

Designing Authority Data Properties Based on Microdata Method and Study of Web Search Engines' Reaction to Them

Seyed Mahdi Taheri^{1✉} | Samira Aghadeh Dehzadeh² | MohammadReza Taghva³ | Monireh Qasemi⁴ | Negin Shokrzadeh Hashtroudi⁵

1. Corresponding author, Associate Professor, Department of Knowledge and Information Science, Faculty of Psychology and Education, Allameh Tabataba'i University, Tehran, Iran. E-mail: taherism@atu.ac.ir
2. MA, Department of Knowledge and Information Science, Faculty of Psychology and Education, Allameh Tabataba'i University, Tehran, Iran. E-mail: aghadeh2.samira@gmail.com
3. Associate Professor, Department of Industrial Management, Faculty of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran. E-mail: taghva@atu.ac.ir
4. MA, Central Library and Documentation Center, Allameh Tabataba'i University, Tehran, Iran. E-mail: ghasemi.mo87@atu.ac.ir
5. Ph.D. Candidate, Department of Information Science, Faculty of Education and Psychology, Alzahra University, Tehran, Iran. E-mail: n.shokrzadeh@alzahra.ac.ir

Article Info

Article type:

Research Article

Article history:

Received 19 October 2022

Received in revised form 24
November 2022

Accepted 28 December 2022

Published online 05 January
2023

Keywords:

Authority data extension,
Metadata records,
Microdata syntax,
Schema.org,
Web search engines

ABSTRACT

Objective: The purpose of this research was to study the Search Engine's responses to authority data properties embedded into schema.org-based metadata on the Microdata syntax.

Methods: The experimental method was used in this research. The research population comprised 400 records of authority metadata based on the Microdata method from the digital library of Allameh Tabataba'i University. The examination group consisted of 200 metadata records, 100 records with authority data extensions embedded into schema.org-based metadata in the Microdata syntax and 100 other similar records in the JSON-LD syntax (50 samples of name authority, and 50 other subject authority) And the control group consisted of 200 Records, including 100 Records related to the description of the book in the Microdata syntax and 100 other similar records in the JSON-LD syntax. The records have been published on the independent website at www.Aghadeh.ir and have been introduced to the Google, Bing, Yahoo, and Yandex search engines as designers of the schema.org standard. Then, through searching the search engines, using the data gathering tool, the checklist provided by the researchers, the indexing and retrieval of the metadata records of the control groups and experimental groups were evaluated in the search results of the selected search engines.

Results: The results of this study showed that search engines were able to index and retrieve all of the metadata records and values of added extensions associated with authority data. Such a possibility had the same status for the name authority records and the subject authority data.

Conclusions: By retrieving each of the variant properties' values of examination group's records, in addition to the authorized values of the name and subject terms, a suitable platform for the comprehensiveness of the retrieve process, and the authority control in the Web search tools will be improved.

Cite this article: Taheri, S. M., Aghadeh, S., Taghva, M., Qasemi, M., & Shokrzadeh Hashtroudi, N. (2022). Designing Authority Data Properties Based on Microdata Method and Study of Web Search Engines' Reaction to them. *Academic Librarianship and Information Research*, 56 (4), 69-88. <https://doi.org/10.22059/jlib.2023.356370.1675>



طراحی صفات داده‌ای مستند مبتنی بر روش داده‌های خرد و بررسی واکنش موتورهای کاوش وب به پیشنهادها مبتنی بر آن

سیدمهدی طاهری^۱ ✉ | سمیرا آقاده‌زاده^۲ | محمدرضا تقوا^۳ | منیره قاسمی^۴ | نگین شکرزاده هشترودی^۵

۱. نویسنده مسئول، دانشیار، گروه علم اطلاعات و دانش‌شناسی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: taherism@atu.ac.ir

۲. کارشناس ارشد، گروه علم اطلاعات و دانش‌شناسی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه:

aghadeh2.samira@gmail.com

۳. دانشیار، گروه مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: taghva@atu.ac.ir

۴. کارشناس ارشد، کتابخانه مرکزی و مرکز اسناد دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: ghasemi.mo87@atu.ac.ir

۵. دانشجوی دکتری، گروه علم اطلاعات و دانش‌شناسی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه الزهراء، تهران، ایران. رایانامه: n.shokrzadeh@alzahra.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	هدف: پژوهش حاضر با هدف طراحی صفات داده‌های مستند مبتنی بر روش داده‌های خرد و تبیین واکنش موتورهای کاوش وب به پیشنهادها مبتنی بر آن انجام شد.
تاریخ دریافت: ۱۴۰۱/۰۷/۲۷	روش پژوهش: این پژوهش به روش تجربی انجام شد. جامعه پژوهش شامل ۴۰۰ پیشنهاد فراداده‌ای مستند مبتنی بر روش داده‌های خرد و بستر نحوی جی‌سان-ال‌دی. از کتابخانه دیجیتالی دانشگاه علامه طباطبائی بود. گروه گواه را ۲۰۰ پیشنهاد، شامل ۱۰۰ پیشنهاد مربوط به توصیف کتاب در قالب داده‌های خرد و ۱۰۰ پیشنهاد مشابه دیگر در قالب جی‌سان-ال‌دی. تشکیل می‌دادند، و گروه آزمون را ۲۰۰ پیشنهاد، دارای ۱۰۰ پیشنهاد با گسترش‌های داده‌های مستند در قالب داده‌های خرد و ۱۰۰ پیشنهاد مشابه دیگر در قالب جی‌سان-ال‌دی. (که ۵۰ نمونه برای مستند نام‌ها و ۵۰ نمونه دیگر مستند موضوعات) بود. پیشنهادها یاد شده بر روی وبگاه مستقل به نشانی www.aghadeh.ir منتشر و به موتورهای کاوش طراحی فرامای تولید داده‌های ساختارمند (اسکیما، آرگ) معرفی گردید. سپس از طریق جستجو در موتورهای کاوش یاد شده با استفاده از ابزار گردآوری داده‌ها یعنی سباهه و ارسی تهیه شده توسط پژوهشگران، چگونگی نمایه‌سازی و بازیافت‌پذیری پیشنهادها فراداده‌ای گروه‌های گواه و آزمون در نتایج جستجوی موتورهای کاوش منتخب مورد ارزیابی قرار گرفت.
تاریخ بازنگری: ۱۴۰۱/۰۹/۰۳	یافته‌ها: یافته‌ها حاکی از آن است که موتورهای کاوش قادر به نمایه‌سازی و بازیافت‌پذیرسازی تمامی پیشنهادها جامعه پژوهش و ارزش‌های گسترش‌های مرتبط با داده‌های مستند طراحی شده بودند. چنین امکانی برای پیشنهادها داده‌های مستند نام و داده‌های مستند موضوعی وضعیت یکسانی داشت.
تاریخ پذیرش: ۱۴۰۱/۱۰/۰۷	نتیجه‌گیری: با ایجاد امکان نمایه‌پذیری و پدیدارنمایی معنایی هر یک از ارزش‌های صفات ناگزیده در پیشنهادها گروه آزمون، افزون بر ارزش‌های مرجع نام و اصطلاحات موضوعی، جامعیت بازیابی در موتورهای کاوش افزایش، و کنترل مستندات در مهمترین ابزارهای کاوش وب بهبود خواهد یافت.
تاریخ انتشار: ۱۴۰۱/۱۰/۱۵	
کلیدواژه‌ها: گسترش داده‌های مستند، پیشنهادها فراداده‌ای، فرامای تولید داده‌های ساختارمند، داده‌های خرد، موتورهای کاوش وب	

استناد: طاهری، سیدمهدی؛ آقاده‌زاده، سمیرا؛ تقوا، محمدرضا؛ قاسمی، منیره؛ و شکرزاده هشترودی، نگین (۱۴۰۱). طراحی صفات داده‌ای مستند مبتنی بر روش

داده‌های خرد و بررسی واکنش موتورهای کاوش وب به پیشنهادها مبتنی بر آن. *تحقیقات کتابداری و اطلاع‌رسانی دانشگاهی*، ۵۶ (۴)، ۶۹-۸۸.

<https://doi.org/10.22059/lib.2023.356370.1675>



مقدمه

امروزه شبکه جهانی وب به عنوان مهمترین و بزرگترین نظام ذخیره و بازیابی اطلاعات که حجم گسترده‌ای از منابع و اشیای محتوایی^۱ را شامل می‌شود، مطرح گردیده است. محور اصلی جستجو در وب را موتورهای کاوش به خود اختصاص داده‌اند و در زمره پر استفاده‌ترین ابزارهای جستجو و بازیابی اطلاعات وب هستند. به نحوی که نود درصد از زمانی که کاربران وب صرف جستجوی اطلاعات می‌کنند، در محیط این موتورهای جستجو (طاهری، ۱۳۹۴). سهولت جستجو در محیط آنها و قابلیت‌های جستجوی متنوعی که برای سطوح گوناگون کاربران فراهم نموده‌اند، بر اهمیت آنها افزوده است (حریری و طاهری، ۱۳۹۳)؛ اما وب علی‌رغم مزیت‌های بی‌شمار، قابلیت محدودی در درک واژگان جستجو و بازیابی اطلاعات مرتبط و دقیق از میان اقیانوس عظیم داده‌ها و اطلاعات برای کاربران دارد.

از جمله راهکارهای بازیابی جامع و مانع، وب معنایی است که در سال ۱۹۹۹ توسط تیم برنرزیلی^۲ مطرح شد. وب معنایی را می‌توان لایه‌ای از فراداده‌های قابل فهم و قابل پردازش توسط ماشین بر روی وب کنونی دانست (جمالی‌مهمویی، ۱۳۸۲). وب معنایی گسترش وب کنونی است که با برقراری ارتباط و پیوند میان داده‌ها و بافت‌های اطلاعاتی، زمینه را برای تولید معنای مستتر در آنها فراهم کرده است. در این میان، مجموعه‌های داده‌ای غنی در سطح وب نیز می‌توانند راه‌گشا باشند. خرده‌های غنی^۳ شامل داده‌های ساختارمند نشانه‌گذاری شده در محتوای وب است که به الگوریتم^۴های موتورهای کاوش کمک می‌کنند تا محتوای وب را بهتر نمایه‌سازی، رتبه‌بندی و تفسیر کنند. کارکرد اصلی آنها ساختارمندسازی یا معنادار کردن داده‌هاست (طاهری، نیکزاد و صمیعی، ۱۳۹۶). خرده‌ها عبارتند از تعداد محدودی از خطاهای متنی با قابلیت فرایبوند که در بین نتایج جستجو پدیدار شده، و به کاربر کمک می‌کنند تا محتوای صفحه‌ها و مرتبط‌ترین اطلاعات با نیاز خود را تشخیص دهد. داده‌های خرده^۵ نیز به عنوان یکی از شیوه‌های تولید خرده‌های غنی، برای لانه‌سازی^۶ معنایی در محتوای موجود بر روی صفحه‌های وب استفاده می‌شود و سبب درک بهتر موتورهای کاوش از محتوای صفحه‌ها و ارائه نتایج مرتبط‌تر به کاربران می‌شوند (طاهری، ۱۳۹۳). در این روش، از واژگان فرامای تولید داده‌های ساختارمند (اسکیما، آرگ)^۷ که به وبگاه‌ها در نشانه‌گذاری انواع داده و محتوا کمک شایانی می‌کند، به منظور افزودن لایه اطلاعاتی اضافی (همان داده‌های ساختارمند) به صفحه‌های وب از برچسب‌ها و عناصر معنایی تعریف شده اچ.تی.ام.ال. ۵ استفاده می‌شود (وان^۸، ۲۰۱۲). در واقع، مدل‌های داده‌ای ساختار یافته با استفاده از فرامای تولید داده‌های ساختارمند جستجوی دقیق و معنایی را تسهیل و تسریع می‌بخشد (ایلیادیس، آکر، استونس، کاواکلی^۹، ۲۰۲۳). ایجاد داده‌های ساختارمند با فرامای تولید داده‌های ساختارمند و نشانه‌گذاری معنایی، می‌تواند در تحقق وب معنایی و ایجاد وب داده‌ها^{۱۰} تأثیر چشمگیری داشته باشد.

موتورهای کاوش به منزله ابزارهایی برای جستجو و بازیابی، جهت دستیابی به حجم بالایی از اطلاعات، نقش ویژه‌ای ایفا می‌کنند؛ اما هنگامی که در میان انبوهی از داده‌های وب به دنبال موجودیتی می‌گردیم موتورهای کاوش، اطلاعاتی را در اختیارمان قرار می‌دهند که خود از آنها بی‌اطلاع هستند (طاهری، ۱۳۹۴). مشکل اصلی این است که اگرچه امکان جستجوی بافتی^{۱۱} در نظام بازیابی موتورهای کاوش امکان‌پذیر است؛ ولی این جستجوها به بخش‌های معدودی از شیء محتوایی (عنوان، متن و...) محدود می‌شوند (حریری و طاهری، ۱۳۹۳). در واقع، به علت استفاده محدود موتورهای کاوش از نظام‌های معنایی، نیز

1. Content object

2. Tim Berners-Lee

3. Rich snippets

4. Algorithms

5. MicroData

6. Nesting لانه‌سازی (تو در تو نویسی) در زبان نشانه‌گذاری، جایگذاری یک شیء در دل شیء دیگری از همان نوع است که دارای برچسب شروع، پایان و مقدار و

محتوای عنصر است. عناصر اچ.تی.ام.ال قابل لانه‌سازی در عناصر دیگر است بدین معنا که عناصر می‌توانند دربردارنده عناصر دیگر باشند (دبلیوتری اسکولز، ۲۰۱۷).

7. Schema.org

8. Van

9. Iliadis, Acker, Stevens, Kavakli

10. Web of Data

11. Contextual search

عدم استفاده از فهرست‌های مستند نام‌ها از یک سو، و صرفاً ثبت شکل پذیرفته شده نام‌ها و اصطلاحات در پیشینه‌های فراداده‌ای کتابشناختی از سوی دیگر، قابلیت بازیابی پیشینه‌ها و نتایج مورد نظر کاربران از طریق دیگر شکل‌های نام‌ها و اصطلاحات وجود ندارد. بنابراین موتورهای کاوش از جامعیتی که نظام‌های معنایی برای فرایند بازیابی فراهم می‌نمایند بهره‌مند نمی‌شوند (حریری و طاهری، ۱۳۹۳).

در همین راستا، داده‌های خرد به عنوان یکی از روش‌های ایجاد خرده‌های غنی که از اجزای سازنده وب معنایی و شیوه‌های نوین برای ساختارمندسازی داده‌ها هستند، مطرح می‌گردند که برای موتورهای کاوش قابل خواندن و قابل فهم هستند (طاهری، نیکزاد و صمیعی، ۱۳۹۶). چالش اساسی موجود در داده‌های خرد عدم کنترل مستندات است که سبب تأثیرگذاری بر عملکرد بهینه موتورهای کاوش هم می‌گردد. این وضعیت با هدف اساسی داده‌های خرد یعنی تأکید بر ساختارمندسازی و درک معناشناختی موجودیت‌هایی مثل کتاب، مقاله، افراد و مکان و جزء آنها توسط موتورهای کاوش وب، در تضاد است. به بیانی دیگر، تنوع بسیار گسترده میان شکل‌های نام‌های خاص (مانند نام افراد و تالکان و عنوان آثار) و شکل‌های گوناگون اصطلاح‌های دال بر یک مفهوم واحد، موجب کاهش دسترسی یکپارچه به اطلاعات وب می‌گردد. از طرفی بازیابی آثار هر پدیدآورنده تحت هر نامی، و گردآوردن همه آثار هر پدیدآورنده در کنار هم تنها در صورتی امکان‌پذیر است که شکل مشخصی از نام و موضوعات به منزله شکل مستند پذیرفته و دیگر شکل‌های نام موضوعات به آن شکل مستند پیوند داده شود (فتاحی، ۱۳۸۵). بنابراین، پژوهش حاضر درصدد پاسخگویی به این پرسش است که آیا می‌توان با استفاده از گسترش‌های دارای عناصر داده‌ای مستند، موجب بهبود بازیابی و اعمال فرایند کنترل مستندات بر روی پیشینه‌های فراداده‌ای مبتنی بر فرمانای تولید داده‌های ساختارمند (اسکیمای آرگ) شد؟ در صورت درج این گسترش‌ها به پیشینه‌های فراداده‌ای یاد شده، واکنش موتورهای کاوش وب به آنها چگونه خواهد بود؟ طراحی صفات^۱ داده‌های مستند مبتنی بر روش داده‌های خرد و واکنش موتورهای کاوش وب به پیشینه‌های مبتنی بر آن چگونه است؟

هدف و پرسش‌های پژوهش

هدف اصلی از انجام این پژوهش، طراحی صفات داده‌های مستند مبتنی بر روش داده‌های خرد و تبیین واکنش موتورهای کاوش وب به پیشینه‌های مبتنی بر آن بود. برای دستیابی به هدف اصلی پژوهش، پرسش‌های زیر طرح‌ریزی شده‌اند:

۱. وضعیت نمایه‌سازی و بازیافت‌پذیری^۲ صفات داده‌ای مستند مندرج در پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و جی‌سان-ال‌دی. توسط موتور کاوش گوگل چگونه است؟
۲. وضعیت نمایه‌سازی و بازیافت‌پذیری صفات داده‌ای مستند مندرج در پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و جی‌سان-ال‌دی. توسط موتور کاوش بینگ چگونه است؟
۳. وضعیت نمایه‌سازی و بازیافت‌پذیری صفات داده‌ای مستند مندرج در پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و جی‌سان-ال‌دی. توسط موتور کاوش یاندکس چگونه است؟
۴. وضعیت نمایه‌سازی و بازیافت‌پذیری صفات داده‌ای مستند مندرج در پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و جی‌سان-ال‌دی. توسط موتور کاوش یا هو^۳ چگونه است؟
۵. انواع و نحوه درج گسترش‌های مرتبط با صفات داده‌ای مستند به صفات فرمانای تولید داده‌های ساختارمند چگونه است؟

مرور پیشینه‌های پژوهش

شبکه جهانی وب دارای حجم عظیمی از اشیای محتوایی است و برای تسهیل و تسریع دسترس‌پذیری به این داده‌ها نیازمند ابزارهای نوینی است. از مهمترین ابزارها می‌توان به فراداده‌ها و موتورهای کاوش اشاره کرد. این ابزارها امکان گردآوری اشیای

1. Properties

2. Retrievability

3. YaHoo

محتوایی مرتبط در کنار یکدیگر، افزایش دسترس‌پذیری، تسهیل جستجو و بازیابی اشیا اطلاعاتی در سطح وب و انجام جست‌وجوهای پیچیده و جامع را فراهم می‌کنند. فراداده‌ها و موتورهای کاوش هر دو برای توصیف و بازنمون روابط اشیای محتوایی کارکردهای گوناگونی دارند. در سال‌های اخیر، واکنش موتورهای کاوش نسبت به فراداده‌ها و افزایش میان‌کنش‌پذیری بین فراداده‌ها و موتورهای کاوش بیش از پیش اهمیت یافته است. با بررسی پژوهش‌های انجام شده در این حوزه به نظر می‌رسد، پژوهش‌های محدودی به ارزیابی نقش غنی‌سازی مفهومی فراداده‌ها، در سطح وب با استفاده از استانداردها و روش‌های نشانه‌گذاری معنایی جدید جهت بررسی واکنش موتورهای کاوش در چگونگی نمایه‌سازی و پیدانمایی^۱ ساختاریافته آنها پرداخته‌اند. اغلب پژوهش‌های مورد نظر، یا با استفاده از یک طرح فراداده‌ای به بررسی سطوح نمایه‌سازی و نمایش اشیای محتوایی در موتورهای کاوش پرداخته‌اند و یا موتورهای کاوش وب را مطابق سنج‌های معینی که برای ارزیابی موتورهای کاوش به کار برده می‌شود، از منظر بازیابی عادی یا معنایی، مزایا و معایب آنها، مورد مقایسه قرار داده‌اند. تعدادی از پژوهش‌های اخیر از روش‌های تولید خرده‌های غنی، داده‌های خرد و گاه روش ترکیبی با دیگر طرح‌های (های) فراداده‌ای به بررسی واکنش موتورهای کاوش به پیشینه‌های فراداده‌ای که با دو یا چند روش متفاوت ساختارمند یافته‌اند، پرداخته‌اند. نیز برخی از آنها با روش‌های ساختارمند یاد شده به طراحی طرحواره‌ها و گسترش‌های مربوط به بافت‌های گوناگون برای توسعه واژگان فرانمای تولید داده‌های ساختارمند (اسکیما، آرگ) اقدام نموده‌اند.

به طور خلاصه، پیشینه‌های موجود را می‌توان در سه گروه کلی تقسیم‌بندی کرد. گروه نخست دربردارنده پژوهش‌هایی است که کیفیت نمایه‌سازی اشیای محتوایی مبتنی بر طرح‌های فراداده‌ای هسته دابلین و مارک ۲۱ مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر ایکس.ام.ال. و نیز تأثیر کاربرد آن زبان در بازیابی و پیدانمایی اشیای محتوایی توسط موتورهای کاوش در محیط وب را مورد بررسی قرار داده‌اند. طاهری، حریری و فتاحی (۱۳۸۸)، طباطبایی‌امیری، طاهری، فرج‌پهلوی، عصاره و معرف‌زاده (۱۳۹۱)، باب‌الحوادثی، طاهری و آقاعابدی (۱۳۹۴)، حریری و طاهری (۱۳۹۳) و (۱۳۹۴)، در این گروه جای می‌گیرند که با استفاده از روش تجربی به ارزیابی تأثیر زبان نشانه‌گذاری گسترش‌پذیر در بازیابی پیشینه‌های فراداده‌ای و به پیروی از آن اشیای محتوایی پرداختند.

طاهری و دیگران (۱۳۸۸) با هدف تبیین کیفیت نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای هسته دابلین و مارک ۲۱ مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (ایکس.ام.ال.) توسط موتورهای کاوش عمومی و مقایسه آنها با یکدیگر پژوهشی به روش تجربی انجام دادند. جامعه پژوهش را ۱۰۰ شیء محتوایی مبتنی بر ایکس.ام.ال. در قالب دو گروه تشکیل می‌داد؛ گروه اول حاوی پیشینه‌های فراداده‌ای هسته دابلین و گروه دوم حاوی پیشینه‌های فراداده‌ای مارک ۲۱ که بر روی دو وبگاه^۲ انتشار یافته بودند. براساس نتایج آزمون آماری، تفاوتی میان کیفیت نمایه‌سازی پیشینه‌های فراداده‌ای دو گروه توسط موتورهای کاوش گوگل و یاهو مشاهده نگردید. یافته‌های پژوهش نشان داد هیچ یک از دو طرح فراداده‌ای هسته دابلین و مارک ۲۱ مبتنی بر ایکس.ام.ال.، از لحاظ دسترس‌پذیری توسط موتورهای کاوش عمومی، نسبت به یکدیگر برتری ندارند. در همین راستا، پژوهش طباطبایی‌امیری و دیگران (۱۳۹۱) با عنوان بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی اشیا محتوایی حاوی عناصر فراداده‌ای هسته‌ای دابلین و مارک ۲۱ مبتنی بر ایکس.ام.ال. در محیط‌های اطلاعاتی پیوسته پویا توسط موتورهای کاوش عمومی، ۱۰۰ شیء محتوایی (کتاب الکترونیکی) به عنوان جامعه پژوهش ایجاد گردید. یافته‌ها حاکی از آن بود که عناصر فراداده‌ای معرفی شده به گوگل بعد از زمان تعیین شده مورد نمایه‌سازی قرار گرفتند؛ اما عناصر فراداده‌ای معرفی شده به یاهو هیچ پاسخی دربرنداشت. به طور کلی، هیچ یک از دو طرح فراداده‌ای هسته دابلین و مارک ۲۱ مبتنی بر ایکس.ام.ال. در محیط اطلاعاتی پیوسته پویا از نظر دسترس‌پذیری توسط گوگل نسبت به دیگری برتری نداشت.

حریری و طاهری (۱۳۹۳) به تبیین کاربرد روش جزیره داده‌ها در تولید پیشینه‌های فراداده‌ای مبتنی بر طرح‌های فراداده‌ای هسته دابلین، مارک ۲۱ و طرح فراداده‌ای توصیف شیء در قالب زبان نشانه‌گذاری گسترش‌پذیر با قابلیت نمایه‌پذیری و پیدانمایی

1. Visibility

2. www.dcmixml.islamicdoc.com , www.marxml.islamicdoc.com

نام‌های برچسب عناصر در محیط موتورهای کاوش پرداختند. جامعه پژوهش را ۶۰۰ پیشینه فراداده‌ای در قالب دو گروه گواه و آزمون تشکیل می‌دادند. گروه گواه دارای ۳۰۰ پیشینه، کاملاً مبتنی بر فرآیندهای استاندارد طرح‌های فراداده‌ای مورد مطالعه، و گروه آزمون، مبتنی بر فرآیندهای یاد شده اما با ویژگی‌های خاص، دارای ۳۰۰ پیشینه جاسازی شده در برچسب زبان نشانه‌گذاری فرامتن (اچ.تی.ام.ال.) بر اساس شیوه جزیره داده‌ها بودند. یافته‌ها نشان داد که نام‌های برچسب تمامی عناصر فراداده‌ای مربوط به پیشینه‌های عضو گروه آزمون که براساس روش جزیره داده‌ها تولید شده بودند، توسط موتورهای کاوش گوگل و یاهو نمایه شده و در نتایج جستجو پدیدار می‌شوند؛ اما گروه گواه فاقد این ویژگی بود. برخلاف گروه گواه، امکان بازیابی پیشینه‌های فراداده‌ای گروه از طریق نام‌های برچسب آنها در موتورهای کاوش وجود دارد.

پژوهش دیگری با هدف تبیین وضعیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دابلین و مارک ۲۱ در بستر نحوی ایکس.ام.ال. توسط باب‌الحوادثی، طاهری و آقاعبدی (۱۳۹۴) با روش تحلیلی-مقایسه‌ای انجام شد. جامعه پژوهش را ۱۰۰ پیشینه فراداده‌ای مبتنی بر ایکس.ام.ال. در قالب دو گروه تشکیل می‌دادند. گروه نخست، دارای عناصر فراداده‌ای هسته دابلین و گروه دیگر، دارای عناصر فراداده‌ای مارک ۲۱ بودند که بر روی ویگاهی^۱ انتشار یافتند. یافته‌ها حاکی از آن بود که تمامی ارزش‌های عناصر پیشینه‌های فراداده‌ای هسته دابلین و مارک ۲۱ در بستر ایکس.ام.ال. توسط موتورهای کاوش گوگل و یاهو نمایه شدند و امکان بازیابی هر یک از پیشینه‌ها براساس ارزش عناصر آنها در محیط موتورهای کاوش فراهم است. پژوهش حریری و طاهری (۱۳۹۴) با هدف تبیین تأثیر حذف خط اعلان زبان نشانه‌گذاری گسترش‌پذیر و تغییر پسوند فایل بر افزایش نمایه‌پذیری و پیدائمی نام‌های برچسب عناصر فراداده‌ای هسته دابلین مارک ۲۱ و طرح فراداده‌ای توصیف شیء در محیط موتورهای کاوش وب انجام شد. جامعه پژوهش را ۶۰۰ پیشینه فراداده‌ای در قالب دو گروه آزمون و گواه تشکیل می‌دادند. گروه گواه دارای ۳۰۰ پیشینه، تولید شده بر مبنای فرآیندهای یاد شده اما دارای ویژگی‌های خاص، بدون سطر اعلان ایکس.ام.ال. و با پسوند فایل هم‌خوان با نام طرح فراداده‌ای مرتبط بودند. نتایج پژوهش نشان داد که موتورهای کاوش گوگل و یاهو نام‌های برچسب تمامی عناصر فراداده‌ای مربوط به پیشینه‌های عضو گروه آزمون را نمایه‌سازی نموده و در نتایج جستجوی خود پدیدار ساختند؛ اما چنین واکنشی را به نام‌های برچسب عناصر پیشینه‌های فراداده‌ای گروه گواه نشان ندادند.

گروه دوم شامل پژوهش‌هایی است که به مقایسه موتورهای کاوش عمومی و معنایی با یکدیگر و ارزیابی موتورهای کاوش بومی پرداخته‌اند. دلیل انتخاب این گروه، و ارتباط آن با پژوهش حاضر، رویکرد مقایسه‌ای پژوهش حاضر در ارزیابی واکنش چهار موتور کاوش طراح فرآیند تولید داده‌های ساختارمند با هدف اعتبارسنجی پیشینه‌های فراداده‌ای گروه آزمون و نشان دادن تغییر رویکرد چهار موتور کاوش مذکور از نحوی به معنایی بود. این گروه از نوشتار بیانگر تأکید پژوهش آنها بر بررسی و ارزیابی قابلیت‌های موتورهای کاوش معنایی و عمومی در مقوله ضریب دقت و بازیافت و نیز همپوشانی اطلاعات بود. پژوهش‌های ترنر، شاه و بیتیریم^۲ (۲۰۰۹)؛ عزیزان^۳ و همکاران (۲۰۱۳)؛ نگی و کومار^۴ (۲۰۱۴)؛ دری (۱۳۹۳)؛ نوکاریزی و زینالی (۱۳۹۶) در این گروه قرار می‌گیرند. ترنر، شاه و بیتیریم (۲۰۰۹) به بررسی عملکرد جستجوی معنایی سه موتور کاوش معمولی گوگل، یاهو، ام.اس.ان.^۵ و یک موتور کاوشی معنایی هاکیا^۶ پرداختند. آنان ده جستجو از موضوعات مختلف و چهار عبارت را که دارای معنای متفاوت ولی املاي مشابه بودند، تعیین کردند و هر جستجو را در تک تک موتورهای جستجو اجرا نمودند. علاوه بر این هر عبارت را که شامل یک جستجو می‌شد، در موتور کاوش معنایی اجرا کردند. سپس بیست نتیجه اول بازیابی شده را در دو گروه مرتبط و غیرمرتبط جای دادند و ضریب دقت و بازیافت آنها را مورد محاسبه قرار دادند. یافته‌ها بیانگر بهترین عملکرد در رابطه با ضریب دقت توسط یاهو بود، این در حالی است که گوگل بهترین بازیافت را نشان داد و در آخر به این نتیجه رسیدند که عملکرد

1. www.aghaabedi.com

2. Turner, Shah & Bitirim

3. Azizan

4. Negi and Kumar

5. MSN

6. Hakia

جستجوی معنایی در هر دو گروه موتور کاوشی کلیدواژه‌ای و معنایی در سطح پایینی از کیفیت قرار دارد. دری (۱۳۹۳) در پژوهشی با عنوان ارزیابی موتورهای جستجوی معنایی به بررسی عملکرد پنج موتور کاوش معنایی: داک‌داک‌گو^۱، کلوز^۲، هاکیا^۳، لکسی^۴ و فکت‌بیتس^۵ پرداخته است. موتورهای کاوش مذکور، با به کارگیری ۴۵ معیار مشخص در قالب سیاهه‌ای متشکل از دو بخش ویژگی‌های عادی و معنایی، مورد مقایسه قرار گرفتند. نتایج پژوهش نشان داد که موتورهای کاوش مورد مطالعه در شاخص ویژگی‌های عادی و معنایی، عملکرد مناسب و کارایی نداشتند. پژوهش دیگری توسط عزیزان و همکاران (۲۰۱۳) پیرامون ویژگی‌های رابط کاربری موتورهای جستجوی معنایی انجام شده است. در این پژوهش ارزیابی مختصری روی هشت موتور کاوش معنایی با هدف شناسایی امکانات عمده و متداولی که در طراحی رابط کاربری موتورهای کاوش نام‌برده مورد استفاده قرار گرفته، انجام شده است. فهرستی از همه امکاناتی که توسط این موتورهای کاوش می‌شود ارائه گردید. یافته‌ها بیانگر این بود که اغلب موتورهای کاوش امکاناتی همچون: کادر جستجوی ساده، صفحه جستجوی شفاف، نتایج تصویری و نمایش برجسته کلیدواژه‌ها را دارند.

پژوهش نگی و کومار (۲۰۱۴) عملکرد جستجوی معنایی دو موتور کاوش معمولی گوگل و یاهو و سه موتور کاوش معنایی هاکیا، داک‌داک‌گو و بینگ همچنین عملکرد آنها در پردازش عبارات جستجوی زبان طبیعی را مورد بررسی قرار داد. ده عبارت جستجو از موضوعات مختلف در موتورهای کاوش صورت گرفت. سپس، ۲۰ نتیجه اول بازیابی شده را به دو گروه مرتبط و غیر مرتبط تقسیم کرده و ضریب دقت آنها را محاسبه کردند. علی‌رغم اینکه موتور کاوش بینگ نتایج مرتبط بیشتری را بازیابی کرد اما نتایج نشانگر عملکرد بهتر موتور کاوش گوگل و هاکیا در پردازش عبارات جستجوی زبان طبیعی بود. آنان در نهایت دریافتند که موتورهای کاوش معنایی، به لحاظ عملکرد جستجوی معنایی از موتورهای جستجوی عادی، بالاتر و بهتر است. و در آخر، نوک‌اریزی و زینالی (۱۳۹۶) با هدف سنجش همپوشانی موتورهای کاوش بومی پارسی‌جو^۶، یوز^۷، پارسیک^۸ و ریسمون^۹ و مقایسه توانمندی‌های آنها در پوشش دادن وب نمایه‌پذیر، پژوهشی از نوع کاربردی ارزیابانه انجام دادند. آنها با روش مبتنی بر کلیدواژه، ابتدا کلیدواژه‌های منتخب را به موتورهای کاوش ارائه و از نتایج بازیابی شده نمونه‌گیری کردند و با توجه به وجود یا عدم وجود این نتایج در موتورهای کاوش، داده‌ها را گردآوری و با آمار استنباطی مورد تجزیه تحلیل قرار دادند. یافته‌ها نشانگر همپوشانی نسبی موتور کاوش پارسیک نسبت به پارسی‌جو و همپوشانی نسبی پارسی‌جو نسبت به یوز به طور متوسط ۲۶ درصد بود که در این میان موتور کاوش پارسیک بیشترین بازیابی را به خود اختصاص داد.

گروه سوم دربردارنده پژوهش‌هایی است که با روش تجربی به استفاده از نشانه‌گذاری‌های معنایی محتوا نظیر روش‌های تولید خرده‌های غنی از جمله داده‌های خرد، گاه به صورت ترکیبی با روش‌های دیگر و نیز طراحی گسترشی برای توسعه واژگان فرمانمای تولید داده‌های ساختارمند (اسکیما.آرگ) به منظور ارزیابی تأثیر داده‌های ساختارمند بر بازیابی کارآمد اطلاعات از شبکه وب (موتورهای کاوش وب) تأکید دارند. پژوهش‌های هگنبوم، فراسینکر و ون‌دیک^{۱۰} (۲۰۱۱)؛ الدائج^{۱۱} (۲۰۱۵)؛ طاهری، نیکزاد، صمیعی (۱۳۹۶)؛ طاهری، ذوالقدر و صمیعی (۱۳۹۷)؛ فردحسینی (۱۳۹۷)، محمدی استانی، چشمه‌سهرابی، طاهری، شعبانی و عاصمی^{۱۲} (۲۰۲۱)، در این گروه جای می‌گیرند.

1. www.duckduckgo.com

2. www.cluuz.com

3. www.haika.com

4. www.lexxe.com

5. www.factbites.com

6. www.parsijoo.ir

7. www.yooz.ir

8. www.parseek.ir

9. www.rismoon.com

10. Hogenboom, Frasinca, & Vandic

11. Aldaej

12. MohammadiOstani, CheshmehSohrabi, Taheri, Shabani, & Asemi

هگنبوم، فراسینکر، ون‌دیک (۲۰۱۱) پژوهشی با نام حاشیه‌نویسی خودکار صفحه‌های وب با استفاده از خرده‌های غنی گوگل انجام دادند. آنان در روشی تجربی، به منظور ایجاد قابلیت بهره‌مندی از خرده‌های غنی گوگل، به کارگیری چارچوب ARROW را پیشنهاد می‌کنند که به طور خودکار به شناسایی و حاشیه‌نویسی خصیصه بازدید صفحه‌های وب با استفاده از روش آر.دی.اف.ای. برای تولید خرده‌های غنی و واژگان گوگل می‌پردازد. این چارچوب متشکل از چهار مرحله: شناسایی نقاط، تجزیه و تحلیل درونی، استخراج اطلاعات و حاشیه‌نویسی صفحه بود. پژوهش طاهری، نیکزاد، صمیعی (۱۳۹۶)، با هدف بررسی واکنش موتورهای کاوش وب به پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی خرده‌های غنی و روش داده‌های کنسرسیوم محتوای ملی، با رویکرد تجربی انجام شد. پژوهشگر با انتخاب ۲۰۰ پیشینه فراداده‌ای از دروازه اطلاعاتی کنسرسیوم محتوای ملی در قالب دو گروه گواه و آزمون، به تهیه ۱۰۰ پیشینه فراداده‌ای با رویکرد ترکیبی مبتنی بر روش داده‌های خرد و داده‌های پیوندی (قالب آر.دی.اف. / ایکس.ام.ال.) در گروه آزمون و ۱۰۰ پیشینه بدون تغییر، در قالب گروه گواه پرداخت و آنها را بر روی وبگاه^۱ منتشر کرد، سپس با روش‌های توصیه شده وبگاه خود را به موتورهای کاوش گوگل و بینگ معرفی کرد، همچنین با استفاده از سیاهه وارسی، وضعیت نمایه‌پذیری و پیدانمایی پیشینه‌های فراداده‌ای مذکور را بررسی کرد. یافته‌ها نشان داد که موتورهای کاوش گوگل، بینگ و یاهو قادر به نمایه‌سازی پیشینه‌های فراداده‌ای بودند و اما هیچ یک از موتورهای کاوش، در نمایش ساختارمند پیشینه‌های فراداده‌ای موفق عمل نکردند. پیشینه‌های فراداده‌ای این پژوهش با روش جی‌سان نیز ترکیب شدند که بلافاصله توسط موتورهای کاوش یاد شده نمایه‌سازی شدند اما فقط برچسب‌های نشانه‌گذاری به مرحله پیدانمایی رسیدند. در واقع، توانایی شناسایی و تشخیص برچسب‌های (عناصر) توصیه شده براساس روش داده‌های خرد را که در قالب چارچوب توصیف منبع (آر.دی.اف.)، به عنوان بستر نحوی روش داده‌های پیوندی، پیاده‌سازی شده بود، نداشتند در نتیجه پیشینه‌ها را در نتایج جستجو به صورت معنادار نمایش ندادند.

الدائج (۲۰۱۵) پژوهشی با نام محیط‌های یادگیری مجازی^۲ پیشرفته مبتنی بر فرانمای تولید داده‌های ساختارمند (اسکیما.آرگ) و رسانه‌های اجتماعی، به روشی تجربی جهت افزودن نشانه‌گذاری معنایی بر اساس واژگان فرانمای تولید داده‌های ساختارمند (اسکیما.آرگ) و نیز یکپارچه‌سازی رسانه‌ها و شبکه‌های اجتماعی به منظور توسعه بافت یادگیری مجازی (وی.ال.ای.) انجام داد. در این پژوهش از نسخه ۷، سیستم مدیریت محتوای معنایی^۳ (اس.سی.ام.اس.) دروپال^۴ استفاده شد، همچنین مفاهیم و واژگان پیشنهادی پروژه خود از جمله نوع: دوره‌های آموزشی، نشست، ارزشیابی، آزمون و تکلیف^۵ را برای توصیف محتوای وی.ال.ای.ها^۶ با بهره‌جویی از مزایای فناوری‌های وب اجتماعی و معنایی، به دلیل عدم پشتیبانی از توصیف منابع آموزشی در واژگان فرانمای تولید داده‌های ساختارمند (اسکیما.آرگ)، ارائه نمود که در نهایت گسترش‌های یاد شده، در سلسله مراتب واژگان فرانمای تولید داده‌های ساختارمند (اسکیما.آرگ) به عنوان زیرنوع، اکشن^۷ قرار گرفتند. یافته‌ها همچنین حاکی از برتری مزیتی اس.سی.ام.اس. دروپال، نسبت به دیگر بسترهای موجود بود.

در تلاشی دیگر طاهری، ذوالقدر و صمیعی (۱۳۹۷) با هدف تبیین نمایه‌سازی و پیدانمایی خرده‌های غنی توسط موتورهای کاوش وب با رویکردی تطبیقی و روشی تجربی، ۲۰۰ پیشینه فراداده‌ای مربوط به کتاب‌های الکترونیکی مبتنی بر زبان نشان‌گذاری فرامتن را تحلیل نمودند که از میان پیشینه‌های کتابخانه دیجیتال دانشگاه علامه طباطبائی به روش نمونه‌گیری تصادفی انتخاب شدند. این پیشینه‌ها در قالب یک گروه آزمون، شامل ۱۰۰ پیشینه فارسی و انگلیسی که براساس روش داده‌های خرد ایجاد شده بودند و یک گروه گواه، حاوی ۱۰۰ پیشینه فارسی و انگلیسی با ساختار معمولی بودند. یافته‌ها حاکی از آن بود که موتور کاوش گوگل پیشینه‌های فراداده‌ای هر دو گروه آزمون و گواه را نمایه‌سازی و پیشینه‌های گروه آزمون را به صورت معنادار

1. www.nikzad.ir

2. Virtual Learning Environments (VLE)

3. Semantic Content Management System (SCMS)

4. Drupal

5. Course, Session, Examination, test, Assignment

6. Virtual Learning Environments (VLEs)

7. Action

در نتایج جستجو پدیدار ساخت؛ اما موتورهای کاوش یاهو و بینگ، پیشینه‌های هر دو گروه را فقط نمایه‌سازی نموده، و قادر به پیدانمایی پیشینه‌های گروه آزمون نبودند. بنابراین، با وجود اعلام موتورهای کاوش معتبر وب مبنی بر نمایش معنادار پیشینه‌های فراداده‌ای مبتنی بر خرده‌های غنی، تنها موتور کاوش گوگل پیدانمایی معنادار آنها را پشتیبانی کرد.

محمدی استانی و همکاران (۲۰۲۱) در پژوهش خود به بومی‌سازی فرانمای تولید داده‌های ساختارمند برای وصیف نسخه‌های خطی در بافت اطلاعاتی ایرانی-اسلامی پرداختند. در این پژوهش از روش کیفی و تحلیل محتوای اسنادی استفاده شد. یافته‌های این پژوهش نشان داد برای بهبود نمایه‌سازی و توصیف معنایی نسخه‌های خطی در موتورهای جستجوی وب لازم است تعداد ۶ صفت به نوع «آثار خلاقانه» افزوده شود. نیز، نتایج ای پژوهش نشان داد واکنش موتورهای کاوش نسبت به پیشینه‌های مبتنی بر طرح‌واره پیشنهادی نسخه‌های خطی مثبت بوده و این ویژگی‌ها منجر به بومی‌سازی طرح‌واره موجود برای سازگاری نوع نسخه خطی با بافت اطلاعاتی ایرانی-اسلامی می‌شوند. آخرین مورد از تلاش‌های صورت گرفته در این زمینه، پژوهش فردحسینی، طاهری، حریری، باب‌الحوائجی و نوشین‌فرد (۱۳۹۹) بود که در آن به بهبود کارکردپذیری صفات و روابط میان موجودیت‌های آثار خلاقانه اسکیمای آرگ با بازنمایی آنها در الگوی مفهومی ال.آر.ام. پرداخته شد. جامعه این پژوهش صفات موجودیت کتاب از مجموعه موجودیت‌های آثار خلاقانه اسکیمای آرگ بود و داده‌های مورد نیاز با استفاده از سیاهه واری گردآوری شدند. یافته‌های پژوهش نشان داد برای صفاتی همانند «نویسنده»، «قالب فایل»، «بستر نحوی رمزگذاری» و «رسانه مرتبط» در صفات فرانمای تولید داده‌های ساختارمند در موجودیت «آثار خلاقانه» و صفات «درخواست‌کننده» و «جنسیت» در موجودیت «شخص» در این فرانما هیچ رابطه یا ویژگی‌ای در الگوی «ال.آر.ام.» مطابقت نداشت. بنابراین، این‌گونه صفات باید در استانداردهای فراداده‌ای و محتوایی لحاظ شوند. از سوی دیگر یافته‌های این پژوهش نشان داد، با بهره‌گیری از پیشنهادها الگوهای مفهومی بافت میراث فرهنگی در تعیین و توسعه ویژگی‌ها کاستی‌های کارکردپذیری این فرانما در محیط‌های داده‌ای رفع می‌شود و نیز کاربردپذیری آن بهبود می‌یابد.

مرور پیشینه‌های مرتبط حاکی از آن است که تاکنون در ایران و خارج از کشور مطالعاتی چند در حوزه میان‌کنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب صورت گرفته است. به‌ویژه پژوهش‌های گروه سوم که از اغلب از روش‌های تولید خرده‌های غنی از جمله داده‌های خرد در ایجاد پیشینه‌های فراداده‌ای یا طراحی طرح‌واره بهره گرفته‌اند. در میان پیشینه‌های داخلی تنها محمدی استانی و همکاران (۲۰۲۱)، به طراحی طرح‌واره با استفاده از روش داده‌های خرد پرداخته است که این طرح‌واره به نسخه‌های خطی در بافت اسلامی-ایرانی اختصاص دارد و در پیشینه‌های خارج از ایران پژوهش الدائج (۲۰۱۵)، جزو معدود مطالعات مشابهی بود که با روش‌های در نظر گرفته شده برای طراحی طرح‌واره در فرانمای تولید داده‌های ساختارمند (اسکیمای آرگ) در زمینه محیط آموزشی و یادگیری برخط، با استفاده از ابزار سیستم مدیریت محتوای دروپال اقدام کرده است که گسترش طراحی شده مورد نظر در پژوهش او در واژگان فرانمای تولید داده‌های ساختارمند (اسکیمای آرگ) به عنوان زیرنوع و نوع فعالیت قرار گرفتند. از سوی دیگر، پژوهشی که به طور خاص طرح‌واره یا گسترش مربوط به داده‌های مستند و با تأکید بر کنترل مستندات به روش داده‌های خرد و جی‌سان-ال.دی.، برای نقاط دسترسی از جمله نام و موضوع، در پیشینه‌های داخل و خارج ایران صورت گرفته باشد، یافت نشد. این مسئله بیانگر وجود خلأ در این زمینه است چرا که اعمال کنترل مستندات در شناسه‌های اصلی نام و موضوع، سبب ایجاد انسجام، یکپارچگی و میان‌کنش‌پذیری معنایی، افزایش ربط و بسط جستجو و بهبود بازیابی در موتورهای کاوش می‌شود و مزایای آن در حوزه سازماندهی اطلاعات و دانش و نیز نقش مؤثر آن در نظام‌های ذخیره و بازیابی بر کارشناسان این امر پوشیده نیست. افزون‌براین، روش داده‌های خرد و اخیراً جی‌سان-ال.دی. شیوه توصیه شده برای نشانه‌گذاری معنایی انواع اشیای محتوایی در سطح وب است می‌باشد و نسبت به دیگر روش‌های تولید فراداده از محبوبیت بیشتری نزد موتورهای کاوش وب برخوردار است.

روش‌شناسی پژوهش

پژوهش حاضر از لحاظ هدف، یک پژوهش توسعه‌ای-کاربردی بود. زیرا به حل مسئله کنترل مستندات و افزایش جامعیت بازیابی پیشینه‌های فراداده‌ای مبتنی بر فرآیند تولید داده‌های ساختارمند (اسکیما.آرگ) با بهره‌گیری از گسترش‌های مرتبط با داده‌های مستند در ابزارها و ساختارهای محبوب موتورهای کاوش وب می‌پرداخت؛ و براساس روش پژوهش تجربی انجام شد. همچنین از نظر رویکرد، یک پژوهش مبتنی بر فناوری بشمار می‌آید. جامعه پژوهش را ۴۰۰ پیشینه فراداده‌ای مبتنی بر فرآیند تولید داده‌های ساختارمند (اسکیما.آرگ) در قالب دو گروه گواه و آزمون، هر گروه ۲۰۰ پیشینه که از کتابخانه دیجیتالی دانشگاه علامه طباطبائی انتخاب و استخراج شدند، تشکیل می‌دادند. این پیشینه‌های فراداده‌ای عبارت بودند از:

الف. ۱۰۰ پیشینه فراداده‌ای مبتنی بر فرآیند تولید داده‌های ساختارمند (اسکیما.آرگ) مربوط به گروه گواه در بستر روش نشانه‌گذاری معنایی داده‌های خرد که به توصیف موجودیت‌های کتابی می‌پرداخت و فاقد گسترش‌های مرتبط با داده‌های مستند بود؛

ب. ۱۰۰ پیشینه فراداده‌ای دیگر مربوط به گروه گواه و مبتنی بر فرآیند تولید داده‌های ساختارمند (اسکیما.آرگ) با این تفاوت که در بستر روش نشانه‌گذاری معنایی و داده‌های پیوندی جی‌سان-ال.دی. بود و ضمن توصیف موجودیت‌های کتابی، همانند پیشینه‌های قبلی فاقد گسترش‌های مرتبط با داده‌های مستند بود؛

ج. ۱۰۰ پیشینه فراداده‌ای مبتنی بر فرآیند تولید داده‌های ساختارمند (اسکیما.آرگ) عضو گروه آزمون پژوهش که به توصیف موجودیت‌های کتابی اختصاص یافته، و بستر نحوی پیاده‌سازی آنها روش نشانه‌گذاری معنایی داده‌های خرد بود. تفاوت این پیشینه‌ها با پیشینه‌های فراداده‌ای مشابه در گروه گواه، درج گسترش‌های مربوط به داده‌های مستند بود. در ۵۰ پیشینه از این گروه، ۵۰ پیشینه فراداده‌ای مبتنی بر طرحواره شخص^۱ که پدیدآورنده اصلی موجودیت کتابی مورد نظر در آن توصیف شده، و دارای گسترش دیگر شکل‌های ناگزیده نام یک پدیدآورنده^۲ انتخاب شده از طرح فراداده‌ای توصیف مستند^۳ بود، بر پایه روش لانه‌سازی جاسازی شدند. این گروه از پیشینه‌ها، کنترل داده‌های مستند مربوط به نام‌های خاص را مد نظر قرار دادند^۴. و ۵۰ پیشینه دیگر، دارای گسترش‌های داده‌های مستند دیگر شکل‌های نامرجح اصطلاحات موضوعی^۵، درج شده در صفت (عنصر) "درباره"^۶ از مجموعه صفات (عناصر) موجودیت کتاب (Book) فرآیند تولید داده‌های ساختارمند بودند؛ و

د. در پایان، ۱۰۰ پیشینه فراداده‌ای مبتنی بر فرآیند تولید داده‌های ساختارمند (اسکیما.آرگ) کاملاً مشابه گروه پیشین، اما این بار در بستر روش نشانه‌گذاری معنایی و داده‌های پیوندی جی‌سان-ال.دی. پیاده‌سازی شدند.

برای اعتبارسنجی و اطمینان از صحت پیشینه‌های فراداده‌ای ساختارمند از ابزار <https://validator.schema.org> وب‌مستر گوگل، استفاده شد و تک تک آنها مورد ارزیابی و تأیید قرار گرفتند^۷. پیشینه‌های فراداده‌ای مذکور از تاریخ ۱ مرداد ماه ۱۳۹۷ در وبگاه www.aghadeh.ir منتشر شدند^۸. سپس، این وبگاه به صورت مستقیم و بر اساس روش‌های توصیه شده، به موتورهای کاوش گوگل، یاهو، بینگ و یاندکس معرفی شد و به مدت دو ماه (تا ۳۱ شهریور ماه) در معرض واکنش موتورهای کاوش انتخابی قرار گرفت. با توجه به استفاده از روش تجربی برای انجام پژوهش، انتشار پیشینه‌های فراداده‌ای مبتنی بر

^۱. Person

^۲. Variants

^۳. Metadata Authority Description Schema (MADS)

^۴. لازم به ذکر است که تعداد ده پیشینه این گروه برای اطمینان از نمایه‌سازی ارزش‌ها و شکل‌های پذیرفته نشده نام در صفت زیر که در اسکیما.آرگ تعریف شده است درج شد. alternate name

^۵. Variants

^۶. About

^۷. پیشینه‌های یاد شده در هنگام انتشار مقاله نیز توسط ابزار اشاره شده اعتبارسنجی، و تأیید شدند.

^۸. گسترش (صلت جدید) انجام شده در پژوهش، هم‌اکنون به طور عملیاتی در ماژول مستندات نرم‌افزار کشف و بازنمود دانش لیبرا مورد استفاده قرار گرفته است و نتایج کاملاً یکسانی با نتایج پژوهش حاضر نشان می‌دهد.

اسکیما.آرگ بر روی یک وبگاه مستقل بدون هیچ پیوند درونی^۱ (دریافتی) و برونی^۲، و معرفی مستقیم آن به موتورهای کاوش، نیز انجام جستجوها از اشتراک فراهم‌کنندگان خدمات اینترنتی (آی.اس.پی.۳) یکسان، به دلیل کنترل تمامی عوامل مداخله‌گر و مؤثر بر نمایه‌سازی و رتبه‌بندی نتایج بود. موتورهای کاوش نام‌برده به این علت که طراحان فرآینمای تولید داده‌های ساختارمند اسکیما.آرگ و سازگار با آن فرآینما هستند، انتخاب شدند. همچنین دلیل انتخاب ۱۰۰ پیشینه فراداده‌ای برای هر بستر نحوی پیاده‌سازی در هر یک از گروه‌های مورد مطالعه، امکان تعمیم و بررسی دقیق‌تر واکنش‌ها بود. اگر چه چهار موتور کاوش مورد مطالعه در این پژوهش طراحان استاندارد فرآینمای تولید داده‌های ساختارمند (اسکیما.آرگ) هستند، و به طور رسمی اعلام نموده‌اند، پیشینه‌های فراداده‌ای مبتنی بر این استاندارد را به صورت معنادار نمایه‌سازی و در نتایج جستجو پدیدار می‌نمایند؛ اما در خصوص بررسی واکنش آنها به پیشینه‌های فراداده‌ای در این پژوهش، ذکر چند نکته ضروری است: نخست (به منزله مهمترین دلیل)، پیشینه‌های گروه آزمون دارای گسترش (صفت افزوده و جدید) برای پشتیبانی از نظام ارجاعات در رابطه هم‌ارز و فرایند کنترل مستند بودند و هدف اصلی از بررسی واکنش موتورهای کاوش، مربوط به نمایه‌سازی و پدیدارنمایی معنایی پیشینه‌های فراداده‌ای دارای صفت جدید بود که باید به صورت عملی مورد مطالعه قرار می‌گرفت. زیرا اعلام موتورهای کاوش مربوط به ساختار و واژگان‌های کنونی فرآینمای داده‌های ساختارمند بود. به عبارت دیگر، هر گسترشی به دلیل تغییر ساختار فرآینما باید مورد بررسی عملی قرار گیرد، دیگر، هر یک از موتورهای کاوش چهارگانه طراح اسکیما.آرگ، دارای خطمشی‌های نمایه‌سازی و رتبه‌بندی متفاوتی برای داده‌های ساختارمند هستند، و باید واکنش آنها با دیگر مقایسه می‌شد تا اثربخشی و سودمندی گسترش انجام شده در این پژوهش به صورت جامع ارزیابی می‌شد. استفاده از ابزار اعتبارسنجی عام و بررسی عملی واکنش موتورهای کاوش در همین راستا بود.

```
<div itemscope itemtype="http://schema.org/Book">
  <span itemprop="name"> متن مجازی </span>
  <span itemprop="about"> نشر الکترونیکی </span>
  <span itemprop="about"> کتابخانه‌های الکترونیکی </span>
  <span itemprop="about"> نشریات ادواری الکترونیکی </span>
  <span itemprop="about"> پایگاه‌های اطلاعاتی </span>
  <span itemprop="about"> فضای مجازی </span>
  <span itemprop="bookFormat"> Paperback </span>
  <span itemprop="datePublished"> ۱۳۸۸ </span>
  <span itemprop="inLanguage"> Persian </span>
  <span itemprop="isbn"> 978-600-5567-15-1 </span>
  <span itemprop="numberOfPages"> ۴۴۰ </span>
  <div itemprop="author" itemscope itemtype="http://schema.org/Person">
    <span itemprop="name"> کتس، ویلیام ای </span>
    Katz, William A. </span>
    <span itemprop="variant"> کتس، ویلیس آرمسترانگ </span>
    Katz, Willis Armstrong </span>
    <span itemprop="variant"> کتس، بیل </span>
    Katz, Bill </span>
    <span itemprop="birthDate"> ۱۹۲۴ </span>
  </div>
  <div itemprop="translator" itemscope itemtype="http://schema.org/Person">
    <span itemprop="name"> سیدسعیدرضا، عاملی </span>
    <span itemprop="variant"> سعیدرضا، زانی </span>
    سعیدرضا </span>
  </div>
</div>
```

^۱. Inlink

^۲. Outlink

^۳. Internet Service Provider

```

<span itemprop="variant">سعیدرضا، عاملی، سعیدرضا</span>
<span itemprop="variant">سعید، عاملی رنانی، سعید</span>
<span itemprop="birthDate">۱۳۴۰</span>
</div>
<div itemprop="publisher" itemscope itemtype="http://schema.org/Organization" >
<span itemprop="name">تهران: پژوهشکده مطالعات فرهنگی و اجتماعی</span>
</div>
</div>

```

نمونه ۱. پیشینه فراداده‌ای گروه آزمون به روش داده‌های خرد برای مستند نام شخص

گردآوری داده‌ها در این پژوهش با روش مشاهده ساختارمند^۱ بود که به شیوه‌ای دقیق و علمی و با استفاده از راهبردهای از پیش تعیین شده به گردآوری داده‌ها می‌پرداخت. مقوله‌های مورد مشاهده، نمایه‌سازی و بازیافت‌پذیری گسترش‌های مربوط به داده‌های مستند (شکل‌های نام انتخاب نشده و شکل‌های اصطلاحات نامرجح) مبتنی بر بستر روش‌های نشانه‌گذاری داده‌های خرد و جی‌سان-ال.دی. توسط موتورهای کاوش وب است. برای انتخاب پیشینه‌ها از روش نمونه‌گیری تصادفی استفاده شد. ۱۰۰ پیشینه فراداده‌ای مربوط به اشیای کتابی از مجموعه کتاب‌های فارسی کتابخانه دیجیتالی دانشگاه علامه طباطبائی که پیشینه‌های مستند نام‌های حقیقی و موضوعی مرتبط به آنها نیز استخراج گردید. پوشش موضوعی پیشینه‌ها در حوزه علوم انسانی از جمله: روانشناسی، ادبیات، تاریخ، علوم سیاسی، مدیریت صنعتی، مدیریت مالی، دین و فلسفه، اقتصاد، علم اطلاعات و دانش‌شناسی، علوم اجتماعی بود. ابزار گردآوری داده‌ها، سیاهه واری (چک لیست) طراحی شده توسط پژوهشگران بود که با توجه به پرسش‌های اساسی پژوهش طراحی شد. از راهبرد زیر برای قابلیت بازیابی پیشینه‌های فراداده‌ای عضو جامعه پژوهش، استفاده شد:

”ارزش برچسب variant“ site: aghadeh.ir

برای تجزیه و تحلیل داده‌ها از آخرین نسخه نرم‌افزار ورد و اکسل استفاده شد. یافته‌ها در قالب جداول و نمودار آماری ارائه شده است. با توجه به اینکه مؤلفه‌های پژوهش شامل نمایه‌سازی و بازیافت‌پذیری بودند، در سیاهه واری با امتیازبندی به هر کدام از این مؤلفه‌ها کد (۰ و ۱) اختصاص یافت بدین ترتیب که به پیشینه‌های فراداده‌ای با گسترش‌های مربوط به داده‌های مستند (شکل‌های نام انتخاب نشده و شکل‌های اصطلاحات نامرجح) مبتنی بر بستر روش‌های نشانه‌گذاری داده‌های خرد و جی‌سان-ال.دی. که به سطح نمایه‌سازی و بازیافت‌پذیری رسیدند کد (۱) و به آن گروه از گسترش‌های مربوط به داده‌های مستند که به سطح نمایه‌سازی و بازیافت‌پذیری نرسیدند کد (۰) اختصاص داده شد. سپس مطابق روش گردآوری داده‌ها (مشاهده ساختارمند)، برای هر یک از مؤلفه‌های مربوط به جامعه پژوهش امتیازی اختصاص یافت.

یافته‌های پژوهش

در پژوهش حاضر پنج پرسش طراحی شد. برای ارائه پاسخ پرسش‌های پژوهش از جدول و نمودار استفاده شد.

پرسش اول پژوهش: وضعیت نمایه‌سازی و بازیافت‌پذیری صفات داده‌ای مستند مندرج در پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و جی‌سان-ال.دی. توسط موتور کاوش گوگل چگونه است؟

^۱. Structured observation

جدول ۱. وضعیت نمایه‌سازی و بازیافت‌پذیری گروه‌های گواه و آزمون مبتنی بر روش داده‌های خرد و جی‌سان‌ال‌دی. در موتور کاوش گوگل

گروه مورد مطالعه	تعداد کل پیشینه‌های فراداده‌ای	وضعیت نمایه‌سازی و بازیافت‌پذیری پیشینه‌های فراداده‌ای در موتور کاوش گوگل	
		تعداد	درصد
آزمون ۱۱	۵۰	۵۰	۱۰۰
آزمون ۲۲	۵۰	۵۰	۱۰۰
آزمون ۳۳	۵۰	۵۰	۱۰۰
آزمون ۴۴	۵۰	۵۰	۱۰۰
گواه ۵۱	۱۰۰	۱۰۰	۱۰۰
گواه ۶۲	۱۰۰	۱۰۰	۱۰۰

براساس داده‌های مندرج در جدول ۱، ۱۰۰ درصد پیشینه‌های فراداده‌ای گروه گواه و آزمون به هر دو روش نشانه‌گذاری معنایی داده‌های خرد و جی‌سان‌ال‌دی. در موتور کاوش گوگل مورد نمایه‌سازی و مورد بازیافت‌پذیری قرار گرفت. پیشینه‌های فراداده‌ای در گروه گواه به دلیل عدم وجود داده‌های مستند فقط از طریق شکل مرجح و پذیرفته شده قابل بازیابی بودند؛ اما در گروه آزمون از طریق تمامی شکل‌های مرجح و نامرجح داده‌های مستند قابل بازیابی هستند. این وضعیت برای داده‌های مستند نام‌ها و مستند موضوعی یکسان بود. به عبارت دیگر، نرم‌افزارهای خزنده-نمایه‌سازی گوگل قادر به تشخیص برچسب‌های (صفت)^۷ در تمامی پیشینه‌های گروه آزمون، و نمایه‌سازی ارزش‌های آنها بوده است. به شیوه‌ای که می‌توان پیشینه‌های گروه‌های آزمون را با استفاده از ارزش‌های داده‌ای مستند نامرجح درج شده در صفت بازیابی نمود؛ اما این قابلیت برای پیشینه‌های گروه‌های گواه وجود ندارد.

پرسش دوم پژوهش: وضعیت نمایه‌سازی و بازیافت‌پذیری صفات داده‌ای مستند مندرج در پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و جی‌سان‌ال‌دی. توسط موتور کاوش بینگ چگونه است؟

جدول ۲. وضعیت نمایه‌سازی و بازیافت‌پذیری گروه‌های گواه و آزمون مبتنی بر روش داده‌های خرد و جی‌سان‌ال‌دی در موتور کاوش بینگ

گروه مورد مطالعه	تعداد کل پیشینه‌های فراداده‌ای	وضعیت نمایه‌سازی و بازیافت‌پذیری پیشینه‌های فراداده‌ای در موتور کاوش بینگ	
		تعداد	درصد
آزمون ۱	۵۰	۵۰	۱۰۰
آزمون ۲	۵۰	۵۰	۱۰۰
آزمون ۳	۵۰	۵۰	۱۰۰
آزمون ۴	۵۰	۵۰	۱۰۰
گواه ۱	۱۰۰	۱۰۰	۱۰۰
گواه ۲	۱۰۰	۱۰۰	۱۰۰

۱. این گروه دارای ۵۰ پیشینه فراداده‌ای مبتنی بر فرانمای شخص به روش داده‌های خرد که گسترش‌های مربوط به داده‌های مستند در آنها درج شده است.
۲. این گروه دارای ۵۰ پیشینه فراداده‌ای مبتنی بر صفت موضوع به روش داده‌های خرد که گسترش‌های مربوط به داده‌های مستند در آنها درج شده است.
۳. این گروه دارای ۵۰ پیشینه فراداده‌ای مبتنی بر فرانمای شخص به روش جی‌سان‌ال‌دی که گسترش‌های مربوط به داده‌های مستند در آنها درج شده است.
۴. این گروه دارای ۵۰ پیشینه فراداده‌ای مبتنی بر صفت موضوعی به روش جی‌سان‌ال‌دی که گسترش‌های مربوط به داده‌های مستند در آنها درج شده است.
۵. این گروه دارای ۱۰۰ پیشینه فراداده‌ای مبتنی بر روش داده‌های خرد، فاقد گسترش‌های مرتبط با داده‌های مستند است.
۶. این گروه دارای ۱۰۰ پیشینه فراداده‌ای دیگر مبتنی بر روش جی‌سان‌ال‌دی، فاقد گسترش‌های مرتبط با داده‌های مستند است.

7. Variant

چنانچه جدول ۲ نشان می‌دهد، همانند موتور کاوش گوگل، نرم‌افزار روپات موتور کاوش بینگ نیز تمامی پیشینه‌های فراداده‌ای گروه‌های گواه و آزمون و ارزش‌های صفت (برچسب) حاوی داده‌های مستند غیرمرجع یا انتخاب نشده، مبتنی بر هر دو روش نشانه‌گذاری معنایی داده‌های خرد و جی‌سان-ال.دی. را نمایه‌سازی، و دسترس‌پذیر نموده است.

پرسش سوم پژوهش: وضعیت نمایه‌سازی و بازیافت‌پذیری صفات داده‌ای مستند مندرج در پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و جی‌سان-ال.دی. توسط موتور کاوش یا هو چگونه است؟

جدول ۳. وضعیت نمایه‌سازی و بازیافت‌پذیری گروه‌های گواه و آزمون مبتنی بر روش داده‌های خرد و جی‌سان-ال.دی در موتور کاوش یا هو

گروه مورد مطالعه	تعداد کل پیشینه‌های فراداده‌ای	وضعیت نمایه‌سازی و بازیافت‌پذیری پیشینه‌های فراداده‌ای در موتور کاوش یا هو	
		تعداد	درصد
آزمون ۱	۵۰	۵۰	۱۰۰
آزمون ۲	۵۰	۵۰	۱۰۰
آزمون ۳	۵۰	۵۰	۱۰۰
آزمون ۴	۵۰	۵۰	۱۰۰
گواه ۱	۱۰۰	۱۰۰	۱۰۰
گواه ۲	۱۰۰	۱۰۰	۱۰۰

همان‌طور که داده‌های مندرج در جدول ۳ نشان می‌دهند، تمامی پیشینه‌های فراداده‌ای گروه‌های گواه و آزمون ایجاد شده بر پایه هر دو روش نشانه‌گذاری معنایی داده‌های خرد و جی‌سان-ال.دی. در موتور کاوش یا هو نیز نمایه‌سازی، و قابل بازیابی هستند و بازیافت‌پذیری را در پی داشته‌اند. بنابراین، یا هو نیز همانند موتورهای کاوش پیشین، به برچسب (صفت) واکنش مثبت نشان داده، و فقط امکان بازیابی پیشینه‌های گروه‌های آزمون از طریق ارزش‌های داده‌ای مستند نامرئج را میسر ساخته است.

پرسش چهارم پژوهش: وضعیت نمایه‌سازی و بازیافت‌پذیری صفات داده‌ای مستند مندرج در پیشینه‌های فراداده‌ای مبتنی بر داده‌های خرد و جی‌سان-ال.دی. توسط موتور کاوش یاندکس چگونه است؟

جدول ۴. وضعیت نمایه‌سازی و بازیافت‌پذیری گروه‌های گواه و آزمون مبتنی

بر روش داده‌های خرد و جی‌سان-ال.دی. در موتور کاوش یاندکس

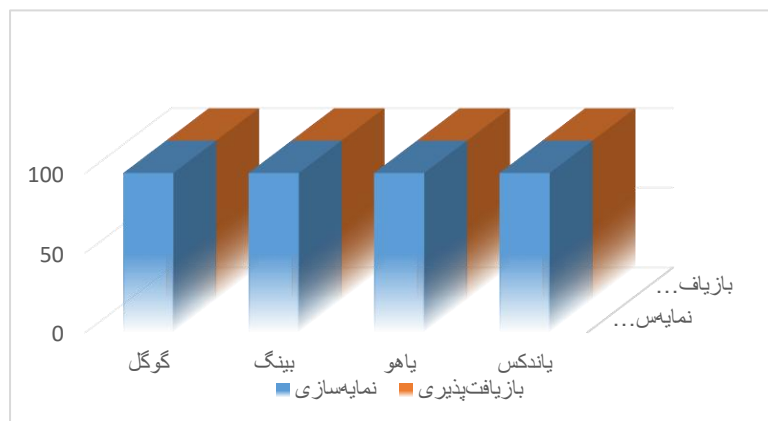
گروه مورد مطالعه	تعداد کل پیشینه‌های فراداده‌ای	وضعیت نمایه‌سازی و بازیافت‌پذیری پیشینه‌های فراداده‌ای در موتور کاوش یاندکس	
		تعداد	درصد
آزمون ۱	۵۰	۵۰	۱۰۰
آزمون ۲	۵۰	۵۰	۱۰۰
آزمون ۳	۵۰	۵۰	۱۰۰
آزمون ۴	۵۰	۵۰	۱۰۰
گواه ۱	۱۰۰	۱۰۰	۱۰۰
گواه ۲	۱۰۰	۱۰۰	۱۰۰

داده‌های ثبت شده در جدول ۴، بیانگر توانایی و واکنش مثبت نرم‌افزار خزنده-نمایه‌ساز موتور کاوش یاندکس همچون سه موتور کاوش پیشین، در فرایند نمایه‌سازی و بازیافت‌پذیری ۱۰۰ درصد پیشینه‌های فراداده‌ای گروه‌های گواه و آزمون، به ویژه گسترش صفت variant در اعمال کنترل مستندات، به هر دو روش نشانه‌گذاری معنایی داده‌های خرد و جی‌سان-ال.دی. است.

پرسش پنجم پژوهش: انواع و نحوه درج گسترش‌های مرتبط با صفات داده‌ای مستند به طرحواره‌های فرانمای تولید داده‌های ساختارمند چگونه است؟

الف. بررسی واکنش موتورهای کاوش مورد مطالعه به پیشینه‌های گروه آزمون مربوط به قالب مستند نام‌ها در هر دو روش داده‌های خرد و جیسان-ال‌دی نشان می‌دهد که برای توسعه طرحواره نام شخص که پدیدآورنده موجودیت کتابی مورد نظر در آن توصیف شده، مطلوب است گسترش شکل‌های ناگزیده نام یک پدیدآورنده در برچسب یا صفت variant برگرفته از طرح فراداده‌ای توصیف مستند (مَدَس)، با رویکرد لانه‌سازی ذیل نوع موجودیت Person و برای گسترش طرحواره نام سازمان برچسب یا صفت variant، ذیل نوع موجودیت Organization، جاسازی شود تا منجر به واکنش مثبت موتورهای کاوش، نمایه‌پذیری و بازیافت‌پذیری داده‌های مستند نام‌های انتخاب نشده پدیدآورنده و سازمان گردد.

ب. برای ایجاد پیشینه‌های مربوط به گروه‌های آزمون مستند مربوط به مستند موضوعی نمایه‌پذیر و بازیافت‌پذیر بر مبنای واکنش مثبت موتورهای کاوش چهارگانه که بر اساس هر دو روش داده‌های خرد و جیسان-ال‌دی، پیاده‌سازی شده بودند، نشان داد که برای طراحی توصیه می‌شود، در طرحواره موضوعی که محتوای موجودیت کتابی مورد نظر در آن توصیف شده، لازم است با لانه‌سازی گسترش شکل‌های ناگزیده اصطلاح‌های موضوعی در برچسب یا صفت variant، هنگامی که موضوع درباره شخص است: ذیل نوع موجودیت Person، و زمانی که موضوع درباره سازمان است ذیل نوع موجودیت Organization، جاسازی گردد، و در مواقعی که درباره هیچ‌یک از این دو نوع موجودیت نیست و موضوع ماهیت مفهومی دارد، گسترش شکل‌های ناگزیده اصطلاح‌های موضوعی نیز مناسب است که به صورت صفت variant، ذیل نوع موجودیت Book، و صفت about که در بردارنده شکل‌های پذیرفته شده موضوع پیشینه فراداده‌ای است، درج شود.



نمودار ۱. وضعیت کلی نمایه‌سازی و بازیافت‌پذیری گروه‌های گواه و آزمون به روش داده‌های خرد و جیسان-ال‌دی. در چهار موتور کاوش مورد مطالعه

برای طرحواره موضوعی که محتوای موجودیت کتاب مورد نظر در آن توصیف شده، با رویکرد لانه‌سازی، گسترش شکل‌های ناگزیده اصطلاح‌های موضوعی در برچسب یا صفت variant جاسازی شود، با این تمایز که همانند آنچه تصویر ۵-۲ نشان می‌دهد برای موضوع‌هایی درباره شخص، شکل‌های نام‌رچ موضوع در برچسب جدید مذکور، ذیل نوع موجودیت Person و برای موضوع‌هایی درباره سازمان‌ها صفت variant، ذیل نوع موجودیت Organization درج گردد. و هنگامی که موضوع در رابطه با هیچ‌یک از این دو نیست و صرفاً ماهیت مفهومی دارد، بهتر است گسترش شکل‌های ناگزیده اصطلاح‌های موضوعی که به صورت صفت variant ذیل نوع موجودیت کتاب و صفت about آن، که حاوی شکل‌های پذیرفته شده موضوع پیشینه‌های فراداده‌ای است، واقع شود.

```
<div itemprop="author" itemscope itemType="http://schema.org/Person">
  <span itemprop="name">عاطلی، سعیدرضا</span>
  <span itemprop="variant">رنانی، سعیدرضا</span>
  <span itemprop="variant">عاطلی، سعیدرضا</span>
  <span itemprop="variant">عاطلی رنانی، سعید</span>
  <span itemprop="birthDate">۱۳۴۰</span>
</div>

<div itemprop="author" itemscope itemType="http://schema.org/Organization">
  <span itemprop="name">بونسکر</span>
  <span itemprop="Variant">سازمان آموزشی، علمی و فرهنگی ملل متحد</span>
  <span itemprop="Variant">سازمان ملل متحد، سازمان آموزشی، علمی و فرهنگی</span>
</div>
```

شکل ۱. گسترش مستند نام پدیدآورنده شخص و نام سازمان به روش داده‌های خرد

```

<div itemscope itemtype="http://schema.org/Book">
  <span itemprop="name">خبریت پرخاشگری (بازنیمه آزمون کنترل خشم)</span>
  <span itemprop="about">پر خاشگری</span>
  <span itemprop="Variant">پر خاش</span>
  <span itemprop="Variant">پر خاشجونی</span>
  <span itemprop="Variant">پر خاشگری (روانشناسی)</span>
  <span itemprop="Variant">تجواز گاری</span>
  <span itemprop="Variant">تو هنر (روانشناسی)</span>
  <span itemprop="Variant">تو نامج (روانشناسی)</span>
  <span itemprop="Variant">تو قار پر خاشجونی</span>
  <span itemprop="Variant">تو و ان‌شناسی پر خاشگری</span>
  <span itemprop="Variant">سنگز دگری (روانشناسی)</span>
</div>

<div itemprop="about" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">طباطبائی، سیدمحمدحسین، ۱۳۶۰ - ۱۳۸۱</span>
  <span itemprop="Variant">طباطبائی، سیدمحمدحسین، ۱۳۶۰-۱۳۸۱</span>
  <span itemprop="Variant">طباطبائی تهریزی، محمدحسین، ۱۳۶۰-۱۳۸۱</span>
  <span itemprop="Variant">علاء طباطبائی، محمدحسین، ۱۳۶۰-۱۳۸۱</span>
  <span itemprop="Variant">قاضی طباطبائی تهریزی، محمدحسین، ۱۳۶۰-۱۳۸۱</span>
</div>

<div itemprop="about" itemscope itemtype="http://schema.org/Organization">
  <span itemprop="name">سازمان اسسکو</span>
  <span itemprop="Variant">اسسکو</span>
  <span itemprop="Variant">اسسکو</span>
  <span itemprop="Variant">سازمان کنفرانس ایلاتی، سازمان آموزشی، علمی و فرهنگی اسلامی</span>
  <span itemprop="Variant">مجلسة الإسلامية للدراسة و العلوم و الثقافة</span>
</div>

```

شکل ۲. گسترش مستند موضوع با ماهیت مفهومی، نام شخص و نام سازمان به روش داده‌های خرد

این شیوه جاسازی با بهره‌گیری از راهبرد لانه‌سازی، ضمن تطابق با معیارها و قواعد فرامای تولید داده‌های ساختارمند (اسکیما.آرگ) زمینه‌ای را برای اعمال کنترل مستندات نام شخص و سازمان و کنترل مستندات موضوعی در سطح وب فراهم ساخته و به تبع آن موتورهای کاوش نیز در واکنشی مثبت به برچسب طراحی شده امکان بازیابی نتایج مرتبط‌تر را بالفعل خواهند کرد، بدین ترتیب که با جستجوی هر یک از شکل‌های مرجع و یا دیگر شکل‌های ناگزیده نام‌های اشخاص حقیقی و حقوقی پیشینه‌های فراداده‌ای مربوط به هر کدام بازیابی خواهند شد. این امر درخصوص گسترش داده‌های مستند موضوعی نیز به همان صورت صادق است، و با جستجوی هر کدام از موضوع‌های پذیرفته شده یا دیگر شکل‌های نامرجحی که ارزش‌های برچسب جدید هستند پیشینه‌های فراداده‌ای مرتبط با هر یک در نتایج جستجو بازیابی می‌گردد.

لازم به ذکر است، برای تأیید اعتبار صفت variant توسط موتورهای کاوش و ابزارهای فرامای تولید داده‌های ساختارمند (اسکیما.آرگ)، به منزله صفت رسماً تعریف شده و معتبر در فرامای مرتبط (مثلاً person یا book)، و نمایش معنادار آن در قالب خرده‌های غنی یا گراف دانش در نتایج موتورهای کاوش وب، باید از فرامای PropertyValue به عنوان ارزش صفتی مانند additionalProperty که با رویکرد لانه‌سازی (تو در تو نویسی) در فرامای اصلی (هدف) مانند book با صفات name و value جاسازی شده است، بهره برد. بدین گونه که در مثال ذیل نمایش داده شده است:

```

<div itemscope itemtype="https://schema.org/book">
  <span itemprop="about">Digital Library</span>
  <div itemprop="additionalProperty" itemscope itemtype="https://schema.org/PropertyValue">
    <span itemprop="name">Variant</span>
    <span itemprop="value">Automatic Library</span>
  </div>
</div>

```

نتیجه‌گیری

موتورهای کاوش به منظور پاسخگویی به نیازهای اطلاعاتی متنوع کاربران در محیط وب ایجاد شده و توسعه یافته‌اند. با وجود به کارگیری روش‌های غنی‌سازی داده‌ها همچون داده‌های خرد و روش توصیه شده جی‌سان-ال.دی، حجم روزافزون محتوای اطلاعاتی وب و داده‌های غیرساختارمند، عدم یکپارچگی و کنترل واژگانی این محیط، تنوع ثبت شکل‌های نام افراد و تنالگان‌ها به علاوه استفاده از نام‌های مستعار و جز آن و تنوع واژگان نماینده مفاهیم موضوعی و... موتورهای کاوش را همچنان با چالش‌هایی در بازیابی اطلاعات مرتبط و دقیق روبرو ساخته، و منجر به کاهش کارایی روش‌ها و استانداردهای نشانه‌گذاری معنایی مذکور گشته است. یافته‌های این پژوهش نشان داد موتور کاوش گوگل قادر است تمامی پیشینه‌های فراداده‌ای گروه‌های گواه و آزمون در هر دو روش نشانه‌گذاری معنایی داده‌های خرد و جیسان-ال.دی. را شناسایی، نمایه‌سازی و بازیابی کند. پیشینه‌های فراداده‌ای در گروه‌های گواه به دلیل عدم وجود داده‌های مستند تنها از طریق شکل مرجع و پذیرفته شده قابل بازیابی بودند. این در حالی است که در گروه آزمون هم از طریق شکل‌های مرجع و هم شکل‌های نامرجح بازیابی امکان‌پذیر

است. به طوری که جستجوی ارزش‌های صفت variant در موتور کاوش گوگل، منجر به بازیابی پیشینه‌های فراداده‌ای مرتبط در صفحه نتایج جستجو گردید. این امر برای داده‌های مستند نام‌ها و مستند موضوعی دارای وضعیت یکسانی بود. بنابراین می‌توان گفت که گسترش طراحی شده داده‌های مستند نام و موضوع از نظر خزنده‌ها یا روبات‌های موتور کاوش گوگل دارای ارزش نمایه شدن^۱ و بازیابی بوده است.

نرم‌افزار روبات موتور کاوش بینگ نیز در واکنش مثبت قادر به شناسایی، نمایه‌سازی و بازیابی تمامی پیشینه‌های فراداده‌ای گروه‌های گواه و آزمون مبتنی بر روش‌های نشانه‌گذاری معنایی داده‌های خرد و جیسان-ال.دی. بود. هر چند که سرعت نرم‌افزارهای خزنده و نمایه‌ساز بینگ در نمایه‌سازی محتوای صفحات وب کمتر از گوگل است. بدین معنی که ارزش‌های صفت variant حاوی داده‌های مستند غیرمرجع یا انتخاب نشده، برای موتور کاوش بینگ قابل شناسایی، و ارزش نمایه‌ای داشته، و قابل بازیابی هستند. همانند موتورهای کاوش گوگل و بینگ، موتور کاوش یاهو، تمامی پیشینه‌های فراداده‌ای گروه گواه و آزمون مبتنی بر هر دو روش نشانه‌گذاری معنایی داده‌های خرد و جیسان-ال.دی. نمایه‌سازی کرده و در نتایج جستجو پدیدار ساخته است. بنابراین موتور کاوش یاهو نیز همانند موتورهای کاوش پیشین، به برچسب (صفت) variant واکنش مثبت نشان داده، و فقط امکان بازیابی پیشینه‌های گروه‌های آزمون از طریق ارزش‌های داده‌ای مستند نام‌رجح را میسر ساخته است. نیز یافته‌ها بیانگر توانایی نرم‌افزار خزنده-نمایه‌ساز موتور کاوش یاندکس به پیشینه‌های فراداده‌ای مبتنی بر فرامای تولید داده‌های ساختارمند، به ویژه گسترش صفت variant و راهبرد لانه‌سازی برای پشتیبانی از فرایند کنترل مستندات، و به طور کلی واکنش مثبت آن موتور کاوش به پیشینه‌های فراداده‌ای مورد مطالعه بوده است. در موتور کاوش یاندکس نیز همچون سه موتور کاوش پیشین نمایه‌سازی و بازیافت‌پذیری ۱۰۰ درصد پیشینه‌های فراداده‌ای گروه‌های گواه و آزمون به هر دو روش نشانه‌گذاری معنایی داده‌های خرد و جیسان-ال.دی صورت گرفت. برآیند این یافته‌ها نشان می‌دهد، افزودن صفت جدید variant به طرحواره‌های (واژگان‌های) فرامای تولید داده‌های ساختارمند، امکان کنترل مستندات در محیط وب، به مثابه یکی از بزرگترین چالش‌های محیط وب (گورمن^۲، ۲۰۰۴)، آن هم در پرکاربردترین ابزار کاوش محتوای آن یعنی موتورهای کاوش، میسر می‌سازد. بر این اساس، افزون بر نمایه‌سازی و بازیابی شکل‌های گزیده نام‌ها و اصطلاحات به منزله ارزش‌های صفات موجود در واژگان‌های اسکیمای آرگ، شکل‌های ناگزیده نام‌ها و اصطلاحات به مثابه ارزش‌های صفت variant که گسترش جدید بشمار می‌آید نیز نمایه‌پذیر و بازیافت‌پذیر شده، افزایش دسترس‌پذیری پیشینه‌های فراداده‌ای در پی خواهد داشت. بدین‌گونه امکان جستجوهای معنایی^۳ در موتورهای کاوش به وجود می‌آید.

به طور کلی می‌توان بیان داشت، موتورهای کاوش مورد مطالعه به روش ایجاد صفت مرتبط با داده‌های مستند به همان شکلی که در پیشینه‌های فراداده‌ای گروه‌های آزمون برای گسترش مستند نام شخص و سازمان، و گسترش مستند موضوعات در هر دو روش نشانه‌گذاری معنایی داده‌های خرد و جیسان-ال.دی. از طریق رویکرد لانه‌سازی صورت پذیرفت واکنش مطلوب و مناسبی نشان می‌دهند. بدین ترتیب طبق آن چه که تصویر ۱ نشان می‌دهد لازم است، ذیل نوع موجودیت Person گسترش شکل‌های ناگزیده نام پدیدآورنده اشیای محتوایی از جمله نویسنده، مترجم، ویراستار و جز آنها در صفت جدید variant (که چنانچه پیشتر بیان شد از طرح فراداده‌ای داده‌های مستند (مدس) گرفته شده است)، با رویکرد لانه‌سازی افزوده شود. به همین ترتیب برای ایجاد گسترش مستند نام سازمان نیز باید زیر نوع موجودیت Organization شکل‌های ناگزیده نام سازمان را در برچسب جدید variant درج کرد.

در فرایند نمایه‌سازی و بازیابی پیشینه‌های فراداده‌ای توسط موتورهای کاوش یافته‌های پژوهش حاضر با نتایج پژوهش طاهری، حریری، فتاحی (۱۳۸۸)، طباطبایی امیری و همکاران (۱۳۹۰)، باب‌الحوائجی، طاهری، آقاعابدی (۱۳۹۴)، حریری و طاهری (۱۳۹۴) و حریری و طاهری (۱۳۹۵) که در این پژوهش‌ها وضعیت نمایه‌سازی پیشینه‌های فراداده‌ای هسته دابلین و

1. Indexable

2. Gorman

3. Semantic search

مارک ۲۱ که مبتنی بر ایکس.ام.ال. بودند را مورد ارزیابی قرار داده بودند و نیز پژوهش طاهری، نیکزاد، صمیعی (۱۳۹۶) و طاهری، ذوالقدر و صمیعی (۱۳۹۷) همسو است. در مقایسه با پژوهش‌های گروه دوم که به ارزیابی موتورهای کاوش عمومی و معنایی با یکدیگر پرداخته‌اند نیز به طور کلی می‌توان چنین بیان کرد که در راستای معنایی شدن موتورهای کاوش، به تدریج موتورهای کاوش عمومی هم از عملکردی معنایی برخوردار می‌شوند و پیشینه‌های مبتنی بر نشانه‌گذاری‌های معنایی فرانمای تولید داده‌های ساختارمند (اسکیما.آرگ) را به صورت معنادار نمایه‌سازی، بازیابی و پدیدار می‌نمایند. به عبارت دیگر، توسعه فرانمای تولید داده‌های ساختارمند برای تبدیل موتورهای کاوش عمومی غیرمعنایی به موتورهای کاوش معنایی بوده است. ارزیابی واکنش چهار موتور کاوش گوگل، یاهو، بینگ و یاندکس با رویکرد مقایسه‌ای نیز برای بیان این مطلب بوده است. در میان پیشینه‌های پژوهش، تنها الدائج (۲۰۱۵) اقدام به ایجاد طرحواره و گسترش کرده است با این تفاوت که او در زمینه محیط آموزشی و یادگیری آنلاین و مجازی، با استفاده از ابزار سیستم مدیریت محتوای معنایی دروپال، همچون پژوهش حاضر دریافت که با به کارگیری استانداردها و رعایت معیارهای فرانمای تولید داده‌های ساختارمند (اسکیما.آرگ) می‌توان طرحواره و گسترش ایجاد کرد تا زمینه ارتقا بازیافت‌پذیری اشیای محتوایی را فراهم ساخت.

در مورد محدودیت‌های پژوهش می‌توان بیان داشت، موتورهای کاوش به دلیل بحث رقابتی، ماهیت تجاری، خطمشی‌های محرمانه و خاصی در نمایه‌سازی، رتبه‌بندی و پیدانمایی اشیای محتوایی دارند و شفاف عمل نمی‌کنند. از این‌رو، در فرایندهای نمایه‌سازی، رتبه‌بندی و پیدانمایی، الگوریتم‌ها و نیز بسیاری از اطلاعات فنی را پنهان می‌نمایند. بدین ترتیب پژوهشگران برای نیل به نتیجه، ناگزیر از مطالعه متون و پژوهش‌های گذشته و در مواردی اجرای آزمون و خطا بودند. این مسئله بر تعیین روش و اجرای پژوهش تأثیرگذار است. از سوی دیگر، استفاده از فراهم‌کنندگان خدمات اینترنتی (آی.اس.پی.) مختلف به دلیل مبتنی بودن موتورهای کاوش بر طیف آی.پی. جغرافیایی و محدودیت‌های فراهم‌کنندگان یاد شده، نتایج متفاوتی در بردارد، پژوهشگران ناچار به انجام پژوهش در یک مکان با آی.اس.پی واحد بودند. از دیگر محدودیت‌های پژوهش، محدودیت تجربیات موجود در زمینه پیاده‌سازی گسترش‌های داده‌های مستند با استفاده از روش‌های توصیه شده بر پایه فرانمای تولید داده‌های ساختارمند است که به نوبه خود محدودیت‌هایی را در اجرای فرایند پژوهش ایجاد می‌کند.

به طور کلی، یافته‌ها حاکی از سودمند واقع شدن اعمال کنترل مستندات نام و موضوع در کنار افزودن معنا به محتوای مذکور با نشانه‌گذاری‌های معنایی (ساختارمندسازی داده‌ها) در فرایند نمایه‌سازی و بازیابی توسط موتورهای کاوش در سطح وب است. به بیانی دیگر با ایجاد امکان بازیابی هر یک از ارزش‌های نامرئج افزون بر ارزش‌های شکل پذیرفته شده نام و اصطلاح موضوعی، بستر مناسبی برای جامعیت فرایند بازیابی، و بسط جستجو و بازیابی در محیط نتایج مرتبط در موتورهای کاوش فراهم می‌شود. انتظار می‌رود استفاده از داده‌های مستند و اعمال کنترل مستندات در شناسه‌های اصلی نام و موضوع و دیگر شناسه‌های مورد استفاده در بازیابی، در کنار آن بهره‌مندی از روش‌های توصیه شده نشانه‌گذاری معنایی مانند روش داده‌های خرد و جی‌سان-ال.دی.، منتج به ایجاد انسجام و یکپارچگی معنایی، افزایش ربط و بسط جستجو و بهبود بازیابی، مؤثرتر واقع شدن استانداردها و در مجموع افزایش میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش در سطح وب گردد. براساس نتایج حاصل از این پژوهش بررسی امکان پیدانمایی معنادار گسترش‌های مرتبط با داده‌های مستند در نتایج موتورهای کاوش وب از جمله گراف دانش گوگل پیشنهاد می‌شود.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده‌اند و این موضوع مورد تأیید همه آنهاست.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

سپاسگزاری

از معاونت محترم پژوهشی دانشگاه علامه طباطبائی به خاطر حمایت معنوی در اجرای پژوهش حاضر سپاسگزاری می‌شود.

منابع

- باب‌الحوائجی، فهیمه؛ سیدمهدی طاهری و زهرا آقاعابدی (۱۳۹۴). بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای هسته دابلین و مارک ۲۱ توسط موتورهای کاوش عمومی. *مطالعات دانش‌شناسی*، ۱(۳)، ۴۳-۵۹
- جمالی مهموئی، حمیدرضا (۱۳۸۲). وب معنایی: شیوه‌های رو به تکامل برای ذخیره و بازیابی کارآمدتر اطلاعات روی اینترنت. *اطلاع‌شناسی*، ۱(۲)، ۴۷-۶۶
- حریری، نجلا؛ طاهری، سید مهدی (۱۳۹۳). بررسی کاربرد روش جزیره داده‌ها در تولید پیشینه‌های فراداده‌ای با قابلیت نمایه پذیری و پیدانمایی نام‌های برچسب عناصر در محیط موتورهای کاوش وب: بهبود دسترسی به اشیای محتوایی. *پژوهشنامه کتابداری و اطلاع‌رسانی*، ۴(۱)، ۹۱-۷۳
- حریری، نجلا؛ طاهری، سید مهدی (۱۳۹۴). بررسی تأثیر حذف خط اعلان زبان نشانه‌گذاری گسترش‌پذیر (XML) و تغییر پسوند فایل بر افزایش نمایه‌پذیری و پیدانمایی پیشینه‌های فراداده‌ای در محیط موتورهای کاوش وب. *پژوهشنامه کتابداری و اطلاع‌رسانی*، ۵(۱)، ۱۰-۲۱.
- دری، راحله (۱۳۹۳). مقایسه و ارزیابی موتورهای جستجوی معنایی، *پردازش و مدیریت اطلاعات*، ۳۰(۲)، ۴۶۷-۴۹۰.
- طاهری، سیدمهدی (۱۳۹۳). کارگاه آموزشی "خرده‌های غنی: داده‌های خرد، قالب‌های خرد، و چهارچوب توصیف منبع (خصایص)". برای موسسه خانه کتاب. تهران: بیست و هفتمین نمایشگاه بین‌المللی کتاب، سرای اهل قلم، ۱۲ اردیبهشت ۱۳۹۳.
- طاهری، سیدمهدی (۱۳۹۴). کارگاه آموزشی "پژوهش در محیط وب (۵)". برای دانشکده روانشناسی دانشگاه علامه طباطبائی (ره). تهران، ۳۰ اردیبهشت ۱۳۹۴.
- طاهری، سیدمهدی؛ حریری، نجلا و فتاحی، رحمت‌الله (۱۳۸۸). بررسی تطبیقی کیفیت نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای هسته دابلین و مارک ۲۱ توسط موتورهای کاوش عمومی. *کتابداری و اطلاع‌رسانی*، ۱۲(۴)، ۱۴۱-۱۶۲.
- طاهری، سید مهدی؛ ذوالقدر، سارا و صمیعی، میترا (۱۳۹۷). مقایسه عملکرد موتورهای کاوش وب در نمایه‌سازی و پیدانمایی پیشینه‌های فراداده‌ای مبتنی بر روش داده‌های خرد. *بازیابی دانش و نظام‌های معنایی*، ۴(۱۶)، ۸۳-۱۰۱.
- 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)، ۶۵۸-۶۳۹.
- طباطبائی امیری، فائزه‌السادات؛ طاهری، سیدمهدی؛ فرج‌پهلوی، عبدالحسین؛ عصاره، فریده و معرف‌زاده، عبدالحمید (۱۳۹۱). موتورهای کاوش وب و نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای دسترس‌پذیر در محیط‌های اطلاعاتی پیوسته پویا. *پژوهشنامه پردازش و مدیریت اطلاعات*، ۲۷(۴)، ۹۰۷-۹۲۰.
- فتاحی، رحمت‌الله (۱۳۸۵). از آرمان‌ها تا واقعیت: تحلیلی از مهمترین چالش‌ها و رویکردهای سازماندهی اطلاعات در عصر حاضر. *کتابداری و اطلاع‌رسانی*، ۴۰(۱۰)، ۲۲-۳۷
- فردحسینی، مهسا؛ طاهری، سیدمهدی؛ حریری، نجلا؛ باب‌الحوائجی، فهیمه و نوشین فرد، فاطمه (۱۳۹۹). بازنمایی صفات و روابط میان موجودیت‌های آثار خلاقانه فرانمای تولید داده‌های ساختارمند مبتنی بر الگوی مرجع کتابخانه‌ای (ال‌آرام). *پژوهشنامه پردازش و مدیریت اطلاعات*، ۳۶(۲)، ۵۳۳-۵۶۲.
- نوکاریزی، محسن، و زینالی تازه‌کندی، مهدی. (۱۳۹۶). ارزیابی همپوشانی و پوشش چهار موتور جستجوی بومی اصلی: پارسی جو، یوز، پارسیک و ریسمون. *تعامل انسان و اطلاعات*، ۴(۳)، ۴۸-۵۹.

References

- Aldaej, A. A. (2015). *An enhanced semantic VLE based on schema.org and social media* (Doctoral dissertation, University of Surrey). Retrieved 28 March 2021 from <https://epubs.surrey.ac.uk/807070/1/Aldaej%20PhD%20Thesis.pdf>
- Azizan, Z. Baker, A. Ismail, N., & Amran M. (2013). Interface Features of Semantic Web Search Engine. *IEEE Conference on e-Learning, e-Management and e-Services*, December 2–4, Sarawak, Malaysia.
- Babolhavaeji, F., Taheri, S. M., & AghaAbedi, Z. (2015). A comparative Study of Indexing Quality and Ranking of Metadata Records of Dublin Core and MARC 21 by Search Engines. *Quarterly of Knowledge Studies*, 1(3). 43-59. (in Persian)
- Dorri, R. (2015). Comparison and Evaluation of Semantic Search Engines. *Iranian Journal of Information Processing and Management*. 30(2), 467-490 (in Persian).
- Fardehosseini, M., Taheri, S.M., Hariri, N., Babalhavaeji, F., & Nooshinfard, F. (2020). Representing Properties and Relationships between Entities of Creative Works in Schema.org Based on Library Reference Model (LRM). *Iranian Journal of Information Processing and Management*, 36(2), 533-562. (in Persian).
- Fattahi, S. R. (2007). From ideals to reality; an analysis of the most important challenges and approaches to organizing information in today's era. *Library and Information Sciences*, 40(10), 22-37. (in Persian).
- Gorman, M. (2004). Authority control in the context of bibliographic control in the electronic environment. *Cataloging & classification quarterly*, 38(3-4), 11-22.
- Hariri, N. & Taheri, S.M. (2014). Study on the Use of "Data Island Method" for Creating Metadata Records with Access Improvement to Content Objects by Web Search Engines. *Library and Information Science Quarterly*, 4(1), 73-91 (in Persian).
- Hariri, N. & Taheri, S. M. (2015). Study of the Effect of Discarding XML Declaration and Changing the File Extension on Increasing the Indexability and Visibility of Metadata Records in the Web Search Engines Environment. *Library and Information Science Quarterly*, 5 (1), 10-21. (in Persian).
- Hogenboom, F. Frasinca, F. & Damir, V. (2011). *Automatically Annotating Web pages using Google Rich Snippets*. Econometric institute, Erasmus University Rotterdam, Amsterdam, the Netherlands.
- Iliadis, A., Acker, A., Stevens, W., & Kavakli, S. B. (2022). One schema to rule them all: How Schema.org models the world of search. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24744>
- MohammadiOstani, M., CheshmehSohrabi, M., Taheri, S. M., Shabani, A., & Asemi, A. (2022). Localization of Schema.org for Manuscript Description in the Iranian-Islamic Information Context. *Knowledge Organization*, 48(5), 345-356.
- Negi, Y., & Kumar, S. (2014). A comparative analysis of keyword- and semantic-based search engines. *Intelligent Computing, Networking, and Informatics*, (243), 727-736.
- Nowkarizi, M., & Zeinali, M. (2017). The overlap and coverage of 4 local search engines: Parsijoo, Yooz, Parseek and Rismoun. *Human Information Interaction*, 4(3), 48-59 (in Persian).
- Tabatabaei Amiri, F.S., Taheri, S.M., & Farajpahlu, H. (2012). Web search engines, indexing and Ranking of content objects including metadata elements in dynamic environment. *Iranian Journal of Information Processing and Management*. 27(4), 907-920. (in Persian)
- Taheri, S. M. (2014). *Rich snippets: Macro data, macro formats, and resource description framework (attributes) Workshop*. For the Khaneeye Ketab Institute. Tehran: 27th International Book Fair, Sarai Ahl Qalam, May, 2014 (in Persian)

- Taheri, S. M. (2015). *Research in the web environment (5)" Workshop*. The Faculty of Psychology of Allameh Tabatabae'i University, Tehran, May, 2015 (in Persian)
- Taheri, S. M.; Hariri, N., & Fattahi, S. R. (2009). A comparative study of the quality of indexing and ranking of content objects including Dublin Core metadata elements and MARC 21 by general search engines. *Library and Information Sciences*, 12(4), 141-162 (in Persian).
- Taheri, S. M., Nikzad, R., & Samiee, M. (2015). Study of Response of web-based search engines to metadata based on combined method of rich snippets and linked data: A case study of national content consortium. *Iranian Journal of Information Processing and Management*, 33(2), 639-658. (in Persian).
- Taheri, S. M., Zolghadr, S., & Hariri, N. (2018). Comparing the function of web search engines in indexing and finding the metadata records based on the microdata method. *Knowledge Retrieval and Semantic Systems*, 5(16), 83-101 (in Persian). <https://doi.org/10.22054/jks.2019.38110.1210>
- Turner, D., M. Shah, A., & Bitirim, Y. (2009). An empirical evaluation on semantic search performance of keyword-based and semantic search engines: Google, Yahoo, MSN and Hakia. In *proceedings of Fourth international conference on internet monitoring and protection*, 24-28. Venice, Mestre. <https://doi.org/10.1109/ICIMP.2009.16>
- Van, T. (2012). *How to implement video microdata for Google & Schema.org*. Retrieved 10 June 2021, from <https://www.searchenginewatch.com/2012/02/28/how-to-implement-video-microdata-for-google-schema-org/>
- W3Schools (2017). *HTML5 History*. Retrieved 15 May 2017, from: <https://www.w3schools.com/html/html5.asp>