

## Domain Ontology Modeling in the Digital Archive of the Islamic Revolution Document Center: A Hybrid Approach of Text Analysis and Ontology Reuse

|                                                                                                                   |                                                                                                                                                        |
|-------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Nader Naghshineh</b>          | Associate Professor, Knowledge and Information Science Dept., Faculty of Management, University of Tehran, Tehran, Iran                                |
| <b>Alireza Entehaee Saray *</b>  | Ph.D. Student in Knowledge and Information Science, Information Retrieval Specification, University of Tehran, Tehran, Iran                            |
| <b>Behrouz Minaei-Bidgoli</b>  | Full Professor, Artificial Intelligence and Robotics Depart., Faculty of Computer Engineering, Iran University of Science and Technology, Tehran, Iran |
| <b>Ali Shabani</b>             | Assistant Professor, Knowledge and Information Science Dept., Faculty of Management, University of Tehran, Tehran, Iran                                |

### Abstract

Organizing information in digital archives, particularly historical ones such as the Islamic Revolution Document Center of Iran, faces challenges like the inefficiency of traditional information retrieval systems. This center, with over 4.5 million document pages, 31 thousand hours of oral history, and millions of news titles and articles, requires innovative approaches such as domain ontology and knowledge graphs to improve semantic access to key entities.

\* Corresponding Author: rentehaee@gmail.com

**How to Cite:** Entehaee Saray, A., Naghshineh, N., Minaei Bidgoli, B., & Shabani, A. (2025). Domain ontology modeling in the digital archive of the Islamic Revolution Document Center: A hybrid approach of text analysis and reuse of ontologies. Journal of Knowledge Retrieval and Semantic Systems, volume(issue), pp. start-end.

The aim of this research is to model a domain ontology for the digital archive of this center using a hybrid approach of text analysis and reuse of existing ontologies. The research method is mixed: semi-automatic and automatic text analysis for entity extraction, model design based on ontology principles, and validation of findings using the nominal group technique. The findings include 535 classes that, after validation with criteria of 1) hierarchical logic, 2) historical/cultural accuracy, 3) alignment and equivalence, and 4) completeness and simplicity, resulted in 457 classes confirmed, 70 classes added, 53 classes relocated in the hierarchy, and 19 classes removed. This ontology provides a foundation for a knowledge graph, enhances semantic retrieval, and serves as a logical basis for artificial intelligence systems in managing Iran's historical documents. This research fills the gap in designing practical ontologies for contemporary Iranian history and is extensible to similar domains.

**Keywords:** Domain ontology, Knowledge graph, Semantic information retrieval, Semantic search, Semantic network, Entity extraction, Digital archive, Islamic Revolution Document Center.



## مدل سازی هستی شناسی دامنه در آرشیو دیجیتال مرکز اسناد انقلاب اسلامی: رویکرد ترکیبی تحلیل متن و بازاستفاده از هستی شناسی ها<sup>۱</sup>

دانشیار گروه علم اطلاعات و دانش شناسی، دانشکده مدیریت، دانشگاه تهران،  
تهران، ایران

ناظر نقشینه 

دانشجوی دکتری رشته علم اطلاعات و دانش شناسی - گرایش بازیابی اطلاعات و  
دانش، دانشکده مدیریت، دانشگاه تهران، تهران، ایران

علیرضا انتهای  
سرای 

استاد گروه هوش مصنوعی و رباتیک، دانشکده مهندسی کامپیوتر، دانشگاه علم و  
صنعت ایران، تهران، ایران

بهروز مینایی  
بیدگلی 

استادیار گروه علم اطلاعات و دانش شناسی، دانشکده مدیریت، دانشگاه تهران،  
تهران، ایران

علی شعبانی 

### چکیده

سازماندهی اطلاعات در آرشیوهای دیجیتال، به ویژه آرشیوهای تاریخی مانند مرکز اسناد انقلاب اسلامی ایران، با چالش هایی مانند ناکارآمدی نظام های سنتی بازیابی اطلاعات مواجه است. این مرکز، با بیش از ۴۵ میلیون برگ سند، ۳۱ هزار ساعت تاریخ شفاهی، و میلیون ها عنوان خبری و مقاله، نیاز به رویکردهای نوین مانند هستی شناسی دامنه و گراف دانش دارد تا دسترسی معنایی به موجودیت های کلیدی را بهبود بخشد. هدف این پژوهش، مدل سازی هستی شناسی دامنه برای آرشیو دیجیتال این مرکز با رویکرد ترکیبی تحلیل متن و بازاستفاده از هستی شناسی های موجود است. روش پژوهش آمیخته است: از تحلیل متن به شکل نیمه خودکار و خودکار برای استخراج موجودیت ها، طراحی مدل بر اساس اصول هستی شناسی و اعتبارسنجی یافته ها با تکنیک گروه اسمی استفاده شده است. یافته ها شامل ۵۳۵ رده است که پس از اعتبارسنجی با معیارهای (۱) منطق سلسله مراتبی، (۲) صحت تاریخی/فرهنگی، (۳) همترازی و معادل سازی و (۴) کامل بودن و سادگی، ۴۵۷ رده تأیید، ۷۰ رده اضافه شده، ۵۳ رده در سلسله مراتب جابجا و ۱۹ رده حذف شدند. این

<sup>۱</sup> مقاله حاضر برگرفته از رساله دکتری رشته علم اطلاعات و دانش شناسی دانشگاه تهران است.

\* نویسنده مسئول: a.entehaee@ut.ac.ir

هستی‌شناسی مبنایی را برای گراف دانش فراهم می‌کند، بازیابی معنایی را ارتقا می‌دهد، و پایه‌ای منطقی برای سیستم‌های هوش مصنوعی در مدیریت اسناد تاریخی ایران است. این پژوهش خلاء طراحی هستی‌شناسی کاربردی در تاریخ معاصر ایران را پر کرده و قابل توسعه برای دامنه‌های مشابه است.

**کلیدواژه‌ها:** هستی‌شناسی دامنه، گراف دانش، بازیابی اطلاعات معنایی، جستجوی معنایی، شبکه معنایی، استخراج موجودیت، آرشیو دیجیتال، مرکز اسناد انقلاب اسلامی

## مقدمه

سازماندهی اطلاعات و دانش یکی از ارکان هر نظام بازیابی اطلاعات است. در واقع، پرس و جوی کاربر با نمایه‌های نظام بازیابی به مدارک و اطلاعات هدایت می‌شود و میزان اثربخشی آن به نوع و روش سازماندهی اطلاعات و کارکرد صحیح آن مرتبط می‌شود. تاکنون روش‌های مختلفی برای سازماندهی اطلاعات و منابع اطلاعاتی تعریف شده و بکار گرفته شده‌اند که از جمله آن‌ها می‌توان به نظام‌های رده‌بندی، نمایه‌سازی، فهرست‌نویسی، نظام‌های کدگذاری و جستجوی متن آزاد اشاره کرد. برای پشتیبانی از این نظام‌ها منابع مختلفی نیز به عنوان طرح‌های رده‌بندی، سرعنوان‌های موضوعی، اصطلاحنامه‌ها، مستندات نام‌های اشخاص، مکان‌های جغرافیایی و سازمان‌ها توسعه پیدا کرده‌اند. همچنین استانداردهای مختلفی برای توصیف، کدگذاری و مدیریت ساختار منابع تدوین شده‌اند تا امکان ذخیره‌سازی و بازیابی اطلاعات به شکل یکدست را فراهم کرده و امکان تعامل‌پذیری بین سامانه‌های اطلاعاتی را فراهم کنند. با ظهور عصر دیجیتال و توسعه سریع استانداردهای حوزه داده و اطلاعات و فناوری‌های جدید کتابخانه‌ها و آرشیوهای دیجیتالی گروه‌های مختلف به دنبال راهکارهای نوینی برای سازماندهی بهینه حجم بالایی از اطلاعات دیجیتال هستند. رویکردهای دهه اخیر در این حوزه متمرکز بر وب معنایی و داده‌های پیوندی، نسل جدید فهرست‌های کتابخانه‌ای، چارچوب دسترسی و توصیف منبع<sup>۱</sup>، سازماندهی و به اشتراک‌گذاری تصاویر دیجیتال، جستجو و مرور چهریزه‌ای و استانداردهای کیفیت فراداده‌ها بوده است (PARK & HOWARTH, 2013).

یکی از رویکردهای نوین برای سازماندهی اطلاعات و دانش استفاده از هستی‌شناسی (یا آن‌تولوژی) و در مفهوم دیگر توسعه گراف‌های دانش است. در هستی‌شناسی تمرکز تنها بر شناسایی مفاهیم به عنوان نقاط دسترسی نیست، بلکه جایگاه مفاهیم در دنیای هستی و نحوه ارتباط و اثرگذاری آنها نیز مورد توجه قرار می‌گیرد. بنابراین آنچه مورد پرس و جو قرار می‌گیرد صرفاً یک نمایه مرتبط با مدارک نیست، بلکه شبکه یا گرافی از دانش مورد کاوش و بازنمایی آن است. گراف دانش<sup>۲</sup> که به عنوان یک حوزه اصلی در بازیابی اطلاعات و پشتیبانی از هوش مصنوعی ظاهر شده است، اولین بار توسط گوگل برای بازنمون دانش از

<sup>۱</sup> RDA

<sup>۲</sup> Knowledge Graphs

حجم وسیعی از اطلاعات مورد استفاده قرار گرفت (Kejriwal, Knoblock, & Szekely, 2021). گراف‌های دانش مجموعه‌ای بزرگ از موجودیت‌های مرتبط به هم هستند که بوسیله برچسب‌های معنایی غنی شده‌اند. در اینجا منظور از موجودیت انواع موجودیت‌های نامدار و غیر نامدار مانند اسامی اشخاص، مکان، سازمان، رویداد، زمان، مفاهیم و ... است (Sajadi & Minaei Bidgoli, 2020).

یکی از چالش‌های کلیدی آرشیوهای دیجیتال، به‌ویژه آرشیوهای تاریخی مانند آرشیو دیجیتال مرکز اسناد انقلاب اسلامی، ناکارآمدی نظام‌های سنتی بازیابی اطلاعات در ارائه دسترسی جامع و معنایی به محتوای اسناد است. این مرکز که در سال ۱۳۶۰ با هدف تدوین تاریخ انقلاب اسلامی و جمع‌آوری اسناد، مدارک و خاطرات مرتبط با نهضت امام خمینی تأسیس شد، آرشیوی غنی شامل ۴۵ میلیون برگ سند مکتوب، ۳۱ هزار ساعت مصاحبه تاریخ شفاهی، ۳۷۱،۹۰۶ عکس، ۱۳۶ میلیون عنوان خبری و ۹۰،۰۰۰ مقاله داخلی و خارجی را در اختیار دارد. بخش قابل توجهی از این اسناد با استفاده از فناوری‌های تشخیص کاراکتر نوری<sup>۱</sup> و تبدیل گفتار به متن<sup>۲</sup> به فرمت دیجیتال تبدیل شده‌اند، اما نظام فعلی بازیابی اطلاعات این مرکز عمدتاً بر فراداده‌های توصیفی اسناد و جستجوی درون متنی متمرکز است. این امر سبب شده تا بسیاری از موجودیت‌های اطلاعاتی کلیدی (مانند اسامی اشخاص، مکان‌ها، رویدادها و مفاهیم) در جستجوهای کاربران مخفی بمانند و دسترسی پژوهشگران به اطلاعات جامع محدود شود. بنابراین، ضرورت بازطراحی نظام بازیابی اطلاعات با استفاده از رویکردهای مبتنی بر گراف دانش و هستی‌شناسی به منظور استخراج و سازماندهی معنایی موجودیت‌ها و بهبود دسترسی به اطلاعات احساس می‌شود.

هدف این پژوهش، مدل‌سازی هستی‌شناسی دامنه برای آرشیو دیجیتال مرکز اسناد انقلاب اسلامی با استفاده از رویکرد ترکیبی تحلیل متن و بازاستفاده از هستی‌شناسی‌های موجود است. این پژوهش با تمرکز بر استخراج خودکار و نیمه‌خودکار موجودیت‌های اطلاعاتی (مانند اشخاص، سازمان‌ها، رویدادها و مفاهیم) و طراحی ساختار سلسله‌مراتبی، به دنبال ایجاد پایه‌ای برای گراف دانش است که بتواند از بازیابی اطلاعات معنایی پشتیبانی کند. این رویکرد ترکیبی، با بهره‌گیری از فناوری‌های پردازش زبان طبیعی و مدل‌های زبانی

<sup>۱</sup> OCR

<sup>۲</sup> ASR

بزرگ و یکپارچه سازی هستی شناسی ها و رده بندی های دیگر مانند دی بی پدیا<sup>۱</sup>، هستی شناسی گراف دانش فارسی<sup>۲</sup> و رده بندی رده DSR تاریخ ایران، امکان سازماندهی بهینه و بازنمایی دانش موجود در آرشیو را فراهم می کند.

در این راستا، پژوهش با استفاده از فناوری های پردازش زبان طبیعی، مدل های زبانی بزرگ و تلفیق هستی شناسی ها و رده بندی های معتبر، به دنبال فراهم سازی بستر سازماندهی و بازنمایی معنایی اطلاعات آرشیوی است. اهداف فرعی پژوهش عبارت اند از:

۱. شناسایی و استخراج خودکار و نیمه خودکار موجودیت های اطلاعاتی (از جمله اشخاص، سازمان ها، رویدادها و مفاهیم) از اسناد آرشیو دیجیتال مرکز اسناد انقلاب اسلامی؛
۲. طراحی ساختار سلسله مراتبی رده ها و روابط میان موجودیت ها در چارچوب یک هستی شناسی دامنه؛
۳. شناسایی و بازاستفاده از هستی شناسی ها و رده بندی های موجود مناسب برای یکپارچه سازی با مدل پیشنهادی.

بدین ترتیب، نتایج این پژوهش می تواند به توسعه یک گراف دانش تاریخی برای ایران و بهبود نظام های بازیابی اطلاعات معنایی در محیط های آرشیوی کمک کند.

### پیشینه پژوهش

پژوهش های انجام گرفته در این حوزه به مقوله های استخراج خودکار یا ترکیبی اطلاعات ساختاریافته مبتنی بر مدل هستی شناسی، توسعه گراف دانش، پاسخگویی به پرس و جوها بر اساس گراف دانش، طراحی هستی شناسی به صورت انسانی، بهینه سازی بازیابی اطلاعات با هستی شناسی، بازیابی اطلاعات و موتورهای جستجوی مبتنی بر پایه گراف دانش، بهینه سازی پرس و جو بر مبنای گراف دانش، طراحی سیستم های اطلاعاتی مبتنی بر هستی شناسی و غنی سازی اصطلاحنامه ها برای ساخت هستی شناسی توجه کرده اند.

امامی (امامی، ۱۴۰۱)؛ مرادی، وزیرنژاد و بحرانی (مرادی، وزیرنژاد و بحرانی، ۱۳۹۴)؛ گوتیرز و همکارانش (Gutierrez et al., 2016) و ویمالاسوریا و دوو

<sup>1</sup> DBpedia

<sup>2</sup> farsbase.net

(Wimalasuriya & Dou, 2010) به حوزه استخراج خودکار یا ترکیبی اطلاعات به صورت ساختاریافته از درون متن پرداخته‌اند.

امامی به دنبال روش جدیدی برای استخراج خودکار اطلاعات ساختاریافته محدود به دامنه از متون فارسی صفحات وب که نیازی به دانش پیش‌زمینه در مورد موجودیت‌ها و ویژگی‌های آن‌ها ندارد، پرداخته و بر اساس آزمایشی کارآیی آن را بهتر از روش‌های دیگر یافته‌است.

مرادی، وزیرنژاد و بحرانی با هدف ساخت هستی‌شناسی برای دانش عرفی زبان فارسی به کمک سه روش مبتنی بر الگو، ترجمه ماشینی و استفاده از منابع ساخت‌یافته به کمک سه پیکره زبان فارسی هفت‌نوع رابطه و در مجموع هفتاد هزار اظهار (رابطه) استخراج کرده و در قالب یک هستی‌شناسی ارائه کرده‌اند. میانگین دقت روابط استخراج‌شده این هستی‌شناسی برابر با ۷۵ درصد برای روش مبتنی بر الگو، ۷۰ درصد برای ترجمه ماشینی و ۱۰۰ درصد برای اطلاعات استخراج‌شده از جعبه اطلاعات بود.

گوتیرز و همکارانش یک سیستم هیبرید را برای توسعه دو افزونه برای افزودن دقت استخراج ماشینی موجودیت‌ها و روابط آنها از متن پیشنهاد کرده‌اند.

ویمالاسوریا و دوو مقدمه‌ای بر استخراج خودکار اطلاعات به کمک مدل هستی‌شناسی پرداخته و جزئیات سامانه‌هایی که تاکنون برای این منظور توسعه یافته‌اند را بررسی کرده‌اند.

در مقوله توسعه گراف دانش، مینایی و سجادی (سجادی و مینایی بیدگلی، ۱۳۹۸)، رجبی و اطمینانی (Rajabi & Etminani, 2022)؛ چویی (Choi, 2022)، تان و همکارانش (Tan, Qiu, Guo, & Li, 2021)، عسگری و همکارانش (Asgari-Bidgoli, 2019)، زکی و همکارانش (Zaki, Bidhendi, Hadian, & Minaei-Bidgoli, 2019)، رحمان و تاکاسو (Tennakoon, Al Ashwal, Al Jaber, & Al Ameri, 2019) و رحمان و تاکاسو (Rahman & Takasu, 2017) به شکل متمرکز به ساختن گراف دانش، توسعه آن یا بررسی ادبیات این مقوله پرداخته‌اند.



مینایی و سجادی به دنبال استخراج گراف دانش زبان فارسی از ویکی‌پدیای فارسی برای تهیه جامع‌ترین پایگاه دانش زبان فارسی به صورت عمومی و چند دامنه‌ای بوده که گرافی مشتمل بر ۵۰۰ هزار موجودیت و ۷ میلیون رابطه میان آن‌ها ارائه کرده‌اند.

رجبی و اطمینانی به مرور نظام‌مند پرداخته و در ۴۰ مقاله موضوع نحوه بکارگیری گراف‌های دانش در هوشمندی درونی سیستم‌های هوش مصنوعی را بررسی کرده‌اند و نتیجه گرفته‌اند که عمده این گراف‌ها به شناسایی و استخراج موجودیت‌ها و روابط آنها و همچنین تفسیر و استنتاج پرداخته‌اند.

چویی در مطالعه خود از خلاصه ادبیات پژوهش یک روش گسترش گراف دانش با درونی سازی گراف پیشنهاد کرده است.

تان و همکارانش به ساخت مدل گراف دانش و استنتاج دانش در حوزه ترافیک شهری پرداخته و نتایج این تحقیق نشان داده است که این مدل در شناسایی روابط پنهان میان موجودیت‌های ترافیک شهری کارآمد هستند.

عسگری و همکارانش از فنون متنوع و ترکیبی برای استخراج و یکپارچه‌سازی دانش از منابع مختلف مانند ویکی‌پدیا، جداول وب و متون ساختار نیافته استفاده کرده‌اند تا فارس‌بیس را که یک گراف دانش باز فارسی است را ایجاد کنند.

زکی و همکارانش به بررسی روش‌هایی پرداخته‌اند که پایگاه‌هایی اطلاعاتی زیست پزشکی را به هم پیوند داده و اثربخشی رویکردهای موجود را که می‌توانند پیوندهای ناشناخته بین موجودیت‌ها را در پایگاه‌های داده به منظور بهینه‌سازی گراف دانش پیش‌بینی کنند را ارزیابی کرده‌اند.

رحمان و تاکاسو دو مدل را ارائه کرده‌اند که روابط موجودیت‌ها برای تولید گراف دانش را پیشنهاد می‌کند.

یانگ و همکارانش (Yang, Guo, Dai, & Chen, 2023) و تانگ، ژانگ و یائو (Tong, Zhang, & Yao, 2019) به حوزه پاسخگویی به پرس و جوها بر اساس گراف دانش پرداخته‌اند.

یانگ و همکاران یک مدل جدید برای تجزیه سوال را بکار گرفتند که از مکانیسم‌های توجه و رمزگذارهای دوگانه استفاده می‌کند تا بازنمایی سوال را بهبود بخشد. نتایج آزمایشی روی مجموعه داده ComplexWebQuestions کارآمدی مدل پیشنهادی را نشان می‌دهد، شاخص F1 به امتیاز ۷۲ درصد دست می‌یابد و اثربخشی این رویکرد را در تجزیه سؤالات پیچیده به زیرمجموعه‌های ساده و بهبود دقت منطق سؤال را نشان می‌دهد.

تانگ و همکارانش در مطالعه خود مدلی را ارائه می‌کنند که به سوال‌ها بر اساس گراف دانش موضوعی پاسخ می‌دهد.

در مقوله طراحی هستی‌شناسی به صورت انسانی، ترابی و همکارانش به طراحی و پیاده‌سازی هستی‌شناسی حوزه پزشکی هسته‌ای (ترابی و دیگران، ۱۴۰۱)؛ نیارکی، برزگر و شاکری (صادقی نیارکی، برزگر و شاکری، ۱۳۹۷) به طراحی و توسعه هستی‌شناسی تجربیات مکانی در مسیریابی شهری؛ تقی‌زاده نائینی، فهم‌نیا و نقشینه (تقی‌زاده نائینی، فهم‌نیا و نقشینه، ۱۳۹۸) به استخراج هستی‌شناسی هویت دیجیتال مبتنی بر تحلیل حوزه و ثروتی، ولوی و حورعلی (ثروتی، ولوی و حورعلی، ۱۳۹۶) به توسعه توسعه یک هستی‌شناسی جغرافیایی برای استفاده در کاربردهای نظامی پرداخته‌اند. همه این پژوهش‌ها بر اساس تجربه انسانی به مدلسازی یک هستی‌شناسی مبتنی بر دامنه پرداخته‌اند.

در حوزه بازیابی اطلاعات و موتورهای جستجوی مبتنی بر پایه گراف دانش، همایونی و همکارانش (Homayouni, Ding, Roy, & Ebeid, 2018)، لی (Li, 2019) و وانگ و همکارانش (Wang, Jiang, Xu, & Zhang, 2019) پژوهش کرده‌اند.

همایونی و همکارانش مطالعه گراف پزشکی<sup>۱</sup>: به عنوان یک چارچوب بازیابی اطلاعات زیست پزشکی معنایی با استفاده از جاسازی گراف دانش برای پابمد<sup>۲</sup> را انجام دادند. نتایج این مطالعه شواهدی را ارائه می کند مبنی بر اینکه رویکردهای معنایی برای جستجو و ارتباط در کتابخانه های دیجیتال زیست پزشکی که بر مدل سازی گراف دانش تکیه دارند، نتایج جستجوی بهتری را در مقایسه با روش های سنتی از نظر معیارهای هدف ارائه می دهند.

لی در مقاله خود به بررسی تئوری ساخت و روش پیاده سازی این نوع موتورهای جستجو در یک حوزه خاص می پردازد، به این امید که از طریق این موتور جستجو بتوان پرس و جوی موجودیت را در یک حوزه خاص ارائه کرد، یعنی از طریق درک عمیق نیازهای کاربر، موجودیت های مربوطه را از پایگاه داده موجودیت بازیابی و به کاربران بازگرداند.

وانگ و همکارانش در پژوهش خود راه حل ساخت یک گراف دانش چندمنبعی در مقیاس بزرگ را از ابتدا برای شرکت سوگو<sup>۳</sup>، شامل معماری، پیاده سازی فنی و کاربردهای آن به منظور پشتیبانی از جستجو در وب و فوآر از آن معرفی می کند.

مطالعه شن و همکارانش بر روی بهینه سازی پرس و جو بر مبنای گراف دانش (Shen, Chen, Chen, Zeng, & Xiao, 2022) بوده است. ایشان یک مولد پرسش پیشنهاد کرده اند که از یک گراف دانش بیرونی به عنوان راهنما استفاده می کند.

در مقوله بهینه سازی بازیابی اطلاعات با هستی شناسی، جعفری پاورسی و همکارانش (پاورسی و دیگران، ۱۳۹۹) به بهینه سازی بازیابی معنایی اطلاعات با استفاده از روش های برچسب گذاری و هستی شناسی پرداخته و به این نتیجه رسیده اند که دو روش پردازش زبان طبیعی و هستی شناسی به ارتقای بازیابی معنایی اطلاعات منجر می شود.

علاوه بر این، اسیم و همکارانش (Asim et al., 2019) به مرور و ارزیابی روش های جدید بازیابی اطلاعات مبتنی بر هستی شناسی برای متون و چندرسانه ای ها و انواع داده های بین زبانی پرداخته است.

<sup>1</sup> MedGraph

<sup>2</sup> PubMed

<sup>3</sup> Sogou Inc

عالیشان کرمی و همکارانش (عالیشان کرمی و دیگران، ۱۳۹۶) در پژوهشی به بررسی ۵ مقاله پرداخته‌اند که در خصوص کاربردها و روش ساخت هستی‌شناسی‌هایی هستند که در نظام‌های بازیابی اطلاعات زیست‌پزشکی مبتنی بر هستی‌شناسی به کار رفته‌اند. یکی از پژوهش‌های یافت شده به طراحی سیستم اطلاعاتی مبتنی بر هستی‌شناسی در حوزه کشاورزی پرداخته است.

لیائو و همکارانش یک سامانه اطلاعاتی کشاورزی یکپارچه بر اساس هستی‌شناسی پیشنهاد کرده‌اند تا اطلاعات جامع و دقیقی برای راهنمایی کارآمد کشاورزان و متخصصان کشاورزی ارائه کند (Liao, Li, & Liu, 2015).

در مقوله غنی‌سازی اصطلاحنامه‌ها برای ساخت هستی‌شناسی، حسینی بهشتی و اژه‌ای (حسینی بهشتی و اژه‌ای، ۱۳۹۴) و وو (Wu, 2018) به ساخت هستی‌شناسی بر اساس اصطلاحنامه موضوعی پرداخته‌اند.

حسینی بهشتی و اژه‌ای به طراحی و پیاده‌سازی هستی‌شناسی علوم پایه بر اساس مفاهیم و روابط موجود در اصطلاحنامه‌های مرتبط تمرکز کرده است.

وو روشی را برای غنی‌سازی اصطلاحنامه با ایجاد روابط متفاوت بین اصطلاح‌های مرتبط بخشی از اصطلاحنامه اریک<sup>۱</sup> ارائه کرده است. نتایج آزمایشی این مطالعه نشان داد که اصطلاحنامه غنی شده ابزار بهتری برای بازیابی اطلاعات است.

مرور ادبیات موجود در حوزه سازماندهی دانش و اطلاعات بر پایه هستی‌شناسی و گراف‌های دانش نشان می‌دهد که بخش قابل توجهی از پژوهش‌های انجام شده در دهه‌های اخیر در چند محور عمده متمرکز بوده‌اند. نخست، پژوهش‌هایی که به استخراج خودکار یا ترکیبی اطلاعات ساختاریافته از متون پرداخته‌اند. دوم، مطالعاتی که به توسعه گراف‌های دانش در دامنه‌های خاص مانند زبان فارسی، زیست‌پزشکی، ترافیک شهری یا سامانه‌های جستجو معطوف بوده‌اند. سوم، پژوهش‌هایی که بر طراحی هستی‌شناسی به صورت انسانی یا نیمه‌خودکار در حوزه‌های محدود علمی و تخصصی مانند پزشکی، جغرافیا، کشاورزی و

<sup>1</sup> ERIC

مسیریابی تمرکز داشته‌اند. علاوه بر این، شماری از تحقیقات نیز به بهینه‌سازی بازیابی اطلاعات مبتنی بر هستی‌شناسی و غنی‌سازی اصطلاحنامه‌ها برای ساخت هستی‌شناسی‌های تخصصی پرداخته‌اند.

بررسی محتوایی این پژوهش‌ها نشان می‌دهد که هرچند همگی در راستای بهره‌گیری از رویکردهای معنامحور برای سازماندهی و بازنمایی دانش حرکت کرده‌اند، اما دامنه داده‌ها و بستر کاربردی آن‌ها محدود و متفاوت از محیط‌های آرشیوی واقعی بوده است. به بیان دیگر، بیشتر این مطالعات بر روی متون عمومی یا تخصصی نسبتاً ساخت‌یافته مانند ویکی‌پدیا، جداول وب، بانک‌های اطلاعاتی علمی و متون فنی انجام شده‌اند؛ در حالی که محیط‌های آرشیوی، به‌ویژه در حوزه اسناد تاریخی، با چالش‌هایی از قبیل ناهمگنی منابع، تنوع قالب‌ها و ابهام معنایی موجودیت‌ها مواجه‌اند. افزون بر این، در اغلب پژوهش‌های پیشین، هدف اصلی ارائه یا ارزیابی روش‌های فنی برای استخراج یا مدل‌سازی هستی‌شناسی بوده و کمتر به کاربرد عملی این مدل‌ها در نظام‌های واقعی بازیابی اطلاعات و مدیریت دانش توجه شده است.

در چنین بستری، پژوهش حاضر جایگاهی متعادل و رویکردی کاربردی دارد. تفاوت بنیادین آن با مطالعات پیشین نه تنها در دامنه داده‌ها، بلکه در رویکرد ترکیبی، مسئله‌محور و عملیاتی آن نهفته است. در این پژوهش، طراحی هستی‌شناسی دامنه با بهره‌گیری از دو مسیر مکمل انجام می‌شود:

۱. تحلیل و استخراج خودکار و نیمه‌خودکار موجودیت‌ها و روابط از اسناد تاریخی آرشیوی؛

۲. بازاستفاده و یکپارچه‌سازی هستی‌شناسی‌ها و رده‌بندی‌های موجود در سطح وب معنایی و منابع ملی.

بدین ترتیب، این پژوهش در پی آن است که برای نخستین بار یک مدل هستی‌شناسی دامنه در حوزه تاریخ ایران را در بستر واقعی آرشیو دیجیتال طراحی و پیاده‌سازی کند. به‌طور کلی، این مطالعه ضمن بهره‌گیری از دستاوردهای نظری پیشین در حوزه وب معنایی،

هستی‌شناسی و گراف دانش، خلأ موجود در ادبیات پژوهش داخلی در زمینه به‌کارگیری فناوری‌های معنامحور برای سازماندهی و بازیابی معنایی اسناد تاریخی ایرانی را برطرف می‌سازد. دستاورد مورد انتظار آن، ایجاد پایه‌ای برای گراف دانش تاریخی ایران است که می‌تواند زمینه‌ساز توسعه سامانه‌های هوشمند پژوهشی، تصمیم‌یار و مدیریت دانش تاریخی کشور باشد.

## روش

برای انجام این پژوهش از روش آمیخته استفاده شده است. برای تهیه هستی‌شناسی دامنه در حوزه تاریخ معاصر ایران، لازم به شناخت موجودیت‌های اطلاعاتی اسناد، طراحی ساختار هستی‌شناسی، استخراج اطلاعات بر اساس مدل هستی‌شناسی و استخراج روابط بین موجودیت‌ها و اعتبارسنجی داده‌های استخراج شده و روابط بین آن‌هاست. برای شناخت موجودیت‌های اطلاعاتی از شیوه تحلیل متن استفاده شده است. انواع موجودیت‌های اطلاعاتی استخراج شده مقوله‌بندی شده و با نظرسنجی از خبرگان به روش تکنیک گروه اسمی<sup>۱</sup> (متخصصان حوزه تاریخ معاصر و مدیران و کارشناسان مرکز اسناد انقلاب اسلامی) اعتبارسنجی شده‌اند. در طراحی هستی‌شناسی دامنه از اصول بهروش‌های<sup>۲</sup> طراحی هستی‌شناسی دامنه بر پایه هستی‌شناسی شکلی<sup>۳</sup> پایه‌ارائه شده توسط رابرت آرپ و همکارانش در مرکز ملی هستی‌شناسی زیست پزشکی<sup>۴</sup> (Arp, Smith, & Spear, 2015) الگو گرفته شده است که شامل مراحل زیر است:

۱. تعیین حدود موضوع هستی‌شناسی.
۲. جمع‌آوری اطلاعات: شناسایی اصطلاحات کلی<sup>۵</sup> استفاده شده در هستی‌شناسی‌های موجود و متون جامعه هدف؛ تجزیه و تحلیل برای حذف افزونگی‌ها.
۳. مرتب‌سازی این اصطلاحات در یک سلسله‌مراتب از اصطلاحات عام‌تر و خاص‌تر.

<sup>۱</sup> Nominal Group Technique

<sup>۲</sup> Best Practices

<sup>۳</sup> Basic Formal Ontology (BFO)

<sup>۴</sup> National Center for Biomedical Ontology

<sup>۵</sup> Universals

۴. نظم‌دهی به نتیجه برای اطمینان از:

الف. انسجام منطقی، فلسفی و علمی،

ب. انسجام و سازگاری با هستی‌شناسی‌های مشابه،

ج. قابلیت فهم برای انسان، به‌ویژه از طریق تدوین تعاریف قابل خواندن برای انسان.

د. رسمی‌سازی دست‌سازه<sup>۱</sup> بازنمایی سازمان‌یافته در یک زبان قابل استفاده توسط کامپیوتر به‌گونه‌ای که نتیجه بتواند در یک چارچوب محاسباتی پیاده‌سازی شود.

بر اساس بهروش‌های<sup>۲</sup> ارائه شده در طراحی هستی‌شناسی دامنه که توسط آرپ و همکارانش (Arp et al., 2015) ارائه شده است، اولین گام در ساخت یک هستی‌شناسی دامنه، تعیین صریح دامنه موردنظر هستی‌شناسی است، یعنی پاسخ به این پرسش: «این هستی‌شناسی، هستی‌شناسی چه بخشی از واقعیت است؟» ارائه یک بیانیه صریح از این دامنه، نشان‌دهنده آن چیزی است که باید در هستی‌شناسی گنجانده شود و آنچه باید از آن حذف شود. بیانیه هستی‌شناسی مرکز اسناد انقلاب اسلامی ایوان به شرح زیر است:

هستی‌شناسی مرکز اسناد انقلاب ایران محدود به تاریخ معاصر ایران شامل دوره‌های تاریخی پهلوی دوم، انقلاب اسلامی و پس از آن، دوره جمهوری اسلامی تا به امروز است. این هستی‌شناسی در رده‌های مربوط به اشخاص سیاسی، احزاب و گروه‌ها، دوره‌های حکومت، ساختار حکومت‌ها، خانواده‌ها و گروه‌های قومی، تشکلات، سازمان‌ها و نهادهای مکان‌ها و زیرساخت‌ها و رویدادهای مرتبط با این دوران دارای سطح دانه‌بندی تا جزئی‌ترین رده‌هاست. دانه‌بندی در هستی‌شناسی به این معنی است که به عنوان مثال، زمانی که در مورد اشخاص صحبت می‌شود صرف شخص بودن، یا فقط سیاستمدار بودن یا هنرمند بودن کافی نیست و نوع مسئولیت سیاسی فرد به تناسب شغل وی از جمله رئیس‌جمهور، وزیر، سفیر، نماینده مجلس سنا و .. مانند آن در هستی‌شناسی مورد توجه قرار گرفته است.

در گام دوم طراحی هستی‌شناسی دامنه، به منظور جمع‌آوری اطلاعات از منابع اسنادی هدف و شناسایی اصطلاحات دامنه مورد بررسی، متون تحلیل شدند و استخراج اصطلاح‌های کاندید به دو روش نیمه خودکار و خودکار برای استخراج اصطلاح‌ها با توجه به محدوده

<sup>1</sup> Artefact

<sup>2</sup> Best practices

تعیین شده هستی‌شناسی انجام گرفت. جامعه پژوهش را تمامی منابع دیجیتال شده موجود در آرشیو دیجیتال مرکز اسناد انقلاب اسلامی تشکیل می‌دهد. جامعه اسنادی این مرکز شامل بیش از ۴/۵ میلیون برگ سند مکتوب، حدود ۳۱ هزار ساعت مصاحبه تاریخی شفاهی، نزدیک به ۳۷۲ هزار قطعه عکس، بیش از ۱۳ میلیون عنوان خبری و حدود ۹۰ هزار مقاله داخلی و خارجی است. بخش قابل توجهی از این منابع با استفاده از فناوری‌های تشخیص کاراکتر نوری و تبدیل گفتار به متن به فرمت دیجیتال تبدیل شده‌اند و از طریق سامانه آرشیو دیجیتال مرکز قابل دسترسی هستند.

با توجه به گستردگی جامعه پژوهش، نمونه مورد بررسی در این مطالعه بر اساس تحلیل محتوایی و با مشورت مدیران و کارشناسان مرکز، به مجموعه‌ای از منابع منتخب محدود شد که به‌طور مستقیم با محورهای زیر مرتبط بودند:

- منابعی که در مورد اشخاص از جمله رجال سیاسی، رجال مذهبی، افراد نظامی، هنرمندان، مشاهیر و شخصیت‌های نامی علمی و ادبی، خانواده‌ها و اعضای دربار پهلوی دوم گردآوری یا تدوین شده‌اند؛
- منابعی که درباره خانواده‌های معروف و خاندان سلطنتی تهیه شده‌اند،
- منابع مرتبط با گروه‌های قومی و اقلیت‌های مذهبی
- منابع مرتبط با رویدادها و گاه‌شمارها
- منابعی که درباره احزاب، تشکلات و سازمان‌ها هستند،
- منابع مرتبط با اماکن، بناها، زیرساخت‌ها و سازه‌ها

سپس، برای شناخت و استخراج موجودیت‌های اطلاعاتی، از دو ابزار اصلی پژوهش استفاده شد:

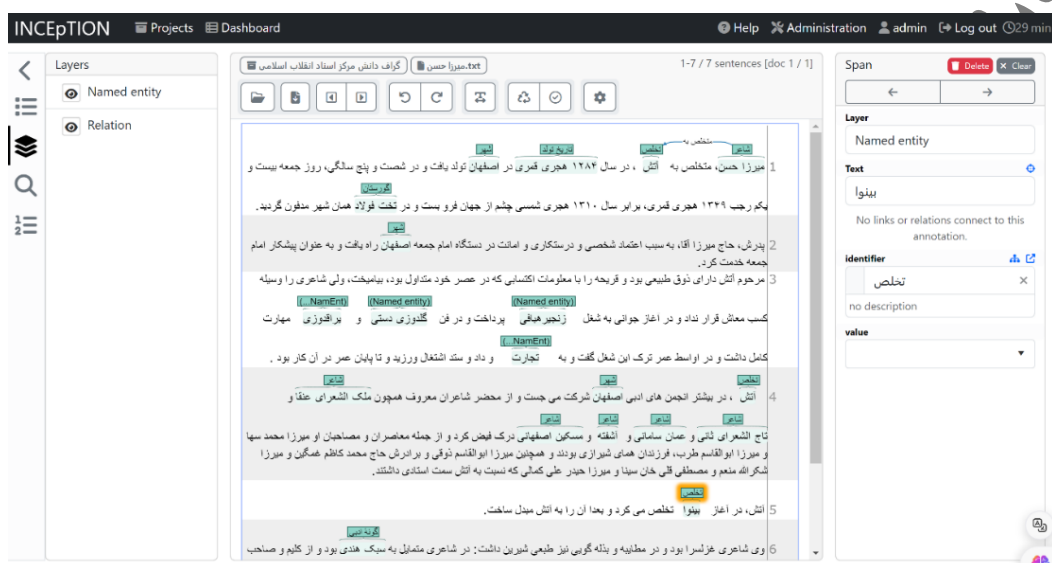
نخست، نرم‌افزار متن‌باز اینسپشن<sup>۱</sup> به عنوان ابزار برچسب‌گذاری به منظور برچسب‌گذاری دستی و نیمه خودکار متون هدف به کار گرفته شد (شکل ۱). این نرم‌افزار امکان تعریف انواع موجودیت‌ها، روابط و کنترل کیفی داده‌های زبانی را فراهم می‌سازد.

<sup>۱</sup> Inception: <https://inception-project.github.io/>



دوم، مدل زبانی بزرگ Groq نسخه ۳/۰ به عنوان ابزار پردازش زبان طبیعی برای استخراج خودکار موجودیت‌ها بر پایه روش تشخیص موجودیت‌های اسمی<sup>۱</sup> مورد استفاده قرار گرفت (شکل ۲).

شکل ۱. برچسب‌گذاری با استفاده از نرم‌افزار متن باز اینسپشن



<sup>1</sup> Named Entity Recognition

شکل ۲. برچسب‌گذاری متن با استفاده از مدل زبانی بزرگ گروک نسخه ۳,۰

- شخص: آبی
- مکان: سبز
- سازمان: بنفش
- رویداد: نارنجی
- مفهوم: خاکستری

[شاپور بختیار]، متولد ۱۲۹۳، ادعا می‌کرد پدرش، یکی از خوانین [بختیاری]، در سال ۱۳۱۳ توسط [رضاخان] اعدام شده است. تحصیلات ابتدایی را در [شهر کرد] گذراند و در سال ۱۳۰۵ به [اصفهان] رفت، جایی که با محافل [فراماسونری] آشنا شد. در سال ۱۳۰۹ به [بیروت] رفت و دیپلم ریاضی از مدرسه فرانسوی‌ها گرفت. در سال ۱۳۱۳ به [ایران] بازگشت و سپس برای ادامه تحصیل به [فرانسه] رفت. در سال ۱۳۱۸ با یک دختر فرانسوی ازدواج کرد. در [جنگ جهانی دوم]، در ارتش [فرانسه] خدمت کرد. در سال ۱۳۲۴ به [ایران] بازگشت و در سال ۱۳۲۸ مدیرکل اداره کار [خوزستان] شد. در سال ۱۳۲۹ به [کمیته فدائیان شاه] پیوست و از [محمد مصدق] حمایت کرد. او به [حزب ایران]، یکی از احزاب زیرمجموعه [جبهه ملی ایران]، پیوست و به یکی از رهبران آن تبدیل شد. [شاپور بختیار] به جاسوسی برای [ساواک]، [انگلیس]، [فرانسه]، و ارتباط دوستانه با [سفارت آمریکا] شهرت داشت. او در روزهای پایانی [رژیم پهلوی]، [نخست‌وزیری ۱۳۵۷] را پذیرفت تا رژیم را نجات دهد، اما با هوشیاری [امام خمینی] و مردم، پس از ۳۷ روز از [ایران] گریخت. در [فرانسه] با حمایت کشورهای غربی و عربی، به‌ویژه [عراق]، علیه [پیروزی انقلاب اسلامی] فعالیت کرد و در ۱۵ مرداد ۱۳۷۰ در [پاریس] کشته شد.

در استخراج اطلاعات به شکل خودکار، که از مدل‌های زبانی بزرگ استفاده شده است خطاها با نظارت انسانی اصلاح شدند. به عنوان مثال در شکل ۲ «بختیاری» موجودیتی از نوع «گروه قومی» یا «ایل» است که مدل به عنوان مکان شناسایی کرده است. علاوه بر تحلیل متون، به منظور شناسایی اصطلاح‌های کلی استفاده‌شده در هستی‌شناسی‌ها و رده‌بندی‌های موجود، از رویکردی ترکیبی شامل تحلیل محتوای دستی و به‌کارگیری ابزارهای نرم‌افزاری استفاده شد. در این مرحله، منابع زیر به‌صورت نظام‌مند مورد بررسی قرار گرفته و اصطلاح‌های مرتبط استخراج شدند. برای تحلیل و تطبیق مفاهیم از نرم‌افزار پروتژه<sup>۱</sup> و ابزارهای اسپارکیوال کوئری<sup>۲</sup> جهت مرور، و استخراج مفاهیم از فایل‌های آر.دی.اف / او.دبلیو.ال.<sup>۳</sup> بهره گرفته شد. منابع مورد استفاده عبارت‌اند از:

- هستی‌شناسی دی‌بی‌پدیا

<sup>۱</sup> Protégé

<sup>۲</sup> SPARQL Query

<sup>۳</sup> RDF/OWL

- هستی‌شناسی گراف دانش فارسی<sup>۱</sup>
- هستی‌شناسی بیبفریم<sup>۲</sup> (مدل کتابشناسی)
- رده‌بندی کتابخانه کنگره: DSR تاریخ ایران

در گام سوم اصطلاحات گردآوری شده با استفاده از رویکرد تحلیل محتوا و طبقه‌بندی مفهومی و بر پایه اصول طراحی هستی‌شناسه دامنه در یک ساختار سلسله مراتبی سازمان‌دهی شده و سپس هر اصطلاح به یک رده (کلی‌ها<sup>۳</sup>) مرتبط شد. این مرحله بخشی از فرآیند مدل‌سازی مفهومی در روش پژوهش ترکیبی حاضر محسوب می‌شود. برای قالب‌بندی اصطلاحات اصول زیر مورد استفاده قرار گرفتند (Arp et al., 2015):

۱. استفاده از اسامی مفرد
۲. استفاده از حروف کوچک ایتالیک برای اسامی عام.
۳. پرهیز از مخفف‌ها
۴. اختصاص شناسه الفبایی-عددی منحصر به فرد به هر اصطلاح.
۵. تضمین یک‌معنایی اصطلاحات.
۶. تضمین یک‌معنایی عبارت‌های رابطه‌ای
۷. پرهیز از اسامی غیر قابل شمارش<sup>۴</sup>

پس از قالب‌بندی، برای ایجاد ساختار سلسله مراتبی از تحلیل تطبیقی هستی‌شناسی‌ها و رده‌بندی‌های موجود بر اساس اصول مربوط به طبقه‌بندی‌ها در هستی‌شناسی استفاده شد (Arp et al., 2015):

۱. ساختاربندی هر هستی‌شناسی حول یک سلسله مراتب «هست\_یک»<sup>۵</sup> به عنوان ستون فقرات
۲. تضمین کامل بودن سلسله مراتب «هست\_یک»
۳. تضمین توارث منفرد (تک والدی) بیان شده

<sup>1</sup> <http://farsbase.net/>

<sup>2</sup> Bibframe

<sup>3</sup> Universals

<sup>4</sup> Mass nouns

<sup>5</sup> is\_a

۴. رعایت فرض دنیای باز<sup>۱</sup>

۵. پایبندی به قاعده عینیت.

بر اساس اصول مربوط به اصطلاح‌شناسی، تعاریف و طبقه‌بندی‌ها ساختار اولیه سلسله‌مراتبی هستی‌شناسی دامنه بر اساس مراحل زیر شکل گرفت:

۱- برای نظم‌دهی به رده‌ها از عام به خاص با توجه به محدوده تعیین شده هستی‌شناسی، ساختار سلسله‌مراتبی هستی‌شناسی دی‌بی‌پدیا به عنوان مبنا قرار گرفت و از سایر رده‌بندی‌ها و هستی‌شناسی‌ها رده‌های مرتبط در جایگاه پیشنهادی اولیه قرار گرفتند. این پیشنهاد بر اساس مقایسه بین ساختار سلسله‌مراتب سایر هستی‌شناسی‌ها و رده‌بندی‌ها و همچنین پیشنهادهای مدل‌های زبانی بزرگ مانند گروک<sup>۲</sup>، دیپ‌سیک<sup>۳</sup> و کیوون<sup>۴</sup> انجام گرفته است. سپس نمونه‌های استخراج شده از تحلیل متون بر اساس شباهت‌ها و ویژگی‌های مشترک خوشه‌بندی شده و به رده مرتبط نگاشت شدند.

۲- در مورد موجودیت‌هایی که دارای رده‌ای دی‌بی‌پدیا یا رده‌بندی تاریخ ایران نبودند رده‌ای با مشورت با خبرگان و به کمک مدل‌های زبانی بزرگ پیشنهاد شد. برای مثال برای موجودیت «دادگاه ویژه روحانیت» و مانند آن، رده «نهاد انقلاب اسلامی» ذیل رده «سازمان» ایجاد شد.

و در نهایت به منظور اعتبارسنجی رده‌های استخراج شده از روش گروه اسمی<sup>۵</sup> استفاده شد. در این گروه ۱۱ نفر از مدیران و خبرگان (شامل اعضای هیأت علمی) مرکز اسناد انقلاب اسلامی شرکت کردند که بر اساس معیارهایی چون سابقه فعالیت پژوهشی و مدیریتی در حوزه اسناد تاریخی، تسلط بر نظام‌های رده‌بندی و سازماندهی، تسلط به مفاهیم تاریخ معاصر ایران و تجربه در تحلیل محتوای اسناد آرشیوی انتخاب شدند.

نشست این گروه در دو جلسه سه‌ساعته انجام شد. با توجه به تعداد بالای رده‌ها، این روش به عنوان جایگزینی سریع و ساختارمند برای روش دلفی به کار گرفته شد؛ روشی که به جای تکرارهای زمان‌بر دلفی، در یک یا دو جلسه ۲ تا ۳ ساعته با حضور ۸ تا ۱۲ نفر از خبرگان و

<sup>1</sup> open-world assumption

<sup>2</sup> GROK

<sup>3</sup> Deepseek

<sup>4</sup> Qwen

<sup>5</sup> Nominal Group Technique (NGT)

تمرکز بر اجماع و اولویت‌بندی نظرات انجام می‌شود. معیارهای اصلی اعتبارسنجی شامل (۱) منطق سلسله‌مراتبی، (۲) صحت تاریخی/فرهنگی، (۳) همترازی و معادل‌سازی مفاهیم، و (۴) کامل بودن و سادگی ساختار در نظر گرفته شدند.

#### یافته‌ها

**سؤال اول پژوهش:** موجودیت‌ها و رده‌های حوزه دامنه موضوعی اسناد آرشیو دیجیتال مرکز اسناد انقلاب اسلامی کدام‌اند و چگونه می‌توان آنها را در قالب هستی‌شناسی دامنه بازنمایی کرد؟

**پاسخ:** بر اساس تحلیل متون آرشیو دیجیتال مرکز اسناد انقلاب اسلامی و رده‌های هستی‌شناسی دی‌بی‌دیا، رده‌های هستی‌شناسی بیفریم و رده‌بندی تاریخ ایران بر اساس کتابخانه کنگره آمریکا، یافته‌های این بخش شامل ۵۳۵ اصطلاح بر اساس محدوده موضوعی تعیین شده هستی‌شناسی بود که بر اساس رده‌های اصلی زیر تنظیم شدند (جدول ۱):

جدول ۱. تعداد رده‌های اصلی و زیررده‌ها هستی‌شناسی دامنه مرکز اسناد انقلاب اسلامی ایران

| رده | زیر رده      | زیر رده | تعداد زیر رده |
|-----|--------------|---------|---------------|
| چیز | عامل         |         | ۵۳۴           |
|     |              |         | ۲۵۳           |
|     |              | شخص     | ۱۷۱           |
|     |              | سازمان  | ۷۵            |
|     |              | خانواده | ۶             |
|     | رویداد       |         | ۳۷            |
|     | رسانه        |         |               |
|     | اثر          |         | ۱۸            |
|     | دوره حکومت   |         | ۵۲            |
|     | مفهوم موضوعی |         | ۱             |
|     | مکان         |         | ۳۷            |
|     | سازه معماری  |         | ۶۸            |
|     | گروه قومی    |         | ۶۷            |
|     |              |         |               |

**سؤال دوم پژوهش:** ساختار سلسله‌مراتبی و روابط میان رده‌های هستی‌شناسی دامنه چگونه باید باشد؟

**پاسخ:** پیش‌نویس ساختار سلسله‌مراتبی اصطلاح‌های استخراج شده در یافته‌های بخش اول، بر پایه رده‌های اصلی و ساختار هستی‌شناسی دی‌بی‌پدیا شکل گرفت. پس از اعتبارسنجی توسط گروه خبرگان و اجماع اکثریت، در ساختار رده‌بندی ۴۵۶ رده تایید شد، ۷۱ رده اضافه شد، جایگاه ۵۳ رده در سلسله‌مراتب تغییر کرد، ۱۹ رده حذف شد، اصطلاح ۵ رده تغییر کرد و ۲ رده هم در جایگاه سلسله‌مراتب تغییر کرده و هم اصطلاحشان عوض شد (جدول ۲).

جدول ۲. نتیجه اعتبارسنجی رده‌های هستی‌شناسی دامنه مرکز اسناد انقلاب اسلامی

| نتیجه اعتبارسنجی                    | تعداد |
|-------------------------------------|-------|
| تایید رده                           | ۴۵۷   |
| اضافه کردن رده                      | ۷۰    |
| تغییر رده در سلسله‌مراتب            | ۵۳    |
| حذف رده                             | ۱۹    |
| تغییر اصطلاح                        | ۵     |
| تغییر اصطلاح / تغییر در سلسله‌مراتب | ۲     |

عمده رده‌های افزوده شده مربوط به رده‌های اصلی عامل ← شخص و مکان هستند. برخی از رده‌ها نیز به عنوان مفهوم موضوعی در نظر گرفته شدند. خبرگان مرکز اسناد افزودن مفاهیمی همچون جریان‌ها مانند جریان‌های سیاسی (نوگرایی، روشنفکری و...)، اقتصادی، اجتماعی و همچنین مکتب‌های فکری و ایدئولوژی‌ها، فعالیت‌های سیاسی و فعالان سیاسی را ضروری دانستند که ذیل مفهوم موضوعی پوشش داده می‌شود. از آنجا که موضوعات و رده‌بندی علم بسیار وسیع است بنابراین در این هستی‌شناسی یک شاخه اصلی به عنوان مفهوم موضوعی قرار داده شده است تا مفاهیم موضوعی بر اساس نظام‌های موضوعی مختلف ذیل آن قرار گیرند که در پوشش این مطالعه نمی‌گنجد. سازه معماری نیز به این بخش منتقل شد و رده‌های ذیل آن به عنوان تقسیم‌بندی‌های مکانی در نظر گرفته شدند. جدول شماره ۳ نتیجه اجماع خبرگان این حوزه برای افزودن رده‌های ضروری به هستی‌شناسی است.

جدول ۳. فهرست رده‌های افزوده شده با تعیین جایگاه در سلسله مراتب

| رده‌های افزوده شده                                                                        | مکان در سلسله مراتب                                 |
|-------------------------------------------------------------------------------------------|-----------------------------------------------------|
| ولیعهد، رئیس شورای شهر، عضو شورای شهر                                                     | چیز ← عامل ← شخص ← سیاستمدار                        |
| عضو حوزه علمیه (استاد یا مدرس)، فرد شبه نظامی، عارف                                       | چیز ← عامل ← شخص                                    |
| مداح، تعزیه‌خوان، قاری قرآن، موزن                                                         | چیز ← عامل ← شخص ← هنرمند ← خواننده ← خواننده آئینی |
| خواننده سنتی، خواننده پاپ، خواننده فولکلور (محلی)، خواننده رپ                             | چیز ← عامل ← شخص ← هنرمند ← خواننده                 |
| راهب بودایی                                                                               | چیز ← عامل ← شخص ← پیشوای مذهبی                     |
| فرد نظامی نیروی زمینی، فرد نظامی نیروی هوایی، فرد نظامی نیروی دریایی، مأمور نیروی انتظامی | چیز ← عامل ← شخص ← فرد نظامی (نیروی مسلح)           |
| مقام امنیتی، کارشناس امنیتی                                                               | چیز ← عامل ← شخص ← فرد امنیتی                       |
| شرکت                                                                                      | چیز ← عامل ← سازمان                                 |
| شرکت دولتی، شرکت عمومی غیر دولتی                                                          | چیز ← عامل ← سازمان ← شرکت                          |
| نافرمانی مدنی                                                                             | چیز ← رویداد ← رویداد اجتماعی                       |
| منطقه                                                                                     | چیز ← مکان ← کشور ← سکونتگاه (مکان مسکونی)          |
| جلگه، خور، آبشار، رود فصلی (مسیل)، تالاب، باتلاق، بیابان، کلت، دشت، چشمه، قله کوه         | چیز ← مکان ← کشور ← مکان طبیعی                      |
| مکان انسان ساخته                                                                          | چیز ← مکان ← کشور                                   |
| سالن موسیقی (کنسرت)، کاباره                                                               | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان هنری    |

| رده‌های افزوده شده                                                                                                                                                                | مکان در سلسله مراتب                                              |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| مکان مذهبی، مکان تاریخی، مکان نظامی، مکان تجاری / اقتصادی، مکان تفریحی / گردشگری، مکان صنعتی، سکوی نفتی، سکوی گازی، مکان کشاورزی (زراعی)، مکان فرهنگی و هنری، مکان خدماتی / عمومی | چیز ← مکان ← کشور ← مکان<br>انسان ساخته                          |
| عبادتگاه (معبد)، زیارتگاه، حسینیه (تکیه)                                                                                                                                          | چیز ← مکان ← کشور ← مکان<br>انسان ساخته ← مکان مذهبی             |
| امامزاده                                                                                                                                                                          | چیز ← مکان ← کشور ← مکان<br>انسان ساخته ← مکان مذهبی<br>زیارتگاه |
| بیمارستان، زایشگاه، تیمارستان                                                                                                                                                     | چیز ← مکان ← کشور ← مکان<br>انسان ساخته ← مکان خدماتی / عمومی    |
| بازار، فروشگاه / مغازه                                                                                                                                                            | چیز ← مکان ← کشور ← مکان<br>انسان ساخته ← مکان تجاری / اقتصادی   |
| مزرعه، شالیزار، نخلستان                                                                                                                                                           | چیز ← مکان ← کشور ← مکان<br>انسان ساخته ← مکان کشاورزی (زراعی)   |
| آرامگاه (بقعه، مرقد، مدفن، مزار یا مقبره)                                                                                                                                         | چیز ← مکان ← کشور ← مکان<br>انسان ساخته ← آرامستان (قبرستان)     |
| ایستگاه درون شهری، ایستگاه برون شهری                                                                                                                                              | چیز ← مکان ← کشور ← مکان<br>انسان ساخته ← ایستگاه                |
| پایانه مسافربری، قنات (کاهریز)، چاه آب، کانال مصنوعی                                                                                                                              | چیز ← مکان ← کشور ← مکان<br>انسان ساخته                          |

علاوه بر رده‌های افزوده شده، برخی رده‌ها در ساختار رده‌بندی جایجا شدند و در جایگاه جدید قرار گرفتند. جدول ۴ فهرست این رده‌ها و جایگاه جدید آنها در سلسله مراتب هستی‌شناسی را نشان می‌دهد.



جدول ۴. فهرست رده‌های جایجا شده در سلسله‌مراتب رده‌بندی

| رده تغییر یافته / مسیر سلسله مراتب قبلی                                                   | مسیر سلسله‌مراتب (محل قرارگیری جدید)                                                            |
|-------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| شهردار / چیز ← عامل ← شخص                                                                 | چیز ← عامل ← شخص ← سیاستمدار ← زیررده (سیاستمدار)                                               |
| رهبر ارکستر / چیز ← عامل ← شخص                                                            | چیز ← عامل ← شخص ← هنرمند ← زیررده (هنرمند)                                                     |
| امام جماعت / چیز ← عامل ← شخص ← پیشوای مذهبی                                              | چیز ← عامل ← شخص ← پیشوای مذهبی ← روحانی ← روحانی مسلمان ← زیررده (روحانی مسلمان)               |
| مقام ساواک، مقام واجار / چیز ← عامل ← شخص ← مأمور امنیتی                                  | چیز ← عامل ← شخص ← فرد امنیتی ← زیررده (مقام امنیتی)                                            |
| بازجو / چیز ← عامل ← شخص ← مأمور امنیتی                                                   | چیز ← عامل ← شخص ← فرد امنیتی ← زیررده (کارشناس امنیتی)                                         |
| روزنامه‌نگار / چیز ← عامل ← شخص                                                           | چیز ← عامل ← شخص ← زیررده (خبرنگار)                                                             |
| شرکت خصوصی / چیز ← عامل ← سازمان                                                          | چیز ← عامل ← سازمان ← شرکت ← زیررده (شرکت)                                                      |
| ائتلاف / چیز ← عامل ← سازمان                                                              | چیز ← عامل ← سازمان ← گروه ← زیررده (گروه)                                                      |
| میدان، خیابان، شهرک، برج / چیز ← سازه معماری ← زیرساخت                                    | چیز ← مکان ← کشور ← سکونتگاه (مکان مسکونی) ← منطقه ← استان ← شهرستان ← بخش ← شهر ← زیررده (شهر) |
| دهستان / چیز ← مکان ← کشور ← سکونتگاه (مکان مسکونی) ← منطقه ← استان ← شهرستان ← بخش ← شهر | چیز ← مکان ← کشور ← سکونتگاه (مکان مسکونی) ← منطقه ← استان ← شهرستان ← بخش ← زیررده (بخش)       |
| چشمه آب گرم / چیز ← مکان ← کشور ← مکان طبیعی                                              | چیز ← مکان ← کشور ← مکان طبیعی ← چشمه ← زیررده (چشمه)                                           |
| سالن سینما، سالن تئاتر / چیز ← سازه معماری ← ساختمان ← محل برگزاری                        | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان فرهنگی و هنری ← زیررده (مکان فرهنگی و هنری)         |

| رده تغییر یافته / مسیر سلسله مراتب قبلی                                                                        | مسیر سلسله مراتب (محل قرارگیری جدید)                                                          |
|----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| موزه، کتابخانه/ چیز ← سازه معماری ← ساختمان                                                                    | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان فرهنگی و هنری ← زیررده (مکان فرهنگی و هنری)       |
| کلیسا، صومعه، مسجد، آتشکده، کنیسه/ چیز ← سازه معماری ← ساختمان ← ساختمان مذهبی                                 | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان مذهبی ← عبادتگاه (معبد) ← زیررده (عبادتگاه)       |
| زیارتگاه/ چیز ← سازه معماری ← ساختمان ← ساختمان مذهبی                                                          | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان مذهبی ← زیررده (مکان مذهبی)                       |
| قلعه/ چیز ← سازه معماری ← زیرساخت ← سازه نظامی                                                                 | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان تاریخی+ ← زیررده (مکان تاریخی)                    |
| ورزشگاه/ چیز ← سازه معماری ← زیرساخت                                                                           | چیز ← مکان ← کشور ← مکان انسان ساخته ← زیررده (مکان انسان ساخته)                              |
| زندان/ چیز ← سازه معماری ← ساختمان                                                                             | چیز ← مکان ← کشور ← مکان انسان ساخته ← زیررده (مکان انسان ساخته)                              |
| معدن/ چیز ← مکان ← کشور ← مکان طبیعی                                                                           | چیز ← مکان ← کشور ← مکان انسان ساخته ← زیررده (مکان انسان ساخته)                              |
| آرامستان (قبرستان)/ چیز ← مکان ← کشور                                                                          | چیز ← مکان ← کشور ← مکان انسان ساخته ← زیررده (مکان انسان ساخته)                              |
| فرودگاه، سد، دیوار دریایی، سکوی پرتاب، بندر، پست برق، آسیاب، بنای یادبود، ایستگاه/ چیز ← سازه معماری ← زیرساخت | چیز ← مکان ← کشور ← مکان انسان ساخته ← زیررده (مکان انسان ساخته)                              |
| رستوران، مرکز خرید (مال)/ چیز ← سازه معماری ← ساختمان                                                          | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان تجاری / اقتصادی ← زیررده (مکان تجاری / اقتصادی)   |
| کازینو/ چیز ← سازه معماری ← ساختمان                                                                            | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان تفریحی / گردشگری ← زیررده (مکان تفریحی / گردشگری) |

| رده تغییر یافته / مسیر سلسله مراتب قبلی                                            | مسیر سلسله مراتب (محل قرارگیری جدید)                                                            |
|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| پارک (بوستان) / چیز ← مکان ← کشور                                                  | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان تفریحی / گردشگری ← زیررده (مکان تفریحی / گردشگری)   |
| محل / چیز ← سازه معماری ← ساختمان                                                  | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان تفریحی / گردشگری ← زیررده (مکان تفریحی / گردشگری)   |
| باغ و محوطه / چیز ← سازه معماری ← زیرساخت                                          | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان تفریحی / گردشگری ← زیررده (مکان تفریحی / گردشگری)   |
| کارخانه / چیز ← سازه معماری ← ساختمان                                              | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان صنعتی ← زیررده (مکان صنعتی)                         |
| نیروگاه / چیز ← سازه معماری ← زیرساخت                                              | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان صنعتی ← زیررده (مکان صنعتی)                         |
| باغ / چیز ← مکان ← کشور                                                            | چیز ← مکان ← کشور ← مکان انسان ساخته ← مکان کشاورزی (زراعی) ← زیررده (مکان کشاورزی (زراعی))     |
| ایستگاه مترو، ایستگاه تراموا، ایستگاه مسیر / چیز ← سازه معماری ← زیرساخت ← ایستگاه | چیز ← مکان ← کشور ← مکان انسان ساخته ← ایستگاه ← ایستگاه درون شهری ← زیررده (ایستگاه درون شهری) |
| ایستگاه راه آهن / چیز ← سازه معماری ← زیرساخت ← ایستگاه                            | چیز ← مکان ← کشور ← مکان انسان ساخته ← ایستگاه ← ایستگاه برون شهری ← زیررده (ایستگاه برون شهری) |
| اقلیت مذهبی / چیز ← گروه قومی                                                      | چیز ← زیررده (چیز)                                                                              |

طی فرآیند اعتبارسنجی، برخی رده‌ها حذف شده و اصطلاح برخی رده‌ها نیز تغییر کرد. در بخش رده‌های اشخاص، «دانشجو، مامور ساواک، افسر پلیس و مامور واجا» حذف شدند. همچنین اصطلاح «شطرنج‌باز» به جای «بازیکن شطرنج» اصلاح شد تا اصطلاح دقیق‌تر و متعارفی داشته باشد.

در حوزه سازمان‌ها و نهادها، اصلاحات اصطلاحی قابل توجهی انجام شد: «موسسه آموزشی» به «نهاد آموزشی» تغییر یافت و به زیررده دانشگاه منتقل شد، «سازمان قانونگذار» به «نهاد قانونگذار»، «سازمان مشورتی» به «نهاد مشورتی» و «موسسه رسانه‌ای» به «سازمان رسانه‌ای» اصلاح شد. برخی رده‌ها نیز حذف شدند، از جمله «گروهک سیاسی و فرقه ضاله» حذف شده و «رسانه» به دلیل همپوشانی با «سازمان رسانه‌ای» نیز از هستی‌شناسی خارج شد.

در بخش مکان‌ها و سازه‌ها، سازه معماری که شامل دو زیررده «ساختمان و زیرساخت» بود حذف شدند و رده‌های مرتبط با آنها شامل «ساختمان مذهبی، معبد، ساختمان، تاریخ، آسمان‌خراش، محل برگزاری، سازه نظامی و قلعه نظامی» حذف شدند و معادل آن‌ها به عنوان مکان ذیل رده مکان اضافه شد. همچنین برخی ویرایش‌های اصطلاحی انجام شد، مانند تغییر «مسیر حمل و نقل» به «مسیر انتقال» و «موتور بادی» به «توربین بادی».

به طور کلی، این بازنگری و اعتبارسنجی باعث افزودن رده‌های لازم، پاکسازی رده‌های غیرضروری، اصلاح اصطلاحات غیرمتعارف و بهبود جایگاه سلسله‌مراتب رده‌ها شد و همزمان انسجام، منطق و دقت هستی‌شناسی را در سطوح مختلف افزایش داد.

**سؤال سوم پژوهش:** چه هستی‌شناسی‌ها و رده‌بندی‌هایی برای بازاستفاده در تهیه هستی‌شناسی دامنه مرکز اسناد انقلاب اسلامی مطلوب هستند؟

**پاسخ:** بر اساس جستجوی موضوعی در داخل کشور هستی‌شناسی‌ای برای تاریخ معاصر ایران یا مرتبط با آن یافت نشد. ولی گسترش رده‌بندی تاریخ ایران بر اساس نظام کتابخانه کنگره آمریکا و رده‌بندی تاریخ ایران بر اساس نظام دهمی دیوئی مورد بررسی تطبیقی قرار گرفتند. با توجه به اینکه گسترش رده‌بندی کنگره کلیه رده‌های دیوئی را پوشش داده است، لذا کنار گذاشته شد. علاوه بر این هستی‌شناسی‌های عمومی داخلی و خارجی شامل هستی‌شناسی‌های گراف دانش فارسی، دی‌بی‌پدیا، ویکی‌دیتا<sup>۱</sup> و یا گو<sup>۲</sup> که حوزه‌هایی از دامنه مورد مطالعه را پوشش داده بودند بررسی شدند. بر اساس مطالعه عسگری و همکاران (Asgari-Bidhendi et al., 2019)، دی‌بی‌پدیا به عنوان هستی‌شناسی هسته در وب معنایی شناخته می‌شود که سایر هستی‌شناسی‌های عمومی را پوشش می‌دهد. هستی‌شناسی

<sup>1</sup> Wikidata.com

<sup>2</sup> <https://yago-knowledge.org>

گراف دانش فارسی نیز تنها در ۱۰ رده با هستی‌شناسی دی‌بی‌پدیا تفاوت دارد که قابل چشم‌پوشی بود. در حوزه رده‌های مربوط به «اثر» و زیررده‌های آن نیز از هستی‌شناسی بیبفریم به عنوان مرجع قابل استناد در این حوزه استفاده شد.

### بحث و نتیجه‌گیری

آنچه در این مطالعه بدست آمد، یک رده‌بندی هستی‌شناسی مبنا برای حوزه تاریخ معاصر ایران بر اساس اسناد و منابع موجود در آرشیو دیجیتال مرکز اسناد انقلاب اسلامی است. این رده‌بندی بر اساس چارچوب هستی‌شناسی دامنه و رعایت اصول آن استخراج شده و قابل توسعه است. بر اساس این رده‌بندی، موجودیت‌های قابل استخراج از منابع آرشیو دیجیتال مرکز اسناد انقلاب اسلامی در قالب رده‌های اصلی شخص، سازمان، خانواده، دوره حکومت، رویداد، مکان، اثر، اقلیت مذهبی، گروه قومی و زیررده‌های آنها (۵۳۵ اصطلاح) مشخص شدند. این نتیجه با پژوهش‌های امامی و مرادی، وزیرنژاد و بحرانی هم‌راستا است که هر دو استخراج خودکار یا نیمه‌خودکار روابط و مفاهیم از متون فارسی را آزموده‌اند. با این حال، تمایز مهم پژوهش حاضر در این است که داده‌های آن از متون واقعی آرشیوی و تاریخی استخراج شده است، در حالی که پژوهش‌های یادشده عمدتاً بر داده‌های ساخت یافته یا متون عمومی تکیه داشته‌اند. به علاوه، گوئیرز و همکاران و ویمالاسوریا و دوو نیز بر استخراج ترکیبی موجودیت‌ها با استفاده از مدل‌های هستی‌شناختی تأکید کرده‌اند که با رویکرد ترکیبی این تحقیق هم‌خوانی دارد. با این حال، پژوهش حاضر به جای تمرکز بر الگوریتم‌های استخراج، بر ساختار مفهومی بومی و تناسب آن با دامنه تاریخی ایران تمرکز کرده است.

نتیجه اعتبارسنجی خبرگان منجر به اصلاح و بازنگری در ساختار نهایی هستی‌شناسی شامل تایید ۴۵۷ رده، افزودن ۷۰ رده، حذف ۱۹ رده، تغییر ۵ اصطلاح، جابه‌جایی ۵۳ رده در سلسله‌مراتب و تغییر اصطلاح و جابه‌جایی همزمان ۲ رده شد. این فرایند نشان داد که ساختارهای عمومی مانند دی‌بی‌پدیا یا گراف دانش فارسی برای پوشش دقیق مفاهیم بومی تاریخ ایران کفایت ندارند و نیاز به بومی‌سازی گسترده دارند. این یافته با نتایج مینایی و سجادی و عسگری و همکاران هم‌راستا است؛ زیرا آن‌ها نیز در توسعه گراف دانش فارسی بر ضرورت اصلاح و تکمیل ساختارهای عمومی تأکید کرده‌اند. در عین حال، با یافته‌های

تان و همکاران که گراف دانش حوزه ترافیک را بر مبنای داده‌های ساختاریافته بنا کردند، تفاوت دارد؛ چراکه در پژوهش حاضر داده‌ها از اسناد تاریخی نیمه‌ساخت‌یافته استخراج شدند و ماهیت پویاتر و نامتجانس‌تری داشتند. افزوده شدن رده‌های مرتبط با جریان‌های فکری، سیاسی و مذهبی نیز بیانگر لزوم درک مفهومی و تاریخی داده‌ها است، موضوعی که در پژوهش‌های پیشین کمتر به آن توجه شده است.

مقایسه این پژوهش با پژوهش‌های مرتبط در حوزه طراحی هستی‌شناسی انسانی و کاربردی نشان می‌دهد که نتایج آن با یافته‌های ترابی و همکاران، نیارکی و همکاران و تقی‌زاده نائینی و همکاران مشابهت دارد، از آن جهت که همه آن‌ها طراحی هستی‌شناسی را مبتنی بر تحلیل دامنه و مشارکت خبرگان دانسته‌اند. اما تمایز پژوهش حاضر در آن است که این طراحی صرفاً مفهومی نبوده، بلکه با بازخورد عملی خبرگان آرشویی و در بستر واقعی مرکز اسناد انجام گرفته است؛ در نتیجه، خروجی آن قابلیت به کارگیری مستقیم در نظام آرشو دیجیتال را دارد.

یافته‌های این پژوهش در زمینه غنی‌سازی ساختار مفهومی و معنایی داده‌های تاریخی، با نتایج جعفری پاورسی و همکاران و و هم‌راستا است که بر نقش هستی‌شناسی در بهبود بازیابی معنایی اطلاعات تأکید دارند. با این حال، تفاوت اصلی در نوع داده و کاربرد نهایی است؛ پژوهش حاضر بر اسناد تاریخی و روابط مفهومی میان شخصیت‌ها، رویدادها و مکان‌ها متمرکز بوده، در حالی که تحقیقات پیشین بیشتر در حوزه‌های پزشکی، آموزشی یا چندرسانه‌ای انجام شده‌اند.

نتایج پژوهش حاضر ضمن تأیید کارآمدی رویکرد ترکیبی در طراحی هستی‌شناسی، بر اهمیت بومی‌سازی مدل‌های مفهومی در آرشیوهای تاریخی تأکید دارد. هم‌راستایی نتایج با مطالعات پیشین در حوزه‌های استخراج خودکار داده‌ها، توسعه گراف دانش، پاسخگویی معنایی به پرس‌وجوها و طراحی سیستم‌های مبتنی بر هستی‌شناسی نشان می‌دهد که استفاده از گراف دانش می‌تواند بستر مؤثری برای سازماندهی معنایی، بازیابی هوشمند و تحلیل محتوای تاریخی فراهم کند. با این حال، یافته‌های بومی این پژوهش شکاف میان الگوهای

جهانی و نیازهای محتوایی آرشیوهای ایرانی را برجسته می‌سازد و راه را برای توسعه مدل‌های دانش‌محور بومی هموار می‌کند.

برای ادامه زنجیره این پژوهش، پیشنهاد می‌شود حوزه‌های زیر برای مطالعات آتی مورد توجه قرار گیرند:

۱. توسعه هستی‌شناسی‌ها دامنه در حوزه‌های موضوعی دیگر با رعایت اصول

تعامل‌پذیری میان هستی‌شناسی‌ها

۲. پیاده‌سازی گراف دانش عملیاتی و ارزیابی کارایی آن در بازیابی اطلاعات

۳. توسعه روش‌های خودکارتر برای استخراج مفاهیم و روابط با استفاده از

مدل‌های زبانی بزرگ

۴. مطالعه الگوریتم‌های وزن‌دهی روابط در گراف دانش

تعارض منافع

تعارض منافع ندارم.

سپاسگزاری

مراتب و قدردانی خود را از رئیس مرکز اسناد انقلاب اسلامی و معاون مرکز اسناد که با حمایت از این پژوهش و فراهم کردن داده‌ها و تأمین زیرساخت‌های مورد نیاز امکان تحقق آن را فراهم کردند به عمل می‌آوریم. همچنین از گروه خبرگان که وقت و دانش ارزشمند خود را در اختیار این مطالعه قرار دادند سپاسگزاریم.

## منابع

- امامی، حجت. (۱۴۰۱). ارائه روشی برای استخراج اطلاعات ساختاریافته محدود به دامنه از صفحات وب فارسی. فصلنامه پردازش علائم و داده ها، ۵۲، ۱۴۶-۱۳۳.  
<http://jsdp.rcisp.ac.ir/article-1-1102-fa.html>
- ترابی، لاله، میرحسینی، زهره، ابادری، زهرا و حسینی بهشتی، ملوک السادات (۱۴۰۱). طراحی و پیاده سازی هستان‌شناسی پزشکی هسته‌ای. مدیریت اطلاعات سلامت، ۱۸(۲)، ۸۱-۸۸.  
<https://doi.org/10.22122/him.v18i2.4357>
- تقی زاده نایینی، جواد، فهیم نیا، فاطمه و نقشینه، نادر (۱۳۹۸). استخراج آنتولوژی هویت دیجیتال مبتنی بر تحلیل حوزه. پژوهشنامه پردازش و مدیریت اطلاعات، ۳۴(۴)، ۱۶۶۹-۱۷۰۰.  
<https://doi.org/10.35050/JIPM010.2019.018>
- ثروتی، لیلا، ولوی، محمدرضا، و حورعلی، مریم (۱۳۹۶). توسعه یک هستان‌شناسی جغرافیایی برای استفاده در کاربردهای نظامی. فصلنامه اطلاعات جغرافیایی (سپهر)، ۲۶(۱۰۲)، ۵-۱۷.  
<https://www.magiran.com/paper/1738409>
- جعفری پاورسی، حمیده، حریری، نجان، علیپورحافظی، مهدی، باب الحوائجی، فهیمه و خادمی، مریم. (۱۳۹۹). ارتقای بازیابی معنایی اطلاعات با استفاده از برچسب گذاری و هستان‌شناسی. نشریه مطالعات کتابداری و سازماندهی اطلاعات، ۳۱(۱)، ۱۸-۳۸.  
<https://www.magiran.com/paper/2127226>
- حسینی بهشتی، ملوک السادات، و اژه‌ای، فاطمه. (۱۳۹۴). طراحی و پیاده‌سازی هستی‌شناسی علوم پایه بر اساس مفاهیم و روابط موجود در اصطلاحنامه‌های مرتبط. پژوهشنامه پردازش و مدیریت اطلاعات، ۶۷۷-۶۹۶.  
[https://jipm.irandoc.ac.ir/article\\_699309.html](https://jipm.irandoc.ac.ir/article_699309.html)
- سجادی، سید محمدباقر، و مینایی بیدگلی، بهروز. (۱۳۹۸). معماری سامانه گراف دانش زبان فارسی. پژوهشنامه پردازش و مدیریت اطلاعات، ۳۵(۲)، ۴۲۵-۴۶۲.  
<https://doi.org/10.35050/JIPM010.2020.057>
- صادقی نیارکی، ابوالقاسم، برزگر، مریم، و شاکری، مریم. (۱۳۹۷). طراحی و توسعه آنتولوژی تجربیات مکانی در مسیریابی شهری. مجله علمی رایانش نرم و فناوری اطلاعات، ۷(۲)، ۱-۱۵.  
[https://jscit.nit.ac.ir/article\\_80956.html](https://jscit.nit.ac.ir/article_80956.html)
- عالیشان کرمی، نادر، حاجی زین العابدینی، محسن، رداد، ایرج، و قاضی میرسعید، سید جواد. (۱۳۹۶). کاربرد و نقش هستان‌شناسی در نظام‌های بازیابی اطلاعات زیست پزشکی. مجله انفورماتیک سلامت و زیست پزشکی، ۴(۴)، ۳۲۷-۳۴۰.  
<http://jhbmj.ir/article-1-222-fa.html>
- مرادی، مهدی، وزیرنژاد، بهرام، و بحرانی، محمد. (۱۳۹۴). ساخت هستان‌شناسی دانش عرفی زبان فارسی با رویکردی تلفیقی. پژوهشنامه پردازش و مدیریت اطلاعات، ۳۱(۱)، ۱۰۹-۱۲۴.  
[https://jipm.irandoc.ac.ir/article\\_699334.html](https://jipm.irandoc.ac.ir/article_699334.html)



## References

- Alishan Karami, N., Haji Zeinolabedini, M., Radad, I., & Ghazi-Mirsaeid, S. J. (2018). Application and role of ontologies in biomedical information retrieval systems. *Journal of Health and Biomedical Informatics*, 4(4), 327–340. <http://jhbmi.ir/article-1-222-fa.html> [In Persian]
- Arp, R., Smith, B., & Spear, A. D. (2015). Building Ontologies with Basic Formal Ontology. Retrieved from <http://www.jstor.org/stable/j.ctt17kk7vw>
- Asgari-Bidhendi, M., Hadian, A., & Minaei-Bidgoli, B. (2019). FarsBase: The Persian knowledge graph. *Semantic Web*, 10(6), 1169–1196. <https://doi.org/10.3233/SW-190369>
- Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, N., & Mahmood, W. (2019). The Use of Ontology in Retrieval: A Study on Textual, Multilingual, and Multimedia Retrieval. *IEEE Access*, 7(c), 21662–21686. <https://doi.org/10.1109/ACCESS.2019.2897849>
- Choi, J. (2022). Graph Embedding-Based Domain-Specific Knowledge Graph Expansion Using Research Literature Summary. *Sustainability (Switzerland)*, 14(19). <https://doi.org/10.3390/su141912299>
- Emami, H. (2022). A method for extracting domain-specific structured information from Persian web pages. *Journal of Signal and Data Processing*, (52), 133–146. <http://jsdp.rcisp.ac.ir/article-1-1102-fa.html> [In Persian]
- Fensel, D., \cSim\csek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., ... Wahler, A. (2020). Knowledge Graphs: Methodology, Tools and Selected Use Cases. Retrieved from <https://books.google.com/books?id=1qnNDwAAQBAJ>
- Gutierrez, F., Dou, D., Fickas, S., Wimalasuriya, D., & Zong, H. (2016). A hybrid ontology-based information extraction system. *Journal of Information Science*, 42(6), 798–820. <https://doi.org/10.1177/0165551515610989>
- Homayouni, R., Ding, D., Roy, S., & Ebeid, I. A. (2018). OPEN ACCESS EDITED BY Ebeid IA ((((( MedGraph: A semantic biomedical information retrieval framework using knowledge graph embedding for PubMed MedGraph: A semantic biomedical information retrieval framework using knowledge graph embedding for PubMed. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/>
- Hosseini Beheshti, M. S., & Ejei, F. (2015). Designing and implementing basic sciences ontology based on concepts and relationships of relevant thesauri. *Iranian Journal of Information Processing and Management*, 30(3), 677–696. [https://jipm.irandoc.ac.ir/article\\_699309.html](https://jipm.irandoc.ac.ir/article_699309.html) [In Persian]
- Jafari Pavarsi, H., Hariri, N., Alipour Hafezi, M., Babolhavaeji, F., & Khademi, M. (2020). Enhancing semantic information retrieval using tagging and ontology. *Quarterly Journal of Knowledge and Information*

- Management, 31(1), 18–38. <https://www.magiran.com/paper/2127226> [In Persian]
- Kejriwal, M., Knoblock, C. A., & Szekely, P. (2021). Knowledge Graphs: Fundamentals, Techniques, and Applications. Cambridge: MIT Press.
- Li, J. (2019). Research on search engine technology of knowledge graph analysis. AIP Conference Proceedings, 2073(February). <https://doi.org/10.1063/1.5090759>
- Liao, J., Li, L., & Liu, X. (2015). An integrated, ontology-based agricultural information system. Information Development, 31(2), 150–163. <https://doi.org/10.1177/0266666913510716>
- Moradi, M., Vazirnezhad, B., & Bahrani, M. (2015). Commonsense knowledge extraction for Persian language: A combinatory approach. Iranian Journal of Information Processing and Management, 31(1), 109–124. [https://jipm.irandoc.ac.ir/article\\_699334.html](https://jipm.irandoc.ac.ir/article_699334.html) [In Persian]
- PARK, J.-R., & HOWARTH, L. C. (2013). NEW DIRECTIONS IN INFORMATION ORGANIZATION.
- Rahman, M. M., & Takasu, A. (2017). Entity Oriented Action Recommendations for Actionable Knowledge Graph Generation. Proceedings of the International Conference on Web Intelligence, 686–693. <https://doi.org/10.1145/3106426.3106546>
- Rajabi, E., & Etminani, K. (2022). Knowledge-graph-based explainable AI: A systematic review. Journal of Information Science. <https://doi.org/10.1177/01655515221112844>
- Sadeghi Niaraki, A., Barzegar, M., & Shakery, M. (2018). Design and development of ontology for spatial experiences in urban routing. Journal of Soft Computing and Information Technology, 7(2), 1–15. [https://jscit.nit.ac.ir/article\\_80956.html](https://jscit.nit.ac.ir/article_80956.html) [In Persian]
- Sajadi, M. B., & Minaei Bidgoli, B. (2020). The Architecture of Farsi Knowledge Graph System. Iranian Journal of Information Processing and Management, 35(2), 425–462.
- Sajjadi, S. M. B., & Minaei Bidgoli, B. (2020). Architecture of Persian language knowledge graph system. Iranian Journal of Information Processing and Management, 35(2), 425–462. <https://doi.org/10.35050/JIPM010.2020.057> [In Persian]
- Servati, L., Valavi, M. R., & Hourali, M. (2017). Development of a geographic ontology for use in military applications. Sepehr Geographical Information Quarterly, 26(102), 5–17. <https://www.magiran.com/paper/1738409> [In Persian]
- Shen, X., Chen, J., Chen, J., Zeng, C., & Xiao, Y. (2022). Diversified query generation guided by knowledge graph. In WSDM 2022 - Proceedings of the 15th ACM International Conference on Web Search and Data Mining (Vol. 1). <https://doi.org/10.1145/3488560.3498431>
- Taghizadeh Nayini, J., Fahimnia, F., & Naghshineh, N. (2019). Ontology extraction of digital identity based on domain analysis. Iranian Journal of Information Processing and Management, 34(4), 1669–1700. <https://doi.org/10.35050/JIPM010.2019.018> [In Persian]

- Tan, J., Qiu, Q., Guo, W., & Li, T. (2021). Research on the construction of a knowledge graph and knowledge reasoning model in the field of urban traffic. *Sustainability* (Switzerland), 13(6). <https://doi.org/10.3390/su13063191>
- Tong, P., Zhang, Q., & Yao, J. (2019). Leveraging Domain Context for Question Answering Over Knowledge Graph. *Data Science and Engineering*, 4(4), 323–335. <https://doi.org/10.1007/s41019-019-00109-w>
- Torabi, L., Mirhosseini, Z., Abazari, Z., & Hosseini-Beheshti, M. S. (2022). Design and construction of nuclear medicine ontology. *Health Information Management*, 18(2), 81–88. <https://doi.org/10.22122/him.v18i2.4357> [In Persian]
- Wang, P., Jiang, H., Xu, J., & Zhang, Q. (2019). Knowledge graph construction and applications for web search and beyond. *Data Intelligence*, 1(4), 333–349. [https://doi.org/10.1162/dint\\_a\\_00019](https://doi.org/10.1162/dint_a_00019)
- Wimalasuriya, D. C., & Dou, D. (2010). Ontology-based information extraction: An introduction and a survey of current approaches. *Journal of Information Science*, 36(3), 306–323. <https://doi.org/10.1177/0165551509360123>
- Wu, Y. (2018). Enriching a thesaurus as a better question-answering tool and information retrieval aid. *Journal of Information Science*, 44(4), 512–525. <https://doi.org/10.1177/0165551517706219>
- Yang, L., Guo, H., Dai, Y., & Chen, W. (2023). A Method for Complex Question-Answering over Knowledge Graph. *Applied Sciences* (Switzerland), 13(8), 1–23. <https://doi.org/10.3390/app13085055>
- Zaki, N., Tennakoon, C., Al Ashwal, H., Al Jaber, A., & Al Ameri, A. (2019). Methods of creating knowledge graph by linking biological databases. In *Advances in Intelligent Systems and Computing* (Vol. 803). [https://doi.org/10.1007/978-3-319-98702-6\\_7](https://doi.org/10.1007/978-3-319-98702-6_7)

**استناد به این مقاله:** انتهای سرای، علیرضا، نقشینه، نادر، مینایی بیدگلی، بهروز و شعبانی، علی. (۱۴۰۴). مدل سازی هستی شناسی دامنه در آرشیو دیجیتال مرکز اسناد انقلاب اسلامی: رویکرد ترکیبی تحلیل متن و بازاستفاده از هستی شناسی ها. *فصلنامه بازیابی دانش و نظام های معنایی*، سال (شماره)، ص آغاز-ص پایان.



**Name of Journal** is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

زود آیند ویراستاری نشده (نشریه بازیابی دانش و نظام‌های معنایی)