

Methodology Explanation of Schema.org and Analysis of its Approach to the Processing and Organization of Web Content Objects

Morteza Mohammadi Ostani

PhD Candidate in Knowledge and Information Science;
University of Isfahan Email: mmohamadiostani@gmail.com

Mozaffar Cheshmesohrabi*

PhD in Communication and Information Science; Associate
Professor; University of Isfahan Email: mo.sohrabi@edu.ui.ac.ir

Ahmad Shabani

PhD in Knowledge and Information Science; Professor;
University of Isfahan Email: shabania@edu.ui.ac.ir

Asefe Asemi

PhD in Knowledge and Information Science; Associate Professor;
University of Isfahan Email: asefi@edu.ui.ac.ir

Seyed Mahdi Taheri

PhD in Knowledge and Information Science; Assistant Professor;
Allameh Tabataba'i University Email: taherismster@gmail.com

**Iranian Journal of
Information
Processing and
Management**

Iranian Research Institute

for Information Science and Technology
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 34 | No. 4 | pp. 1767-1798

Summer 2019



Received: 28, Jul. 2018 | Accepted: 05, Dec. 2018

Abstract: The purpose of this conceptual research was to explain the capabilities, semantic platform and view point of Schema.org to processing and organization of web content objects (data entities) by analytical approach. To collect data documentary analysis was used. The research community included texts and researches related to the field of “structured data” and “Schema.org”. A total of 43 sources, as well as the official website of the “Schema.org” were selected using a purposive sampling method for analysis. The results of the survey showed that Schema.org is a common vocabulary that is used to describe and markup web content objects and create structured data for better processing and organization. It has a certain structure and semantic platform. Its structure is like an ontology for naming the types and properties of content objects, the relationships between types and properties, and the capabilities of describing these properties and relationships. Its semantic platform is adapted by semantic markups such as microformat, microdata, RDFa 1.1, and JSON-LD. The

* Corresponding Author

results of the research showed that there are three major approaches to the processing and organizing of content objects in the Schema.org: the ontological, context-oriented, and nesting approaches. Overall research results showed existence of different approaches to Schema.org represents a comprehensive view of the Web content objects while paying attention to improving interoperability with search engines. Also, the production of structured data with such schemata is an important contribution to the realization of semantic web or web of data.

Keywords: Schema.org, Semantic Web, Semantic Markup, Processing, Web Content Objects, Structured Data, Ontological Approach, Context-oriented Approach, Nested Approach



تبیین فرانمای تولید داده‌های ساختارمند (اسکیمادات آرک) و تحلیل رویکرد آن به پردازش و سازماندهی اشیای محتوایی وب

مرتضی محمدی استانی

دانشجوی دکتری علم اطلاعات و دانش‌شناسی؛
گروه علم اطلاعات و دانش‌شناسی؛ دانشگاه اصفهان؛
mmohamadiostani@gmail.com

مظفر چشمه‌سهرابی

دکتری علوم اطلاعات و ارتباطات؛ دانشیار؛
گروه علم اطلاعات و دانش‌شناسی؛ دانشگاه اصفهان؛
پدیده‌آور رابط mo.sohrabi@edu.ui.ac.ir

احمد شعبانی

دکتری علم اطلاعات و دانش‌شناسی؛ استاد،
گروه علم اطلاعات و دانش‌شناسی؛ دانشگاه اصفهان؛
shabania@edu.ui.ac.ir

عاصفه عاصمی

دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛
گروه علم اطلاعات و دانش‌شناسی؛ دانشگاه اصفهان؛
asemi@edu.ui.ac.ir

سید مهدی طاهری

دکتری علم اطلاعات و دانش‌شناسی؛ استادیار؛ گروه
علم اطلاعات و دانش‌شناسی؛ دانشگاه علامه طباطبائی؛
taherismster@gmail.com



دریافت: ۱۳۹۷/۰۵/۰۶ | پذیرش: ۱۳۹۷/۰۹/۱۴ | مقاله برای اصلاح به مدت یک روز نزد پدیدآوران بوده است.

چکیده: هدف از پژوهش مفهومی حاضر، تبیین قابلیت‌ها، بستر معناشناختی و نگرش فرانمای تولید داده‌های ساختارمند بر روی وب به پردازش و سازماندهی اشیای محتوایی وب (موجودیت‌های داده‌ای) با رویکرد تحلیلی بود. برای گردآوری داده‌ها از روش تحلیل اسنادی استفاده گردید. نتایج بررسی نشان داد که فرانمای تولید داده‌های ساختارمند بر روی وب، مجموعه‌ی واژگان مشترکی است که برای توصیف و نشانه‌گذاری اشیای محتوایی وب و ایجاد داده‌های ساختارمند به‌منظور پردازش و سازماندهی بهتر استفاده می‌شود. این فرانما دارای ساختار و بستر معناشناختی خاصی است. ساختار آن نظیر یک هستی‌نگاشت برای نام‌گذاری نوع‌ها و صفات اشیای محتوایی، روابط میان نوع‌ها و صفات، و قابلیت‌های چگونگی توصیف

فصلنامه | علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

شاپا (چاپی) ۲۲۵۱-۸۲۲۳

شاپا (الکترونیکی) ۲۲۵۱-۸۳۳۱

نماینده در SCOPUS، ISI، و LISTA

www.jipm.irandoc.ac.ir

دوره ۳۴ | شماره ۴ | صص ۱۷۶۷-۱۷۹۸
تابستان ۱۳۹۸



این صفات و روابط است. بستر معناشناختی آن نیز با نشانه‌گذارهای معنایی همچون ریزفرمت، ریزداده، RDFa 1.1، و JSON-LD سازگار است. دیگر نتایج پژوهش نشان داد که سه رویکرد عمده در پردازش و سازماندهی اشیای محتوایی در فرآیند تولید داده‌های ساختارمند بر روی وب به شرح: رویکرد هستی‌شناسانه، بافت‌مدار، و آشیانه‌ای (تودرتو) وجود دارد. نتایج کلی پژوهش نشان داد که وجود رویکردهای مختلف در فرآیند تولید داده‌های ساختارمند بر روی وب، نشانگر دیدی جامع به اشیای محتوایی وب در عین توجه به بهبود میانکنش‌پذیری با موتورهای کاوش است. همچنین، تولید داده‌های ساختارمند با چنین فرآیندهایی کمک شایان توجهی به تحقق وب معنایی یا وب داده‌ها دارد.

کلیدواژه‌ها: فرآیند تولید داده‌های ساختارمند بر روی وب (اسکیما دات آرگ)، وب معنایی، نشانه‌گذارهای معنایی، پردازش، اشیای محتوایی وب، داده‌های ساختارمند، رویکرد هستی‌شناسانه، رویکرد بافت‌مدار، رویکرد آشیانه‌ای (تودرتو)

۱. مقدمه

گسترش روزافزون وب و حجم آن از یک سو و تنوع اشیای محتوایی دسترس‌پذیر بر روی آن از سوی دیگر، به صورت مسئله‌ای بغرنج در پردازش، سازماندهی، و بازیابی مطرح شده است. پیش‌بینی شده که حجم داده‌های وب تا سال ۲۰۲۰ میلادی به ۴۰ زتابایت برسد و در دهه بعدی، هر ساله ۴۰ درصد به این میزان افزوده شود (Nguyen and Song 2016). با توجه به رشد فزاینده داده‌ها مبحث اصلی در پردازش، سازماندهی، و بازیابی این حجم از داده‌ها، ساختار و تنوع آن است. از آنجا که قابلیت‌های محیط داده‌های وب با محیط‌های داده‌های سنتی متفاوت است، در نتیجه پردازش، سازماندهی، و بازیابی آن نیز نیازمند ابزار و استانداردهای جدیدی است.

اغلب طراحان و مدیران وب‌سایت‌ها با برچسب‌های زبان نشانه‌گذاری فرامتن (اچ‌تی‌ام‌ال) برای طراحی وب‌سایت‌ها و راه‌های بهینه‌سازی آشنا هستند. برچسب‌های «اچ‌تی‌ام‌ال» در اصل شیوه‌نمایش داده‌ها را به مرورگرها اعلام می‌کنند. مثلاً برچسب علم اطلاعات به مرورگر اعلام می‌کند که رشته حرفی «علم اطلاعات» را به صورت ضخیم (پررنگ) نشان دهد؛ اما برچسب‌های «اچ‌تی‌ام‌ال» هیچ‌گونه اطلاعاتی درباره معنای این رشته‌های حرفی اعلام نمی‌کنند یا نشان نمی‌دهند. «علم اطلاعات» ممکن است اسم، نام فیلم، نام رشته موضوعی، و یا هر چیزی دیگر باشد. این مورد

در مواجهه با موتورهای کاوش وب به مراتب با مشکلات بیشتری در پردازش و بازیابی روبه‌رو خواهد شد. موتورهای کاوش وب در برخورد با این برچسب‌ها هیچ‌گونه فهم و درکی از معنای این رشته‌های حرفی نخواهند داشت. از این رو، در هنگام بازیابی با حجم انبوهی از اطلاعات روبه‌رو خواهند شد که ارتباطی با مطالب درخواستی ندارد. موتورهای کاوش وب برای برطرف کردن این مشکل راه‌های متفاوتی در پیش گرفته‌اند که یکی از این موارد معرفی طرحی موسوم به «فرآینمای تولید داده‌های ساختارمند بر روی وب (اسکیمادات آرگ)»^۱ است.

به‌علت ساختارمند نبودن داده‌های وب کنونی، کار بر روی هستی‌نگاشت‌ها و ساختارمندسازی داده‌ها در قالب‌هایی نظیر «آردی‌اف» به‌منظور تحقق وب معنایی از ضروریات است. از این رو، به‌جای تمرکز بر هستی‌شناسی‌های مفهومی و لزوم ساختارمندی داده‌ها، تنها ایجاد یک هستی‌شناسی یا فرآینمای یکپارچه برای تمام انواع اشیای محتوایی موجود در وب ایده‌ای مناسب خواهد بود. فرآینمای تولید داده‌های ساختارمند بر روی وب در پی تحقق این ایده توسط موتورهای کاوش وب است. با تحقق این طرح، برای هر شخصی در هر زمانی و از هر مکانی، امکان انتشار داده‌های ساختارمند اشیای محتوایی موجود در وب نظیر شخص، رویداد، مکان، سازمان و جز این‌ها، که برنامه‌های کاربردی نیز از آن‌ها پشتیبانی خواهند کرد، فراهم می‌شود.

لازم به ذکر است که فناوری ایجاد داده‌های ساختارمند و استفاده از آن‌ها، قبل از معرفی فرآینمای تولید داده‌های ساختارمند بر روی وب مطرح بوده است و موتورهای کاوش وب، مزایای استفاده از چنین فناوری را در جهت پردازش، سازماندهی، و بازیابی بهتر درک کردند؛ اما فراهم‌کنندگان داده‌های پیوسته و تولیدکنندگان محتوا از این فناوری‌ها کمتر استفاده می‌کردند. آن‌ها محتواهای خود را به‌صورت ساختارمند منتشر نمی‌کردند یا از واژگان مختلفی برای بیان داده‌های مشابه استفاده می‌کردند که این امر باعث بروز مشکلاتی در هنگام بازیابی، تفسیر، و دسترس‌پذیری می‌شد (Guha, Brickley and Macbeth 2015; Freidrich 2015). موتورهای کاوش وب با درک این وضعیت و سازگاری پایین‌شان با استانداردهای جدید بافت میراث فرهنگی (کتابخانه‌ها، موزه‌ها، و آرشیوها)، چنین فرآینمای یکپارچه و مشترکی را طراحی و معرفی کردند. استفاده از این فرآینماها در هنگام ایجاد و

۱. در پژوهش حاضر از عبارت «فرآینمای تولید داده‌های ساختارمند بر روی وب» برای اشاره به Schema.org استفاده شده است.

سازماندهی اشیای محتوایی، میانکنش‌پذیری با موتورهای کاوش و تحقق‌بازیابی معنایی را نیز بهبود می‌بخشد که نشان از اهمیت آن‌ها دارد. از سویی نیز رویکردهایی که فرامای تولید داده‌های ساختارمند بر روی وب نسبت به پردازش و سازماندهی اشیای محتوایی دارد، نیازمند تبیین و تحلیل است که پژوهش حاضر در راستای پاسخ به این مسئله شکل گرفت. از این رو، پژوهش حاضر ضمن تبیین فرامای تولید داده‌های ساختارمند بر روی وب، به تحلیل رویکردهای آن به پردازش و سازماندهی اشیای محتوایی وب (موجودیت‌های داده‌ای) نیز می‌پردازد. لذا، مقاله حاضر در صدد پاسخگویی به پرسش‌های زیر است:

۱. مفهوم و واژگان‌های (طرحواره‌های) فرامای تولید داده‌های ساختارمند چگونه هستند؟
۲. در فرامای تولید داده‌های ساختارمند از چه مدل داده‌ای استفاده شده و این مدل دارای چه ساختار و بستر معناشناختی است؟
۳. فرمت‌های نشانه‌گذاری معنایی در فرامای تولید داده‌های ساختارمند کدام‌اند؟
۴. رویکردهای فرامای تولید داده‌های ساختارمند در پردازش و سازماندهی اشیای محتوایی (موجودیت‌های داده‌ای) وب کدام‌اند؟

۲. پیشینه پژوهش

بررسی متون و پیشینه فرامای تولید داده‌های ساختارمند در وب نشان داد که مطالعات و پژوهش‌های این حوزه را می‌توان در دو بخش کلی ارائه کرد. بخش اول، پژوهش‌های مبتنی بر فرامای تولید داده‌های ساختارمند در وب و بخش دوم، پژوهش‌هایی که بر نشانه‌گذارهای معنایی متمرکز هستند. برخی از پژوهش‌ها نیز با واژگان فرامای تولید داده‌های ساختارمند در وب به نشانه‌گذاری معنایی اشیای محتوایی در فرمت‌های مختلف پرداخته‌اند.

بخش اول: این بخش مشتمل بر مطالعات و پژوهش‌های پیرامون فرامای تولید داده‌های ساختارمند بر روی وب است. بخش عمده این مطالعات به معرفی، ساختار و کاربرد، مطالعات آماری انتشار در وب، و به کارگیری واژگان این فرامای پرداخته است. مطالعاتی نظیر (Rosati and Mayernik (2013)، (Hawkser, Barker and Compbell (2013)، (Medeiros (2014)، (Aldaej and Aldaej (2015)، (Guha, Brickley and Macbeth (2015)، (Difranzo and et al. (2014)، (Krause (2015)، (Meusel, Bizer and Paulheim (2015)، (Friedrich (2016)، (Bizer, Meusel and Primpeli (2017) و (Newcomer (2018) از این نوع هستند. برخی از این پژوهش‌ها نیز به

انطباق نوع و صفات فرآیند تولید داده‌های ساختارمند بر روی وب با سایر استانداردهای فراداده‌ای مبادرت کرده‌اند. (2015) Gracy به انطباق صفات این فرآیند با عناصر مارک و (2017) Col and et al. نیز به انطباق با عناصر استاندارد فراداده‌ای توصیف شیء پرداخته‌اند. حال به برخی از این پژوهش‌ها اشاره می‌شود.

«روزاتی و مایرنیک» در پژوهشی داده‌های ساختارمند وب معنایی در فرمت «آردی‌اف/ایکس‌ام‌ال» را با به‌کارگیری دو واژگان «طرح تبادل و استفاده مجدد اشیای آرشیوهای باز»^۱ و «فرآیند تولید داده‌های ساختارمند بر روی وب» برای نشانه‌گذاری صفحات وب به‌کار بردند. نتایج نشان داد که هر کدام از رویکردها دارای مشخصه‌هایی از جمله سهولت استفاده، در دسترس بودن استانداردها و واژگان، سهولت میان‌کنش‌پذیری، و ارتباط میان روش‌ها و ابزارهای استاندارد داده‌ها هستند (2013) Rosati and Mayernik. «آلداج» و «آلداج و کروز» در پژوهش‌های خود اشیای محتوایی موجود در محیط‌های یادگیری مجازی را با کمک واژگان مربوطه در فرآیند تولید داده‌های ساختارمند بر روی وب، نشانه‌گذاری و مورد سنجش قرار دادند. به‌زعم آن‌ها، واژگان این فرآیند توانایی معنایی کمی در پشتیبانی از اشیای محتوایی در بافت یادگیری دارد؛ از این رو، طرح فراداده‌ای منابع یادگیری^۲ را پیشنهاد کردند که به‌طور رسمی در آوریل ۲۰۱۳ به فرآیند تولید داده‌های ساختارمند بر روی وب اضافه گردید (2015) Aldaej و (2015) Aldaej and Krause.

«فردریش» در پژوهشی نشان داد که به‌کارگیری نشانه‌گذاری معنایی بعد از معرفی فرآیند تولید داده‌های ساختارمند بر روی وب در سال ۲۰۱۱ رشد تصاعدی داشته است. عمده صفحات با فرمت ریز داده (۸۰ درصد)، نشانه‌گذاری شده است؛ زیرا این فرمت در ابتدا به‌عنوان فرمت توصیه‌شده فرآیند تولید داده‌های ساختارمند بر روی وب بود. وی همچنین مدل جست‌وجوی مبتنی بر موجودیتی را ارائه کرد که از این فرآیندها و شیوه داده‌های پیوندی بهره‌جویی می‌کند (2016) Friedrich.

بخش دوم: پژوهش‌های این بخش عمدتاً متمرکز بر نشانه‌گذاری معنایی نظیر ریز داده، «آردی‌اف‌ای» و «جی‌سون-ال‌دی» و به‌کارگیری آن‌هاست. برخی از پژوهش‌های این بخش به بررسی تأثیر استفاده از نشانه‌گذاری معنایی در بازیابی موتورهای کاوش نیز

1. Open Archives Initiative Object Re-use and Exchange (OAI-ORE)
2. Learning Resource Metadata Initiative (LRMI)

مبادرت کرده‌اند. پژوهش‌های (Wetherill, Matosevic (2014)، (Rasool, Iqbal and Malik (2013)، (Clark and Young (2015)، (2014)، (Bohne-Lang (2016) از این نوع هستند. در سال‌های اخیر پژوهش‌های داخلی نیز در این بخش صورت گرفته که از جمله آن‌ها می‌توان به «ذوالقدر» (۱۳۹۵)، «طاهری، نیکزاد بهله و صمیعی» (۱۳۹۶)، و «آقاده» (۱۳۹۷) اشاره کرد. «ذوالقدر» به بررسی تطبیقی نمایه‌سازی و پیدانمایی خرده‌های غنی در موتورهای کاوش وب با روش تجربی پرداخت. یافته‌های وی نشان داد که موتور کاوش «گوگل» پیشینه‌های فراداده‌ای هر دو گروه آزمون و گواه را نمایه‌سازی، و پیشینه‌های گروه آزمون را به‌صورت معنادار در نتایج جست‌وجو پدیدار می‌کند. اما موتورهای کاوش «ياهو» و «بینگ»، فقط پیشینه‌های هر دو گروه را نمایه‌سازی کرده، ولی قادر به پیدانمایی پیشینه‌های گروه آزمون نیستند (۱۳۹۵).

«طاهری، نیکزاد بهله و صمیعی» در پژوهشی به بررسی واکنش نمایه‌پذیری و پیدانمایی پیشینه‌های فراداده‌ای مبتنی بر روش ترکیبی ریزداده و داده‌های پیوندی در موتورهای کاوش وب (گوگل، بینگ و ياهو) پرداختند. نتایج این پژوهش نشان داد که موتورهای کاوش وب قادر به نمایه‌سازی و نمایش پیشینه‌ها در نتایج کاوش بودند و در مقابل، هیچ‌یک از موتورهای کاوش وب، قادر به پیدانمایی و نمایش ساختارمند پیشینه‌های فراداده‌ای نبودند. پیشینه‌های فراداده‌ای پژوهش، پس از ترکیب با قالب «جی‌سون-ال‌دی» و معرفی به موتورهای کاوش، قادر به نمایه‌سازی پیشینه‌های فراداده‌ای بودند، اما با توجه به روش ترکیبی خرده‌های غنی و داده‌های پیوندی، قادر به پیدانمایی و نمایش ساختارمند نبوده، در صورتی که قادر به پیدانمایی برجسب‌های تعریف شده بود؛ ولی نه به‌صورت آنچه که در این پژوهش مد نظر است (۱۳۹۶).

بررسی و مرور تحقیقات نشان می‌دهد که ایجاد داده‌های ساختارمند با استفاده از فرمانای تولید داده‌های ساختارمند و نشانه‌گذارهای معنایی از سال ۲۰۱۵ به بعد جهش قابل ملاحظه‌ای داشته است. اکثر این پژوهش‌ها خارجی است و تحقیق مشابه داخلی در بین آن‌ها دیده نمی‌شود. همچنین، اهم این پژوهش‌ها به تأثیر استفاده از فرماناهای موجود پرداخته‌اند که بهبود نمایه‌پذیری، کشف‌پذیری، و پیدانمایی را به دنبال داشته است. برخی از پژوهش‌ها نیز به توسعه فرماناهای موجود پرداخته که نشان از روزآمدی این فرماناها و میان‌کنش‌پذیری مناسب با موتورهای کاوش وب دارد. پژوهش‌هایی که بر نشانه‌گذاری معنایی متمرکز بودند، نیازمند تقویت و توسعه جهت کاربرد بیشتر

نشانه‌گذارها در صفحات وب هستند؛ زیرا این نشانه‌گذارها قابلیت جای‌گذاری در صفحات «اچ‌تی‌ام‌ال» را دارند که این امر کاربردپذیری آن‌ها را آسان‌تر می‌کند. در این بخش به‌تازگی پژوهش‌های چندی در داخل صورت گرفته است که نشان از اهمیت این نوع از موضوعات در حوزه وب معنایی و جایگاه آن در بازیابی و نمایش ساختارمند یا معنایی در بازیابی اشیای محتوایی دارد. پژوهش‌هایی از این قسم، به‌نوعی به‌دنبال بازیابی دانش‌اند. آنچه در پژوهش‌های داخلی مشخص است، به‌کارگیری نشانه‌گذارهای معنایی بدون استفاده از واژگان فراماهاست. این مورد می‌تواند یک نقص در پژوهش‌های داخلی باشد، زیرا اکثر پژوهش‌های خارجی در این زمینه همراه با واژگان فراماهاست. این امر موجب یکدستی در استفاده از واژگان برای نشانه‌گذاری معنایی اشیای محتوایی می‌شود. با توجه به مطالب پیش‌گفته، پژوهش مفهومی حاضر در عین تبیین نشانه‌گذارهای معنایی سازگار با فرآینمای تولید داده‌های ساختارمند در وب، در بخش اول جای می‌گیرد؛ زیرا به تبیین و تحلیل فرانما و رویکردهای آن به پردازش و سازماندهی اشیای محتوایی وب (موجودیت‌های داده‌ای) می‌پردازد که تاکنون صورت نگرفته است.

۳. روش پژوهش

پژوهش حاضر، به‌دلیل تبیین مفهوم «فرآینمای تولید داده‌های ساختارمند بر روی وب» با رویکرد تحلیلی و بررسی رویکرد آن به سازماندهی و پردازش اشیای محتوایی وب (موجودیت‌های داده‌ای)، و رعایت ساختار این گونه از پژوهش‌ها، پژوهشی مفهومی^۱ به‌شمار می‌آید. به‌منظور گردآوری داده‌ها از روش تحلیل اسنادی استفاده شده است. همچنین، به‌منظور انجام پژوهش، متون و پژوهش‌های صورت‌گرفته و مرتبط با حوزه‌های «داده‌های ساختارمند»، «فرآینمای تولید داده‌های ساختارمند بر روی وب»، و «نشانه‌گذاری‌های معنایی» مورد تحلیل و بررسی قرار گرفت.

۴. یافته‌های پژوهش

پرسش اول: مفهوم و واژگان‌های (طرحواره‌های) فرآینمای تولید داده‌های ساختارمند چگونه هستند؟

در پاسخ به پرسش اول پژوهش مبنی بر چیستی فرمانای تولید داده‌های ساختارمند بر روی وب و هدف اصلی از طراحی و ایجاد آن، تبیین برخی مقدمات و پیشینه آن لازم است. «تیم برنرزیلی» به منظور تحقق وب معنایی و ایده داده‌های پیوندی، طرح سیستم رتبه‌بندی پنج ستاره‌ای را برای داده‌ها ارائه داد که بر طبق ساختاردهی به داده‌ها حرکت می‌کند و حالت تجمیعی دارد. به‌زعم وی هرچه داده‌های وب ساختارمندتر باشد، تحقق وب معنایی و ایده داده‌های پیوندی تسریع می‌شود (نقل در 2014 Dourado). داده‌های ساختارمند، اسمی عام برای همه نشانه‌گذارهایی است که با مجموعه‌ای از قوانین از پیش تعریف شده مشخص می‌شوند که می‌توانند به‌طور مستقیم به نشانه‌گذار «اچ‌تی‌ام‌ال» صفحه اضافه شوند (2014 Bohatch). این نوع داده‌ها طیف وسیعی از وظایف خودکار را نسبت به داده‌های بدون ساختار فراهم کرده، به قسمی که آن‌ها را می‌توان بسیار کارآمدتر از فایل‌های متنی ساده مورد پردازش قرار داد (2017a Sikos). Twittercards و Google Authorship نمونه‌هایی از داده‌های ساختارمند یا داده‌های غنی در دنیای واقعی وب هستند که به ترتیب برای «تویتر» و «گوگل پلاس» طراحی شده است (2014 Bohatch). فرمانای تولید داده‌های ساختارمند بر روی وب برای تحقق این امر، نقشی بارز و مؤثر خواهد داشت. داده‌هایی که با واژگان چنین فرماناهایی نشانه‌گذاری شده‌اند، داده‌های ساختارمندی هستند که امکان پردازش، سازماندهی و بازیابی آن‌ها بسیار تسهیل می‌گردد. فرمانای تولید داده‌های ساختارمند بر روی وب توسط موتورهای کاوش «گوگل»، «یاهو» و «بینگ» در دوم ژوئن ۲۰۱۱ طراحی و معرفی شد (2017b Sikos). از اول نوامبر ۲۰۱۱، «یاندکس»، بزرگ‌ترین موتور کاوش روسیه نیز به این تلاش مشترک پیوست. هدف از این طرح، ایجاد فرمانایی منفرد برای طیف وسیعی از موضوعات شامل افراد، مکان‌ها، رویدادها، محصولات، فروش، و جز این‌ها بود. وجود یک فرمانای یکپارچه این موضوعات را پوشش می‌دهد (2015 Guha, Brickley and Macbeth). فرمانای تولید داده‌های ساختارمند بر روی وب با پوشش تقریبی تعاریف ۳۰۰ مفهوم، یکی از متداول‌ترین طرح‌های نشانه‌گذاری داده‌های ساختارمند است (2017b Sikos).

اطلاعات فرمانای تولید داده‌های ساختارمند بر روی وب در وب‌سایت رسمی^۱ آن ذکر شده است. درک هسته این فرمانا مربوط به درک واژگان آن است که معمولاً با عنوان

1. Yandex

2. www.schema.org

فرآیند تولید داده‌های ساختارمند بر روی وب شناخته می‌شود (Yu 2014). این فرآیند، واژگانی است که اصطلاحات (نوع^۱، صفت^۲، و ارزش‌های شمارش شده^۳) را تعریف می‌کند (How we work 2015). فرآیند تولید داده‌های ساختارمند بر روی وب شامل تعاریف ماشین‌خوان از رایج‌ترین مفاهیم کاربردی است که امکان حاشیه‌نویسی فعالیت‌ها، آثار خلاق، رویدادها، خدمات، موجودیت‌های پزشکی، سازمان‌ها، افراد، مکان‌ها، و محصولات را می‌دهد (Sikos 2017b). از دیدگاه وب معنایی، این فرآیند هستی‌نگاشتی است که جنبه‌های عمومی زندگی روزمره را پوشش داده و توسط دنیای برنامه‌های کاربردی اصلی نیز پشتیبانی می‌شود (Yu 2014). برخی فرآیند تولید داده‌های ساختارمند بر روی وب را طرحی می‌دانند که با اهداف مشابه طرح فراداده‌ای هسته‌دابلین (یعنی ایجاد، نگهداری و ارتقای فرآیندها برای داده‌های ساختارمند در اینترنت) آغاز شد (Fenner 2017). بنابراین، این طرح یک واژگان یا هستی‌نگاشت واحد است که برای توصیف صفات و بازنمون روابط میان اشیای محتوایی وب معرفی و طراحی شده که برای نشانه‌گذاری اشیای محتوایی وب به شیوه‌ای قابل درک توسط موتورهای کاوش مناسب است.

بیشتر واژگان فرآیند تولید داده‌های ساختارمند بر روی وب از فرمت‌های قدیمی‌تر نظیر ریزفرمت‌ها، FOAF^۴، GoodRelations^۵، و OpenCyc و جز این‌ها الهام گرفته شده است؛ به قسمی که گستره آن در سطح وب بین سال‌های ۲۰۱۲ تا ۲۰۱۴ شدیداً افزایش یافت (Schema.org 2014). واژگان فرآیند تولید داده‌های ساختارمند بر روی وب در اکثر واژگان‌های مشهور نظیر (GoodRelations (2012)، W3C BibExtend (2014)، MuscicBrainz، Vocabulary (2015)، و Automotive Ontology (2015) نیز گنجانده شده است (Paulheim 2015). در واقع، فرآیند تولید داده‌های ساختارمند بر روی وب با جذب سایر هستی‌نگاشت‌ها رشد می‌کند. به عنوان مثال، در ۸ نوامبر ۲۰۱۲، هستی‌نگاشت GoodRelations به فرآیند تولید داده‌های ساختارمند بر روی وب متصل شد. به عبارت دیگر، اصطلاحات موجود در

1. type

2. property

3. enumerated value

۴. Friend of a Friend (FOAF): یک هستی‌نگاشت ماشین‌خوان است که اشخاص، فعالیت‌ها، و روابطشان با دیگر اشخاص و اشیاء را توصیف می‌کند. هر شخصی می‌تواند از «فواف» برای توصیف خود استفاده کند.

۵. هستی‌نگاشت معمولی (متوسطی) برای تجارت الکترونیکی در وب است که برای توصیف تجارت، کسب‌وکار، سفارش، قیمت‌گذاری، شیوه‌های پرداخت، ساعات شروع و غیره مناسب است. این هستی‌نگاشت به‌طور گسترده توسط «یاهو» و BestBuy استفاده می‌شود (Sikos 2017b).

هستی‌نگاشت GoodRelations به‌طور مستقیم از درون فرمانای تولید داده‌های ساختارمند بر روی وب در دسترس هستند. بنابراین، کاربران می‌توانند از تمام نوع‌ها و صفات GoodRelations به‌طور مستقیم از فضای نام فرمانای تولید داده‌های ساختارمند بر روی وب در نشانه‌گذاری استفاده کنند (Yu 2014).

پرسش دوم: در فرمانای تولید داده‌های ساختارمند از چه مدل داده‌ای استفاده شده و این مدل دارای چه ساختار و بستر معناشناختی است؟

در پاسخ به پرسش دوم پژوهش مبنی بر مدل داده‌ای استفاده‌شده در فرمانای تولید داده‌های ساختارمند وب و ساختار و بستر معناشناختی آن مشخص شد که مدل داده‌ای مورد استفاده در این فرامنا بسیار عمومی و مشتق از «آردی‌افاس»^۱ (که خود از CysL مشتق شده) است. این فرامنا، دارای مجموعه‌ای از «نوع‌ها» است که هر نوع، موجودیت‌های شناسایی‌شده در گستره وب را تعریف می‌کند. هر کدام از این نوع‌ها دارای «صفات» است که ممکن است با سایر صفات نوع‌های دیگر مشترک یا مرتبط باشند. این نوع‌ها در یک ساختار سلسله‌مراتبی چندگانه به‌همراه واژگان توصیفی آن‌ها (صفات) نظیر یک هستی‌نگاشت مرتب شده‌اند. به‌طور کلی، نظیر بسیاری از استانداردهای فراداده‌ای، این فرامنا نیز دارای ساختار (مدل داده‌ای) و بستر معناشناختی خاصی است. هر کدام از فرامنا‌های موجود در این طرح دارای دو جزء اصلی است که یکی مربوط به ساختار و دیگری مربوط به بستر معناشناختی است. این دو جزء به‌ترتیب عبارت‌اند از:

الف. هستی‌نگاشت نظیر واژگانی برای نام‌گذاری نوع‌ها و صفات موجودیت‌ها، روابط با یکدیگر، و قیود چگونگی توصیف این صفات و روابط؛

ب. بیان اطلاعات مزبور به فرمت‌های ساختارمند (ماشین‌خوان) نظیر ریزداده‌ها، «آردی‌افای ۱/۱»، و «جی‌سون-ال‌دی» (Barker and Campbell 2015).

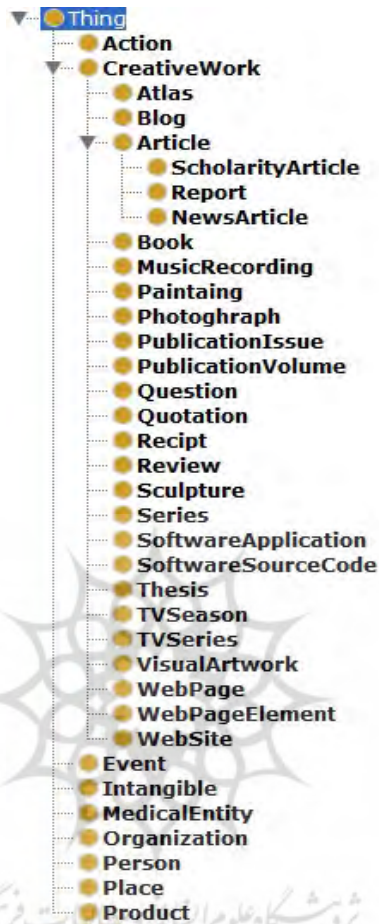
در وب‌سایت رسمی فرمانای تولید داده‌های ساختارمند بر روی وب، مدل داده‌ای به‌صورت زیر تبیین شده است:

الف. مجموعه‌ای از نوع‌ها، که به‌صورت سلسله‌مراتبی توارثی چندگانه تنظیم شده‌اند؛ جایی که هر نوع ممکن است زیرکلاسی از نوع‌های چندگانه باشد.

ب. مجموعه‌ای از صفات:

۱. هر صفت ممکن است یک یا چند نوع به‌عنوان دامنه^۱ داشته باشد. این صفت ممکن است برای نمونه‌هایی^۲ از هر کدام از این نوع‌ها مورد استفاده قرار گیرد.
 ۲. هر صفت ممکن است یک یا چند نوع به‌عنوان طیف^۳ داشته باشد. ارزش صفات باید نمونه‌هایی از حداقل یکی از این نوع‌ها باشد (Data model Schema.org 2015).
- همان‌طور که ملاحظه می‌شود، وجود ساختار سلسله‌مراتبی نوع‌ها، بازنمون روابط، صفات و مشخصات آن‌ها نظیر یک هستی‌نگاشت است. از این رو، ساختار و رویکرد هستی‌شناسانه در فرآینمای تولید داده‌های ساختارمند بر روی وب وجود دارد. حال، به توضیح هر کدام از بخش‌ها مبادرت می‌شود.
- ساختار:** مرحله اول در توصیف مدرک، طبقه‌بندی نوع خاص اشیای محتوایی است که به‌نحو طبیعی نوع آن را شناسایی می‌کند. در این فرآینما برای هر نوع، مجموعه‌ای از صفات که ممکن است برای توصیف مشخصه‌ها و خصیصه‌های مدرک به کار آید، تعریف می‌شود. در این قسمت انواع اشیای محتوایی و صفات آن در ساختار مناسبی تعریف می‌گردد (Barker and Campbell 2015). ساختار اصلی فرآینمای تولید داده‌های ساختارمند بر روی وب به‌صورت سلسله‌مراتبی چندلایه است. بالاترین سطح سلسله‌مراتب که عام‌ترین و کلی‌ترین آن است، نوع «thing» است که تمام زیرنوع‌ها و موجودیت‌های دیگر زیرمجموعه آن هستند. در سطح دوم سلسله‌مراتب، هفت نوع کلی نظیر آثار خلاقانه^۴، رویداد^۵، (آثار) لمس‌ناشدنی^۶، سازمان^۷، شخص^۸، مکان^۹، و محصول^{۱۰} وجود دارد که هر یک از اشیای محتوایی می‌توانند در یکی از نوع‌ها، تعریف و با سایر نوع‌های موجود در سلسله‌مراتب رابطه داشته باشد که این رابطه نیز نیازمند تعریف است. البته، در گسترش‌های بعدی دو زیرنوع اصلی دیگر در سطح دوم سلسله‌مراتب یعنی عمل^{۱۱} و موجودیت پزشکی^{۱۲} به هفت مورد پیشین اضافه شده است. در سطح سوم سلسله‌مراتب، موجودیت‌های وب کنونی قرار دارند. شکل ۱، نمایی کلی از سطوح موجود در فرآینمای تولید داده‌های ساختارمند بر روی وب را نشان می‌دهد.

1. domain	2. instance	3. range	4. Creative Work
5. event	6. intangible	7. organization	8. person
9. place	10. product	11. action	12. medical entity



شکل ۱. نمایشی از سطوح نوع‌ها در فرآیند تولید داده‌های ساختارمند بر روی وب

نوع «CreativeWork» دارای بیشترین زیرنوع است که در شکل ۱ نیز مشخص است و حتی برخی از زیرنوع‌های آن نظیر «Article» دارای زیرنوع‌های دیگری است. هر نوع در سلسله‌مراتب، دارای صفاتی است و به‌واسطه ساختار سلسله‌مراتب گونه، از نوع پدر (مافوق) خود ارث می‌برد. به عبارت دیگر، هر کدام از صفات یک نوع بر تمام زیرنوع‌های خود نیز کاربرد دارد. به‌طور مثال، صفاتی که توسط نوع‌های دیگر از صفات اصلی «thing»، به‌عنوان بالاترین سطح ارث می‌برند، عبارت‌اند از:

- description (text): a short description of the item
- image (URL): URL of an image of the item
- name (text): the name of the item
- url (URL): URL of the item

همچنین، نوع «CreativeWork» تمام صفات نوع «thing» را به‌عنوان بالاترین سطح به ارث می‌برد، و خود نیز دارای صفات انحصاری است؛ به‌طوری که حدود ۴۰ صفت را برای نوع خود تعریف می‌کند، برخی از آن‌ها عبارت‌اند از:

- about (thing): the subject matter of the content
- author (person or organization): the author of this content
- dateCreated (date): the date on which the CreativeWork was created
- publisher (organization): the publisher of the creative work

عبارت‌های «text»، یا «URL» داخل پرانتز در عبارت‌های مذکور بیانگر «نوع داده»^۱ برای ارزش آن صفت است؛ یعنی ارزش صفات «description» و «name» باید از نوع «text» و ارزش صفت «image» از نوع «URL» باشند. تبیین و تعیین نوع داده در سایر طرحواره‌ها و هستی‌نگاشت‌ها نظیر «آردی‌اف‌اس» نیز مرسوم است. همچنین، «نوع مورد انتظار»^۲ نیز در فرآیند تولید داده‌های ساختارمند بر روی وب وجود دارد که نشان از نوع مورد انتظار از سایر نوع‌های سلسله‌مراتب است که ارتباط بین نوع‌های مختلف را نشان می‌دهد. به‌عبارت دیگر، ارزش صفت «author» باید از نوع مورد انتظار «person or organization» باشد که خود دارای صفتی نظیر «name» است که ارزش نوع داده آن «text» است. به‌عنوان نمونه، نوع داده‌هایی که در نوع «CreativeWork» پشتیبانی و قابل تعریف است، عبارت‌اند از Text, Number, Date و Boolean (Schema.org types and properties tutorial 2018). همچنین، نوع «text» می‌تواند به‌صورت «URL» نیز باشد. این فرآیند دارای مجموعه‌ای کوچک از نوع‌های داده اولیه^۳ برای اعداد، متن و غیره نیز است (Schema.org 2014).

هر یک از نوع‌های موجود در فرآیند تولید داده‌های ساختارمند بر روی وب دارای یک فرآیند^۴ مجزا است. صفات هر کدام از فرآیندها مشتمل بر چند بخش است. بخشی از این صفات، مختص نوع مورد توصیف و بخش‌های دیگر، صفات به ارث برده‌شده از ابرنوع‌های آن نوع است. به‌عنوان نمونه بخشی از فرآیند نوع «Book»، که در جدول ۱، ارائه شده، مشتمل بر صفات مخصوص «Book» و صفات ابرنوع‌های آن در سلسله‌مراتب یعنی «CreativeWork» و «thing» هست که از آن‌ها ارث برده است. همچنین، نوع داده و نوع

-
1. data type
 2. expected type
 3. primitive data types
 4. schemata
 5. super type

مورد انتظار برای هر کدام از صفات به همراه توضیح آن‌ها در فرانما ارائه شده است. در حال حاضر، تعداد صفات ارث برده شده از نوع «CreativeWork»، هفتادوهفت و از نوع «thing»، یازده مورد است.

جدول ۱. بخشی از فرانمای نوع «Book» (Schema.org 2014)

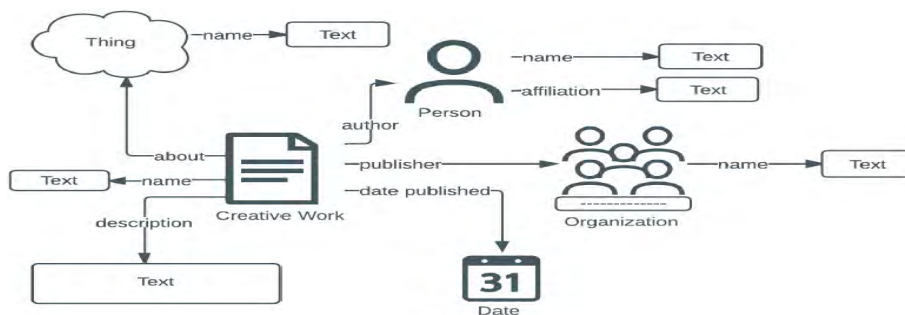
Book: Schema
Canonical URL: <http://schema.org/Book>

Thing > **CreativeWork** > **Book**

A book.
Usage: Between 10,000 and 50,000 domains

Property	Expected Type	Description
Properties from Book		
bookEdition	Text	The edition of the book.
bookFormat	BookFormatType	The format of the book.
illustrator	Person	The illustrator of the book.
isbn	Text	The ISBN of the book.
numberOfPages	Integer	The number of pages in the book.
Properties from CreativeWork		
about	Thing	The subject matter of the content.
author	Organization or Person	The author of this content or rating....
contributor	Organization or Person	A secondary contributor to the CreativeWork or Event.
dateCreated	Date or DateTime	The date on which the CreativeWork was created or the item was added to a DataFeed .
isPartOf	CreativeWork	Indicates a CreativeWork that this CreativeWork is (in some sense) part of. Inverse property: hasPart .
.....
Properties from Thing		
description	Text	A description of the item.
image	ImageObject or URL	An image of the item. This can be a URL or a fully described ImageObject .
name	Text	The name of the item.
url	URL	URL of the item.

همچنین، لازم به ذکر است که علاوه بر تعیین سلسله مراتب نوع‌های موجود، روابط میان نوع‌ها نیز حائز اهمیت است. شکل ۲، نمونه‌ای از روابط پیرامون نوع «CreativeWork» را که امکان دارد با فرانمای تولید داده‌های ساختارمند بر روی وب ایجاد شود، نشان می‌دهد:



شکل ۲. برخی روابط نوع «CreativeWork» با سایر نوع‌ها (Barker and Campbell 2015)

همان‌طور که در شکل ۲، مشخص است، نوع «CreativeWork» دارای صفاتی نظیر «name, description, author, publisher, date published» است که صفت «author» از نوع «Person» است که آن نیز دارای صفاتی همچون «name» و «affiliation» است. در شکل ۲، روابطی که میان نوع‌های «CreativeWorks, Person, Organization» و «Date» وجود دارد، همراه با برخی از صفات آن‌ها به تصویر کشیده شده است.

همان‌طور که هر نوع یک یا چند ابرنوع دارد، هر صفت نیز ممکن است دارای یک یا چند ابرصفت باشد. به عنوان نمونه «partOfSeason» یک زیرمجموعه از «isPartOf» است. در این فرآیند صفات نیز می‌تواند به عنوان مقلوب یکدیگر نظیر «alumni/alumniOf» یا «isPartOf/hasPart»، تعریف و نشانه‌گذاری شوند (How we work 2015). فرآیند تولید داده‌های ساختارمند بر روی وب با ۲۹۷ کلاس (نوع) و ۱۸۷ صفت شروع شد و طی چهار سال به تعداد ۶۳۸ کلاس (نوع) و ۹۶۵ صفت رسید (Guha, Brickley and Macbeth 2015; Paulheim 2015). لازم به ذکر است که صفات هر نوع در فرآیند یا ویژگی‌ها در بافت میراث فرهنگی همانند نیست، بلکه ممکن است شامل هم صفات و هم نشانگرهای روابط (رابطه) باشد. همان‌طور که در جدول ۱، قابل مشاهده است، یکی از صفات نوع «Book»، «isPartOf» است که نشانگر رابطه‌ای است تا صفت یک نوع. همچنین، صفت مقلوب آن یعنی «hasPart» نیز ذکر شده است.

بستر معناشناختی: واژگان فرآیند تولید داده‌های ساختارمند بر روی وب در حال حاضر با نشانه‌گذارهای معنایی نظیر ریزفرمت‌ها، ریزداده‌ها، «آردی‌اف‌ای ۱/۱»، و «جی‌سون-ال‌دی» برای افزودن یا حاشیه‌نویسی اشیای محتوایی وب سازگار است که بستر معناشناختی آن را شکل می‌دهد (Sack 2014; Yu 2014; Data model Schema.org 2014; Barker and Campbell

2014). در ابتدا، ریزداده‌ها، بستر معناشناختی مورد نظر و توصیه‌شده فرامای تولید داده‌های ساختارمند بر روی وب، به‌علت پیچیدگی کم و انعطاف‌پذیری بیشتر قرار گرفت. اخیراً فرامای تولید داده‌های ساختارمند بر روی وب، بستر معناشناختی «جی‌سون-ال‌دی» را به‌عنوان بستر معناشناختی توصیه‌شده معرفی کرده است (Schema.org 2018). دربارهٔ بستر معناشناختی در بخش نشانه‌گذارهای معنایی بیشتر صحبت خواهد شد.

داده‌های فرامای تولید داده‌های ساختارمند بر روی وب را می‌توان در همان صفحه یا سندی که انواع دیگر داده‌های ساختارمند نظیر ریزفرمت‌ها، ریزداده‌ها، «آردی‌اف‌ای ۱/۱»، و «جی‌سون-ال‌دی» وجود دارند و از واژگان مستقل استفاده می‌کنند، منتشر کرد (How we work 2015) به‌عبارت دیگر، ممکن است در یک سند از چند نوع داده ساختارمند یا نشانه‌گذار با واژگان مختلف استفاده کرد. مدل داده‌ای این فرامای انعطاف‌پذیری بالایی دارد. به‌عنوان نمونه، می‌توان از آن برای توصیف اشیای محتوایی در دو نوع مختلف که شباهت‌های بیشتری با یکدیگر دارند یا ممکن است زیرنوع یک نوع عام‌تر نظیر «Book» و «Article» باشند، از صفات مشترکی که این صفات از یکدیگر ارث می‌برند، استفاده کرد. مدل داده‌ای فرامای تولید داده‌های ساختارمند بر روی وب به‌طور طبیعی انعطاف‌پذیرند و پایه‌ای قابل‌گسترش برای تولید داده‌های ساختارمند ایجاد می‌کنند.

همچنین، می‌توان در توصیف اشیای محتوایی در یک نوع از چند واژگان نیز استفاده کرد. برای توصیف یک کتاب، نیازمند واژگانی برای تعریف نوع «Book» و صفات خاص آن هستیم. در صورتی که خواهان صفات توصیفی بیشتری نسبت به صفت «name» از فرامای تولید داده‌های ساختارمند بر روی وب باشیم، می‌توان صفت «title» را از واژگان «دابلین‌کور» استفاده نمود، که در نتیجه در دو فضای نام تبیین و توصیف می‌گردد. کدهای این توصیف شبیه به کدهای مثال زیر است:

```
...  
xmlns:schema="http://schema.org/"  
xmlns:dc="http://purl.org/dc/terms/"  
...  
<schema:Book>  
<dc:title>Web Standards: Mastering HTML5, CSS3, and XML</dc:title>  
</schema:Book>
```

مثال ۱. توصیف یک سند در دو فضای نام متفاوت

در مثال بالا، «Schema:Book» اختصار فضای نام <http://schema.org/Book>، تعریف ماشین‌خوان از نوع یا کلاس کتاب، و «dc:title» اختصار فضای نام <http://purl.org/dc/terms/>

title، تعریف ماشین‌خوان از صفت عنوان است (Sikos 2017b).

پرسش سوم: فرمت‌های نشانه‌گذاری معنایی در فرمانمای تولید داده‌های ساختارمند کدام‌اند؟

در پاسخ به پرسش سوم، فرمت‌های نشانه‌گذاری معنایی در فرمانمای تولید داده‌های ساختارمند بر روی وب، همان‌طور که تا اینجا اشاره شد، فرمانمای تولید داده‌های ساختارمند بر روی وب تنها ساختار و واژگان توصیف (مشمول بر نوع و صفات) را ارائه داده که می‌توان برای نشانه‌گذاری اشیای محتوایی استفاده کرد. اولین گام فرایند نشانه‌گذاری، انتخاب واژگان مناسب است. به لحاظ نظری، برای نشانه‌گذاری محتوای صفحه وب، امکان انتخاب و استفاده از هر هستی‌نگاشتی وجود دارد (Yu 2014). گام بعدی، انتخاب بستر معناشناختی یا قالب حاشیه‌نویسی ماشین‌خوان یا نشانه‌گذار است. چهار قالب حاشیه‌نویسی ماشین‌خوان برای وب‌سایت‌ها به ترتیب ظهور وجود دارد:

۱. ریزفرمت‌ها، که از طریق ویژگی‌های اصلی «ایکس‌اچ‌تی‌ام‌ال» داده‌های ساختارمند، مفاهیم اساسی نظیر افراد، مکان‌ها، رویدادها، دستورالعمل‌ها، و مواد شنیداری را منتشر می‌کند؛

۲. «آردی‌اف‌ای»، که «آردی‌اف» را در ویژگی‌های نشانه‌گذاری بیان می‌کند و بخشی از واژگان هسته «ایکس‌اچ‌تی‌ام‌ال» نیست؛

۳. ریزداده‌ها، که نشانه‌گذار «اچ‌تی‌ام‌ال ۵»^۱ را با فراداده ساختارمند گسترش می‌دهد؛ و

۴. «جی‌سون-ال‌دی»، که داده‌های ساختارمند را به‌عنوان کد «جاوا اسکریپت»^۲ نشانه‌گذاری می‌کند (Sikos 2017b).

«آردی‌اف‌ای» و «جی‌سون-ال‌دی» را می‌توان در اکثر نسخه‌های متنوع زبان نشانه‌گذاری استفاده کرد؛ در حالی که ریزداده را فقط در «ایکس‌اچ‌تی‌ام‌ال ۵»، می‌توان استفاده کرد. تمام فرمت‌های حاشیه‌نویسی دارای نحو خاص خود هستند (Sikos 2017a). هر کدام از این موارد برای نشانه‌گذاری اشیای محتوایی نیازمند ویژگی‌هایی هستند. ویژگی‌های مورد استفاده در ریزفرمت‌ها عبارت‌اند از: rel, class, rev و title (Jangra, Shashikumara and Kikon 2018). «آردی‌اف‌ای ۱/۱» لایت، پنچ ویژگی دارد: typeof, vocab, resource, property و prefix (Yu 2014; Sikos 2017b). ریزداده نیز دارای پنچ ویژگی عمومی

Jangra, Shashikumara and Kikon) است itemid و itemref، itemprop، itemtype، itemscope، id، type، context، های ویژگی‌های (2018; Microdata (HTML) container و vocab است (Kellogg 2017). برخی از این ویژگی‌ها در فرمت‌های مختلف دارای کاربردهای یکسان هستند؛ مثلاً، ویژگی «typeof» در «آردی اف‌ای»، ویژگی «itemtype» در ریزداده و ویژگی «@type» در «جی‌سون-ال‌دی» برای بیان نوع موجودیت مورد توصیف به کار می‌رود. به کمک این ویژگی‌ها و با واژگان فرانمای تولید داده‌های ساختارمند بر روی وب یا هر هستی‌نگاشت دیگر، صفحات و محتواهای وب نشانه‌گذاری می‌شود. در ادامه، یک مثال با چهار نشانه‌گذار معرفی شده در جدول ۲، ذکر می‌شود. این مثال صفحه شخصی و تصویر موجودیتی با نام Leslie Sikos را نشانه‌گذاری می‌کند.

جدول ۲. داده‌های ساختارمند در فرمت‌های ریزفرمت، ریزداده، «آردی اف‌ای» و «جی‌سون-ال‌دی»
(منبع: Sikos 2017a)

Semantic Markup	Example
Markup without Semantic Annotation	Leslie Sikos Leslie's web site: lesliesikos.com
Markup with the hCard microformat	<link rel="profile" href="http://microformats.org/profile/hcard" /> ... <div class="vcard"> Leslie Sikos Leslie's web site: lesliesikos.com </div>
Markup with HTML5 Microdata	<div itemscope itemtype="http://schema.org/Person"> Leslie Sikos Leslie's web site: lesliesikos.com </div>

Semantic Markup	Example
Markup with RDF	<pre><div vocab="http://schema.org" typeof="Person"> Leslie Sikos Leslie's web site: lesliesikos.com </div></pre>
Markup with JSON-LD	<pre><script type="application/ld+json"> { "@context": "http://schema.org", "@type": "Person", "image": "lesliesikos.jpg", "name": "Leslie Sikos", "url": "http://www.lesliesikos.com" } </script></pre>

دربارهٔ این که کدام فرمت را برای نشانه‌گذاری محتوا باید به کار برد، بحث‌های زیادی شده است. ریزفرمت‌ها دارای کمترین پیچیدگی و کمترین قدرت بیان هستند، «آردی‌اف‌ای» انعطاف‌پذیری و پیچیدگی بالایی دارد، و به نظر می‌رسد ریزداده‌ها نوعی توازن بین آن‌هاست (Friedrich 2015). همچنین، ریزداده‌ها اغلب به علت دسترسی به مستندات، ابزارهای پشتیبان، انگیزه‌های تجاری، طراحان‌های انعطاف‌پذیر و چابکی سازگاری بیشتری با فرآینمای تولید داده‌های ساختارمند بر روی وب استاندارد دارند (Paulheim 2015). از سوی دیگر، ریزداده‌ها و «آردی‌اف‌ای» برای مدت‌زمان طولانی در کنار هم کار می‌کردند. لازم به ذکر است که قبلاً توصیهٔ «گوگل» به استفاده از ریزداده‌ها بود، اما در حال حاضر «گوگل»، «جی‌سون-ال‌دی» را برای نشانه‌گذاری با این فرآینما توصیه می‌کند.

در این بخش، دو مثال از توصیف دو موجودیت رویداد (مثال ۲) و کتاب (مثال ۳) با استفاده از واژگان فرآینمای تولید داده‌های ساختارمند بر روی وب در دو بستر نحوی ریزداده و «جی‌سون-ال‌دی» در یک سند «اچ‌تی‌ام‌ال» ارائه می‌شود.

```
<div itemscope itemtype="http://schema.org/Event">
<a itemprop="url" href="event_site.html">
<span itemprop="name">A space odyssey</span>
</a>
<meta itemprop="startDate" content="2001-07-05T13:37">
<span>Thu, 07/05/01 1:37 p.m.</span>
```

```
<div itemprop="offers" itemscope itemtype="http://schema.org/AggregateOffer">
  <span itemprop="offerCount">42</span> tickets left
</div>
</div>
```

مثال ۲. توصیف موجودیت رویداد در بستر معناشناختی ریزداده در سند «اچ تی ام ال» (Friedrich 2015)

مثال بالا سند وبی است که به توصیف موجودیت رویداد می پردازد. در خط اول با توجه به بستر معناشناختی ریزداده به ترتیب، با ویژگی های «itemtype» و «itemprop»، نوع و صفات موجودیت را مشخص می کند. صفات این موجودیت شامل «name, url, startDate» و «offeres» از واژگان فرامای تولید داده های ساختارمند بر روی وب است.

```
<script type="application/ld+json">
```

```
{
  "@context": "http://schema.org",
  "@type": "Book",
  "title": "Shahnamah",
  "author": {
    "@type": "Person",
    "name": "Ferdowsi"
  },
  "about": "Persian poetry",
  "isbn": "964-7608-06-3"
}
```

مثال ۳. توصیف موجودیت کتاب در بستر معناشناختی «جی سون-ال دی» در سند «اچ تی ام ال»

مثال بالا سند وبی است که به توصیف موجودیت کتاب می پردازد. در سطح اعلان نوع بستر معناشناختی یعنی «جی سون-ال دی» مشخص شده است. ویژگی های «@context» و «@type» به ترتیب، واژگان و نوع موجودیت را مشخص کرده و سایر صفات به ترتیب ذکر شده اند. صفات موجودیت کتاب در مثال بالا شامل «author, title, about» و «isbn» از واژگان فرامای تولید داده های ساختارمند بر روی وب مشخص است. در هر دو مثال، یک صفت از موجودیتی دیگر با صفات خاص خود با رویکرد آشیانه ای در درون سند اصلی توصیف شده است. صفت «offers» در موجودیت رویداد و صفت «author» در موجودیت کتاب در ضمن این که خود به عنوان موجودیتی مجزا شناخته می شوند، با رویکرد آشیانه ای (تودرتو) در درون یک موجودیت دیگر نشانه گذاری شده اند.

پرسش چهارم: رویکردهای فرامای تولید داده های ساختارمند در پردازش و سازماندهی

اشیای محتوایی (موجودیت‌های داده‌ای) وب کدام‌اند؟

با توجه به حجم فزاینده اطلاعات در سطح وب، لزوم پردازش و سازماندهی ساختارمند این حجم از اطلاعات دوچندان شده است. محیط وب نیازمند ابزارهایی متناسب با آن است. ایجاد داده‌های ساختارمند با استفاده از فرآیند تولید داده‌های ساختارمند بر روی وب، کمک شایان توجهی به امر پردازش و سازماندهی معنادار داده‌ها در محیط وب کنونی خواهد کرد که مرور پیشینه‌ها نیز حکایت از اهمیت و سودمندی آن دارد. حال در این بخش به رویکردهایی که فرآیند تولید داده‌های ساختارمند بر روی وب در جهت پردازش و سازماندهی اشیای محتوایی وب (موجودیت‌های داده‌ای) دارد، پرداخته می‌شود. با بررسی و تبیین ساختار و بستر معناشناختی این فرآیند می‌توان در هر بخش، رویکرد خاصی را مشاهده کرد. به‌طور کلی، سه رویکرد عمده در این فرآیند شناسایی شده که به آن‌ها اشاره می‌شود:

الف. رویکرد هستی‌شناسانه: همان‌طور که بیان شد، ساختار اصلی فرآیند تولید داده‌های ساختارمند بر روی وب به‌صورت سلسله‌مراتبی چندلایه است. بالاترین سطح سلسله‌مراتب، «thing» است. «thing» به‌معنای «شیء» است و منطقاً همه چیز را در بر می‌گیرد. نظیر چنین رویکردی در استناداردهای بافت میراث فرهنگی نظیر «مدل مرجع کتابخانه‌ای (ال‌آرام)»^۱ مشاهده شده است. در این مدل، یک موجودیت سطح بالای انفرادی با نام «رس»^۲ ایجاد و هشت موجودیت دیگر به‌صورت مستقیم یا غیرمستقیم زیرمجموعه آن هستند (Riva, Bœuf and Zumer 2017). «رس» در مدل مرجع کتابخانه‌ای همانند «thing» در فرآیند تولید داده‌های ساختارمند بر روی وب است. در سطح دوم سلسله‌مراتب، تقسیم‌بندی کلی از موجودیت‌ها ارائه شده و در سطح سوم، سایر موجودیت‌ها به‌صورت ریز و جزئی مبتنی بر مصادیق اشیای محتوایی موجود در وب ذکر شده است. موجودیت‌هایی که توسط موتورهای کاوش وب شناسایی و نمایه‌سازی می‌شوند، در سطح سوم سلسله‌مراتب قرار دارند. سؤالی که در اینجا مطرح می‌شود این است که آیا ممکن است در وب چیزهای دیگری باشد که در این موجودیت‌ها ذکر نشده است؟ پاسخ به این سؤال مثبت است؛ زیرا ممکن است از لحاظ منطقی بتوان موجودیت دیگری شناسایی و به آن اضافه کرد. دلیل این مدعا نیز وجود گسترش‌ها در فرآیند تولید داده‌های ساختارمند بر روی وب و سازوکارهای آن

است.^۱ طرح‌های رده‌بندی کنگره نیز بر اساس کتاب‌های موجود در «کتابخانه کنگره» ایجاد شد و با ورود کتاب‌هایی با موضوعات جدیدتر، به طرح اصلی افزوده شدند. از این رو، رویکرد فرامای تولید داده‌های ساختارمند بر روی وب در توصیف و بازنمایی اشیای محتوایی وب (موجودیت‌های داده‌ای)، رویکردی هستی‌شناسانه و فلسفی است.

فرامای تولید داده‌های ساختارمند بر روی وب برای هر نوع در ساختار سلسله‌مراتبی، مجموعه‌ای از صفات را برای توصیف ویژگی‌های آن نوع شیء تعریف کرده است. این نوع نگاه به شیء اطلاعاتی و بررسی صفات آن و ارتباط با سایر نوع‌ها، ساختار هستی‌شناسانه‌ای را ایجاد می‌کند. بر اساس رویکرد هستی‌شناسانه، هر موجودیتی در ارتباط با سایر موجودیت‌ها به مرحله هستی می‌رسد و معنا پیدا می‌کند (طاهری ۱۳۹۲). این است که رویکرد فرامای تولید داده‌های ساختارمند بر روی وب در تعیین صفات یک شیء محتوایی و ارتباط با سایر اشیای محتوایی رویکردی هستی‌شناسانه است. کشف و ثبت روابط میان موجودیت‌ها در محیط اطلاعاتی نیز بر مبنای رویکرد هستی‌شناسانه است که به ایجاد ارزش افزوده منتهی می‌شود (همان). از این رو، صفات مطرح در این فرامای با صفات و ویژگی‌های مورد نظر استانداردهای فراداده‌ای بافت میراث فرهنگی کمی متفاوت است. صفاتی که برای توصیف یک موجودیت در فرامای تولید داده‌های ساختارمند بر روی وب به کار می‌رود، ممکن است، هم صفت و هم نشانگر رابطه‌ای نظیر هستی‌نگاشت‌ها باشند. به‌طور مثال، در جدول ۱، که بخشی از طرحواره «Book» را نشان می‌دهد، صفاتی نظیر نویسنده، تاریخ ایجاد، درباره (موضوع)، توصیف و جز این‌ها ذکر شده که همانند صفات مورد توصیف در استانداردهای فراداده‌ای بافت میراث فرهنگی است. اما صفتی نظیر «isPartOf»، صفت یا خصیصه یک شیء محتوایی مورد نظر استانداردهای فراداده‌ای بافت میراث فرهنگی نیست، بلکه نشانگر رابطه‌ای است که بین دو موجودیت برقرار می‌شود. این صفت حتی دارای صفت مقلوب «hasPart» نیز است و امکان تعریف آن وجود دارد (روابط قرینه). در هستی‌نگاشت‌ها روابط بین کلاس (موجودیت‌های مختلف را با مشخصاتی از قبیل تک‌ارزشی، تک‌ارزشی مقلوب، متعددی،

۱. در فرامای تولید داده‌های ساختارمند بر روی وب دو نوع گسترش عمده وجود دارد: گسترش اصلی و گسترش بیرونی. همچنین، در این فرامای به غیر از دو گسترش فوق، تغییرات پیوسته اصطلاحات در release، گسترش pending و ایده اصلاح دسترسی‌های اولیه جهت تسریع در روزآمدی مورد توجه است (Schema.org extensions 2015).

متمقارن، نامتمقارن، بازتابی، و غیربازتابی مشخص می‌کنند (محمدی استانی، آذرگون و چشمه‌سهرابی ۱۳۹۷) که در فرانمای تولید داده‌های ساختارمند بر روی وب نیز این روابط تحت عنوان صفات هر نوع، امکان تعریف دارند. یکی از دلایل رویکرد هستی‌شناسانه در فرانما می‌تواند همین نوع رویکرد یعنی در نظر گفتن روابط با سایر موجودیت‌ها، علاوه بر تعیین صفات باشد. لازم به ذکر است که صفت هر نوع شیء محتوایی در بستر معناشناختی فرانمای تولید داده‌های ساختارمند بر روی وب، نظیر روابط میان کلاس (موجودیت)‌ها در بافت هستی‌نگاشت است که هر دو ناظر بر یک مفهوم هستند.

در مدل داده‌ای مورد استفاده در فرانمای تولید داده‌های ساختارمند بر روی وب هر نوع در سلسله‌مراتب، مجموعه صفات خاصی را برای توصیف تعریف کرده‌اند. بنابراین، به‌واسطه ساختار سلسله‌مراتب گونه از صفات سطوح بالاتر، همانند رابطه پدر و فرزندی، ارث می‌برد. به‌عبارت دیگر، تمام صفات بالاترین و عام‌ترین سطح یعنی «thing»، برای تمام زیرنوع‌ها در سطوح دیگر قابل کاربرد است و از آن ارث می‌برند. این نوع رویکرد یعنی ارث‌بردن یا اشتراک صفات در فرانما برجسته است. از این رو، رویکرد فرانمای تولید داده‌های ساختارمند بر روی وب در ارث‌بردن یا اشتراک صفات در توصیف و سازماندهی اشیای محتوایی وب نیز رویکردی هستی‌شناسانه است. لازم به ذکر است که در رویکردهای شیء‌گرا^۱ نیز بحث ارث‌بری و اشتراک صفات مطرح است.

ب. رویکرد بافت‌مدار: بحث دیگری که در این فرانما مطرح است، تعیین نوع داده و نوع مورد انتظار است. به‌عبارت دیگر، برای هر نوع و هر صفت در فرانما، نوع داده‌ای اعم از متن، تصویر، URL و جز این‌ها و نوع موجودیت مورد انتظار اشیای محتوایی تعریف شده است. تعیین نوع داده و نوع مورد انتظار اشیای محتوایی در این فرانما، به‌نوعی روابط میان اشیای محتوایی و موجودیت‌ها را مشخص و تبیین می‌کند که این امر باعث ایجاد بافت خاصی خواهد شد. در متون، بافت را مجموعه‌ای از موجودیت‌های مرتبط به‌هم (مانند خانواده یا جامعه) تعریف کرده‌اند (طاهری ۱۳۹۲، ۴) که در این فرانما تعیین نوع داده و نوع اشیای محتوایی، بافت آن موجودیت را مشخص می‌کند (Microdata (HTML) (2016). از این رو، رویکرد فرانمای تولید داده‌های ساختارمند بر روی وب در توصیف و سازماندهی اشیای محتوایی وب، رویکرد بافت‌مدار است. به‌طور کلی، رویکرد بافت‌مدار

1. object-oriented approach

2. context-oriented approach

نسبت به اشیاء محتوایی، زمینه‌های تعامل میان انسان و ماشین و ماشین با ماشین را فراهم می‌کند که می‌تواند منجر به ارزش افزوده در داده‌ها از جمله ایجاد دانش و افزایش نقاط دسترس‌پذیری شود (طاهری و همکاران ۱۳۹۴).

ج. رویکرد آشیانه‌ای (تو در تو): رویکرد آشیانه‌ای را قرار گرفتن در درون چیزی از همان نوع تعریف کرده‌اند که به صورت مجزا نیز قابل شناسایی و پردازش است (Rouse 2005). هر یک از چهار نشانه‌گذار سازگار با این فرانما می‌تواند درون برچسب‌های «اچ‌تی‌ام‌ال» صفحات وب جای‌گذاری (جاسازی)^۲ شود. همچنین، امکان نشانه‌گذاری و ارتباط میان یک یا چند شیء محتوایی درون یک سند وجود دارد. فرانمای تولید داده‌های ساختارمند بر روی وب برای تعیین ارتباط میان اشیای محتوایی در بستر معناشناختی و فرمت‌های نشانه‌گذاری از رویکرد آشیانه‌ای بهره می‌گیرد. این فرانما با استفاده از همین رویکرد می‌تواند موجودیت‌های مختلف را در درون یک سند وبی، توصیف و ارتباطات میان آن‌ها را تعیین کند. موتورهای کاوش با تجزیه هوشمند این تگ‌ها، نسبت به پردازش و بازیابی موجودیت اصلی به همراه سایر موجودیت‌ها و روابطشان اقدام خواهند کرد. رویکرد متداول در برقراری ارتباط میان اشیای محتوایی از طریق صفات ارتباطی است؛ اما در رویکرد آشیانه‌ای به جای تعیین صفات ارتباطی، ارتباط میان موجودیت‌ها در درون یک سند ایجاد شده است. نمونه‌ای از این رویکرد در مثال‌های ۲ و ۳ به کار رفته است. در مثال ۳، صفت «author» با موجودیتی از نوع «Person» و صفت ویژه خود در درون سند موجودیت «Book»، نشانه‌گذاری شده است. به عبارت دیگر، سند اصلی مورد توصیف، موجودیتی از نوع «Book» است که سایر موجودیت‌ها از جمله «Person» با صفات خود قابل تعریف است. در مثال ۳، صفت «publisher» را که از نوع «organization» است نیز می‌توان تعریف کرد و ارتباط میان آن‌ها را با همین رویکرد نشان داد. همچنین، امکان تعریف چندین نوع موجودیت در درون اسناد وبی با رویکرد آشیانه‌ای وجود دارد.

پیاده‌سازی رویکرد آشیانه‌ای نیازمند استفاده از ابزار یا بسترهای معناشناختی است. همان‌طور که بیان شد بستر معناشناختی این فرانما مبتنی بر چهار نشانه‌گذار ریز فرمت، ریز داده، «آردی‌افای» و «جی‌سون-ال‌دی» برای نشانه‌گذاری و حاشیه‌نویسی اشیای محتوایی به کار می‌رود. بنابراین، فرانمای تولید داده‌های ساختارمند بر روی وب برای اجرای بستر معناشناختی

1. nesting approach

2. embed

خود از زبان‌های نشانه‌گذاری استفاده می‌کند. زبان نشانه‌گذاری، زبانی رایانه‌ای است که از برجسب‌ها برای تعریف عناصر درون یک سند استفاده می‌کند. این زبان انسان-فهم (Christensson 2011) در اسناد وبی به‌منظور نوشتن اطلاعات اضافی یا نشانه‌گذاری و تسهیل پردازش خودکار به کار می‌رود. نمونه‌هایی از زبان‌های نشانه‌گذاری در نحوهای مختلف در جدول ۲، ارائه شده است.

۵. نتیجه‌گیری

پژوهش مفهومی حاضر با رویکرد تحلیلی به معرفی و تبیین فرآیند تولید داده‌های ساختارمند بر روی وب - به‌عنوان یک هستی‌نگاشت فراداده‌ای به‌منظور ایجاد داده‌های ساختارمند - پرداخته است. همچنین مدل داده‌ای، ساختار، بستر معناشناختی، نشانه‌گذارهای معنایی، و تحلیل رویکردهای این فرآیند به پردازش و سازماندهی اشیای محتوایی (موجودیت‌های داده‌ای) در محیط وب نیز تحلیل گردید. واژگان فرآیند تولید داده‌های ساختارمند بر روی وب همراه با فرمت‌های نشانه‌گذاری برای حاشیه‌نویسی اشیای محتوایی وب به کار می‌رود. این طرحواره به جهت احراز هستی‌نگاشت واحد و یکدستی، تحولی در داده‌های ساختارمند قلمداد می‌گردد که می‌تواند کمک شایان توجهی به تحقق وب معنایی یا وب داده‌ها کند. در پاسخ به سؤال اول پژوهش، در مجموع می‌توان گفت که فرآیند تولید داده‌های ساختارمند بر روی وب یک واژگان یا هستی‌نگاشت واحد است که برای توصیف صفات و بازنمون روابط میان اشیای محتوایی وب و ایجاد داده‌های ساختارمند به‌منظور پردازش، سازماندهی و بازیابی مناسب و مرتبط توسط موتورهای کاوش معرفی و طراحی شده است.

در پاسخ به پرسش دوم پژوهش، نتایج پژوهش نشان داد که مدل داده‌ای مورد استفاده در این فرآیند بسیار عمومی و مشتق از «آردی‌اف‌اس» است. این مدل دارای انعطاف‌پذیری بالا و شامل مجموعه‌ای از نوع‌ها و صفات است. ساختار اصلی فرآیند تولید داده‌های ساختارمند بر روی وب به‌صورت سلسله‌مراتبی چندلایه و بستر معناشناختی آن بر نشانه‌گذارهای معنایی ریزفرمت‌ها، ریزداده‌ها، «آردی‌اف‌ای ۱/۱»، و «جی‌سون-ال‌دی» استوار است که در پرسش سوم پژوهش، فرمت‌های نشانه‌گذاری مورد بحث قرار گرفته است. هر یک از این فرمت‌ها با ویژگی‌ها و نحو خاص خود به نشانه‌گذاری اشیای محتوایی وب می‌پردازد.

در پاسخ به پرسش چهارم پژوهش، تحلیل رویکردهای فرانمای تولید داده‌های ساختارمند بر روی وب در پردازش و سازماندهی اشیای محتوایی حکایت از سه رویکرد دارد. رویکرد این فرانما در توصیف و بازنمایی اشیای محتوایی و ایجاد ساختار سلسله‌مراتبی چندلایه رویکردی هستی‌شناسانه است. تعیین صفات و روابط میان موجودیت‌ها و تعیین مشخصه‌های روابط نیز رویکردی هستی‌شناسانه است. ایجاد ساختار سلسله‌مراتبی و تعیین صفات باعث می‌شود که نوع‌های سلسله‌مراتب پایین‌تر از صفات سطوح بالاتر ارث ببرند. این ویژگی یعنی اشتراک و ارث‌بردن صفات نیز نشان از رویکرد هستی‌شناسانه در فرانمای تولید داده‌های ساختارمند بر روی وب دارد. تعیین نوع داده و نوع مورد نیاز اشیای محتوایی در این فرانما نشانگر بهره‌جویی از رویکرد بافت‌مدار دارد. در این فرانما برای تعیین ارتباط میان اشیای محتوایی در بستر معناشناختی و فرمت‌های نشانه‌گذاری، از رویکرد آشیانه‌ای (تودرتو) نیز استفاده شده است. برای پیاده‌سازی رویکرد آشیانه‌ای به بستر معناشناختی مناسب نیاز هست که زبان‌های نشانه‌گذاری این مهم را تحقق می‌بخشند. وجود رویکردهای مختلف در فرانمای تولید داده‌های ساختارمند بر روی وب نشانگر داشتن دیدی جامع نسبت به پردازش و سازماندهی اشیای محتوایی وب در ضمن بهبود میان‌کنش‌پذیری با موتورهای کاوش است. وجود رویکرد هستی‌شناسانه در این فرانما برجسته‌تر از سایر رویکردهاست.

ایجاد داده‌های ساختارمند در محیط وب موجبات مزایای مثبتی هم برای دارندگان و عرضه‌کنندگان محتوا و هم برای ارائه‌دهندگان برنامه‌های کاربردی وب و عامل‌های نرم‌افزاری فراهم می‌آورد و توجه به آن در طی سالیان اخیر رشد فزاینده‌ای داشته است. طراحان وب‌سایت‌ها جهت افزایش یافت‌پذیری و دسترس‌پذیری و نمایه‌پذیری محتواهای خود، مشتاق به استفاده از این فرامناها خواهند بود. بسیاری از دارندگان محتوا با صرف هزینه‌های گزاف و مشاوره با شرکت‌های فعال در این زمینه، خواهان قرار گرفتن در رتبه‌های بالای نتایج موتورهای کاوش وب هستند؛ در حالی که با رعایت و به‌کارگیری این فرامناها و نشانه‌گذاری اشیای محتوایی می‌توانند در نتایج موتورهای کاوش رتبه بهتری کسب کنند. در خصوص سایر مزایای استفاده از این فرامناها برای ایجاد داده‌های ساختارمند می‌توان به جست‌وجوی معنایی، سیستم‌های خبره، کشف دانش، مدیریت دانش، بازیابی اطلاعات، بازیابی دانش، و افزایش نرخ کلیک (Newcomer 2018) اشاره کرد. همچنین، وب‌سایت DataCite به‌عنوان یکی از نمایندگی‌های ثبت‌شاسگر اشیای

دیجیتالی^۱ برای اشیای محتوایی، اخیراً از فرآیند تولید داده‌های ساختارمند بر روی وب با نشانه‌گذار «جی‌سون-ال‌دی» برای ایجاد داده‌های ساختارمند در هنگام ثبت و تخصیص شناسگر اشیای دیجیتال استفاده می‌کند (Fenner 2017).

منظور از وب داده‌ها، ایجاد داده‌های ساختارمندی است که روابط میان آن‌ها بازنمون شده است. معنا از تعامل و ارتباط میان داده‌ها حاصل می‌شود. این نوع از داده‌ها، داده‌هایی هوشمند بوده که ایده‌های پیوندی در پی تحقق آن است. ایجاد داده‌های ساختارمند با چنین فرآیندهایی در ترکیب با نشانه‌گذارهای معنایی و ارتباط داده‌ها با یکدیگر، بازیابی دانش در محیط وب را در پی خواهد داشت. اشیای محتوایی که مبتنی بر این فرآیندها هستند و با نشانه‌گذارهای معنایی غنی شده‌اند، هنگام بازیابی در ترکیب با سایر داده‌های نشانه‌گذاری شده و ارتباطات میان آن‌ها علاوه بر بازیابی داده، دانش را نیز برای کاربران به ارمغان می‌آورند؛ زیرا موارد بازیابی شده، داده‌هایی است که با سایر داده‌ها در ارتباط است و ماحصل آن دانشی است که امکان دارد قبلاً تعریف نشده و در هنگام بازیابی این امر ایجاد شده است.

پیشنهاد می‌گردد دارندگان محتوا در بافت میراث فرهنگی به چنین فرآیندهایی نیز توجه داشته باشند؛ زیرا نوع‌های موجود در این بافت در فرآیند تولید داده‌های ساختارمند بر روی وب وجود دارد و می‌توان انواع اشیای محتوایی بافت میراث فرهنگی را با آن‌ها ساختارمند و نشانه‌گذاری کرد و در محیط وب منتشر ساخت. با این عمل یافت‌پذیری و دسترس‌پذیری اشیای محتوایی بافت میراث فرهنگی در محیط وب افزایش خواهد یافت. برای نمونه، می‌توان به ایجاد طرحواره‌های مرتبط با حوزه پزشکی بر اساس مدل سرعنوان‌های موضوعی پزشکی (مش)^۲ اشاره کرد که از طرف «کتابخانه ملی پزشکی آمریکا»^۳ نیز پشتیبانی می‌شود (Friedrich 2015). همچنین، پیشنهاد می‌شود پژوهش‌هایی جهت بررسی میان‌کنش‌پذیری واژگان فرآیند تولید داده‌های ساختارمند در وب با موتورهای کاوش صورت گیرد. در این صورت، در عمل، این استانداردهای باز که از سوی موتورهای کاوش ارائه شده‌اند، مورد سنجش و بررسی قرار می‌گیرند. اضافه بر آنچه اشاره شد، این پژوهش می‌تواند باعث مطالعات و پژوهش‌های بیشتر در حوزه فرآیند تولید داده‌های ساختارمند و نشانه‌گذارهای معنایی جهت انتشار در محیط وب گردد.

1. digital object identifier (DOI) 2. Medical Subject Heading (MESH) 3. U.S. National Library of Medicine

فهرست منابع

- آقاده، سمیرا. ۱۳۹۷. امکان‌سنجی درج عناصر فراداده‌ای مستند در پیشینه‌های داده‌های خرد و بررسی واکنش موتورهای کاوش وب به آن‌ها. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبائی، تهران.
- ذوالقدر، سارا. ۱۳۹۵. بررسی تطبیقی نمایه‌سازی و پیدانمایی خرده‌های غنی در موتورهای کاوش وب. پایان‌نامه کارشناسی ارشد، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران.
- طاهری، سید مهدی ۱۳۹۲. داده‌های پیوندی: مفاهیم و استانداردها: با تکیه بر جامعه میراث فرهنگی (قسمت اول). فصلنامه داخلی کنسرسیوم محتوای ملی ۱۲: ۴-۱۳.
- _____، ناهید گویلی، مریم شگفته، و مریم کازرانی. ۱۳۹۴. رویکرد بافت‌مدار در طراحی پروفایل کاربردی فراداده‌ای و توسعه وب معنایی. *مطالعات دانش‌شناسی* ۲ (۵): ۴۵-۵۷.
- طاهری، سید مهدی، رضا نیکزاد بهله، و میترا صمعی. ۱۳۹۶. بررسی واکنش موتورهای کاوش وب به فراداده‌ای مبتنی بر روش ترکیبی خرده‌های غنی و روش داده‌های پیوندی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۳ (۲): ۶۳۹-۶۵۸.
- محمدی استانی، مرتضی، مریم آذرگون، و مظفر چشمه‌سهرابی. ۱۳۹۷. روش‌شناسی ساخت و طراحی هستی‌نگاشت‌ها: مورد پژوهی حوزه علم‌سنجی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۳ (۴): ۱۷۹۳-۱۸۲۲.

References

- 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.
- _____, and P. Krause. 2015. *An Enhanced approach to semantic markup of VLEs content based on Schema.org*. In 4th International Workshop on Learning and Education with the Web of Data co-located with 13th International Semantic Web Conference (ISWC 2014) Riva del Garda, Italy, October 20, 2014.
- Barker, P. and L. M. Campbell. 2015. What is Schema.org? A Cetus briefing paper for LRMI. *LRMI*, 21. <http://publications.cetus.org.uk/wp-content/uploads/2014/06/schemaBriefing.pdf>. (accessed Mar. 10, 2018)
- Bizer, C., R. Meusel, and A. Primpeli. 2017. Web data common- RDFa, Microdata, and Microformat data sets: Extracting structured data from the common web crawl. University of Mannheim. <http://webdatacommons.org/structureddata/#results-2017-1>. (accessed May 24, 2018)
- Bohatch, M. 2014. Structured data, Schema.org & Rich Snippets explained. <http://www.simpleinbound.com/structured-data-schema-org-rich-snippets-explained/> (accessed May 18, 2018)
- Bohle-Lang, A. 2016. Semantic metadata for a library website. *GMS Medicine Library Information* 16 (3): 1-11.
- Christensson, P. 2011. *Markup Language Definition*. https://techterms.com/definition/markup_language# (accessed Jul. 2, 2018)
- Clark, J. A. and S. W. H. Young. 2015. Building a better book in the browser (using semantic web technologies and HTML5). *Code 4 lib Journal*. 29. <https://journal.code4lib.org/articles/10668>. (accessed Jun. 10, 2018)

- Col, T. W., M. K. Han, M. J. Sarol, M. Biel, and D. Maus. 2017. Using linked open data to enhance the discoverability, functionality and impact of Emblematica Online. *Library Hi Tech* 35 (1): 159-178. DOI: 10.1108/LHT-11-2016-0126.
- Data Model Schema.org. 2015. <http://schema.org/docs/datamodel.html>. (accessed May 17, 2018)
- DiFranzo, D., J. S. Ericson, M. J. K. T. Gloria, J. S. Luciano, D. L. McGuinness, and J. Hendler. 2014. *The Web Observatory Extension: Facilitating Web Science Collaboration through Semantic Markup*. In *Proceeding of 23rd International Conference on World Wide Web*, Seoul (P.475-480).
- Dourado, A. M. T. 2014. An approach to publish a data warehouse content as Linked Data. M.S. thesis, Department of Computer Science, Instituto Superior de Engenharia do Porto (isep), Porto. http://recipp.ipp.pt/bitstream/10400.22/5607/1/DM_AntonioDourado_2014_MEI.pdf (accessed May 22, 2018)
- Fenner, M. 2017. Using Schema.org for DOI Registration. <https://doi.org/10.5438/0000-00cc>. (accessed May 25, 2018)
- Friedrich, C. 2015. What search engines can't do: Holistic entity search on web data. Ph.D. dissertation. Department of Computer Science. Technische Universität Braunschweig.
- Gracy, K. F. 2015. Archival description and linked data: A preliminary study of opportunities and implementation challenges. *Arch Sci.* 15: 239-294. DOI: 10.1007/s10502-014-9216-2.
- Guha, R. V., D. Brickley, and S. Macbeth. 2015. Schema.org: Evolution of structured data on the web: Big data makes common schemas even more necessary. *Acmqueue* 13 (9). <http://queue.acm.org/detail.cfm?id=2857276>. (accessed May 8, 2018)
- Hawksey, M., P. Barker, and L. M. Compbell. 2013. *New approaches to describing and discovering open educational resources*. In *Proceedings of OER 13: Creating a Virtuous Circle*. Nottingham, Englan. http://publications.cetis.org.uk/wp-content/uploads/2013/04/OER13_resourcediscovery.pdf. (accessed Jun. 12, 2018)
- How We Work [Schema.org]. 2015. <http://schema.org/docs/howwework.html>. (accessed May 13, 2018)
- Kellogg, G. 2017. *JSON-LD Update*. https://www.w3.org/2017/02/wot-f2f/slides/JSON-LD_Update.pdf. (Accessed May 29, 2018)
- Matosevic, G. 2014. The adoption of semantic annotations of products in web shops. *International Journal of Computer and Communication Engineering* 3 (1): 6-10. DOI: 10.7763/IJCE.2014.V3.282.
- Medeiros, I. 2014. *Linked data in use: Schema.org, JSON-LD, and hypermedia APIs* (PowerPoint Slides). <https://www.slideshare.net/icaromedeiros/linked-data-in-use-front-in-bahia-2014>. (accessed Jun. 3, 2018)
- Microdata (HTML). 2016. [https://en.wikipedia.org/wiki/Microdata_\(HTML\)](https://en.wikipedia.org/wiki/Microdata_(HTML)). (accessed Mar. 29. 2018)
- Meusel, R., C. Bizer, and H. Paulheim. 2015. *A web-scale study of the adoption and evolution of the Schema.org vocabulary over time*. In *Proceeding of the 5th International Conference on WIMS*, July 13-15, Laranca, Cyprus. New York: ACM. DOI: 10.1145/2797115.2797124.
- Newcomer, C. 2018. *How to add Schema and rich snippets in WordPress*. <https://bloggingwizard.com/add-schema-and-rich-snippets-in-wordpress/> (accessed May 18, 2018)
- Nguyen, T. T. and I. Song. 2016. Big data visualization: Application in visualizing learning activities. In: Kuinam, J. K., Nikolai, J. (editors). *Information Science and applications (ICISA) 2016* (p. 399-408). Singapore: Springer.
- Jangra, R., A. A. Shashikumara, and P. Kikon. 2018. *Metadata standards for content description: Microdata, Microformats and JSON-LD*. <https://doi.org/10.6084/m9.figshare.6022008.v2> (accessed May 29, 2018)
- Paulheim, H. 2015. *What an adoption of Schema.org tells about Linked Open*

- Data. In The 2nd International Workshop on Dataset PROFiling & Federated Search for Linked Data. <https://www.slideshare.net/heikopaulheim/what-the-adoption-of-schemaorg-tells-about-linked-open-data>. (accessed May 9, 2018)
- Riva, P., P. L. Bœuf, and M. Zumer. 2017. *IFLA Library Reference Model: A conceptual model for bibliographical information*. <https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf>. (accessed Jun. 20, 2018)
- Rasool, S., F. Iqbal, and A. A. Malik. 2013. *Referencing Qur'an Verses and Ahadith on World Wide Web using Microdata*. In Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences. Saudi Arabia. 22-25 Dec. 2013. (P.298-303). DOI: 10.1109/NOORIC.2013.66.
- Rosati, A. and M. Mayernik. 2013. *Facilitating Data Discovery by Connecting Related Resources*. <http://www.semantic-web-journal.net/system/files/swj484.pdf>. (accessed Mar. 11, 2018)
- Rouse, M. 2005. *Nested definition*. <https://whatis.techtarget.com/definition/nested> (accessed July 8, 2018)
- Sack, H. 2014. *Knowledge Engineering with semantic web technologies*. Presented in Hasso Plattner Institute (HPI). Germany.
- Schema.org. 2014. *Frequently asked question*. <http://schema.org/docs/faq/html>. (accessed Mar. 8, 2018)
- Schema.org Extensions. 2015. <http://schema.org/docs/extension.html>. (accessed Mar. 10, 2018)
- Schema.org types and properties tutorial. 2018. <http://www.w3resource.com/schema.org/types-and-properties.php>. (accessed May 25, 2018)
- Sikos, L. F. 2017a. Knowledge Representation with semantic web knowledge. In: *Description logics in multimedia reasoning*. New York: Springer International Publishing. P11-49. DOI: 10.1007/978-3-319-54066-5_2.
- _____. 2017b. Knowledge Representation. In: *Mastering Structured Data on the Semantic Web: From HTML5 microdata to Linked Open Data*. New York: Apress. P13-57. DOI: 10.1007/978-1-4842-1049-9.
- Wetherill, B. 2014. *RDFa and Microdata: Is One Better Than The Other?* A Research Proposal Submitted in Partial Fulfillment of the Requirements for LIBR 281 – Metadata. San Jose State University.
- Yu, L. 2014. Schema.org and semantic markup. In: *A developer's guide to the semantic web*. Berlin: Springer. P.475-516. DOI: 10.1007/978-3-662-43796-4_10.

مرتضی محمدی استانی

متولد ۱۳۶۴، دانشجوی دکتری علم اطلاعات و دانش‌شناسی دانشگاه اصفهان و مدرس گروه علم اطلاعات و دانش‌شناسی دانشگاه قم است. وب معنایی، هستی‌شناسی، سازماندهی اطلاعات، فراداده و مدیریت دانش از جمله علایق پژوهشی وی است.



مظفر چشمه‌سهرابی

متولد ۱۳۵۳، دارای مدرک دکتری در رشته علوم اطلاعات و ارتباطات است. ایشان هم‌اکنون دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است.



علم‌شناسی، سنجش و ارزشیابی علم و پژوهش، اخلاق علمی، بازیابی معنایی، وب معنایی و هستی‌شناسی از جمله علایق پژوهشی وی است.

احمد شعبانی

متولد ۱۳۳۵، دارای مدرک دکتری در رشته علم اطلاعات و دانش‌شناسی است. ایشان هم‌اکنون استاد گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است.



مدیریت دانش و روش تحقیق از جمله علایق پژوهشی وی است.

عاصفه عاصمی

متولد ۱۳۴۸، دارای مدرک دکتری در رشته علم اطلاعات و دانش‌شناسی - تکنولوژی اطلاعات از پونا، هندوستان است. ایشان هم‌اکنون دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است. سیستم‌های اطلاعاتی از جمله علایق پژوهشی وی است.



سید مهدی طاهری

متولد ۱۳۵۷، دارای مدرک دکتری در رشته علم اطلاعات و دانش‌شناسی از دانشگاه آزاد واحد علوم و تحقیقات است. ایشان هم‌اکنون استادیار گروه علم اطلاعات و دانش‌شناسی دانشگاه علامه طباطبائی است.



فرا داده، سیستم‌ها و وب معنایی، موتورهای کاوش، و کتابخانه‌های دیجیتالی از جمله علایق پژوهشی وی است.