

بررسی همخوانی فرآینمای تولید داده‌های ساختارمند برای توصیف حوزه نشر بر اساس
هستی‌شناسی‌های اسپار

فائزه السادات طباطبایی امیری

دانشجوی دکتری علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید چمران، اهواز، ایران

faeze.tabatabai@gmail.com

عبدالحسین فرج‌پهلوی*

استاد گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید چمران اهواز، ایران

farajpahlou@gmail.com

شهناز خادمی‌زاده

دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید چمران اهواز، ایران

s.khademi@scu.ac.ir

سیدمهدی طاهری

دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبایی، تهران، ایران

taherismster@gmail.com

مقدمه

در سال‌های اخیر، استفاده از استانداردهای فراداده‌ای نظیر فرآینمای تولید داده‌های ساختارمند (Schema.org) برای انتشار داده‌های ساختارمند در وب گسترش زیادی یافته است. این استاندارد به‌عنوان یکی از ابزارهای کلیدی برای نشانه‌گذاری داده‌های وب شناخته می‌شود و به موتورهای کاوش کمک می‌کند تا محتوای وب را بهتر درک و جستجو کنند. با این حال، به دلیل نیازهای تخصصی و پیچیده حوزه‌های مختلف، از جمله حوزه نشر، این استاندارد نیازمند گسترش و غنی‌سازی برای بهبود کاربردپذیری آن در این حوزه است. از سوی دیگر، در حوزه نشر، هستی‌شناسی‌های اسپار به‌عنوان یکی از مهم‌ترین ابزارهای توصیف فرایندها و موجودیت‌های مرتبط با نشر علمی شناخته می‌شوند. این هستی‌شناسی‌ها با ارائه واژگان دقیق و ساختارمند برای توصیف داده‌های مرتبط با نشر، امکان بهبود جستجو، کشف و یکپارچگی داده‌ها را فراهم می‌کنند. پژوهش حاضر با هدف بررسی همخوانی میان فرآینمای تولید داده‌های ساختارمند با هستی‌شناسی‌های اسپار انجام شده است. این پژوهش تلاش می‌کند تا با ایجاد نگاهی (جدول تطبیقی) میان این دو استاندارد، نقاط ضعف فرآینمای حوزه

نشر را شناسایی کرده و پیشنهادهایی برای غنی‌سازی آن ارائه دهد. هدف نهایی پژوهش، بهبود کاربردپذیری و دقت فرآیند تولید داده‌های ساختارمند در توصیف موجودیت‌ها و فرایندهای مرتبط با حوزه نشر است.

روش‌شناسی

پژوهش حاضر از نوع کاربردی بوده و از روش تحلیل محتوا برای بررسی همخوانی میان فرآیند تولید داده‌های ساختارمند با هستی‌شناسی‌های اسپار استفاده شد. براساس روش، «واحدهای ثبت» بررسی شده، صفات موجودیت‌های فرآیند تولید داده‌های ساختارمند و واحدهای معنا، کلاس‌ها و صفات موجود در هستی‌شناسی‌های اسپار هستند. جامعه آماری پژوهش شامل موجودیت‌ها و صفات فرآیند تولید داده‌های ساختارمند و کلاس‌ها و صفات هستی‌شناسی‌های اسپار بود. ابزار گردآوری داده‌ها، سیاهه واری بوده که براساس موجودیت‌ها و صفات هر دو استاندارد طراحی شد. روش گردآوری داده‌ها از طریق مشاهده ساختارمند توسط تیم پژوهشی انجام شد. در این مرحله، تمامی موجودیت‌ها و صفات فرآیند تولید داده‌های ساختارمند با کلاس‌ها و صفات هستی‌شناسی‌های اسپار تطبیق داده شدند. برای هر مورد از تطابق یا عدم تطابق، اطلاعات ثبت و تحلیل شدند. در نهایت، با استفاده از تکنیک‌های تحلیل کیفی، داده‌ها بررسی و تحلیل شدند. هدف اصلی این روش، شناسایی نقاط ضعف فرآیند در پوشش نیازهای حوزه نشر بود. با این تطابق، عناصر متناظر و عناصر فاقد تطابق شناسایی شدند تا میزان همخوانی بین این دو استاندارد مشخص شود. سپس، براساس نتایج به‌دست‌آمده، پیشنهادهایی برای غنی‌سازی فرآیند تولید داده‌های ساختارمند ارائه گردید.

یافته‌های اصلی

یافته‌های پژوهش نشان داد که میان فرآیند تولید داده‌های ساختارمند و هستی‌شناسی‌های اسپار تطابق چندانی وجود ندارد. بطوریکه از میان ۷۰۰ کلاس و صفت موجود در هستی‌شناسی‌های اسپار، ۴۸۷ کلاس و صفت معادل مشخصی در فرآیند تولید داده‌های ساختارمند نداشتند. این عدم تطابق به دلیل تفاوت در اهداف و دامنه دو استاندارد است. فرآیند تولید داده‌های ساختارمند به‌عنوان استاندارد عمومی برای توصیف داده‌های مختلف در وب طراحی شده است، در حالی که هستی‌شناسی‌های اسپار به‌طور خاص برای توصیف فرایندها و موجودیت‌های مرتبط با نشر علمی طراحی شده‌اند. این تفاوت در دامنه و هدف، منجر به عدم پوشش کامل نیازهای حوزه نشر توسط فرآیند شده است. به‌ویژه، کلاس‌ها و صفاتی مانند «اجزای سند»، «نقش‌های علمی» و «توصیف عناصر گفتمان» که برای توصیف دقیق داده‌های نشر ضروری هستند، در فرآیند تولید داده‌های ساختارمند به‌درستی تعریف نشده‌اند. با این حال، یافته‌ها نشان داد که برخی از کلاس‌ها و صفات مشترک نیز میان این دو استاندارد وجود دارد. به‌طور کلی، پژوهش نشان داد که برای بهبود کاربردپذیری فرآیند تولید داده‌های ساختارمند در حوزه نشر، غنی‌سازی آن با استفاده از هستی‌شناسی‌های اسپار ضروری است.

بحث و نتیجه‌گیری

نتایج این پژوهش نشان می‌دهد که غنی‌سازی فرمانمای تولید داده‌های ساختارمند با استفاده از هستی‌شناسی‌های اسپار می‌تواند به بهبود دقت و کیفیت داده‌های مرتبط با حوزه نشر منجر شود. با استفاده از کلاس‌ها و صفات تخصصی هستی‌شناسی‌های اسپار، می‌توان فرمانمای تولید داده‌های ساختارمند را به گونه‌ای توسعه داد که قادر به پوشش جامع‌تر نیازهای حوزه نشر باشد. این غنی‌سازی می‌تواند به بهبود میان‌کنش‌پذیری داده‌ها در پلتفرم‌های مختلف نشر کمک کند و قابلیت جستجو و کشف محتوای علمی را افزایش دهد. همچنین، این پژوهش نشان داد که ایجاد نگاشت (جدول تطبیقی) میان این دو استاندارد می‌تواند به یکپارچگی داده‌ها و بهبود قابلیت‌های ماشین‌خوانی آن‌ها منجر شود. با بهبود دقت در توصیف موجودیت‌ها و فرایندهای نشر، موتورهای کاوش نیز قادر خواهند بود تا نتایج دقیق‌تری ارائه دهند و تجربه کاربری بهتری ایجاد شود. در نهایت، غنی‌سازی فرمانمای تولید داده‌های ساختارمند نه تنها به بهبود وضعیت فعلی نشر کمک خواهد کرد، بلکه می‌تواند به استانداردسازی و توسعه روش‌های نوین انتشار معنایی در حوزه نشر دیجیتال منجر شود. یکی از پیشنهاد‌های پژوهش آن است که نگاشت‌های انجام‌شده، در پژوهش‌های آتی به صورت عملی ارزیابی شوند تا تأثیر آن‌ها بر بهبود کیفیت انتشار داده‌های معنایی در حوزه نشر بررسی شود.

کلیدواژه‌ها: هستی‌شناسی‌های اسپار، فرمانمای تولید داده‌های ساختارمند، داده‌های ساختارمند، نشر دیجیتال،

انتشار معنایی

A study of the alignment of Schema.org for describing the publishing domain based on SPAR ontologies

Faezeh Sadat Tabatabai Amiri

Ph.D. Candidate, Knowledge and Information Science, Shahid Chamran University, Ahvaz, Iran. faeze.tabatabai@gmail.com

Abdolhossein Farajpahlou

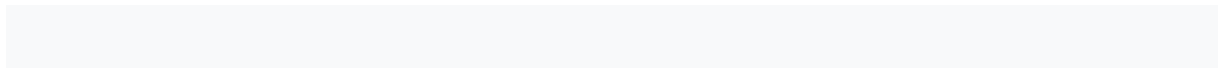
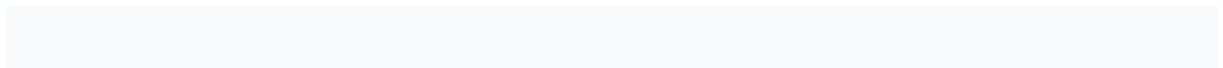
Professor Emeritus, Department of Knowledge and Information Science, School of Education & Psychology, Shahid Chamran University of Ahvaz, Ahvaz, Iran. farajpahlou@gmail.com

Shahnaz Khademizadeh

Associate Professor, Knowledge and Information Science, Department of Knowledge and Information Science, Faculty of Psychology and Educational Sciences, Shahid Chamran University of Ahvaz, Ahvaz, Iran. s.khademi@scu.ac.ir

Sayyed Mahdi Taheri

PhD in Knowledge and Information Science; Associate Professor; Faculty of Psychology and Educational Sciences; Allameh Tabataba'i University; Tehran, Iran. Email: taherismster@gmail.com



abstract

Introduction: In recent years, the utilization of metadata standards such as the Schema.org for publishing structured data on the web has expanded significantly. This standard is recognized as a key tool for marking up web data and assists search engines in better understanding and searching web content. However, due to the specialized and complex needs of various domains, including the publishing domain, this standard requires expansion and enrichment to improve its usability in this area. On the other hand, in the publishing domain, SPAR ontologies are recognized as one of the most important tools for describing processes and entities related to scholarly publishing. These ontologies, by providing precise and structured vocabularies for describing publishing-related data, enable improved search, discovery, and data integration. This research aims to investigate the alignment between the Schema.org and SPAR ontologies. This study seeks to identify the weaknesses of the schema.org in the publishing domain by creating a Crosswalk between the two standards and to provide suggestions for enriching it. The ultimate goal of the research is to improve the usability and accuracy of the schema.org in describing entities and processes related to the publishing domain.

Methodology: This study is applied research that employs content analysis to investigate the alignment between the Schema.org and SPAR ontologies. According to the methodology, the "units of analysis" examined are the attributes of entities in the Schema.org and the semantic units, classes, and properties in SPAR ontologies. The study population consists of entities and properties of the Schema.org and classes and properties of SPAR ontologies. The data collection tool is a checklist designed based on the entities and properties of both standards. The data collection method was conducted through structured observation by the research team. At this stage, all entities and properties of the Schema.org were matched with classes and properties of SPAR ontologies. For each case of match or mismatch, information was recorded and analyzed. Finally, using qualitative analysis techniques, the data was examined and analyzed. The main objective of this method was to identify the weaknesses of the Schema.org in covering the needs of the publishing domain. Through this matching, corresponding elements and elements without a match were identified to determine the level of alignment between the two standards. Subsequently, based on the obtained results, suggestions were provided to enrich the Schema.org

Main findings: The research findings revealed a significant lack of alignment between the Schema.org and SPAR ontologies. Of the 700 classes and properties in SPAR ontologies, 487 did not have clear equivalents in the Schema.org. This mismatch is attributed to the differences in the objectives and scope of the two standards. Schema.org is designed as a general standard for describing various data on the web, while SPAR ontologies are specifically designed to describe processes and entities related to scholarly publishing. This difference in scope and purpose has resulted in the schema.org inability to fully cover the needs of the publishing domain. In particular, classes and properties such as "document components," "scholarly roles," and "discourse element descriptions," which are essential for accurately describing publishing data, are not adequately defined in schema.org. However, the findings showed that there are also some common classes and properties between the two standards. Overall, the research indicated that to improve the usability of schema.org in the publishing domain, it is necessary to enrich it using SPAR ontologies.

Discussion and conclusions: The results of this research demonstrate that enriching the Schema.org using SPAR ontologies can enhance the accuracy and quality of data related to the

publishing domain. By leveraging the specialized classes and properties of SPAR ontologies, Schema.org can be expanded to more comprehensively address the needs of the publishing domain. This enrichment can improve data interoperability across various publishing platforms and enhance the searchability and discoverability of scholarly content. Furthermore, this research indicated that creating a Crosswalks between these two standards can lead to data integration and improve their machine-readability. By improving the accuracy of describing publishing entities and processes, search engines will be able to provide more accurate results and create a better user experience. Ultimately, enriching Schema.org will not only improve the current state of publishing but can also lead to the standardization and development of new semantic publishing methods in the digital publishing domain. One of the research suggestions is that the conducted mappings be evaluated practically in future research to assess their impact on improving the quality of semantic data publishing in the publishing domain.

Keywords: SPAR ontologies, Schema.org, Structured data, Digital publishing, Semantic publishing