

مقایسه عملکرد موتورهای کاوش وب در نمایه‌سازی و پیدانمایی پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد

سید مهدی طاهری^{۱*}، سارا ذوالقدر^۲، نجلا حریری^۳

مطالعات دانش‌شناسی

سال چهارم، شماره ۱۶، پاییز ۹۷، ص ۸۳ تا ۱۰۱

تاریخ دریافت: ۹۷/۰۷/۱۲

تاریخ پذیرش: ۹۷/۰۹/۱۵

چکیده

پژوهش حاضر با هدف تبیین عملکرد موتورهای کاوش وب در نمایه‌سازی و پیدانمایی پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد به صورت تطبیقی و با روش تجربی انجام شده. جامعه پژوهش را ۱۰۰ پیشینه فراداده‌ای (مربوط به کتاب‌های الکترونیکی) مبتنی بر زبان نشانه‌گذاری فرامتن (اچ‌تی‌ام‌ال) در قالب یک گروه آزمون، دارای ۵۰ پیشینه به زبان فارسی و ۵۰ پیشینه به زبان انگلیسی که بر اساس روش داده‌های خرد ایجاد شده بودند و یک گروه گواه، دارای یک‌صد پیشینه فارسی و انگلیسی با ساختار معمولی، تشکیل می‌دادند. این پیشینه‌ها با روش نمونه‌گیری تصادفی از میان پیشینه‌های فراداده‌ای کتابخانه دیجیتال دانشگاه علامه طباطبائی انتخاب گردیدند. پیشینه‌های فراداده‌ای مورد مطالعه بر روی یک وب‌سایت مستقل با نشانی <http://www.szolghadr.ir> منتشر و بر اساس بهترین روش‌های معرفی شده به موتورهای کاوش گوگل، یاهو و بینگ معرفی شدند. سپس با استفاده از سیاهه واری، عملکرد موتورهای کاوش مورد مطالعه در نمایه‌سازی و پیدانمایی پیشینه‌های فراداده‌ای هر دو گروه مورد بررسی قرار گرفت. یافته‌ها حاکی از آن است که موتور کاوش گوگل پیشینه‌های فراداده‌ای هر دو گروه آزمون و گواه را نمایه‌سازی و پیشینه‌های گروه آزمون را به صورت معنادار در نتایج جستجو پدیدار نمود. اما موتورهای کاوش یاهو و بینگ، پیشینه‌های هر دو گروه را فقط نمایه‌سازی نموده و قادر به پیدانمایی معنادار پیشینه‌های گروه آزمون نبودند. بنابراین با وجود اعلام موتورهای کاوش معتبر وب، به عنوان طراحان فرامتن تولید داده‌های ساختارمند (schema.org)، مبنی بر نمایه‌سازی و نمایش معنادار پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد، تنها موتور کاوش گوگل پیدانمایی معنادار آن‌ها (به صورت خرده‌های غنی) را پشتیبانی نمود.

واژگان کلیدی: پیدانمایی، پیشینه‌های فراداده‌ای، روش داده‌های خرد، موتورهای کاوش وب، نمایه‌سازی

۱. * استادیار گروه علم اطلاعات و دانش‌شناسی دانشگاه علامه طباطبائی، taherismster@gmail.com

۲. کارشناس ارشد علم اطلاعات و دانش‌شناسی، کتابخانه مرکزی و مرکز اسناد دانشگاه علامه طباطبائی

sara.zolghadr432@gmail.com

۳. استاد گروه علم اطلاعات و دانش‌شناسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران

NadjlaHariri@gmail.com

مقدمه

با گسترش روزافزون وب، به‌عنوان مهم‌ترین فناوری و رسانه شبکه اینترنت، این شبکه از رسانه‌ای صرفاً ارتباطی به رسانه‌ای اطلاعاتی تبدیل شده که از قابلیت‌های منحصر به فردی نسبت به محیط‌های اطلاعاتی سنتی برخوردار است. بسیاری از سازمان‌ها، ناشران، مراکز اطلاعاتی و پدیدآورندگان به انتشار و ارائه اطلاعات و خدمات خود بر روی این رسانه تمایل پیدا کرده‌اند. این تمایل که با رشد تصاعدی همراه است، به یک‌باره حجم وسیعی از اطلاعات را بر روی شبکه وب دسترس‌پذیر نموده است. بنابراین همانند سایر محیط‌های اطلاعاتی، طراحی و ایجاد ابزارها، فنون و استانداردهایی به‌منظور تسریع و تسهیل دسترسی به این حجم وسیع از اشیای محتوایی^۱ ضرورت دارد (طاهری، ۱۳۸۷). موتورهای کاوش که از مهم‌ترین ابزارهای نمایه‌سازی و بازیابی منابع وب هستند، کاستی‌ها و نقص‌های فراوانی دارند و ناکارآمدی آن‌ها در نظام‌های اطلاعاتی از وجوه گوناگون مشخص شده است. ناتوانی آن‌ها در نمایه‌سازی همه انواع اشیای محتوایی وب، ناپایداری و بی‌ثباتی آن‌ها، افت‌وخیز و تغییرات شدید نتایج بازیابی شده آن‌ها طی زمان و در کل، اثربخشی پایین آن‌ها در بازیابی منابع نمونه‌هایی از نقص و ناکارآمدی این ابزارها به‌عنوان نظام‌های بازیابی اطلاعات پیوسته هستند، به‌خوبی در پژوهش‌های گوناگون بیان شده‌اند. نمایه‌سازی اشیای محتوایی وب به‌وسیله موتورهای کاوش و استفاده نکردن از داده‌های توصیفی در فرایند نمایه‌سازی، از محدودیت‌های آشکار موتورهای کاوش به‌شمار می‌آید. بنابراین، اضافه کردن داده‌های توصیفی و ساختارمند فراتر از آنچه در محتوای اشیای وب موجود است، ضروری است. از سوی دیگر و به‌موازات گسترش ابزارهای کاوش وب، توسعه فراداده^۲ به‌عنوان ابزاری برای توصیف و سازمان‌دهی اشیای محتوایی وب انجام شده است (حریری و دیگران، ۱۳۹۲). فراداده، ابزار سازمان‌دهی دانش در محیط و رسانه‌های اطلاعاتی جدید، به‌ویژه شبکه وب است. استانداردهای فراداده‌ای^۳ که مجموعه‌ای از عناصر با ساختارهای

1. content object
2. metadata
3. metadata standards

معناشناختی^۱ خاص هستند به منظور توصیف، شناسایی، کشف، نگهداری و مدیریت اشیای محتوایی در محیط جدید گسترش یافته‌اند. استانداردهای فراداده‌ای که از کارکردهای شناسایی و کشف اشیای محتوایی پشتیبانی می‌کنند، فراداده توصیفی^۲ نام گرفته‌اند. سازمان‌دهی و توصیف اشیای محتوایی وب بر اساس این نوع فراداده انجام می‌گیرد. راهبرد برون‌رفت از مشکلات جستجو در سال ۱۹۹۶ توسط تیم برنرزلی، با همکاری دو تن دیگر از پژوهشگران با ارائه و معرفی یک زبان آسان برای توصیف اطلاعات در قالب صفحه‌های اچ‌تی‌ام‌ال^۳ و گسترش وب کنونی به وب معنایی^۴ ارائه شد (طاهری، ۱۳۹۲ الف). وب معنایی داده‌ها را از اشیای محتوایی مختلف گردآوری و یکپارچه‌سازی نموده و پردازش می‌نمایند. بدین منظور ممکن است یک نظام اطلاعاتی نیازمند تعامل با سایر نظام‌ها و به عبارت دیگر میانکنش‌پذیری^۵ بین نظام‌ها باشد. رسالت اصلی وب معنایی توانا تر کردن وب برای افزایش درک معنایی ماشین‌ها است به نحوی که عوامل نرم‌افزاری می‌توانند معنای داده‌های پراکنده در وب را درک کرده و استفاده از وب را برای انسان آسان‌تر کنند. در وب معنایی نرم‌افزارها قادرند به صورت جهانی، سازوکاری برای تبادل و به اشتراک گذاری داده‌ها میان گره‌های متصل به شبکه را ایجاد نمایند (نوروزی و طاهریان، ۱۳۹۱). فرمانای تولید داده‌های ساختارمند^۶ یکی از روش‌های ایجاد وب معنایی است که در سال‌های اخیر توسط طراحان موتورهای کاوش وب توسعه یافته است (دی‌مرس^۷، ۲۰۱۶). در این روش، با استفاده از نشانه‌گذاری معنایی^۸ یا ساختارمند کردن محتوای اشیای وب، به ویژه در بسترهای نحوی گوناگون، به وسیله شیوه‌هایی مانند افزودن ویژگی‌ها^۹ به نام برجسب‌ها، داده‌ها را برای موتورهای کاوش قابل فهم نموده، موجب پیدانمایی معنادار اشیای نمایه‌سازی شده در نتایج جستجو می‌شوند. اکنون فرمانای تولید داده‌های ساختارمند، استاندارد فراداده‌ای جدیدی

1. semantic structure
2. descriptive metadata
3. Hyper Text Markup Language (HTML)
4. semantic web
5. interoperability
6. schema.org
7. DeMers
8. semantic markup
9. attributes

است که پیشینه‌های فراداده‌ای ایجاد شده بر پایه آن، برای نرم‌افزارهای خزنده-نمایه‌ساز موتورهای کاوش وب قابل فهم و معنادار هستند (طاهری و دیگران، ۱۳۹۶؛ محمدی استانی و دیگران، ۱۳۹۷).

فرانمای مذکور با فراهم نمودن واژگانی متعدد متناسب با ویژگی‌های هر یک از موجودیت‌های داده‌ای (اشیای محتوایی) دسترس‌پذیر بر روی وب و با بافت اطلاعاتی که موجودیت‌های موردنظر در آن تولید و مورد استفاده قرار می‌گیرند، امکان ایجاد داده‌های ساختارمند (پیشینه‌های فراداده‌ای) بر اساس روش‌ها و بسترهای نحوی چند فراهم می‌نماید (<http://www.schema.org>). گروه کاری فناوری کاربردی فرامتن وب^۱ در آوریل ۲۰۰۷ پنجمین نسخه اچ‌تی‌ام‌ال ۲.۵ را به کنسرسیوم وب پیشنهاد داد. مجموعه‌ای کوچک از گسترش‌های اچ‌تی‌ام‌ال که برای معنابخشی به صفحات وب طراحی شده است. نشانه‌گذاری جدید برای ایجاد بستر معنانشاخی در اچ‌تی‌ام‌ال برای تقویت بعد معنایی داده‌های وب بیش از ویژگی‌های نمایشی و ارائه اطلاعات است. نسخه جدید اچ‌تی‌ام‌ال امکان ایجاد ویژگی‌های معنایی برای قابل فهم کردن محتوای اشیای وب را به روبات‌های موتورهای کاوش می‌دهد (<https://whatwg.org>).

عناصر فراداده‌ای فرانمای تولید داده‌های ساختارمند، خدمت جدیدی از موتورهای کاوش وب هستند که محتوای موجودیت‌های داده‌ای را به نحوی توصیف و بازنمایی می‌نمایند که کاربران هنگام مشاهده نتایج جستجو محتوای شیء را به صورت معنادار و در ارتباط با اشیای مرتبط درک خواهند کرد. اهمیت استفاده از این فرانما، تولید داده‌های ساختارمند و نمایش معنادار موجودیت‌های داده‌ای در نتایج جستجو است. به کمک این روش، رتبه وبسایت‌ها در نتایج نیز به میزان قابل توجهی بهبود می‌یابد. خرده‌های غنی (نمایش معنادار داده‌های ساختارمند تولید شده بر اساس فرانما) مربوط به موجودیت‌های داده‌ای وب، افزون بر اطلاعات متنی، اطلاعات اضافی مانند تصاویر، وزن‌دهی، اطلاعات پدیدآورندگان و جز آن را برای نمایش در نتایج جستجو پالایش و آماده می‌نمایند.

1. Web Hypertext Application Technology Working Group (WHATWG)
2. HTML5

به عبارت دیگر، از آنجا که خرده‌های غنی پیشنهادی‌های فراداده‌ای را به صورت ساختارمند توصیف می‌کنند، این ویژگی آن‌ها باعث شده که موتورهای کاوش این پیشنهادی‌های فراداده-ای^۱ را به صورتی معنادار نمایه‌سازی نموده و به سطح پیدانمایی^۲ برسانند (طاهری و دیگران، ۱۳۹۶). یکی از روش‌های متداول و کارآمد برای تولید داده‌های ساختارمند بر اساس فرامای schema.org، روش داده‌های خرد^۳ است (محمدی استانی و دیگران، ۱۳۹۷) که به کمک ویژگی‌های اچ‌تی‌ام‌ال نسخه ۵، محتوای اشیای محتوایی را نشانه‌گذاری نموده و برای موتورهای کاوش قابل درک می‌نماید (<http://www.schema.org>).

اکنون مسئله‌ای که مطرح می‌شود آن است که با وجود توصیه همه موتورهای کاوش وب به استفاده از فرامای تولید داده‌های ساختارمند (schema.org) برای تولید پیشنهادی‌های فراداده‌ای، به ویژه بر اساس روش داده‌های خرد به دلیل وابستگی به بستر نحوی زبان نشانه‌گذاری فرامتن (اچ‌تی‌ام‌ال)، به ویژه موتورهای کاوشی که این فرانما را طراحی نموده‌اند، آیا این موتورها به پیشنهادی‌های فراداده‌ای مبتنی بر فرامای یادشده و روش داده‌های خرد واکنش مثبت نشان داده و آن‌ها را به صورت معنادار نمایه‌سازی و در نتایج جستجو نمایش می‌دهند؟ آیا واکنش همه موتورهای کاوش به این نوع از پیشنهادی‌های فراداده‌ای یکسان است و همه آن‌ها داده‌های ساختارمند مبتنی بر روش داده‌های خرد را با عملکردی مشابه به صورت معنادار نمایه‌سازی و پدیدار می‌نمایند؟ پژوهش حاضر برای پاسخ به این مسئله طرح‌ریزی شده است. برای پاسخ به این مسئله، تحقق هدف ذیل مدنظر بوده است:

تبیین عملکرد موتورهای کاوش وب در نمایه‌سازی و پیدانمایی معنادار پیشنهادی‌های

فراداده‌ای مبتنی بر روش داده‌های خرد

همچنین این پژوهش دارای یک فرضیه بود:

موتور کاوش گوگل نسبت به دیگر موتورهای وب به پیشنهادی‌های فراداده‌ای مبتنی بر

داده‌های خرد واکنش مناسب‌تری نشان می‌دهد.

1. metadata records
2. visibility
3. microdata

فرانمای تولید داده‌های ساختارمند (schema.org) با هدف ایجاد فراداده برای انواع موجودیت‌های داده‌ای (اشیای محتوایی) شبکه وب متناسب با نیازهای محیط‌های داده‌ای گوناگون و نیازهای محلی فراهم‌کنندگان اطلاعات پیوسته گسترش یافته است. اقبال به ویژگی‌های بومی موجودیت‌های داده‌ای دسترس‌پذیر بر روی وب، توسعه واژگانی برای توصیف انواع متعدد موجودیت‌های موجود و پشتیبانی از کارکردهای مورد انتظار افراد حقیقی و حقوقی منتشرکننده داده‌ها، از مهم‌ترین قابلیت‌های این فرانما به شمار می‌آید. ارزیابی اثربخشی و ارائه راهکارهایی مناسب برای توسعه و بهبود فرانمای یادشده، اشتیاق پژوهشگران برای انجام مطالعات بر روی فرانما را برانگیخته است. اگرچه تعداد این پژوهش‌ها، به دلیل نو بودن فرانما (از سال ۲۰۱۱)، محدود است، اما تعداد آن‌ها رو به فزونی است و بخشی از اولویت‌های پژوهشی حوزه مدیریت داده‌های وب را به خود اختصاص داده است. به‌منظور مرور پژوهش‌های انجام شده و یا در حال اجرای مرتبط با فرانمای تولید داده‌های ساختارمند، به برخی از مهم‌ترین آن‌ها با رویکرد تحلیلی پرداخته خواهد شد.

بررسی گرایش و تمرکز موضوعی پژوهش‌های پیشتر صورت گرفته و در حال انجام مربوط به فرانما بیانگر آن است که پژوهش‌های این حوزه در سه گروه کلی قرار می‌گیرند. گروه نخست را پژوهش‌هایی تشکیل می‌دهند که به ارزیابی کارکردپذیری و کاربردپذیری واژگان‌های فرانما برای توصیف و پردازش موجودیت‌های داده‌ای در بافت‌های اطلاعاتی متفاوت پرداخته‌اند؛ گروه دیگر، طراحی و تدوین گسترش‌های فرانما در دو سطح واژگان جدید و افزودن صفاتی خاص به‌منظور توسعه دامنه کاربردی فرانما متناسب با موجودیت‌های داده‌ای وب را مدنظر قرار داده‌اند و در پایان، مطالعات مربوط به بررسی میانکنش‌پذیری پیشینه‌ها یا عناصر فراداده‌ای تولید شده بر مبنای فرانما تولید داده‌های ساختارمند با موتورهای کاوش وب برای درک میزان کارایی بهره‌گیری از فرانما برای سازمان‌دهی اشیای محتوایی وب هستند.

گروه نخست، شامل پژوهش‌های هاوسکی^۱ و دیگران (۲۰۱۳)، فریر^۲ و دیگران (۲۰۱۶)، نوگالس^۳ و دیگران (۲۰۱۶)، ویلاس^۴ و دیگران (۲۰۱۷)، بالکی^۵ و دیگران (۲۰۱۸) است که اثربخشی پیشینه‌های فراداده‌ای مبتنی بر فرانما را از دو بعد کارکردپذیری و کاربرپذیری ارزیابی نمودند. هاوسکی و دیگران (۲۰۱۳) با تمرکز بر روی سه رویکرد: بهینه‌سازی موتورهای کاوش، جاسازی فراداده‌های مبتنی بر فرانما در بستر داده‌های خرد و اشتراک داده‌های فراسوی درباره این که منابع اطلاعاتی چگونه مورد استفاده قرار می‌گیرند، بر کارایی و تأثیر مثبت فراداده‌های تولید شده بر اساس فرانما در بهبود رتبه موجودیت‌های داده‌ای مورد توصیف در نتایج جستجو، کاربرپذیری و افزایش کشف‌پذیری آن‌ها تأکید نمودند. فریر و همکارانش (۲۰۱۶) کارآمدی فراداده‌های مبتنی بر فرانمای تولید داده‌های ساختارمند به‌عنوان قالب فراداده‌ای مشترک برای گردآوری داده‌ها از مؤسسات میراث فرهنگی عضو شبکه یورویانا^۶ مورد بررسی قرار دادند. آن‌ها بیان داشتند، استفاده از فرانما هیچ مانع و محدودیتی برای تولید فراداده‌های مشترک توسط مؤسسات عضو، به‌منزله فراهم‌کنندگان داده‌ها، با کیفیت معناشناختی مطلوب و متناسب با ملزومات شبکه مذکور وجود ندارد. در ادامه پژوهش پیشین، والیس و همکاران (۲۰۱۷) توصیه‌های برای استفاده از فراداده‌های مبتنی بر فرانما به‌مثابه قالب فراداده‌ای مشترک برای گردآوری داده‌ها با هدف افزایش ربط و پیدانمایی اشیای فرهنگی شبکه یورویانا در نتایج موتورهای کاوش وب ارائه نمودند.

تحلیل همخوانی فرانمای تولید داده‌های ساختارمند با وب داده‌های پیوندی (معنایی) و ایجاد پیوند میان واژگان‌های هر دو استاندارد، موضوع دیگری بود که توسط نوگالس و دیگران (۲۰۱۶) انجام گرفت. یافته‌ها همخوانی مناسب و سطح بالایی را میان اصطلاحات (صفات) فرانما با واژگان‌های پیوندی باز^۷ نشان می‌دادند. پژوهش پایانی گروه نخست، به تحلیل کاربرد فرانمای تولید داده‌های ساختارمند در حوزه گردشگری بین سال‌های ۲۰۱۳ تا

1. Hawskey
2. Freire
3. Nogales
4. Willas
5. Balci
6. europeana network
7. Linked Open Vocabularies (LOV)

۲۰۱۶ اختصاص داشت. بالکی و دیگران (۲۰۱۸) نتیجه گرفتند، بهره‌مندی از فرانما در حوزه گردشگری، افزون بر رشد قابل توجه در طول سال‌های مورد مطالعه، به آمادگی مناسب موجودیت‌های داده‌ای این حوزه برای دسترس‌پذیری و ارائه خدمات به عوامل خودکار و هوشمند وب انجامیده است.

گروه دیگر، طراحی و تدوین واژگان‌های جدید یا افزودن صفاتی دیگر به واژگان‌های موجود برای افزایش کاربردپذیری فرانمای تولید داده‌های ساختارمند را پایه پژوهش‌های خود قرار داده‌اند. پژوهش‌های الدائج^۱ (۲۰۱۵)، آقاده (۱۳۹۷) و پژوهش‌های در حال انجام محمدی استانی (۱۳۹۷) و فردحسینی (۱۳۹۷) مربوط به این گروه هستند. الدائج (۲۰۱۵) در رساله دکترای خود و بر اساس طرح فراداده‌ای منابع یادگیری (ال آرام‌آی)^۲، پیشنهادهایی برای گسترش طرح‌واره منابع آموزشی فرانمای تولید داده‌های ساختارمند در قالب صفاتی چند با هدف بهبود توصیف منابع آموزشی دسترس‌پذیر در محیط‌های یادگیری مجازی به صورت معنایی توصیه کرد. پژوهش دیگری در ایران توسط آقاده (۱۳۹۷) اجرا شد. در این پژوهش، با شناسایی و طراحی چند صفت مرتبط با داده‌های مستند، چگونگی افزودن شکل‌های انتخاب‌نشده اصطلاحات نقاط دسترسی به پیشینه‌های فراداده‌ای مبتنی بر فرانما مورد بررسی قرار گرفت. یافته‌های پژوهش، افزایش جامعیت در بازیابی موتورهای کاوش و بهبود کنترل مستندات در محیط وب را بیان داشت.

در پژوهش دیگری که توسط محمدی استانی (۱۳۹۷) در حال انجام است، طرح‌واره توصیف نسخ خطی بافت اطلاعاتی ایرانی-اسلامی بر اساس فرانمای تولید داده‌های ساختارمند برای نخستین بار طراحی شد و ضمن تأیید اعتبارسنجی آن توسط ابزارهای سنجش اعتبار مرتبط که توسط موتورهای کاوش وب توسعه یافته‌اند، نمایه‌پذیری و پیدانمایی پیشینه‌های نمونه مبتنی بر این طرح‌واره مورد آزمایش قرار گرفت. نتایج پژوهش تا کنون، اثربخشی و کارایی طرح‌واره گسترش یافته را تأیید می‌کنند. همچنین فرد حسینی (۱۳۹۷) در پژوهش مرتبط با رساله دکترای خود، تأثیر استانداردها و الگوهای مفهومی بافت

1. Aldaej
2. Learning Resource Metadata Initiative (LRMI)

کتابخانه‌ای، مانند استاندارد توصیف و دسترسی به منبع (آردی‌ای) و الگوی مرجع کتابخانه‌ای (ال‌آرام)، بر بهبود عملکرد فرآیند تولید داده‌های ساختارمند را مورد بررسی قرار داد. یافته‌های پژوهش وی تا زمان تدوین گزارش حاضر بیانگر آن است که طرح‌واره‌های آثار خلاقانه فرآیند تولید داده‌های ساختارمند، برای پشتیبانی از کارکردهای مدنظر بافت کتابخانه‌ای از فراداده و افزایش کاربردپذیری این طرح‌واره‌ها در بافت یاد شده، نیاز به افزودن صفات پرشمار طرح‌واره‌های کنونی فرانما و حتی توسعه طرح‌واره‌های جدید کاملاً محسوس است.

گروه دو دیگر، پژوهش‌هایی هستند که میانکنش‌پذیری پیشینه‌های فراداده‌های ایجاد شده بر پایه فرآیند تولید داده‌های ساختارمند با موتورهای کاوش وب را مورد مطالعه قرار دادند. به عبارت دیگر، هدف این پژوهش‌ها، تبیین واکنش نرم‌افزارهای خزنده-نمایه‌ساز موتورهای کاوش وب به فراداده‌های مبتنی بر فرانما به منظور آگاهی از اثربخشی آن‌ها در افزایش و بهبود دسترس‌پذیری اشیای محتوایی مورد توصیف است. اگرچه در پژوهش‌های آقاده (۱۳۹۷)، محمدی استانی (۱۳۹۷) و فرد حسینی (۱۳۹۷) نیز به منظور اعتبارسنجی و تصریح میزان کارکردپذیری گسترش‌های صورت گرفته بر روی فرانما، واکنش موتورهای کاوش به پیشینه‌های فراداده‌ای نمونه مطالعه شد، اما پژوهش طاهری و دیگران (۱۳۹۶) به صورت ویژه و متمرکز بر روی حوزه میانکنش‌پذیری فراداده‌های مبتنی بر فرانما با موتورهای کاوش وب صورت گرفت. پژوهش یاد شده اذعان داشت، تولید پیشینه‌های فراداده‌ای بر اساس فرآیند تولید داده‌های ساختارمند و پیاده‌سازی آن‌ها در روش داده‌های خرد، به عنوان یکی از مهم‌ترین بسترهای نحوی فرانما و بازنمایی این پیشینه‌ها در قالب چارچوب توصیف منبع (آردی‌اف)، به عنوان یکی از استانداردهای روش داده‌های پیوندی، اگرچه از ویژگی نمایه‌پذیری تمامی ارزش‌های عناصر فراداده‌ای برخوردار است، اما به دلیل عدم شناسایی و تشخیص صفات (عناصر) پیشینه‌های فراداده‌ای مورد بررسی توسط روبات

موتورهای کاوش، پیشینه‌های مذکور به صورت معنادار در نتایج بازیابی پدیدار نمی‌شوند. اگرچه به نظر می‌رسد، توسعه بستر نحوی جی‌سان‌ای دی به حل مشکل یاد شده بیانجامد^۱. همان‌طور که مرور پیشینه‌های انجام شده بر روی فرمانای تولید داده‌های ساختارمند نشان می‌دهد، پژوهش‌های این حوزه کوشش نموده‌اند، تمامی جنبه‌های مرتبط با فرمانای یاد شده را برای آگاهی از و بهبود عملکرد فرمانا در توصیف و پردازش معنایی موجودیت‌های داده‌ای وب مورد مطالعه قرار دهند. ارزیابی کارکردپذیری و کاربردپذیری طرح‌واره‌های فرمانا در محیط‌ها و بافت‌های اطلاعاتی گوناگون که در پژوهش‌های گروه نخست مرور شدند، نشانگر کارآمدی و قابل استفاده بودن آن‌ها در توصیف و بازنمایی اشیای محتوایی و بهبود پردازش این اشیاء توسط نرم‌افزارهای روبات موتورهای کاوش و عوامل هوشمند به صورت معنادار بود. با وجود این کارآمدی و اثربخشی، پژوهش‌های گروه دیگر ابراز داشتند که برای افزایش کاربردپذیری طرح‌واره‌های فرمانا و سازگاری بیشتر آن‌ها با ویژگی‌های بومی و نیازهای محلی بافت‌های داده‌ای متفاوت، نیاز به گسترش طرح‌واره‌های جدید و یا غنی‌سازی صفات طرح‌واره‌های موجود است. همچنین کنترل مستندات در محیط فراداده‌های مبتنی بر فرمانا، ضروری و امکان‌پذیر است. بعلاوه پژوهش‌های گروه دیگر که به میانکنش‌پذیری فراداده‌های ایجاد شده بر اساس فرمانا با نرم‌افزارهای عنکبوت-نمایه‌سازی که به منظور سنجش اعتبار گسترش‌های انجام شده، پرداختند، پژوهش گروه دو دیگر، کاملاً متمرکز بر روی مسئله میانکنش‌پذیری شد و عدم همخوانی فرمانا با استانداردهای روش داده‌های پیوندی، به منزله مهم‌ترین روش تحقق وب معنایی را اعلام نمود. چنان‌که گفته شد، بستر نحوی جی‌سان‌ال‌دی برای رفع مشکل مذکور توسعه یافته است. با وجود بررسی جنبه‌ها و جریان‌های مرتبط با فرمانای تولید داده‌های ساختارمند پژوهش‌های پیشین و با توجه به این‌که طراحی و تدوین فرمانا توسط چهار موتور کاوش پرکاربرد وب یعنی گوگل، بینگ، یاهو و یاندکس صورت گرفته است، اما جای خالی

۱. برای اطلاعات بیشتر در خصوص میانکنش‌پذیری نظام‌های فراداده‌ای با موتورهای کاوش وب، به منبع ذیل مراجعه کنید: حریری، نجلا؛ طاهری، سید مهدی؛ فتاحی، رحمت‌الله؛ نوشین فرد، فاطمه (۱۳۹۲). میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب: چالش‌ها و رویکردهای جاری. پژوهشنامه، سال ۳، شماره ۲.

پژوهشی که به مطالعه عملکرد موتورهای کاوش به‌ویژه طراحان فرانما، آن‌هم به‌طور مقایسه‌ای و تطبیقی، در نمایه‌سازی و پیدانمایی معنایی موجودیت‌های داده‌ای وب اختصاص یابد، محسوس است. نتایج چنین پژوهش می‌تواند میزان سازگاری با و پشتیبانی موتورهای کاوش وب از پیشینه‌های مبتنی بر فرانما را تبیین نماید.

روش

پژوهش حاضر از نظر نوع پژوهش، یک پژوهش کاربردی بود، چراکه به تبیین عملکرد موتورهای کاوش وب در نمایه‌سازی و پیدانمایی معنادار پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد در محیط شبکه وب در شرایط واقعی و عملی می‌پردازد. و از لحاظ روش، از روش تجربی برای رسیدن به اهداف پژوهش استفاده نمود. جامعه پژوهش حاضر شامل ۲۰۰ پیشینه فراداده‌ای (مربوط به کتاب‌های الکترونیکی) به زبان‌های فارسی و انگلیسی (هر زبان ۱۰۰ پیشینه) در قالب دو گروه آزمون، دارای ۱۰۰ پیشینه فراداده‌ای مبتنی بر روش داده‌های خرد و گروه گواه، دارای ۱۰۰ پیشینه با ساختار و برچسب‌های معمولی اچ‌تی‌ام‌ال تشکیل می‌دادند که بر روی وب‌سایتی مستقل با نشانی <http://szolghadr.ir> انتشار یافتند. پیشینه‌های فراداده‌ای یاد شده از مجموعه اشیای الکترونیکی کتابخانه دیجیتال دانشگاه علامه طباطبائی^۱، به‌صورت تصادفی و از حوزه‌های موضوعی علم اطلاعات و علم رایانه انتخاب گردیدند. با توجه به این که پژوهش حاضر قصد داشت عملکرد موتورهای کاوش وب در نمایه‌سازی و پیدانمایی معنادار پیشینه‌های فراداده‌ای را مورد مطالعه قرار دهد، بدیهی است باید بهترین موتورهای کاوش وب که خود در طراحی فرانمای تولید داده‌های ساختارمند مشارکت داشته‌اند و می‌توانستند اهداف این پژوهش را محقق سازند، انتخاب می‌شدند.

بر اساس منابع معتبر وب و از دیدگاه کاربران که به سطح تواتر رسیده است و رتبه موتورهای کاوش در پایگاه‌هایی که به رتبه‌بندی آن‌ها می‌پردازند، موتورهای کاوش گوگل، یاهو و بینگ پرکاربردترین و مهم‌ترین موتورهای کاوش وب به شمار می‌آیند که در زمره

طراحان اصلی فرانمای تولید داده‌های ساختارمند نیز به شمار می‌آیند. به همین دلیل، هر سه موتور مذکور برای نیل به اهداف پژوهش برگزیده شدند. لازم به ذکر است، اگرچه موتورهای کاوش یاهو و بینگ با یکدیگر ائتلافی ایجاد نموده‌اند و روبات موتور کاوش یاهو حذف شده است (<https://searchengineland.com/library/features/microsoft-yahoo-merger>) و نتایج هر دو موتور کاوش در بیشتر موارد یکسان است، اما رتبه‌بندی نتایج و نیز نمایش معنادار آن‌ها که مهم‌ترین ویژگی برای بررسی خرده‌های غنی است، با یکدیگر متفاوت است و هر یک دارای خط‌مشی‌های ویژه‌ای هستند. پس از فرایند انتخاب، مطالعه و بررسی ساختار، روش‌های نمایه‌سازی و پیدانمایی موتورهای کاوش، شیوه‌های تعامل آن‌ها با اشیای محتوایی وب، میانکنش‌پذیری با پیشینه‌های فراداده‌ای به‌ویژه پیشینه‌های مبتنی بر فرانمای تولید داده‌های ساختارمند، از سوی سه موتور در این زمینه مدنظر قرار گرفت.

روش گردآوری داده‌ها در این پژوهش از طریق مشاهده طراحی شده^۱ تماشای دقیق به شیوه‌ای علمی و نظام‌مند با استفاده از راهبردهای از پیش تعیین شده^۲ که اغلب در مطالعه‌های کمی برای توصیف‌های نظام‌مند یا آزمون فرضیه‌های علمی به کار می‌رود - انجام گرفته است. در این پژوهش این مقوله‌ها عبارت‌اند از: نمایه‌سازی و پیدانمایی معنادار پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و با ساختار و برچسب‌ها و فرابرجسب‌های معمولی توسط موتورهای کاوش. چهارچوب مرجع مشاهده‌گر واکنش نرم‌افزارهای نمایه‌سازی موتورهای کاوش و الگوریتم‌های و خط‌مشی‌های نمایه‌سازی و پیدانمایی آن‌ها نسبت به پیشینه‌های فراداده‌ای منتشر شده (جامعه پژوهش) است. لازم به ذکر است برای اعتبارسنجی و صحت پیشینه‌های فراداده‌ای عضو جامعه پژوهش، از ابزار آزمون داده‌های ساختارمند^۳ گوگل استفاده شد. ابزار گردآوری داده‌ها، سیاهه واریسی بود که با استفاده از نرم‌افزار آماری اس‌پی‌اس‌اس - به‌عنوان ابزار ثبت مکانیکی، بر اساس پرسش‌های پژوهش طراحی شده بود.

1. structured observation

۲. در این پژوهش منظور راهبردهای جستجوی (Search strategies) تعیین شده بر اساس قوانین و توصیه‌های ارائه شده از سوی موتورهای کاوش گوگل، بینگ و یاهو است.

3. google structured data testing tools

این سیاهه در معرض قضاوت و نظرخواهی متخصصان در این حوزه بازیابی اطلاعات در موتورهای کاوش وب قرار گرفته و روایی آن تأیید شد و مؤلفه‌های آن وزن گرفتند. سپس داده‌های گردآوری‌شده در اسفند ۱۳۹۶ در قالب جداول و نمودارهایی با استفاده از نرم‌افزار تحلیل آماری مورد استفاده، تهیه گردیدند.

```
<div itemscope itemtype="http://schema.org/Book">
<d1>
<dt>عنوان:</dt>
<dd itemprop="name">فرا داده‌ها، موتورهای کاوش وب و میانکنش پذیری آن‌ها</dd>
<dt>مؤلف</dt>
<dd itemprop="author">سید مهدی طاهری، نجلا حریری و رحمت‌الله فتاحی</dd>
<dt>توصیفگر</dt>
<dd itemprop="description">میانکنش پذیری نظام‌های فراداده‌ای</dd>
<dt>تعداد صفحه</dt>
<dd><span itemprop="numberOfPages">۲۰۱</span>P.</dd>
<dt>شابک</dt>
<dd itemprop="ISBN">9786002410948</dd>
<dt>محل نشر</dt>
<dd itemprop="location" itemscope
temtype="http://schema.org/Organization"><span itemprop="name">:
تهران</span></dd>
<dt>ناشر</dt>
<dd itemprop="publisher" itemscope
itemtype="http://schema.org/Organization" itemscope=""><span
itemprop="name">کتابدار</span></dd>
<dt>سال نشر</dt>
<dd itemprop="datepublished">۱۳۹۴</dd>
<dt>زبان</dt>
<dd itemprop="inLanguage">فارسی</dd>
<dt>ژانر</dt>
<dd itemprop="genre">کتاب</dd>
```

<dt>صاحب محتوا</dt>

<dd itemprop="copyrightholder">کتابخانه دیجیتال دانشگاه علامه طباطبائی</dd>

</dl></div>

نمونه‌ای از یک پیشینه فراداده‌ای مبتنی بر داده‌های خرد

یافته‌ها

هدف اصلی پژوهش: تبیین عملکرد موتورهای کاوش وب در نمایه‌سازی و پیدانمایی معنادار پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد.

جدول ۱. وضعیت کلی عملکرد موتورهای کاوش در نمایه‌سازی و پیدانمایی معنادار پیشینه‌های فراداده‌ای گروه‌های آزمون و گواه

موتور کاوش	نمایه‌سازی				پیدانمایی			
	گروه آزمون		گروه گواه		گروه آزمون		گروه گواه	
	بله	خیر	بله	خیر	بله	خیر	بله	خیر
گوگل	۱۰۰	-	۱۰۰	-	۱۰۰	-	-	۱۰۰
بینگ	۱۰۰	-	۱۰۰	-	۱۰۰	-	۱۰۰	-
ياهو	۱۰۰	-	۱۰۰	-	۱۰۰	-	۱۰۰	-

همان‌گونه که داده‌های مندرج در جدول ۱ نشان می‌دهند، ۱۰۰٪ پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد مربوط به گروه آزمون به‌طور کامل توسط موتورهای کاوش گوگل، یاهو و بینگ نمایه‌سازی شده‌اند. این وضعیت در مورد پیشینه‌ها مربوط به گروه گواه نیز یکسان بود و پیشینه‌های این گروه نیز به‌طور کامل توسط موتورهای کاوش مورد مطالعه نمایه‌سازی شده‌اند. همچنین داده‌های گردآوری‌شده بیانگر آن هستند که موتورهای کاوش سه‌گانه قادر به نمایش معنادار پیشینه‌های فراداده‌ای گروه گواه نبودند. این مسئله به دلیل ساختارمند نبودن پیشینه‌های یاد شده بر اساس فرآیند تولید داده‌های ساختارمند است. اما در خصوص پیدانمایی (نمایش معنادار) پیشینه‌های فراداده‌ای گروه آزمون وضعیت تا حدودی متفاوت بود. از میان سه موتور کاوش مورد بررسی، فقط موتور کاوش گوگل عملکرد مثبتی به پیدانمایی معنادار پیشینه‌های گروه آزمون نشان داده است و توانسته این

پیشینه‌ها را به صورت خرده‌های غنی به نمایش درآورد. درحالی که دو موتور کاوش دیگر، فاقد نمایش معنادار پیشینه‌های گروه آزمون در نتایج جستجوی خود بودند. آزمون فرضیه پژوهش: موتور کاوش گوگل نسبت به دیگر موتورهای وب به پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد واکنش مناسب‌تری نشان می‌دهد.

جدول ۲. نتیجه آزمون فرضیه پژوهش

موتور کاوش	نمایه‌سازی	پیدانمایی	خی دو	درجه آزادی	P	نتیجه آزمون
گوگل	٪۱۰۰	٪۱۰۰				
بینگ	٪۱۰۰	عدم پیدانمایی	۲۹۹	۲	۰/۰۰۰۱	رد فرض صفر
ياهو	٪۱۰۰	عدم پیدانمایی				

جدول ۲ بیانگر نتیجه آزمون تنها فرضیه پژوهش است. بر اساس نتایج آزمون کروسکال والیس و سطح معناداری به دست آمده (۰/۰۰۰۱) که از مقدار بحرانی آن یعنی ۰/۰۵ کوچک‌تر است می‌توان این‌گونه نتیجه گرفت که بین نمره پیدانمایی (نمایش معنادار) پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد مربوط به گروه آزمون و پیشینه‌های فراداده‌ای گروه گواه توسط سه موتور کاوش گوگل، بینگ و یاهو تفاوت معناداری وجود داشته است و موتور کاوش گوگل با ۱۰۰ درصد پیدانمایی پیشینه‌های گروه آزمون به‌طور معناداری وضعیت بهتری نسبت به دو موتور کاوش دیگر داشته است. بنابراین فرضیه صفر رد شده و فرضیه خلاف یا همان فرضیه پژوهش تأیید می‌گردد. این بدان معناست که موتور کاوش گوگل نسبت به دیگر موتورهای کاوش وب موردبررسی عملکرد مناسب‌تری در خصوص نمایش معنادار پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد داشته است.

بحث و نتیجه‌گیری

بهره‌گیری از فرآینمای تولید داده‌های ساختارمند برای ایجاد پیشینه‌های فراداده‌ای موجب درک بهتر نرم‌افزارهای خزنده-نمایه‌ساز موتورهای کاوش وب از موجودیت‌های داده‌ای شده و زمینه را برای نمایه‌سازی و پیدانمایی معنادار این موجودیت‌ها فراهم می‌نماید. بر اساس این فرآینما، ساختاری بر روی فراداده‌های هر یک از موجودیت‌های داده‌ای (اشیای

محتوایی) وب اعمال می‌گردد که قابل‌شناسایی معنادار عناصر فراداده‌ای برای موتورهای کاوش شده و این قابلیت، رفتار و واکنش متفاوتی از سوی موتورهای کاوش نسبت به موجودیت‌های داده‌ای در مقایسه با داده‌های معمولی (بدون استفاده از فرامای تولید داده‌های ساختارمند) را در پی خواهد داشت (طاهری و دیگران، ۱۳۹۶؛ محمدی استانی و دیگران، ۱۳۹۷).

همان‌طور که داده گردآوری‌شده در فرایند اجرای پژوهش و نتیجه آزمون فرضیه پژوهش نشان می‌دهند، موتور کاوش گوگل نسبت به موتورهای کاوش بینگ و یاهو عملکرد مناسب‌تری در پیدانمایی پیشینه‌های فراداده‌ای گروه آزمون که بر اساس روش داده‌های خرد، یکی از روش‌های اصلی تولید داده‌های ساختارمند بر اساس فرامای [schema.org](http://www.schema.org) داشته است. به عبارت دیگر، هر سه موتور کاوش مورد مطالعه در رویارویی با پیشینه‌های فراداده‌ای گروه آزمون واکنش یکسان و مثبتی در نمایه‌سازی آن‌ها داشته‌اند، اما فقط یکی از آن‌ها (گوگل) توانسته آن‌ها را به صورت معنادار نیز در نتایج جستجوی خود پدیدار نماید. این در حالی است که هر سه موتور کاوش در گروه طراحان فرامای تولید داده‌های ساختارمند قرار دارند و بر پشتیبانی کامل از فرایندهای نمایه‌سازی و پیدانمایی معنادار پیشینه‌های تولید شده بر اساس فرامای مذکور اذعان داشته‌اند (<http://www.schema.org>). بنابراین با وجود این که هر سه طراح فرامای تولید داده‌های ساختارمند، در نمایه‌سازی فراداده‌های مبتنی بر فرانما خط‌مشی تقریباً یکسانی دارند، اما در پیدانمایی، گوگل همانند بسیاری دیگر از زمینه‌ها، توانمندی بیشتری از خود نشان داد.

در مقایسه با دیگر پژوهش‌های انجام شده، به ویژه پژوهش‌های گروه‌های دیگر و دو دیگر، موتور کاوش گوگل واکنش مناسب‌تری نسبت به دیگر موتورهای کاوش طراح فرانما نشان داده است. مقایسه یافته‌های پژوهش، یعنی نمایه‌سازی پیشینه‌ها توسط هر سه موتور کاوش گوگل، یاهو و بینگ و پیدانمایی معنادار آن توسط گوگل، با یافته‌های پژوهش‌های گروه نخست بیانگر قابلیت فرانما برای توصیف انواع موجودیت‌های داده‌ای بافت‌های اطلاعاتی گوناگون و تسهیل پردازش معنایی آن‌ها توسط ابزارهای گوناگون از

جمله عوامل هوشمند و روبات‌های موتورهای کاوش وب است. در مقایسه با پژوهش‌های گروه دیگر نیز واکنش موتورهای کاوش به پیشینه‌های فراداده‌ای منتشر شده یکسان بوده است. بدین معنی که نتایج پژوهش حاضر، نمایه‌پذیر شدن ارزش‌های تمامی عناصر فراداده‌ای توسط نرم‌افزارهای نمایه‌ساز موتورهای کاوش را نشان داده، بر بهبود میانگین پذیرش نظام‌های فراداده‌ای با محبوب‌ترین ابزارهای کاوش وب تأکید می‌کند. بدین ترتیب، موتورهای کاوش وب نقش یک دروازه اطلاعاتی را برای محیط‌های اطلاعاتی مختلف ایفا نموده و کاربران را به سوی این مجموعه‌های ارزشمند هدایت می‌کنند.

یافته‌های پژوهش طاهری و دیگران (۱۳۹۶) که به نمایه‌شدن تمامی پیشینه‌های فراداده‌ای جامعه پژوهش توسط دو موتور کاوش گوگل و بینگ اذعان داشت، از بعد پیدانمایی نتیجه متفاوتی با این پژوهش را می‌نمایاند. استفاده از رویکرد ترکیبی برای ایجاد پیشینه‌های فراداده‌ای در راستای بهره‌مندی حداکثری از مزایا و قابلیت‌های هر دو روش داده‌های خرد و داده‌های پیوندی، بر پیدانمایی معنادار آن‌ها در فهرست نتایج جستجو تأثیر منفی گذاشته بود. این به دلیل عدم سازگاری بسترهای نحوی فرامای تولید داده‌های ساختارمند با و فقدان توجه کافی به استانداردهای روش داده‌های پیوندی، به‌عنوان روش اصلی تحقق وب معنایی که از کنسرسیوم وب جهانی توصیه شده، بوده است. در اینجا مجدداً تأکید می‌گردد، طراحی بستر نحوی جی‌سان‌ال‌دی برای حل مسئله مذکور بوده است. در حالی که یافته‌های پژوهش حاضر حاکی از آن است که این حداقل موتور کاوش گوگل در پیدانمایی معنایی پیشینه‌های موردبررسی موفق عمل نموده است. از این رو، تقویت شیوه‌های پیدانمایی و سازگاری بیشتر با بسترهای نحوی توصیه شده برای پیاده‌سازی فرامای تولید داده‌های ساختارمند که توسط خود موتورهای کاوش طراحی و توصیه شده، در دو موتور بینگ و یاهو ضروری به نظر می‌رسد. مقایسه واکنش موتورهای کاوش مختلف نسبت به بسترهای نحوی گوناگون توصیه شده از سوی فرانما، به‌ویژه از بعد پیدانمایی می‌تواند زمینه‌ای برای پژوهش‌های آینده باشد.

منابع

- آقاده، سمیرا (۱۳۹۷). طراحی طرح‌واره داده‌های مستند مبتنی بر روش داده‌های خرد و بررسی واکنش موتورهای کاوش وب به پیشینه‌های مبتنی بر آن. پایان‌نامه کارشناسی ارشد. دانشکده روانشناسی و علوم تربیتی. دانشگاه علامه طباطبائی.
- حریری، نجلا؛ طاهری، سید مهدی؛ فتاحی، رحمت‌الله؛ نوشین‌فرد، فاطمه (۱۳۹۲). میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب: چالش‌ها و رویکردهای جاری، پژوهشنامه، سال ۳، شماره ۲.
- طاهری، سید مهدی (۱۳۹۲ الف). داده‌های پیوندی: مفاهیم و استانداردها: با تأکید بر بافت میراث فرهنگی (۱). نشریه تخصصی کنسرسیوم محتوای ملی، شماره ۱۲.
- طاهری، سید مهدی؛ نیکزاد بهله، رضا؛ صمیعی، میترا (۱۳۹۶). بررسی واکنش موتورهای کاوش وب به پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی داده‌های خرد و روش داده‌های پیوندی. پژوهشنامه پردازش و مدیریت اطلاعات، ۳۳، ۲: ۶۳۹-۶۵۸.
- فرد حسینی، مهسا (۱۳۹۷). بررسی تأثیر استانداردها و الگوهای مفهومی بافت کتابخانه‌ای بر بهبود کارکردپذیری خرده‌های غنی. رساله دکترا. دانشکده علوم انسانی. دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران (پژوهش در حال انجام).
- محمدی استانی، مرتضی؛ چشمه سهرابی، مظفر؛ شعبانی، احمد؛ عاصمی، عاطفه؛ طاهری، سید مهدی (۱۳۹۷). تبیین فرآیند تولید داده‌های ساختارمند (اسکیما ارگ) و تحلیل رویکرد آن به پردازش و سازمان‌دهی اشیاء محتوایی وب. پژوهشنامه پردازش و مدیریت اطلاعات (زودآیند).
- محمدی استانی، مرتضی (۱۳۹۷). طراحی طرح‌واره قالب داده‌های خرد نسخه‌های خطی اسلامی-ایرانی و واکنش موتورهای کاوش وب به پیشینه‌های مبتنی بر آن. رساله دکترا. دانشکده روانشناسی و علوم تربیتی. دانشگاه اصفهان.
- نوروزی، مرتضی و طاهریان، محسن (۱۳۹۱). اصول و مبانی وب معنایی. تهران: فرس.
- Aldaej, A.A. (2015). *An Enhanced semantic VLE based on Schema.org and social media*. Submitted for the degree of Doctor of Philosophy, University of Surrey.

- Balci, B.T.; Simsek, U.; Karle, E.; Fensel, D. (2018). *Analysis of Schema.org Usage in the Tourism Domain*. In e-Review of Tourism Research (eRTR), Research Notes in ENTER 2018 Conference, Vol. 9.
- DeMers, Jayson (2016). You're Guide to Structured Snippets for SEO. Retrieved 05 Feb. 2019 from: <https://searchenginewatch.com/sew/how-to/2382954/your-guide-to-structured-snippets-for-seo>.
- Freire, Nuno; Charles, Valentine; Isaac; Antoine (2016). "Evaluation of Schema.org for Aggregation of Cultural Heritage Metadata". Retrieved 07 Feb. 2019 from: <https://www.inesc-id.pt/publications/13560/pdf>.
- Hawkey, M.; Barker, P.; Compbell, L.M. (2013). *New approaches to describing and discovering open educational resources*. In Proceedings of OER 13: Creating a Virtuous Circle. Nottingham, England. Retrieved 12 Feb. 2019 from: http://publications.cetis.org.uk/wp-content/uploads/2013/04/OER13_resourcediscovery.pdf
- Nogales, A.; Sicilia, M.A.; Sanchez-Alonso, S.; Garcia-Barriocanal, E. (2016). Linking from Schema.org Microdata to the web of Linked Data: An empirical assessment. *Computer Standards & Interfaces*, 45: 90-99.
- Wallis, R.; Isaac, A.; Charles, V.; Manguinhas, H. (2017). Recommendations for the application of Schema.org to aggregated Cultural Heritage metadata to increase relevance and visibility to search engines: the case of Europeana. *Code4Lib Journal*, 36. Retrieved 07 Feb. 2019 from: <http://journal.code4lib.org/articles/12330#appendix1>