شیوه مطلوب تولید فراداده‌های معنادار برای موتورهای کاوش، استفاده از روش ترکیبی برای تولید پیشنهادهاست تا حداکثر بهره‌وری از قابلیت‌های معنایی تلاش‌های جدید صورت گیرد. بدین جهت، استانداردها و ابزارهایی چند از سوی کنسرسیوم وب جهانی برای تبدیل پیشنهادها تولید شده بر اساس روش‌های تولید خرده‌های غنی به روش داده‌های پیوندی طراحی و ارائه شده است. با توجه به این که هر دو راهکار بر معنابخشی به فرایندهای نمایه‌سازی و بازیابی اطلاعات تاکید می‌کنند، مسئله‌ای که اینجا مطرح می‌گردد آن است که واکنش موتورهای کاوش وب به پیشنهادها فراداده‌ای تولید شده بر اساس روش ترکیبی (داده‌های پیوندی و خرده‌های غنی) چگونه خواهد بود؟ آیا نرم‌افزارهای خزنده-نمایه‌ساز موتورهای کاوش قادر به تشخیص عناصر فراداده‌ای به طور کامل (نام و ارزش برچسب) و ارائه (پیدانمایی)¹ ساختارمند و معنادار آنها در نتایج خود هستند؟ پژوهش حاضر، برای پاسخگویی به این مسئله طرح‌ریزی شده است. نمایه‌سازی و پیدانمایی معنادار اشیای محتوایی توسط موتورهای کاوش، به دلیل معنابخشی و نمایش معنادار اطلاعات مرتبط با یکدیگر، موجب افزایش رضایت کاربران نهایی و درک بیشتر آنها از محتوای بازیابی شده می‌گردد. این مهم همواره دغدغه طراحان موتورهای کاوش وب بوده است. طراحی تمهیدات گوناگون و انجام اقدامات گسترده در این خصوص از سوی طراحان یاد شده، دلیلی بر این مدعاست. از سوی دیگر، وب در حال تجربه نسل جدید خود، یعنی وب معنایی (یا وب 3.0) است. برای تحقق وب معنایی، روش داده‌های پیوندی از سوی مخترع وب پیشنهاد شده و مورد استقبال گسترده فراهم‌کنندگان اطلاعات پیوسته واقع شده است. این در حالی است که موتورهای کاوش نیز خرده‌های غنی را به عنوان گام معنابخشی توصیه می‌نمایند. استفاده از هر دو روش به صورت ترکیبی، زمینه را برای هم‌راستا نمودن هر دو تلاش برای تحقق وب معنایی، هموار می‌نماید. در عین حال، پاسخ مثبت موتورهای کاوش وب به پیشنهادها ایجاد شده بر پایه روش داده‌های پیوندی، بیانگر اقبال موتورهای کاوش به روش داده‌های پیوندی و همسویی و ایفای نقش فعال در تحقق وب معنایی خواهد بود.

¹ . Visibility

زود آید
ویدئو آپیش نشده

هدف و پرسش‌های پژوهش

هدف اصلی این پژوهش، تبیین واکنش موتورهای کاوش وب به پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و روش داده‌های پیوندی بود. به منظور دستیابی به این هدف چهار پرسش طراحی شد:

۱. وضعیت نمایه‌سازی پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی توسط موتور کاوش گوگل چگونه است؟
۲. وضعیت نمایه‌سازی پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی توسط موتور کاوش بینگ چگونه است؟
۳. وضعیت پیدانمایی (نمایش ساختارمند) پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی توسط موتور کاوش گوگل چگونه است؟
۴. وضعیت پیدانمایی (نمایش ساختارمند) پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی توسط موتور کاوش بینگ چگونه است؟

مرور پیشینه‌های پژوهش

بهبود عملکرد موتورهای کاوش و واکنش آنها نسبت به فراداده‌ها، موجب افزایش و سهولت دسترسی به اطلاعات موجود در شبکه وب خواهد شد. به منظور نیل به این هدف، همواره اقبال ویژه‌ای از سوی طراحان موتورهای کاوش و طراحان نظام‌ها و استانداردهای فراداده‌ای به افزایش میانگین پذیرایی این دو ابزار بازیابی محتوای وب بوده است. در همین راستا، تاکنون پژوهش‌های گوناگونی در حوزه میانگین‌پذیری موتورهای کاوش و نظام‌های فراداده‌ای انجام شده است. هدف مشترک تمامی این پژوهش‌ها، تبیین وضعیت نمایه‌سازی و پیدانمایی پیشینه‌های فراداده‌ای توسط موتورهای کاوش و ارائه راهکارهایی برای بهبود وضعیت موجود بوده است. با مرور این پژوهش‌ها، می‌توان آنها را به سه گروه کلی تقسیم نمود. مبنای این گروه‌بندی، عناصر (برچسب‌های) مورد استفاده برای توصیف و بازنمون خصایص اشیای محتوایی هستند. زیرا تعامل موتورهای کاوش با پیشینه‌های فراداده‌ای و به پیروی از آن اشیای محتوایی مرتبط، از طریق این عناصر صورت می‌گیرد. گروه نخست، پژوهش‌هایی را دربرمی‌گیرد که به مطالعه واکنش موتورهای کاوش به برچسب‌ها و فرابرجسب‌های زبان نشانه‌گذاری فرامتن (اچ‌تی‌ام‌ال) پرداخته‌اند. گروه دیگر، تعامل موتورهای کاوش با نظام‌های فراداده‌ای دارای پیشینه‌های ایجاد شده بر مبنای استانداردهای فراداده‌ای را مورد مطالعه قرار داده‌اند و گروه پایانی، تأثیر استفاده از روش‌های

تولید خرده‌های غنی بر بهبود نمایه‌پذیری و پیدانمایی پیشینه‌های فراداده‌ای توسط موتورهای کاوش را ارزیابی نموده است.

در گروه نخست که بستر نحوی عناصر فراداده‌ای آنها زبان نشانه‌گذاری فرامتن (اچ‌تی‌ام‌ال) بود، صفحات یا اشیای محتوایی جامعه پژوهش، برای توصیف یا سازماندهی محتوای خود از همان برجسب‌ها و فرابرجسب‌های بستر نحوی استفاده نمودند. زبان نشانه‌گذاری فرامتن، به عنوان بهترین قالب نمایش صفحات وب که از اقبال ویژه‌ای در بین فراهم‌کنندگان اطلاعات پیوسته برخوردار است، افزون بر برجسب‌هایی برای ذخیره (بیشتر برای قابلیت‌های نمایشی) محتوا، فرابرجسب‌هایی برای توصیف محتوای ذخیره شده در هر صفحه ارائه می‌کند. پژوهش‌های این گروه به مطالعه تأثیر بکارگیری برجسب‌ها و فرابرجسب‌های اچ‌تی‌ام‌ال بر بهبود نمایه‌پذیری و بازیابی‌پذیری اشیای محتوایی پرداختند. برخی از این پژوهش‌ها مانند پژوهش‌های ترنر^۱ و برکییل^۲ (۱۹۹۸)، کوئودو-توررو^۳ (۲۰۰۴)، و ژانگ^۴ و دیمیتروف (۲۰۰۵a)، بر اثربخشی برجسب‌ها و فرابرجسب‌های زبان نشانه‌گذاری یاد شده بر بهبود بازیابی و پیدانمایی اشیای محتوایی وب متمرکز بودند. نتایج این پژوهش‌ها واکنش مثبت موتورهای کاوش به برجسب‌ها و فرابرجسب‌های زبان نشانه‌گذاری فرامتن را نشان می‌داد، و بهبود بازیابی و پیدانمایی اشیای محتوایی مورد مطالعه قابل مشاهده بود.

گروه دیگر پژوهش‌هایی بودند که کارآمدی استفاده از عناصر یا پیشینه‌های فراداده‌ای مبتنی بر طرح‌های فراداده‌ای استاندارد مانند طرح فراداده‌ای هسته دوبلین، طرح فراداده‌ای توصیف شیء و قالب فراداده‌ای مارک در بهبود بازیابی‌پذیری و رتبه‌بندی اشیای محتوایی مرتبط در موتورهای کاوش را مورد مطالعه قرار دادند. گرایش به بررسی تأثیر عناصر فراداده‌ای استاندارد، جامعیت و بستر معناشناختی استانداردهای فراداده‌ای برای سازماندهی اشیای محتوایی وب در مقابل محدودیت‌های برجسب‌ها و فرابرجسب‌های اچ‌تی‌ام‌ال بود. بستر نحوی پیشینه‌های فراداده‌ای مورد مطالعه در این گروه زبان‌های نشانه‌گذاری فرامتن و گسترش‌پذیر (ایکس‌ام‌ال) بودند. یافته‌های این پژوهش‌ها به دلیل استفاده از رویکردهای

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

گوناگون در ایجاد پیشینه‌های فراداده‌ای مورد بررسی متفاوت بود. برخی به مقایسه برچسب‌ها و فرابرجسب‌های اچ‌تی‌ام‌ال با عناصر فراداده‌ای استاندارد توجه داشتند. نتایج این پژوهش‌ها تأثیر بهتر برچسب‌ها و فرابرجسب‌های اچ‌تی‌ام‌ال نسبت به عناصر فراداده‌ای استاندارد بر بهبود بازیابی اشیای محتوایی توسط موتورهای کاوش را نشان می‌داند (هنشاو^۱ و والاسکاس^۲ ۲۰۰۱؛ ژانگ و دیمیتروف^۳، ۲۰۰۴؛ ژانگ و دیمیتروف^۴، ۲۰۰۵b؛ محمد^۵، ۲۰۰۶؛ شریف^۶، ۱۳۸۶). برخی دیگر، بررسی بازیابی‌پذیری و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای استاندارد، مربوط به طرح فراداده‌ای هسته دوبلین، مبتنی بر زبان نشانه‌گذاری فرامتن را انجام دادند. این پژوهش‌ها اذعان داشتند بهره‌گیری از عناصر فراداده‌ای استاندارد جاسازی شده در اشیای محتوایی مبتنی بر اچ‌تی‌ام‌ال، تأثیر چندانی بر بهبود بازیابی و رتبه‌بندی در محیط موتورهای کاوش وب ندارد (سوکویتن^۵ ۲۰۰۰؛ صفری^۶، ۲۰۰۵). و دسته پایانی در این گروه، پژوهش‌هایی بودند که بر پیاده‌سازی پیشینه‌های فراداده‌ای مورد مطالعه در بستر نحوی اصلی طرح‌های فراداده‌ای استاندارد، یعنی زبان نشانه‌گذاری گسترش‌پذیر (ایکس‌ام‌ال) متمرکز بودند. باور این پژوهش‌ها بر آن بود که بررسی درست واکنش موتورهای کاوش به عناصر و پیشینه‌های فراداده‌ای باید نسبت به بستر نحوی منتخب استانداردهای فراداده‌ای مورد ارزیابی قرار گیرد. نتایج این پژوهش‌ها بیانگر بهبود نمایه‌پذیری عناصر و پیشینه‌های فراداده‌ای مبتنی بر ایکس‌ام‌ال توسط نرم‌افزارهای خزنده-نمایه‌سازی موتورهای کاوش، و پیدانمایی آنها در نتایج جستجو بود (طاهری، حریری، و فتاحی، ۱۳۸۸؛ طباطبائی امیری و دیگران، ۱۳۹۱؛ باب‌الحوائجی، طاهری، و آقاعابدی، ۱۳۹۴). به ویژه پژوهش‌های طاهری و دیگران (۲۰۱۴a) و (۲۰۱۴b) نشان داد که پیاده‌سازی پیشینه‌های فراداده‌ای در بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر (براساس ساختاری خاص)، موجب نمایه‌پذیری نام‌های برچسب عناصر فراداده‌ای و امکان جستجوی اشیای محتوایی از طریق آنها (جستجوی عنصر-پایه) خواهد شد.

1 . Henshaw, Robin
2 . Valauskas
3 . Dimitroff
4 . Mohamed
5 . Sokvitne
6 . Safari

مطالعه کارایی و اثربخشی ایجاد پیشینه‌های فراداده‌ای براساس

جدیدترین روش‌های توصیه شده از سوی طراحان موتورهای کاوش، رویکرد پژوهش‌های گروه پایانی بوده است. در این گروه از پژوهش‌ها، پژوهشگران قصد داشتند وضعیت نمایه‌پذیری و نمایش معنادار برچسب‌های (عناصر) فراداده‌ای مبتنی بر روش‌های تولید خرده‌های غنی را ارزیابی نمایند. روش‌های ایجاد خرده‌های غنی عبارتند از داده‌های خرد، قالب‌های خرد، و چهارچوب توصیف منبع (خصایص) که در بین آنها روش داده‌های خرد بیش از دیگر روش‌ها توصیه شده است. برچسب‌ها (عناصر)، خصایص، و ارزش‌های آنها در روش‌های تولید خرده‌های غنی با نمونه‌های معادل خود در طرح‌های فراداده‌ای استاندارد بافت میراث فرهنگی (کتابخانه، موزه و آرشیو) متفاوت هستند و در اصل گسترشی از جنبه معناشناختی نسخه جدید (نسخه ۵٫۰) زبان نشانه‌گذاری فرامتن به شمار می‌آیند. با این وجود، از کارکردهای مشابه طرح‌های فراداده‌ای استاندارد پشتیبانی می‌کنند و می‌توان آنها را استانداردهای فراداده‌ای جدید دانست. نتایج پژوهش‌های این گروه حاکی از آن است که درج برچسب‌های (عناصر) فراداده‌ای مبتنی بر روش‌های داده‌های خرد، قالب‌های خرد، و چهارچوب توصیف منبع (خصایص) افزون بر بهبود نمایه‌پذیری آنها توسط موتورهای کاوش، آن هم به صورت معنایی، نمایش معنادار پیشینه‌های فراداده‌ای، و به پیروی از آن اشیای محتوایی را در پی خواهد داشت. به عبارت دیگر، به دلیل آن که این روش‌ها از سوی موتورهای کاوش وب توصیه می‌شوند و برخی مانند روش داده‌های خرد توسط توسعه‌دهندگان موتورهای کاوش طراحی شده‌اند، از اقبال بیشتری از سوی این موتورها برخوردارند (ایساکسن^۱، ۲۰۱۱؛ میکستر^۲، ۲۰۱۳؛ طاهری، ذوالقدر و حریری، ۱۳۹۵). اگر چه پژوهش‌های طاهری و دیگران (۲۰۱۴a و ۲۰۱۴b) اعمال تغییراتی محدود با رعایت جنبه‌های فنی استاندارد بر روی پیشینه‌های فراداده‌ای استاندارد مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر را نسبت به پیشینه‌های تولید شده بر پایه روش‌های تولید خرده‌های غنی در بهبود بازیابی‌پذیری اشیای محتوایی در محیط وب کارآمدتر بیان می‌دارد.

مرور پیشینه‌های مربوط به حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش، نشانگر آن است که به طور کلی موتورهای کاوش وب، با وجود مزایای عناصر یا پیشینه‌های

¹ Isaksen

² Mixer

فرا داده‌ای ایجاد شده براساس طرح‌های فراداده‌ای استاندارد که حاصل تلاش‌ها و تجربه بیش از نیم قرن بافت میراث فرهنگی است، واکنش بهتری به برجسب‌های (عناصر) یا پیشینه‌های فراداده‌ای مبتنی بر روش‌های تولید فراداده یا توصیه شده توسط طراحان موتورهای کاوش نشان می‌دهند. اگر چه پیاده‌سازی عناصر یا پیشینه‌های فراداده‌ای استاندارد که بستر نحوی ایکس‌ام‌ال را برای پیاده‌سازی برگزیدند، پاسخ مناسب‌تری نسبت به اچ‌تی‌ام‌ال از سوی نرم‌افزارهای خزنده-نمایه‌ساز موتورهای کاوش دریافت نمودند. به ویژه وقتی ساختاری خاص بر روی آن پیشینه‌ها اعمال می‌شد. از سوی دیگر، مخترع وب (تیم برنرز-لی) و کنسرسیوم وب جهانی روشی (راهبردی) به نام داده‌های پیوندی را برای تحقق وب معنایی توصیه می‌کند که بیشتر به بستر نحوی و ساختار پیشینه‌های فراداده‌ای استاندارد نزدیک است. نیز کنسرسیوم وب جهانی ابزارهایی برای تبدیل پیشینه‌های مبتنی بر روش‌های تولید خرده‌های غنی به بستر نحوی روش داده‌های پیوندی، یعنی آردی‌اف طراحی نموده است. بدیهی است اگر موتورهای کاوش تمهیدات مناسبی برای نمایه‌سازی و پیدانمایی معنادار پیشینه‌ها تولید شده مبتنی بر هر دو روش تولید خرده‌های غنی و روش داده‌های پیوندی بیاندیشند، گام مؤثر و بزرگی در تحقق وب معنایی برداشته‌اند. اما تاکنون پژوهشی برای بررسی وجود چنین تمهیداتی صورت نگرفته است و گمان می‌رود انجام چنین پژوهشی رویکرد کنونی موتورهای کاوش نسبت به استانداردهای وب معنایی را بازنمایاند.

روش‌شناسی

پژوهش حاضر از نظر نوع، یک پژوهش کاربردی به محسوب می‌شود. زیرا به حل مسئله ایجاد پیشینه‌های فراداده‌ای مبتنی بر رویکرد ترکیبی روش تولید خرده‌های غنی و داده‌های خرد با قابلیت نمایه‌پذیری و پیدانمایی ساختارمند (معنادار) در محیط موتورهای کاوش وب پرداخته است. بر انجام پژوهش از روش تجربی استفاده شده است. جامعه پژوهش دارای ۲۰۰ پیشینه فراداده‌های مربوط به موجودیت‌های کتابی موجود در دروازه اطلاعاتی کنسرسیوم محتوای ملی بود که از کتابخانه دیجیتالی دانشگاه علامه طباطبائی (ره)، کتابخانه دیجیتالی دانشگاه تهران، سازمان کتابخانه‌ها، موزه‌ها و مرکز اسناد آستان قدس رضوی و کتابخانه دیجیتالی دفتر تبلیغات اسلامی حوزه علمیه قم در دامنه‌های موضوعی "علم اطلاعات و دانش‌شناسی"، "روانشناسی"، "پزشکی"، و "حقوق" بنا بر سلیقه

پژوهشگران انتخاب شدند. این پیشینه‌ها به دو گروه گواه و آزمون با تعداد

پیشینه‌های یکسان تقسیم گردیدند. پیشینه‌های گروه نخست (گواه)، بر اساس روش داده‌های خرد ایجاد شدند. اما پیشینه‌های گروه آزمون، بر مبنای رویکرد ترکیبی، دو روش داده‌های خرد و داده‌های پیوندی تولید گردیدند. بدین صورت که نخست، پیشینه‌های فراداده‌ای انتخاب شده بر مبنای توصیه‌ها و ساختارهای الگوی داده‌های خرد، به پیشینه‌های مبتنی بر این روش تبدیل شدند (نمونه ۱)، و سپس با استفاده از ابزار بیان شده در ذیل، به قالب پیشنهادی روش داده‌های پیوندی (قالب آردی.اف./ایکس.ام.ال) تغییر بستر نحوی دادند. بنابراین پیشینه‌های جدید، بستر نحوی و قالب داده‌های خرد را رعایت نمودند. همچنین برای بررسی واکنش موتورهای کاوش وب به پیشینه‌های ایجاد شده، دو موتور کاوش گوگل و بینگ به دلیل گستردگی کاربرد و جایگاه آنها در میان دیگر موتورهای کاوش وب انتخاب شدند.

به منظور اعتبارسنجی تمامی پیشینه‌های فراداده‌ای ایجاد شده از بعد رعایت قابلیت‌های فنی روش داده‌های خرد، از ابزار Structured Data Testing Tools^۱ بخش tools webmaster موتور کاوش گوگل استفاده شد. پس از اطمینان از صحت و درستی هر یک از پیشینه‌های مورد مطالعه، برای تبدیل پیشینه‌های فراداده‌ای گروه آزمون از روش داده‌های خرد به روش داده‌های پیوندی (قالب آردی.اف./ایکس.ام.ال)، ابزار Microdata to RDF Distiller^۲ توسعه یافته و دسترس پذیر در وب سایت کنسرسیوم جهانی وب^۲ مورد استفاده قرار گرفت. سپس پیشینه‌های هر دو گروه به صورت مستقیم و بر اساس بهترین روش‌های توصیه شده (XML sitemap) به موتورهای کاوش وب برگزیده معرفی شدند. نمونه‌هایی از پیشینه‌های مربوط به هر دو گروه آزمون و گواه در ذیل ارائه شده‌اند:

تصویر ۱. تصویری از نمایش ساختارمند پیشینه‌های فراداده‌ای توسط موتور کاوش گوگل به عنوان هدف تولید و انتشار پیشینه‌های فراداده‌ای مورد مطالعه در پژوهش حاضر

^۱ ابزاری برای بررسی و اعتبارسنجی داده‌های ساختارمند تولید شده بر اساس الگوی داده‌های خرد که توسط موتور کاوش گوگل طراحی شده است. برای اطلاعات بیشتر در مورد این امکان، به نشانی اینترنتی ذیل مراجعه شود:

<https://developers.google.com/search/docs/guides/prototype>

^۲ <http://www.w3.org/2012/pyMicrodata>

Little Water Cantina - Eastlake - Seattle, WA
www.yelp.com Restaurants Mexican
 ★★★★★ 90 reviews • Price range: \$\$
 90 Reviews of **Little Water Cantina** "Three things are on my list when I eat out: great food, atmosphere, and **Vegetarian Vegan Pizza No Cheese** Recipe - Food.com - 248865
www.food.com/recipe/vegetarian-vegan-pizza-no-c...
 ★★★★★ 2 reviews - 1 hr 32 mins - 242.9 cal
 Aug 26, 2007 - This is from my dad, who developed some **vegan recipes** doesn't have any cheese, and you

Leonard Cohen – Free listening, videos, concerts, stats, & pictures at...
www.last.fm/music/Leonard+Cohen
 Watch videos & listen to **Leonard Cohen**: Suzanne, Hallelujah & more, plus 132 pictures. **Leonard Cohen**, (born September 21, 1934 in Montréal, Quebec, ...

Track	Duration
Suzanne	3:48
The Darkness	4:29
Going Home	3:51
Hallelujah	6:12

نمونه ۱. نمونه‌ای از یک پیشینه فراداده‌ای مبتنی بر الگوی داده‌های خرد

```
<div itemscope itemtype="http://schema.org/Book"><meta
itemprop="bookFormat" content="
اخلاق /><dt><dt>:عنوان</dt><dd itemprop="name">
پزشکی</dd>
<dt>:نویسنده</dt><dd itemprop="author">
منصور</dd><dt>:مترجم</dt><dd
itemprop="translator"></dt><dt>:توصیفگر</dt><dd
itemprop="description">
اخلاق</dd><dt>:تعداد صفحه</dt><dd>span itemprop="number of
pages">328</span>pages</dd><dt>:شابک</dt><dd
itemprop="ISBN"></dd>
<dt>:ناشر</dt><dd itemprop="publisher">
دانشگاه آزاد اسلامی تبریز</dd>
<dt>:سال نشر</dt><dd
itemprop="copyrightyear">1367</dd><dt>:صاحب محتوا</dt>
<dd itemprop="copyrightholder">
سازمان کتابخانه‌ها، موزه‌ها و مرکز اسناد آستان
قدس رضوی</dd>
<dt>:تاریخ ایجاد</dt><dd itemprop="creator">
نیکزاد</dd><dt>:پیشینه</dt>
```

<dd
itemprop="datecreated">1394/3/20</dd><dt>:زبان</dt><dd
itemprop="inLanguage">فارسی</dd><dt>:دسته</dt><dd
itemprop="genre">کتاب</dd>
</dl></div>

نمونه ۲. نمونه‌ای از یک پیشینه فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی (آردی اف/ایکس ام‌ال)

```
<?xml version="1.0" encoding="utf-8"?><rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:schema="http://schema.org/"> <schema:Book>
<schema:copyrightyear>1367</schema:copyrightyear><schema:author
r></schema:author><schema:ISBN></schema:ISBN>
<schema:bookFormat>الکترونیکی</schema:bookFormat><schema:descr
iption> اخلاق </schema:description><schema:datecreated>1394/3/20</schema
:datecreated>
<schema:copyrightholder> سازمان کتابخانه‌ها، موزه‌ها و مرکز اسناد آستان قدس
</schema:copyrightholder><schema:genre>کتاب</schema:genre>
<schema:name> اخلاق </schema:name><schema:creator> نیکزاد</schema:creator>
<schema:inLanguage>فارسی</schema:inLanguage><schema:translator
>
</schema:translator></schema:Book></rdf:RDF>
```

در مرحله اجرای پژوهش، روش گردآوری داده‌ها، مشاهده طراحی شده^۱ و ابزار آن سیاهه واریسی بود که بر پایه پرسش‌های پژوهش و توسط پژوهشگران طراحی گردید. گردآوری داده‌ها در خرداد ماه ۱۳۹۵ صورت گرفت. به منظور استخراج داده‌ها از راهبرد جستجوی site:rnikzad.ir “keyphrase” در موتورهای کاوش مورد مطالعه استفاده شد. برای تجزیه و تحلیل داده‌های گردآوری شده نیز بر مبنای آمار توصیفی، از نرم افزار اس‌پی‌اس‌اس بهره‌گیری

^۱ structured observation

شد و برای نمایش آنها با توجه به پرسش‌های پژوهش، جدول‌هایی متناسب با داده‌های درج شده در سیاهه واریسی استخراج شدند.

یافته‌های پژوهش

پژوهش حاضر برای نیل به اهداف خود، چهار پرسش مطرح نمود. برای پاسخ به پرسش‌های اول و دوم، جدول شماره ۱، و برای پاسخ به پرسش‌های سوم و چهارم، جدول شماره ۲ طراحی شدند. داده‌های گردآوری شده برای پاسخ به پرسش‌های پژوهش تصادفی بودند. پیشنهادی‌های فراداده‌ای هر دو گروه گواه و آزمون به مدت شش ماه روی وب سایت ایجاد شده دسترس پذیر گردیدند، و فرایند گردآوری داده‌ها برای اطمینان از صحت آنها یک ماه به طول انجامید.

پرسش‌های اول و دوم پژوهش:

وضعیت نمایه‌سازی پیشنهادی‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی توسط موتور کاوش گوگل چگونه است؟

وضعیت نمایه‌سازی پیشنهادی‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی توسط موتور کاوش بینگ چگونه است؟

جدول ۱. وضعیت نمایه‌سازی پیشنهادی‌های فراداده‌ای گروه‌های گواه و آزمون توسط موتورهای کاوش گوگل و بینگ

پیشینه‌های فراداده‌ای نمایه شده توسط موتورهای کاوش				تعداد کل پیشنهادها	گروه مورد مطالعه
درصد		تعداد			
بینگ	گوگل	بینگ	گوگل		
٪۱۰۰	٪۱۰۰	۱۰۰	۱۰۰	۱۰۰	گواه
٪۱۰۰	٪۱۰۰	۱۰۰	۱۰۰	۱۰۰	آزمون

چنان که داده‌های درج شده در جدول ۱ نشان می‌دهد، تمامی پیشنهادی‌های فراداده‌ای هر دو گروه گواه و آزمون توسط موتورهای کاوش گوگل و بینگ نمایه‌سازی شده‌اند. به عبارت دیگر، نرم‌افزارهای خزنده-نمایه‌ساز دو موتور کاوش مورد مطالعه، نسبت به عناصر (برچسب‌های)

پیشینه‌های فراداده‌ای گروه‌های گواه و آزمون واکنش مثبت نشان داده‌اند. بنابراین

پیاپی سازی پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد در بستر نحوی آردی‌اف، به عنوان چارچوب توصیه شده در روش داده‌های پیوندی، از ارزش نمایه‌ای برابری در مقایسه با پیشینه‌های فقط مبتنی بر روش داده‌های خرد برخوردار هستند. اگر چه ویژگی نمایه شدن معنادار پیشینه‌های فراداده‌ای مورد مطالعه بدون بازیابی آنها و نمایش ساختارمند در نتایج جستجو امکان پذیر نیست. بنابراین داده‌ها بیانگر نمایه شدن ارزش‌های عناصر پیشینه‌های فراداده‌ای گروه آزمون هستند، و درک معنادار عناصر و صفات^۱ آنها توسط موتورهای کاوش مشخص نیست.

پرسش‌های سوم و چهارم پژوهش:

وضعیت پیدانمایی (نمایش ساختارمند) پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی توسط موتور کاوش گوگل چگونه است؟

وضعیت پیدانمایی (نمایش ساختارمند) پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و داده‌های پیوندی توسط موتور کاوش بینگ چگونه است؟

جدول ۲. وضعیت پیدانمایی (نمایش ساختارمند) پیشینه‌های فراداده‌ای گروه‌های گواه و آزمون توسط موتورهای کاوش گوگل و بینگ

پیشینه‌های فراداده‌ای با نمایش ساختارمند در نتایج جستجو				تعداد کل پیشینه‌ها	گروه مورد مطالعه
درصد		تعداد			
گوگل	بینگ	گوگل	بینگ		
۱۰۰٪	۱۰۰٪	۱۰۰	۱۰۰	۱۰۰	گواه
۰٪	۰٪	۰	۰	۱۰۰	آزمون

بر اساس داده‌های جدول ۲، وضعیت پیدانمایی پیشینه‌های فراداده‌ای گروه آزمون با پیشینه‌های فراداده‌ای گروه گواه کاملاً متفاوت است. موتورهای کاوش گوگل و بینگ پیشینه‌های فراداده‌ای گروه گواه را به صورت ساختارمند و معنایی در نتایج جستجو نشان می‌دهند، اما پیشینه‌های

^۱ . Properties

فرا داده‌ای گروه آزمون را با وجود نمایه‌سازی کامل، به صورت معمولی در فهرست نتایج پدیدار می‌نمایند. بدین معنی که واکنش موتورهای کاوش گوگل و بینگ نسبت به دو گروه از پیشینه‌ها یکسان نیست، و استفاده از استانداردهای روش داده‌های پیوندی بر روی نمایش ساختارمند پیشینه‌های فرا داده‌ای مبتنی بر خرده‌های غنی تاثیر منفی می‌گذارد. به عبارت دیگر، نرم‌افزارهای خزنده-نمایه‌ساز موتورهای کاوش گوگل و بینگ فقط عناصر و صفات پیشینه‌های مبتنی بر روش داده‌های خرد که از سوی این موتورها طراحی و توصیه شده است را به صورت معنادار تشخیص داده و نمایه‌سازی نموده‌اند، اما به قدر به درک عناصر و صفات پیشینه‌های گروه آزمون نبوده‌اند.

نتیجه‌گیری

همان‌طور که برنرزی (۲۰۰۶) بیان داشت، داده‌های منتشر شده کنونی بر روی وب، داده‌های خام هستند و صرفاً توصیف ویژگی‌های آنها، بدون برقراری ارتباط میان داده‌ها پردازش کاملی نخواهد بود. پیوند میان داده‌های مرتبط موجب ایجاد بافت، و به دنبال آن تولید معنا خواهد شد. برای تغییر وب کنونی به وب معنایی، تلاش‌های گوناگونی از سوی طراحان نظام‌های مبتنی بر وب انجام شده است. افزون بر روش داده‌های پیوندی که توسط تیم برنرزی پیشنهاد گردید، توسعه نسخه ۵٫۰ زبان نشانه‌گذاری فرامتن، و به ویژه روش‌های تولید خرده‌ها غنی نیز حرکتی در راستای تولید معنا در محیط وب به شمار می‌آیند. بدیهی است تحقق وب معنایی وابسته به سازگاری این روش‌ها با یکدیگر و کوشش در این راستا خواهد بود. یافته‌های پژوهش حاضر که به منظور تبیین واکنش پرکاربردترین ابزارهای جستجو و دسترسی به داده‌های وب، یعنی موتورهای کاوش، نسبت به پیشینه‌های فرا داده‌ای ایجاد شده بر اساس روش ترکیبی داده‌های خرد و داده‌های پیوندی انجام شد، حاکی از آن است که موتورهای کاوش علاقه بیشتری به نمایه‌سازی و پیدانمایی معنادار پیشینه‌های فرا داده‌ای مبتنی بر داده‌های خرد، که از سوی طراحان آنها گسترش و توصیه شده، دارند و الگوریتم‌های مناسبی برای شناسایی عناصر و صفات پیشینه‌های فرا داده‌ای به صورت معنادار، و به پیروی از آن نمایه‌سازی و نمایش ساختارمند آنها در نتایج جستجو طراحی نموده‌اند. این مهم در پیدانمایی ساختارمند پیشینه‌های فرا داده‌ای تولید شده بر پایه روش داده‌های خرد (گروه گواه) نمایان گردید. چنین واکنش مثبتی بیشتر در یافته‌های پژوهش‌های ترنر و برکیل (۱۹۹۸)، کوئودو-توررو (۲۰۰۴)، و ژانگ و دیمیتروف (۲۰۰۵a)، هنهاو و والاسکاس (۲۰۰۱)،

¹ . Raw data

ژانگ و دیمیتروف (۲۰۰۴)، ژانگ و دیمیتروف (۲۰۰۵b)، محمد (۲۰۰۶) و

شریف (۱۳۸۶) نسبت به روش‌های (یا فرابرجسب‌های) توصیه شده از سوی موتورهای کاوش ذکر گردیده بود. اما در خصوص پیشینه‌های فراداده‌ای تولید شده بر پایه روش ترکیبی، وضعیت نمایه‌سازی و پیدانمایی توسط موتورهای کاوش متفاوت بود. با وجود این که نمایه‌پذیری پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی به دلیل استفاده از بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر (ایکس‌ام‌ال) در سطح نمایه‌پذیری ارزش‌های عناصر فراداده‌ای شبیه پیشینه‌های فراداده‌ای بر اساس روش داده‌های خرد بود، و پژوهش‌های طاهری و دیگران (۱۳۸۸)، طباطبائی امیری و دیگران (۱۳۹۱)، باب‌الحوائجی و دیگران (۱۳۹۴)، و طاهری و دیگران (۲۰۱۴a) و (۲۰۱۴b) نیز یافته‌های یکسانی داشتند، اما این پیشینه‌ها در فهرست نتایج جستجو به صورت ساختارمند و معنادار پدیدار نشدند. به عبارت دیگر، استفاده از روش داده‌های خرد برای ایجاد عناصر یا پیشینه‌های فراداده‌ای در بستر استانداردهای روش داده‌های پیوندی باعث تولید خرده‌های غنی در نتایج جستجوی موتورهای کاوش نخواهد شد، و موتورهای یاد شده این پیشینه‌ها را به صورت معنادار نمایش نمی‌دهند. این یافته بیانگر عدم سازگاری دو روش مهم تولید معنا در محیط وب است. بدین معنا که نوعی واگرایی در تلاش‌های صورت گرفته از سوی طراحان اصلی استانداردها و خدمات وب به چشم می‌خورد. این واگرایی سردرگمی توسعه‌دهندگان اطلاعات پیوسته در وب را موجب می‌شود. در حال حاضر بسیاری از این طراحان در حال پیاده‌سازی روش داده‌های پیوندی که از سوی کنرسیوم وب توصیه شده در نظام‌های اطلاعاتی خود هستند. اگر داده‌های این نظام‌ها حتی با بهره‌گیری از روش‌های تولید خرده‌های غنی، به سطح پدیدارنمایی معنایی در نتایج جستجوی موتورهای کاوش نرسد، چالشی بزرگ در دسترس‌پذیری محتوای وب ایجاد خواهد شد. زیرا عدم سازگاری میان روش‌های توصیه شده از سوی کنرسیوم وب، به عنوان مرجع اصلی طراحی و تدوین استانداردهای وب، و الگوها و روش‌های توصیه شده از سوی موتورهای کاوش، به عنوان پرکاربردترین و محبوب‌ترین ابزارهای کاوش وب را نشان می‌دهد. به نظر می‌رسد طراحی الگوریتم‌های مناسب برای شناسایی استاندارد روش داده‌های پیوندی یعنی چارچوب توصیف منبع به ویژه در بستر نحوی ایکس‌ام‌ال از سوی طراحان موتورهای کاوش، برای نمایش ساختارمند پیشینه‌های فراداده‌ای نمایه شده، راهکار مناسبی در جهت همراستایی تلاش‌های توسعه وب معنایی خواهد بود. بر اساس نتایج پژوهش حاضر، انجام پژوهش‌هایی بر روی تبیین واکنش موتورهای کاوش به داده‌های مستند جاسازی شده در پیشینه‌های مبتنی بر روش‌های تولید خرده‌های غنی، و بررسی استفاده از الگوهای مفهومی و

استانداردهای مدیریت اطلاعات و دانش بافت میراث فرهنگی برای بهبود ساختار و اثربخشی روش‌های تولید خرده‌های غنی در طراحی پیشینه‌های فراداده‌ای با نمایه‌پذیری و پیدانمایی معنادار در محیط موتورهای کاوش وب پیشنهاد می‌گردد.

منابع

آزاد، اسدالله، و عاطفه شریف. ۱۳۸۶. وب معنایی در پیوند با سایبرنتیک. مجله مطالعات تربیتی و روان‌شناسی، ۸ (۳): ۱۳۱-۱۴۷.

باب‌الحوایجی، فهیمه؛ سید مهدی طاهری، و زهرا آقاآبادی. ۱۳۹۴. بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی. فصلنامه مطالعات دانش‌شناسی (۳۴): ۴۳-۵۹.

جمالی مهمویی، حمیدرضا. ۱۳۸۲. وب معنایی: شیوه‌ای روبه تکامل برای ذخیره و بازیابی کارآمدتر اطلاعات روی اینترنت. فصلنامه اطلاع‌شناسی (۲): ۴۷-۶۶.

شریف، عاطفه (۱۳۸۶). بررسی میزان اثر بخشی عناصر ابرداده‌ای بر رتبه‌بندی صفحات وب توسط موتورهای کاوش. فصلنامه کتابداری و اطلاع‌رسانی. شماره ۳۸ (تابستان ۱۳۸۶).

طاهری، سید مهدی، نجلا حریری، و سید 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).

طاهری، سید مهدی. ۱۳۹۲. داده‌های پیوندی: مفاهیم و استانداردها، با تأکید بر جامعه میراث فرهنگی. فصلنامه داخلی کنسرسیوم محتوای ملی (۱۲): ۴-۱۳.

طاهری، سید مهدی. ۱۳۹۴. فراداده‌ها، موتورهای کاوش و میانکنش‌پذیری آنها. تهران: کتابدار.

طاهری، سید مهدی، ناهید گویلی، مریم شکفته، و مریم کازرانی. ۱۳۹۴. رویکرد بافت-مدار در طراحی پروفایل کاربردی فراداده‌ای و توسعه وب معنایی. فصلنامه مطالعات دانش‌شناسی (۵): ۴۵-۵۷.

طاهری، سید مهدی، سارا ذوالقدر، و نجلا حریری. ۱۳۹۵. بررسی تطبیقی نمایه‌سازی و پیدانمایی پیشینه‌های فراداده‌ای مبتنی بر الگوی داده‌های خرد توسط موتورهای کاوش وب. (مقاله در دست داوری).

طباطبایی امیری، فائزه السادات، سید مهدی طاهری، عبدالحسین فرج‌پهلوی، فریده عصاره، و عبدالحمید معرف‌زاده. ۱۳۹۱. موتورهای کاوش وب و نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای دسترس‌پذیر در محیط‌های اطلاعاتی پیوسته پویا. فصلنامه پردازش و مدیریت اطلاعات (۲۷): ۹۰۷-۹۲۰.

Berners-Lee, Tim. 2006. Linked Data. Retrieved 05 Feb. 2016 from <http://www.w3.org/DesignIssues/LinkedData.html>.

Henshaw, Robin, and Edward J Valauskas. 2001. Metadata as a Catalyst: Experiments with Metadata and Search Engines in the Internet Journal. First Monday. www.librijournal.org/pdf/1999-3pp125-131.pdf (accessed 14 July 2016).

Isaksen, Leif. 2011. Archaeology and the Semantic Web. http://eprints.soton.ac.uk/196571/1.hasCoversheetVersion/y_gao_PhD_thesis_0111.pdf (accessed 14 July 2016).

Mixer, Jeffrey. 2013. Linked Data in VRA Core 4.0: Converting VRA XML Records into RDF/XML. <http://jmixter.s3-website-us-east-1.amazonaws.com/thesis/LinkedDataInVRACore4.pdf> (accessed 14 July 2016).

Mohamed, Khaled A.f. 2006. The impact of metadata in web resources discovering. *Online Information Review*. 30 (2): 155-167.

Quevedo-Torrero, and Jesus Ubaldo. 2004. Improving web retrieval by mining the html tags for keywords and exploring the hyperlink structures web pages. Ph. D. Dissertation. Department of Computer Science, University of Houston. <http://wwwlib.umi.com/dissertations/fullcit/3156028>. (accessed 14 July 2016).

Safari, Mehdi. 2005. Search Engine and Resource Discovery on the Web: Is Dublin Core an Impact Factor. www.webology.ir/2005/v2n2/a13.html. (accessed 14 July 2016).

Sokvine, Lloyd. 2000. An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval. www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF (accessed 14 July 2016).

Taheri, S. M., Nadjla Hariri, and S. R. Fattahi. ۲۰۱۴a. Does discarding XML Declarations and Changing File Extensions improve the Indexability and Visibility of Metadata Tag Names in Web Search Engines?. *Journal of Information Science*. Vol.: 40; No.: 6.

Taheri, S. M, Nadjla Hariri, and S. R Fattahi. 2014b. Using Data Island Method for Creating Metadata Records with Indexability and Visibility of Element Tag Names in the Web Search Engines. *Library High Tech*. Vol.:32; Issue: 1.

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. <http://cat.inist.fr/?aModele=afficheN&cpsid=1748620> (accessed 14 July 2016).

Van der Meer, J., F. Boon, F. Hogenboom, F. Frasinca, and U. Kaymak. 2011. A framework for automatic annotation of web pages using the Google rich snippets vocabulary. In *Proceedings of the 2011 ACM Symposium on Applied Computing* 765-772. ACM.

Zhang, Jin, and Alexandra Dimitroff. 2004. Internet search engine's response to metadata Dublin Core implementation. <http://portal.acm.org/citation.cfm?id=1142111> (accessed 14 July 2016).

Zhang, Jin, and Alexandra Dimitroff. 2005a. The impact of metadata implementation on Webpage visibility in search engine result (Part II). 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 14 July 2016).

Zhang, Jin, and Alexandra Dimitroff. 2005b. The impact of Webpage content characteristics on webpage visibility in search engine result (Part I). 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=33927751b92200b392f8c79b950dcdb1 (accessed 14 July 2016).

Study on Search Engines' Reaction to the Metadata Records Created Based on Combined Method of Rich snippets and Linked Data

Seyyed Mahdi Taheri: PhD in Knowledge and Information Science;
Assistant Professor at Department of Knowledge and Information Science,
Allameh Tabataba'i University; taherismster@gmail.com

Reza Nikzad Bahle : M.A. in Knowledge and Information Science in Allameh
Tabataba'i University; r.nikzad20@gmail.com

Mitra Samiee : PhD in Knowledge and Information Science ;Assistant
Professor at Department of Knowledge and Information;
Science, Allameh Tabataba'i University samiei.mitra66@gmail.com

Abstract

The purpose of this research was to find out the reaction of Web Search Engines to Metadata records created based on the combined method of Rich Snippets and Linked Data. 200 metadata records in two groups (100 records as the control group with the normal structure and, 100 records created based on microdata and implemented in RDF/XML as experimental group) extracted from the information gateway of Iranian Content National Consortium (INCN) were analyzed through an experimental approach. The metadata records of two groups were published on an independent website (www.rnikzad.ir), and were introduced to Google and Bing search engines directly. The using a checklist, the status of index ability and visibility of the published metadata records in two search engines were examined. Findings show all the element values of the metadata records of two groups indexed by Google and Bing, and were visible in the search results; but the records were not displayed in the search results semantically like the metadata records created in rich snippets methods only. In the other words, the spider-indexer software of Web search engines are not able to identify and index the metadata elements (tags) created based on microdata method and implemented in RDF (as the syntax of linked data method) and, the search engines cannot display the records in the search results semantically. So Web search engine are not adaptable to the standards of the semantic web.



واکنش موتورهای کاوش وب به پیشینه‌های فراداده‌ای **مبتنی** بر روش ترکیبی داده‌های خرد و روش داده‌های پیوندی |
سیدمهدی طاهری، رضا نیکزاد بهله، میترا صمیعی

Keywords: Indexing; Visibility; Metadata Records; Rich snippets; Linked data; Web search engines