

A Framework for Transforming the Persian Subject Headings into Linked Data

Zeinab Sabbaghi Bidgoli 

M.Sc. Student in Knowledge and Information Science, Tarbiat Modares University, Tehran, Iran

Atefeh Sharif *

Assistant Professor, Knowledge and Information Science, Tarbiat Modares University, Tehran, Iran

Fatemeh Zandian 

Associate Professor, Knowledge and Information Science, Tarbiat Modares University, Tehran, Iran

Abstract

The semantic web which facilitated information retrieval is based on linked data. Libraries that have an important role in authority data can affect future information systems. Therefore, publishing library data as linked data can be a step forward in the semantic web. This paper aims to design a framework for transforming the National Library of Iran Subject Headings into linked data and publish it on the web. This paper is an applied study that used the library method and review the related literature to design a linked data framework for the National Library of Iran Subject Headings. The designed framework was implemented to ensure that its publishing is possible. Thus, first, the data (i.e., Persian topical terms) were cleaned. Then using Open Refine software, they were transformed into an RDF version and linked with the Library of Congress Subject Headings dataset. The transformed version of Persian Subject Headings was imported into the Jena Fuseki RDF Store and finally, the subject headings were displayed through skosmos interface. The designed framework can be used for publishing the National Library of Iran Authority files as linked data. In this framework, linking with similar datasets was considered, and publishing a part of subject headings was successfully reported.

* Corresponding Author: atefeh.sharif@modares.ac.ir

How to Cite: Sabbaghi Bidgoli, Z., Sharif, A. & Zandian, F. (۲۰۲۳). A Framework for Transforming the Persian Subject Headings into Linked Data. *Journal of Knowledge Retrieval and Semantic Systems*, ۱۰(۳۷), ۱-۲۸. DOI: ۱۰,۲۲۰۵۴/JKS.۲۰۲۳,۷۲۵۳۸,۱۵۶۵

Keywords: Linked Data, the National Library of Iran Subject Heading, Open Refine, Data Publication.

۱. Introduction

The emergence of the web facilitated the retrieval of information. This made libraries as one of the most important centers of information considering the web for the information retrieval process. However, the fast change of the web leads to the transformation of library functions. The semantic web is an opportunity for libraries to change their functions. Linked data as a method in the semantic web can make a major change in library functions. It can improve the discoverability, visibility, and interoperability of the resources. For example, all libraries use authority controls for organizing their information. But using authority controls in a traditional way can be challenging. Therefore, using the web can help libraries tackle these potential challenges and problems. Transforming authority data into linked data which seems an innovative and faster way for finding the resources can be a step forward for libraries and users. This paper aims to design a framework for transforming the National Library of Iran Subject Headings into linked data and publish them on the web.

۲. Literature Review

Designing and proposing a framework for linking the data was the topic of some research papers. Linking the university data (Behkamal et al., ۲۰۱۱) linking and visualizing medicine information (Sekhavati, Farahi, & Jalali, ۲۰۱۱) web objects (Hosseini, ۲۰۲۰), table data (Mulwad et al., ۲۰۱۰), Industrial Data (Graube et al., ۲۰۱۲), and government data (Villazón-Terraza, Vilches-Blázquez, Corcho, & Gómez-Pérez, ۲۰۱۱; Mulwad, Finin, & Joshi, ۲۰۱۱) were the topics for some reviewed studies. The results of their studies indicated that in general, linked data could improve information retrieval. Implementing a linked data method in library data was discussed in some papers. Kar & Das (۲۰۲۰) designed a methodology for linking bibliographic information in a digital repository. Similarly, Ryan et al. (۲۰۱۵) examined the linking of place names in a dataset, transferring them into RDF and linking them with other similar datasets. Summers, et al (۲۰۰۸) provide a methodology for transferring subject headings into linked data. their results showed that transferring LCSH into SKOS affects information retrieval. The linking and publishing

National Library of Iran data were also investigated by Eslami & Vaghefzadeh (۲۰۱۳). Fathian Dastgerdi et al (۲۰۲۰) tried to make a pattern for linking data in library systems. They examined the components which are needed for implementing the linked data method in library systems. Their result showed that using linked data in library systems affects the visibility of bibliographic metadata. Based on the reviewed studies, many international papers discussed publishing library linked data in theoretical and practical ways. Whereas studies done in Iran focusing on linked data mostly developed patterns and models for linking data (e.g., Fathian Dastgerdi; ۲۰۲۰). Few Persian studies were done for publishing bibliographic data (e.g., Eslami & Vaghefzadeh, 2۰۱۳; Sekhavati, ۲۰۱۱). Although there is a significant number of papers discussing linked data, the technical aspect for publishing and linking library data was rarely examined. To fill this gap, this study aims to develop a framework for publishing National Library of Iran subject headings which is unlike Fathian Dastgerdi et al., (۲۰۲۰) paper considers the technical tools and aspects and unlike Sekhavati's (۲۰۱۱) paper examines the Persian subject headings.

۳. Methodology

This research is an applied study that utilizes a library method for designing a publishing framework. Linked data was implemented to ensure the possibility of publishing the research data. First, Persian subject headings which are represented in Iran MARC format were obtained in Marc XML files From the National Library of Iran. Then the method for transferring and publishing the data was applied.

۴. Results

The framework developed in this research collected National Library of Iran subject headings randomly. The selected data were first cleaned by Microsoft Excel and MarcEdit. In the next step, cleaned data were converted into RDF Using OpenRefine. The study's project was imported to Open Refine software, linked with external datasets, and saved in a triple store. Finally, the linked subject headings were displayed through the Skosmos interface.

۵. Discussion

Publishing library data as linked data is an example of utilizing Web ۳


in library systems. National libraries worldwide have tried linking their data including subject headings with other datasets. However, there remains a gap in publishing linked Persian subject headings and to the best of the authors' knowledge it seems that no paper has pointed to technical aspects of implementing Persian subject headings.


۶. Conclusion


The current paper has transformed the Persian subject headings into a linked dataset in an RDF turtle format. Then, it visualized the linked data in the Skosmos interface. But there can be some limitations to this study. Using OpenRefine was reported successfully in this paper, but it seems that there may be a problem in data with larger sizes. In conclusion, since this framework improve the retrieval of authority data in this research, it can be used for publishing National library of Iran subject headings.



چارچوبی برای تبدیل و انتشار سرعنوان‌های موضوعی فارسی به داده‌های پیوندی

زینب صباغی بی‌لی  دانشجوی کارشناسی‌ارشد، علم اطلاعات و دانش‌شناسی، دانشگاه تربیت مدرس، تهران، ایران

عاطفه شریف  * استادیار، گروه علم اطلاعات و دانش‌شناسی، دانشگاه تربیت مدرس، تهران، ایران

فاطمه زندیان  دانشیر، گروه علم اطلاعات و دانش‌شناسی، دانشگاه تربیت مدرس، تهران، ایران

چکیده

ظهور وب معنایی در جهت تحقق بازیابی معنایی اطلاعات است و در حال حاضر در داده‌های پیوندی تجلی یافته است. کتابخانه‌ها در تولید و مدیریت داده‌های مستند و معتبر فراوانی نقش دارند و می‌توانند نقشی مؤثر در نظام‌های اطلاعاتی پیش‌رو ایفا کنند و می‌توانند با اجرایی کردن داده‌های پیوندی گامی در این مسیر بردارند. هدف از انجام این پژوهش ارائه چارچوبی برای انتشار و تبدیل سرعنوان‌های موضوعی فارسی مورد استفاده کتابخانه ملی ایران به صورت داده‌های پیوندی و ایجاد پیوند با مجموعه داده‌ای مشابه است. پژوهش حاضر از نوع کاربردی است؛ با استفاده از روش کتابخانه‌ای به طراحی چارچوبی برای انتشار سرعنوان‌های موضوعی پرداخته و برای اطمینان از امکان انتشار داده‌ها، روش مورد نظر مورد پیاده‌سازی قرار گرفته است. بدین ترتیب ابتدا داده‌های موضوعی فارسی مورد پاک‌سازی و ویرایش قرار گرفتند، سپس با نرم‌افزار اپن‌ریفاین به آر.دی.اف تبدیل شدند و با سرعنوان‌های موضوعی کتابخانه کنگره پیوند دریافت کردند. داده‌های مورد مطالعه پس از نگاشت به اسکاس به یک فایل آر.دی.اف در قالب ترنل تبدیل شدند. فایل تبدیل‌شده ابتدا وارد مخزن آر.دی.اف جینا فوسکی شد و سپس در رابط کاربری اسکاسموس در محیط وب نمایش داده شد. به‌طور کلی این چارچوب می‌تواند در فرایند انتشار داده‌های مستند کتابخانه ملی در قالب داده‌های پیوندی مورد استفاده قرار گیرد. در این چارچوب امکان برقراری پیوند با مجموعه داده‌های مشابه نیز در نظر گرفته شده است و پیاده‌سازی نمونه‌ای از داده‌ها با موفقیت انجام

* نویسنده مسئول: atefeh.sharif@modares.ac.ir

پذیرفت.

کلیدواژه‌ها: داده‌های پیوندی، سرعنوان‌های موضوعی فارسی، نرم‌افزار اپن‌ریفاین، کتابخانه ملی ایران، چارچوب انتشار.



مقدمه

ظهور وب در جهان، دسترسی انسان به اطلاعات را سرعت و سهولت بخشیده اما چالش‌ها و مشکلاتی را نیز به همراه داشته است؛ برای مثال ممکن است فردی در یافتن اطلاعات مرتبط و مناسب با مشکلاتی مواجه شود و نتواند به آن‌ها دسترسی پیدا کند (Goswami & Biswas, ۲۰۱۱). همگام با ظهور وب و تحولات بعدی آن، کتابخانه‌ها نیز برای بهبود فرایند بازیابی اطلاعات از آن استفاده کردند. با پیشرفت و استفاده از فناوری اطلاعات و ارتباطات در کتابداری و علم اطلاعات، کارکرد کتابخانه‌ها نیز تغییر کرده است. وب معنایی فرصت مناسبی را برای تحول در کارکردهای کتابخانه فراهم کرده است (Kamila, ۲۰۰۸). داده‌های پیوندی را می‌توان یکی از مصادیق وب معنایی دانست. به‌طور کلی، داده‌های پیوندی رویکردی است که ساخت و استفاده مجدد از محتوای معنایی را سهولت و ارتقا می‌بخشد (Candela et al., ۲۰۲۰). در واقع می‌توان داده‌های پیوندی را به‌عنوان روش‌ها و الگوهای دانست که برای اتصال داده‌های ماشین‌خوان به یکدیگر از طریق وب معنایی به کار گرفته می‌شود و با استفاده از شناسه‌گر واحد منبع (یو.آر.آی)^۱ و طرح فراداده‌ای چارچوب توصیف منبع (آر.دی.اف)^۲ داده‌ها را در معرض وب قرار می‌دهند، آن‌ها را به اشتراک می‌گذارند و متصل می‌کنند (Park & Kipp, ۲۰۱۹).

بحث استفاده و انتشار داده‌های پیوندی در کتابخانه‌ها نیز دارای اهمیت است؛ از یک طرف بهبود بخشیدن به کنترل مستندات که به‌عنوان دارایی‌های اطلاعاتی کتابخانه‌ها هستند، موجب می‌شود فرایند بازیابی و بازنمایی آسان‌تر شود و از سوی دیگر، تعداد کتابخانه‌هایی که به توسعه و انتشار داده‌های خود به‌صورت پیوندی می‌پردازند و مجموعه‌های موجود در کتابخانه‌ها را به وب داده‌ها وارد می‌کنند به‌طور روزافزون بیشتر می‌شود (Lampert & Southwick, ۲۰۱۳). ارتباط دادن داده‌ها به یکدیگر به کاربر اجازه می‌دهد تا به‌طور یکپارچه در محیط وب جستجو کند و صرف‌نظر از این که داده‌ها توسط

^۱. Uniform Resource Identifier (URI)

^۲. Resource Description Framework (RDF)

چه کسی و در کجا منتشر شده‌است به نتایج مرتبطی دست پیدا کند. همچنین این کار باعث حذف ابهام می‌شود و فرایند جستجو را دقیق‌تر می‌کند (Southwick, ۲۰۱۵).

هم‌زمان با افزایش تمایل به انتشار داده‌های پیوندی در جهان، کنسرسیوم وب جهانی مجموعه‌ای از توصیه‌ها را برای انتشار داده‌های کتابخانه‌ها نیز توسعه داد (Bushman et al., ۲۰۱۵). در این راستا، بسیاری از کتابخانه‌های ملی مانند کتابخانه ملی بریتانیا، فرانسه، اسپانیا، کتابخانه ملی پزشکی آمریکا، کتابخانه ملی کشاورزی و دیگر کتابخانه‌ها داده‌های کتابشناختی خود را به‌صورت مجموعه داده‌های پیوندی منتشر کرده‌اند (Bushman et al., ۲۰۱۴). کتابخانه کنگره نیز فهرست سرعنوان‌های موضوعی کنگره را در قالب اسکاس در ابر داده منتشر کرد (شریف، ۱۳۹۳). کتابخانه کنگره در زمینه داده‌های پیوندی گام‌های بیشتری نیز برداشته است؛ برای مثال در سال ۲۰۱۱ اعلام کرد که قصد دارد از چارچوب کتابشناختی جدید به نام بیب‌فریم^۱ استفاده کند (Halla, ۲۰۱۳ as cited in Library of Congress, ۲۰۱۱) که در واقع نقشه راهی برای انتقال از فهرست‌های مبتنی بر مارک به فهرست‌های یکپارچه و مبتنی بر داده‌های پیوندی است (Halla, ۲۰۱۳). اگرچه پیوندی‌شدن داده‌های کتابخانه‌ای در خارج از ایران به‌طور چشمگیری انجام پذیرفته است، اما تاکنون داده‌های کتابخانه ملی در قالب قابل‌تبدیل و انتشار به‌صورت داده‌های پیوندی منتشر نشده‌است. در نتیجه نمی‌توان به‌طور دقیق چارچوب انتشار داده‌های پیوندی کتابخانه ملی را پیش‌بینی کرد. پژوهش حاضر قصد دارد چارچوبی برای انتشار بخشی از سرعنوان‌های موضوعی کتابخانه ملی ایران به‌صورت داده‌های پیوندی ارائه کند که به‌صورت عملی قابل پیاده‌سازی باشد تا گامی برای برطرف کردن خلأ موجود در این قسمت بردارد. به‌طور کلی، هدف پژوهش حاضر ارائه یک روش‌شناسی برای تبدیل، پیاده‌سازی و انتشار سرعنوان‌های موضوعی فارسی به‌صورت داده‌های پیوندی است تا از این طریق گامی برای بازیابی هر چه بهتر این داده‌ها در محیط وب برداشته شود.

^۱. bibframe

پیشینه پژوهش

بهکمال و همکاران (۱۳۹۰) در پژوهش خود برای انتشار داده‌های فارسی به صورت داده‌های پیوندی، چارچوبی را مطرح کردند. آن‌ها در این چارچوب ابتدا داده‌ها را به آر.دی.اف تبدیل کردند. سپس داده‌های تبدیل شده را با تبدیل فرمت به اچ.تی.ام.ال در مخزن ذخیره کردند و در نهایت لغات وید^۱ آن منبع داده را منتشر کردند. برای انجام این کار قسمتی از داده‌های دانشگاه فردوسی مشهد به منظور اعمال چارچوب، انتخاب و مشکلات پیش‌رو برای انتشار داده‌ها بیان شد. یافته‌های پژوهش نشان داد پیش از انتشار داده باید کیفیت داده‌ها بازنگری شوند؛ زیرا با کیفیت بودن داده‌ها پیش‌نیاز پیوند خوردن آن‌ها است. پس از اطمینان از کیفیت داده‌ها، آن‌ها باید به طور مناسبی با یکدیگر پیوند برقرار کنند. بحث روزآمدی داده‌ها و پیوندهای میان آن‌ها و انتخاب بهترین پیوند با توجه به تفاوت‌های زبانی و فرهنگی از دیگر مشکلات اشاره شده در این پژوهش بود.

سختوتی (۱۳۹۰) هدف از پژوهش خود را ارائه چارچوبی برای انتشار اطلاعات کتابخانه به طور کلی بر پایه اصول داده‌های پیوندی دانست. او برای ارائه چارچوب، ابتدا به هر پیشینه یک یو.آر.آی اختصاص داد؛ برای منحصر بودن شناسه‌ها از شبکه کتاب‌ها استفاده شد. برای تبدیل داده‌ها به آر.دی.اف، برنامه‌ای با کد جاوا نوشته شد. داده‌ها پس از تبدیل شدن به آر.دی.اف در مخزن آر.دی.اف Sesame ذخیره شدند. برای پیوند داده‌ها، شرایط ایجاد پیوند بررسی شد و نمونه داده‌هایی که امکان ایجاد پیوند را دارند جستجو شد.

سختوتی و همکاران (۱۳۹۰) چارچوبی برای تبدیل اطلاعات داروها به داده‌های پیوندی را در پژوهش خود پیشنهاد دادند. بدین صورت که داده‌های رابطه‌ای داروها پس از تبدیل به آر.دی.اف و ذخیره در پایگاه، با واژگان مرتبط پیوند می‌خورند و در نهایت منتشر می‌شوند. در این پژوهش میزان مانعیت و جامعیت و دقت داده‌های پیوندی نیز مورد بررسی قرار گرفت.

^۱. VOid

نخستین تلاش برای انتشار داده‌های کتابخانه‌ای توسط اسلامی و واقف‌زاده (۲۰۱۳) و با هدف انتشار داده‌های کتابخانه ملی ایران به صورت داده پیوندی مطرح شد. ایشان با استفاده از مدل اف.آر.بی.آر.^۱ به تعریف مدلی برای ایجاد نسخه داده‌های پیوندی فایل مستندات کتابخانه ملی پرداختند. به همین منظور محدوده کوچکی از داده‌های مستند نام‌های موجود در پایگاه داده رسا انتخاب و با توجه به نگاهت ایجادشده توسط متخصصان به آر.دی.اف تبدیل و ذخیره شد. با این حال، پژوهش آن‌ها در محیط وب اجرایی نشده است.

اکبری داریان و همکاران (۲۰۱۷) پژوهشی برای در دسترس ساختن داده‌های اصطلاح‌نامه‌های فرهنگی سه‌زبانۀ کتابخانه ملی ایران به صورت داده‌های پیوندی و در قالب اسکاس انجام دادند. بدین منظور نگاهتی برای تبدیل مارک به اسکاس گسترش داده شد، یو.آر.ال به آن‌ها اختصاص داده شد و فرایند تبدیل با استفاده از جاوا اسکریپت انجام شد. ارائه چهارچوبی برای پیوندی کردن اشیا موجود در وب، موضوع پژوهش حسینی (۱۳۹۹) بود. در این پژوهش پس از مرحله جمع‌آوری داده‌ها که شامل دریافت و پاک‌سازی داده‌ها می‌شد داده‌ها که به صورت جدولی بود با استفاده از نرم‌افزار اپن‌ریفاین^۲ به آر.دی.اف تبدیل شد. در آخرین مرحله پژوهش نیز داده مورد بازیابی معنایی قرار گرفتند. بر اساس نتایج این پژوهش داده‌های پیوندی می‌تواند باعث بهبود بازیابی معنایی شود.

فتحیان دستگردی و همکاران (۱۳۹۹) مؤلفه‌های موردنیاز برای پیاده‌سازی روش داده‌های پیوندی در نظام‌های کتابخانه‌ای را بررسی و الگویی را بر اساس آن مطرح کردند. نتایج پژوهش بیان کرد در پیشینه‌هایی که در قالب یونی‌مارک/اکس.ام.ال^۳ نمایش داده شدند ارتباط وجود ندارد. از این رو، استفاده از داده‌های پیوندی در نظام کتابخانه تأثیر بسیاری در نمایش فراداده‌های کتابشناختی دارد.

^۱. Functional Requirements for Bibliographic Records (FRBR)

^۲. OpenRefine

^۳. UNIMARC/XML

پژوهش‌هایی در خارج از ایران نیز به ارائه چارچوب و مدل برای پیوندی کردن داده‌های متفاوت پرداخته است. سامرز و همکاران^۱ (۲۰۰۸) روشی را برای تبدیل سرعنوان‌های موضوعی کتابخانه کنگره به داده‌های پیوندی معرفی کردند. هدف آن‌ها توصیف فرایند تبدیل سرعنوان‌های موضوعی - که در قالب مارک اکس.ام.ال هستند - به اسکاس / آر.دی.اف^۲ بود. رویکردی برای در دسترس قرار دادن سرعنوان‌های موضوعی کتابخانه کنگره نیز در برنامه کاربردی وب مطرح شد. نتایج نشان داد که تبدیل سرعنوان‌های موضوعی کنگره به عنوان اسکاس در سطوح متفاوت ارزشمند بوده و قدمی روبه‌جلو برای تحقق بخشیدن به توسعه وب معنایی است.

مولود و همکاران^۳ (۲۰۱۰) چارچوبی خودکار برای تبدیل داده‌های جدولی به داده‌های پیوندی ارائه دادند و بر اساس جدول‌های موجود در وب و تفسیر آن‌ها آر.دی.اف ایجاد کردند.

داده‌های صنعتی موضوع پژوهش گراوب و همکاران^۴ (۲۰۱۲) بود که بر تبدیل داده‌های مربوط به کارخانه‌ها به داده‌های پیوندی تمرکز داشتند. داده‌ها به صورت اکس.ام.ال استخراج شد و تبدیل به آر.دی.اف با استفاده از متن پایتون انجام و داده‌های تبدیل شده در سرور ویرچوسو^۵ ذخیره شدند. در نهایت این داده‌ها به صورت اچ.تی.ام.ال نمایش داده شدند.

در فضای داده‌های کتابشناختی و به‌طور کلی کتابخانه‌ای، داده‌های متفاوتی وجود دارد که می‌تواند به داده‌های پیوندی تبدیل شود. در این حوزه نیز پژوهش‌هایی به انجام رسیده است. انتشار فراداده‌های کتابخانه‌ای در پژوهش هالو و همکاران^۶ (۲۰۱۴) مورد توجه قرار گرفت. روش ارائه‌شده آن‌ها برای پیوندی کردن فراداده‌های کتابخانه‌ای

^۱. Summers et al.

^۲. SKOS/RDF

^۳. Mulwad et al.

^۴. Graube et al.

^۵. virtuoso

^۶. Hallo et al.

شامل استخراج فراداده از کتابخانه، تبدیل آن به آر.دی.اف و انتشار آن با استفاده از اندپوینت اسپارکل بود.

رایان و همکاران^۱ (۲۰۱۵) نیز به توصیف فرایند پیوندی کردن مجموعه داده Logainm (یک پایگاه داده برخط که شامل فهرست اسامی مکان‌های کشور ایرلند به انگلیسی و ایرلندی است) انتقال آن از اکس.ام.ال به آر.دی.اف و ساختن پیوند به دیگر مجموعه داده‌های جغرافیایی پرداختند. ولی روش آن‌ها با روش کار و داس^۲ (۲۰۲۰) متفاوت بود. مفاهیم موجود در پایگاه داده Logainm به شکل یک XML dump تبدیل شد که شامل تمام نام‌های مکان به ایرلندی و انگلیسی، نوع مکان و موقعیت جغرافیایی آن مکان‌ها می‌شد. تبدیل dump پایگاه داده Logainm (که به زبان اکس.ام.ال بود) به آر.دی.اف با استفاده از XSPARQL انجام شد. به این معنا که پرس‌وجویی توسعه داده شد تا داده‌های به زبان اکس.ام.ال را به آر.دی.اف تبدیل کند. مجموعه داده آر.دی.اف Logainm در وبسایت Logainm مستقر شد و پایانه^۳ اسپارکل آن نیز ایجاد شد.

بوشمن و همکاران^۴ (۲۰۱۵) در پژوهش خود به تبدیل سرعنوان‌های موضوعی پزشکی به داده‌های پیوندی پرداختند؛ بنابراین در یک طرح آزمایشی به ایجاد و توسعه یک زیرساخت برای تبدیل، ذخیره و انتشار داده‌های پیوندی کتابخانه پزشکی آمریکا پرداختند. تمام اطلاعات به آر.دی.اف تبدیل شد. سپس داده‌ها در گزاره‌های سه‌گانه با استفاده از یو.آر.آی نمایش داده شد. نویسندگان این پژوهش نتیجه گرفتند که انتشار مش^۵ به‌عنوان داده‌های پیوندی این فرصت را برای کتابخانه ملی پزشکی آمریکا فراهم می‌کند تا داده‌های بیشتری از جمله داده‌های پابمد/مدلاین^۶، داده‌های کتابشناختی و غیره را به‌صورت آر.دی.اف منتشر کنند.

^۱. Ryan et al.

^۲. Kar & Das

^۳. endopint

^۴. Bushman et al.

^۵. Mesh

^۶. Pubmed/Medline

پژوهش کار و داس (۲۰۲۰) یک روش‌شناسی برای پیوندی کردن اطلاعات کتابشناختی و مهندسی آن‌ها در مخازن دیجیتال به‌عنوان یک منبع اطلاعاتی مطرح کرد. در این پژوهش روشی برای طراحی اطلاعات کتابشناختی یک مخزن دیجیتال به‌صورت داده‌های پیوندی ارائه شد. طبق روش‌شناسی ارائه‌شده در این پژوهش، برای تبدیل داده‌ها از RDFizer موجود در اپاچی جینا فوسکی و برای پرس‌وجو قرار دادن نیز از اسپارکل استفاده شد.

کار و داس (۲۰۲۱) روشی را برای انتقال فراداده‌های مستند نام افراد در مارک ۲۱ به داده‌های پیوندی ارائه کردند. به‌منظور انجام این کار، بیش از ۵۰۰ داده مستند از فهرست پیوسته کتابخانه دانشگاه ویدیاشاگور^۱ انتخاب و در پایگاه داده کوها^۲ درج شدند، داده‌های انتخاب‌شده را با نرم‌افزار این‌ریفاین پاک‌سازی کردند و به آن‌ها یو.آر.آی اختصاص دادند. سپس یو.آر.آی‌های ایجادشده را به‌صورت نیمه‌خودکار در نرم‌افزار این‌ریفاین قرار دادند و اتصال آن را با دو مجموعه داده دیگر برقرار کردند. مجموعه داده‌های پیوندی شامل گزاره‌های آر.دی.اف تولیدشده و در قالب‌های جیسون^۳، آر.دی.اف / اکس.ام.ال، تریبل و ان‌تریبل^۴ ساخته شد. در آخر، پیاده‌سازی ابزارها و فن‌های تبدیل قالب پیشینه‌های نام مستندات موفقیت‌آمیز بود.

انتشار داده‌های مستند کتابخانه ملی کشور لاتویا به‌صورت داده‌های پیوندی در پژوهش آپنیت و بوجار^۵ (۲۰۲۱) مورد توجه قرار گرفت. به‌منظور انجام این کار، دادگان با استفاده از mc^۲skos به اسکاس تبدیل و با استفاده از رابط کاربری اسکاسموس^۶ دادگان اسکاس به‌عنوان داده‌های پیوندی منتشر شدند.

اگرچه مشاهده می‌شود که در پژوهش‌های بین‌المللی از مطالعه ابعاد نظری گذر شده

^۱. Vidyasagar University

^۲. Koha

^۳. JSON

^۴. N_Triple

^۵. Apenite & Bojars

^۶. skosmos

و انتشار داده‌های پیوندی محقق شده است اما اندک پژوهش‌های انجام شده در ایران بیشتر به ارائه الگو و در موارد معدودی به انتشار داده‌های کتابشناختی پرداخته‌اند (برای مثال Eslami & Vaghefzadeh, ۲۰۱۳؛ سخاوتی، ۱۳۹۰؛ فتحیان دستگردی و همکاران، ۱۳۹۹). به نظر می‌رسد اهمیت این موضوع در ایران تا حدودی درک شده اما جنبه‌های فنی و عملیاتی که مراحل انتشار داده‌های کتاب‌شناختی را منعکس نماید، کمتر مورد توجه بوده است. به دیگر سخن این پژوهش درصدد است چارچوبی را برای انتشار داده‌های موضوعی کتابخانه ملی ارائه دهد که برخلاف پژوهش فتحیان دستگردی و همکاران (۱۳۹۹) ابزارها و جوانب فنی را در نظر می‌گیرد و برخلاف پژوهش سخاوتی (۱۳۹۰) سرعنوان‌های موضوعی فارسی را مورد پژوهش قرار می‌دهد.

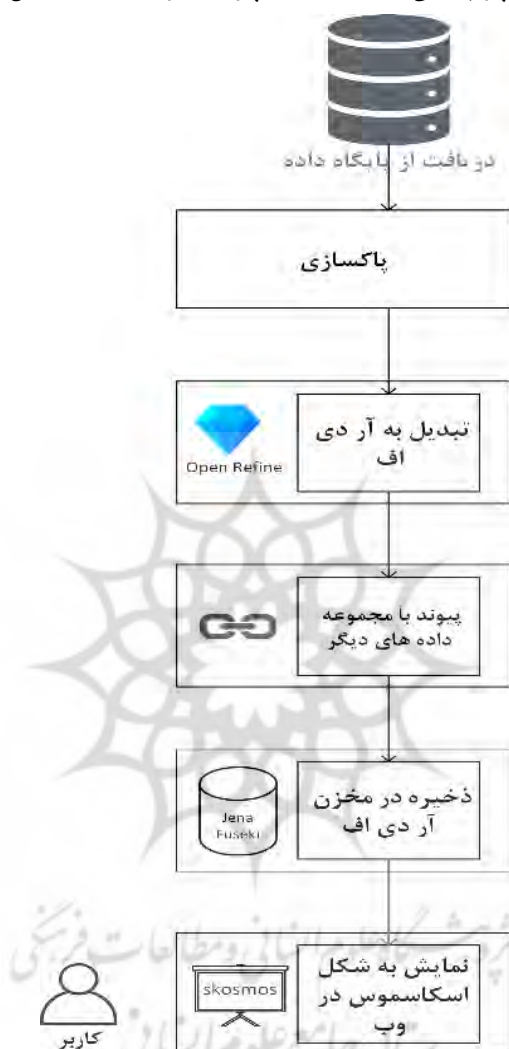
روش پژوهش

با توجه به هدف این مطالعه که ارائه چارچوبی برای تبدیل سرعنوان‌های موضوعی فارسی به یک مجموعه داده پیوندی است، می‌توان این پژوهش را از نظر هدف کاربردی دانست. برای شناسایی چارچوب مطلوب انتشار از روش کتابخانه‌ای استفاده شد. روش پژوهش شامل مراحل جمع‌آوری داده و پیاده‌سازی می‌شود. ابتدا سرعنوان‌های موضوعی با ارسال درخواستی از طریق کتابخانه ملی ایران جمع‌آوری شد. در مرحله پاک‌سازی نیز از نرم‌افزارهای مارک‌ادیت و اکسل استفاده شد. مرحله پیاده‌سازی نیز که شامل تبدیل به آر.دی.اف، ذخیره در مخزن و مصورسازی داده‌ها بود به ترتیب با استفاده از نرم‌افزارهای این‌ریفاین و طرح آپاچی جینا فوسکی و رابط کاربری اسکاسموس صورت گرفت. به این ترتیب در اولین گام برای پیوندی کردن سرعنوان‌های موضوعی، سرعنوان‌های موضوعی فارسی که در قالب مارک ایران تدوین شده و بر اساس مارک جهانی است در قالب مارک اکس.ام.ال از کتابخانه ملی ایران دریافت شد. در ادامه این نوشتار، چارچوب به دست آمده برای انتشار این داده‌ها هم‌زمان با بیان جزئیات مرتبط با آن آمده است.

چارچوب انتشار

چارچوب مورد استفاده در این پژوهش در شکل ۱ آورده شده است. سرعنوان‌های موضوعی استفاده شده ابتدا با ابزارهای لازم از جمله نرم افزار اکسل و نرم افزار اپن ریفاین که در ادامه به صورت مفصل توضیح داده می شود پاک سازی می شوند. در مرحله پاک سازی داده برچسب‌هایی که برای انتشار داده‌های پیوندی سرعنوان‌های موضوعی کاربرد نداشتند با توجه به پژوهش‌های پیشین، مورد پالایش قرار گرفتند. همچنین علامت‌ها و شناساگرهای زائد نیز حذف شدند. سپس داده‌ها مورد تعدیل و ویرایش قرار می گیرند. داده‌های آماده شده با نرم افزارهای تبدیل آر.دی.اف به گزاره‌های سه گانه آر.دی.اف تبدیل می شوند. در هنگام تبدیل به سه گانه‌های آر.دی.اف، طرح مورد نظر انتخاب می شود و پیوند با داده‌های خارجی برقرار می شود و در یک مخزن آر.دی.اف ذخیره می شود. در نهایت، برای در دسترس قرار دادن داده‌ها برای کاربران، داده‌ها در رابط کاربری اسکاسموس نمایش داده می شوند.

شکل ۱. چارچوب کلی انتشار داده‌های پیوندی مورد استفاده در این پژوهش



برای تبدیل داده‌ها بسته به نوع داده ابزارهای گوناگونی برای پیوندی کردن وجود دارد. از آنجایی که اصطلاحات موضوعی دریافت شده در این پژوهش در قالب مارک اکس.ام.ال بودند، روش کار کار و داس (۲۰۲۱) مینا قرار گرفت و فایل اکس.ام.ال به یک فایل جدولی تبدیل و با استفاده از نرم‌افزار اپن‌ریفاین طرح کلی برای ایجاد روابط معنایی

میان موجودیت‌ها تعیین شد. هانسون^۱ (۲۰۱۴) نیز در پژوهش خود داده‌های مارک اکس.ام.ال را با استفاده از فایل XSLT به آر.دی.اف اکس.ام.ال تبدیل کرد. نگارش یک فایل اکس.اس.ال.تی نیاز به دانش زبان برنامه‌نویسی داشت. همچنین، در پیشینه‌ها، فایل اکس.اس.ال.تی که مارک ایران را به آر.دی.اف تبدیل کند یافت نشد؛ بنابراین از این روش استفاده نشد. فرایند پیوند با داده‌های خارجی مطابق با (Tian et al., ۲۰۱۴, □□□□□□ □□□□□□) با استفاده از امکان تطبیق داده‌ها در نرم‌افزار اپن‌ریفاین انجام شد. مرحله ذخیره فایل آر.دی.اف نیز با توجه به هانمان و کت^۲ (۲۰۱۰) در انبار ذخیره Jena TDB انجام شد. در نهایت نمایش داده‌ها برای کاربر بر اساس پژوهش‌های پیشین (مانند سخاوتی، ۱۳۹۰؛ Zengenene et al., ۲۰۱۴) به صورت اچ.تی.ام.ال انجام شد. با این حال، از آنجایی که داده‌های این مطالعه، مستندات موضوعی هستند، مشابه پژوهش آپنیت و بوجار (۲۰۲۱) از اسکاسموس برای نمایش داده‌های موضوعی استفاده شد. در ادامه هم‌زمان با توضیح بیشتر چارچوب کلی، گزارشی مختصر از پیاده‌سازی ارائه می‌شود.

جمع‌آوری و پاک‌سازی داده‌های پژوهش

برای پیوندی کردن داده‌های موضوعی نیاز است تا صحیح بودن داده‌ها مورد بررسی قرار بگیرد و در صورت وجود ناهمگونی میان داده‌ها، مشکلات برطرف شود چراکه مغایرت میان داده‌ها باعث ایجاد خطا در مراحل تبدیل به گزاره‌های آر.دی.اف و برقراری پیوند می‌شود. در نتیجه، پس از دریافت نمونه‌ای تصادفی از سرعنوان‌های موضوعی فارسی در قالب مارک اکس.ام.ال از کتابخانه ملی ایران، داده‌ها وارد نرم‌افزار مارک ادیت^۳ و با استفاده از این نرم‌افزار تبدیل به فایلی در قالب جدولی^۴ شد که مشابه روش کار و داس (۲۰۲۱) است. برچسب‌های مارک برای استفاده در داده‌های پیوندی بر اساس پژوهش‌های پیشین (Akbaridaryan et al., ۲۰۱۷; Kar & Das, ۲۰۲۱; Summers et al., ۲۰۰۸)

^۱. Hanson, E. M.

^۲. Hannemann & Kett

^۳. MarcEdit

^۴. tab delimited file (tsv)

انتخاب و شناسایی شد و تعدادی از آن‌ها که دارای استفاده ملی است (به این معنا که تنها در مارک ایران وجود دارد) و آن‌هایی که برای انتشار سرعنوان‌های موضوعی کاربردی نیست، حذف و سایر برچسب‌ها برای استفاده در نسخه آر.دی.اف انتخاب شدند. در نتیجه برچسب‌های شناسه‌گر پیشینه (۰۰۱)، داده‌های کلی پردازش (۱۰۰)، مدخل اصلی (۲۵۰)، یادداشت کلی توضیحی (۳۲۰)، یادداشت دامنه (۳۳۰)، اصطلاح پذیرفته‌نشده (۴۵۰)، اصطلاحات اعم، اخص و مرتبط (۵۵۰)، منابع یافت‌شده (۸۱۰) و یافت‌نشده (۸۱۵) و یادداشت کلی فهرست‌نویس (۸۳۰) برای نگاشت در فایل آر.دی.اف هر سرعنوان موضوعی انتخاب شدند. پس از حذف داده‌های غیرضروری در این پژوهش از داده‌ها به صورت یک فایل سی‌اس‌وی^۱ خروجی گرفته شد و فایل خروجی وارد نرم‌افزار اپن‌ریفاین شد.

نگاشت به آر.دی.اف

سرعنوان‌های موضوعی فارسی وارد شده در نرم‌افزار اپن‌ریفاین فاقد ارتباط معنایی با یکدیگر هستند؛ بنابراین از امکان ویرایش اسکلت نرم‌افزار استفاده شد تا ارتباط معنایی هر داده برقرار شود. نخستین گام برای ایجاد طرح آر.دی.اف در پیشینه سرعنوان‌های موضوعی، انتخاب یو.آر.آی برای هر پیشینه است. به همین منظور، یو.آر.آی example.org/persiansh به عنوان یو.آر.آی پایه طرح انتخاب شد. از آنجایی که این کار بر اساس نمونه صورت می‌گرفت نام دامنه [example](http://example.org) مورد استفاده قرار گرفت و مسیر فرعی persiansh برای اشاره به سرعنوان‌های موضوعی فارسی انتخاب شد. برای نگاشت بین موجودیت‌های هر سرعنوان از فضای نام‌های اسکاس و دی سی^۲ ترمز^۲ استفاده شد. سپس به تمام برچسب‌های انتخاب‌شده بر اساس فضای نام‌ها، خصیصه‌ای اختصاص یافت (جدول ۱).

^۱. CSV

^۲. dcterms

جدول ۱. برچسب‌های مارک مورد استفاده و خصیصه‌های اختصاص داده شده به آن‌ها

برچسب مارک ایران	خصیصه آر‌دی‌اف	توضیح	کاربرد
۲۵۰.\$a	skos: preflabellang="fa"	if \$v=""	اصطلاح موضوعی فارسی
۲۵۰.\$a	skos: preflabellang="en"	if □v="□□"	اصطلاح موضوعی معادل انگلیسی
۴۵۰.\$a	skos: altlabel		نقطه دسترسی موضوع ارجاعی - نگاه کنید به
۳۳۰.\$a	skos: scopenote		یادداشت کلی دامنه
۵۵۰	skos: broader	if \$e=g	نیز نگاه کنید - اصطلاح اعم
۵۵۰	skos: narrower	if \$e=h	نیز نگاه کنید - اصطلاح اخص
۵۵۰	skos: related	if \$e=۹	نیز نگاه کنید - اصطلاح مرتبط
۳۲۰.\$a	skos: definition		یادداشت ارجاعی توضیحی
۸۳۰	skos: note		یادداشت فهرست‌نویس
۸۱۵	skos:editorialnote		منبع یافت نشده
۰۰۱	skos:concept		شناسه‌گر رکورد
۸۱۰	Dcterms:source		منبع یافت شده
۱۰۰	Dcterms:identfier		داده‌های کلی پردازش

شناسه پیشینه (فیلد ۰۰۱) به‌عنوان گره ریشه^۱ در نظر گرفته می‌شود. به این صورت که با استفاده از زبان پالایش گوگل، یو.آر.آی پایه با مقدار موجود در فیلد ۰۰۱ ادغام می‌شود. فیلد ۲۵۰ که به اصطلاح مرجع اشاره دارد در دو زبان فارسی و انگلیسی آورده شده است. برای اضافه کردن برچسب زبان به این فیلد پس از انتخاب این فیلد در پنجره اسکلت آر.دی.اف گزینه زبان انتخاب می‌شود و نام اختصاری زبان فارسی نوشته و تأیید می‌شود. علاوه بر فیلد ۲۵۰ فارسی یک فیلد ۲۵۰ انگلیسی نیز وجود دارد که با انتخاب آن، نام اختصاری زبان انگلیسی برای آن نوشته می‌شود. همچنین برای فیلد ۴۵۰ که به اصطلاحات جایگزین یا پذیرفته‌نشده اشاره دارد برچسب زبان فارسی منسوب می‌شود. علاوه بر فیلد

^۱. root node

۰۰۱ به فیلهای ۵۵۰ نیز یو.آر.آی اختصاص می‌یابد. فیلهای ۵۵۰ دارای مقادیر عدد و حروف است. به این معنا که در این فیلدها علاوه بر اصطلاح اعم، اخص یا مرتبط، شماره هر پیشینه نیز ذخیره شده‌است؛ بنابراین با استفاده از زبان پالایش گوگل تنها شناسه‌های هر پیشینه جدا و به یو.آر.آی پایه اضافه می‌شود. دیگر موجودیت‌های هر برجسب نیز با توجه به خصیصه‌ای که برای آن انتخاب شده‌است در نرم‌افزار تعریف می‌شود (شکل ۲). پس از ایجاد ساختار معنایی و اتمام فرایند اتصال به داده‌های مستند کتابخانه کنگره از فایل ایجادشده در قالب ترنل خروجی گرفته می‌شود.

شکل ۲. ایجاد طرح آر دی اف برای سرعنوان‌های موضوعی فارسی یا گزینه ویرایش

اسکلت این ریفاین

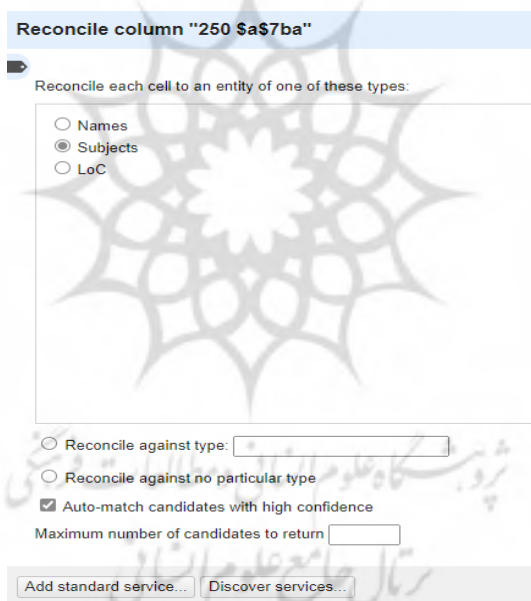
The screenshot shows the 'RDF Schema alignment' interface. It includes a description of the skeleton, a Base URI, and a list of available prefixes and their corresponding URIs. The 'Available prefixes' section shows 'rdf owl skos dcterms rdfs' with an 'Add Manage' button. Below this, there are several rows of prefixes and URIs, including '001 URI', 'skos:Concept', '100\$a Cell', '250 \$a Cell', 'Wikidata-', 'fa-uri2 URI', 'en_wikidata_uri_2...', 'YSO-uri', 'loc_uri URI', and 'fast_uri'. Each row has a checkbox and a right-pointing arrow. At the bottom, there are 'Add another root node', 'Save', 'OK', and 'Cancel' buttons.

پیوند با مجموعه داده‌های خارجی

به‌منظور غنی‌شدن داده‌های موجود نیاز است تا داده‌های مشابه شناسایی شوند و در مجموعه داده حاضر به‌صورت پیوندی اضافه شوند. برای انطباق و تلفیق داده‌های پژوهش حاضر می‌توان مجموعه داده‌ای را انتخاب کرد که شامل مستندات موضوعی باشد.

شبه‌ترین مجموعه داده برای پیوند با سرعنوان‌های موضوعی مجموعه سرعنوان‌های موضوعی کتابخانه کنگره است؛ بنابراین با استفاده از گزینه انطباق^۱ در نرم‌افزار اپن‌ریفاین و یک کد نوشته‌شده به زبان پایتون^۲ اصطلاحات فارسی با اصطلاحات موضوعی کتابخانه کنگره پیوند دریافت می‌کنند. به‌طور کلی در این سرویس امکان تلفیق داده‌ها با سرعنوان‌های موضوعی، مستند نام افراد و سازمان‌های کتابخانه کنگره وجود دارد اما از آنجایی که این پژوهش به دنبال ایجاد ارتباط با سرعنوان‌های موضوعی است، ستون حاوی معادل‌های سرعنوان‌های موضوعی فارسی (فیلد ۲۵۰) تنها با موضوعات کتابخانه کنگره پیوند می‌خورد. در نتیجه تنها موضوعات انتخاب می‌شوند (شکل ۳).

شکل ۳. انتخاب سرعنوان‌های موضوعی کتابخانه کنگره برای انطباق با داده‌های موضوعی فارسی



پس از تأیید فرایند، پیوند با سرعنوان‌های موضوعی کتابخانه کنگره به‌صورت ابرپیوند در اصطلاحات موضوعی تعبیه می‌شوند که با استفاده از زبان پالایش گوگل موجود در

^۱. reconcile

^۲. <https://github.com/cmharlow/lc-reconcile>

نرم‌افزار از یکدیگر جدا شد. در واقع با استفاده از زبان پالایش گوگل، پیوندهای برقرار شده با کتابخانه کنگره در ستون جدیدی قرار می‌گیرند. پس از برقراری پیوند برای ایجاد رابطه معنایی میان پیوندها و اصطلاحات فارسی در بخش ویرایش اسکلت نرم‌افزار این‌ریفاین بین پیوندهای کتابخانه کنگره و اصطلاحات، رابطه `skos:closeMatch` تعیین می‌شود.

ذخیره در مخزن آر.دی.اف و انتشار در وب

جینا فوسکی تی دی بی^۱ به‌عنوان مخزن آر.دی.اف انتخاب شد. برای راه‌اندازی جینا فوسکی، بسته آپاچی جینا از طریق رابط کاربری از وب‌سایت^۲ بارگیری و با اجرای یک دستور در خط فرمان محل نصب جینا متصل شد. پس از آن نیز فایل سرعنوان‌های موضوعی در قالب ترتل در این مخزن ذخیره شد. پس از ذخیره مجموعه داده در مخزن برای مصورسازی نیاز به یک رابط کاربری است. اسکاسموس یک رابط کاربری است که برای مرور، جستجوی داده‌ها و ترسیم سلسله‌مراتب آن‌ها استفاده می‌شود. با استفاده از این ابزار که به‌صورت منبع باز منتشر شده است می‌توان واژگان کنترل‌شده را به دو صورت ماشین‌خوان و انسان‌خوان منتشر کرد (Suominen et al., ۲۰۱۵). در نتیجه پس از ورود به آدرس اختصاص داده‌شده به اسکاسموس در مرورگر، سرعنوان‌های موضوعی فارسی در یک رابط گرافیکی قابل مشاهده است (شکل ۴).

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

^۱. Jena TDB

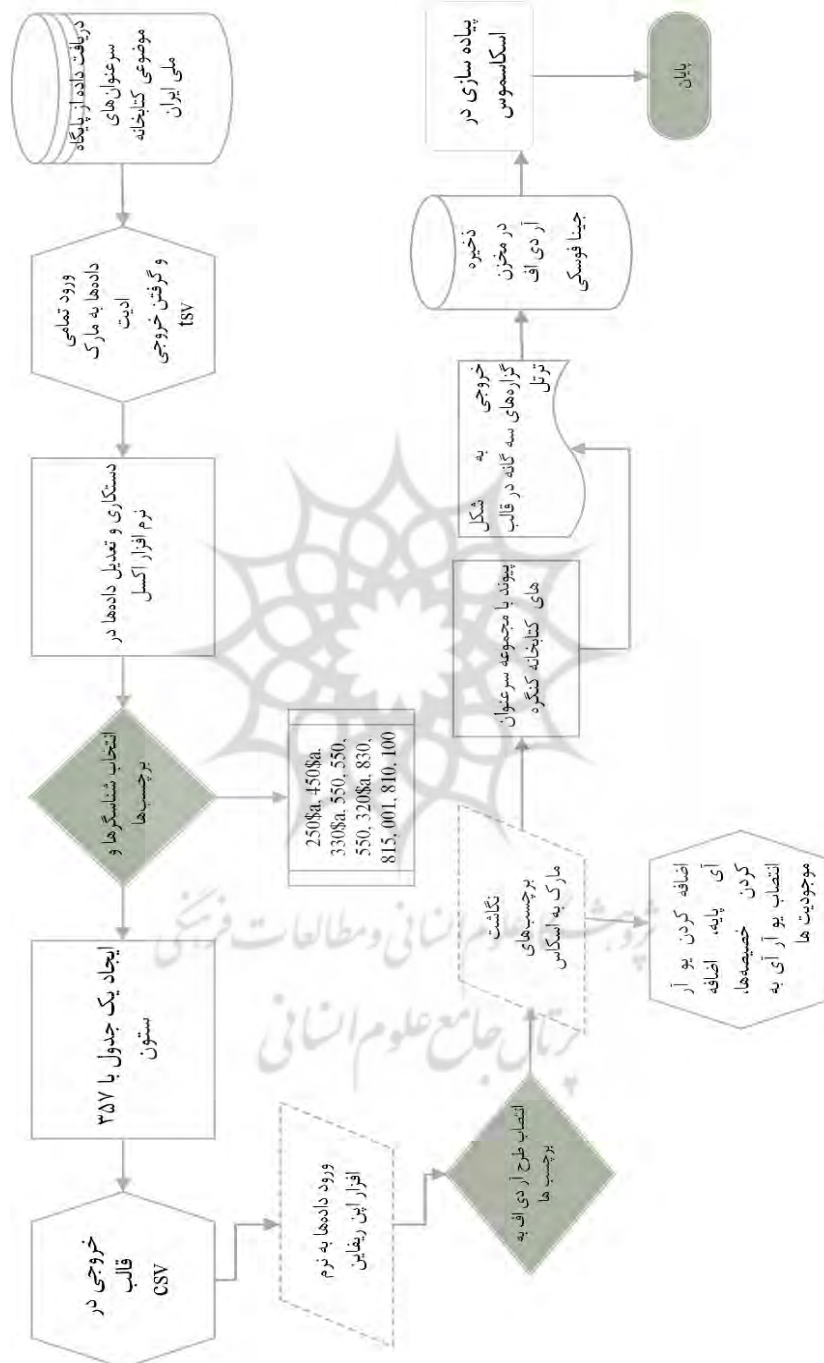
^۲. <https://jena.apache.org/download/>

شکل ۴. نمایش سرعنوان‌های موضوعی در اسکاسموس



در یک جمع‌بندی کلی، برای انتشار داده‌ها سرعنوان‌های موضوعی فارسی ابتدا داده‌ها در نرم‌افزار اکسل پاک‌سازی شدند. در واقع در این مرحله، شناسه‌گرها شناسایی و در جدولی به‌صورت ستون قرار گرفتند. پس از آن داده‌ها آماده تبدیل شدن به آر.دی.اف شدند. طرح کلی آر.دی.اف داده‌ها با استفاده از نرم‌افزار این‌ریفاین ترسیم شد و داده‌ها با یکدیگر روابط معنایی پیدا کردند و سرانجام از داده‌ها در نحو ترنل خروجی گرفته شد. داده‌های تبدیل شده به آر.دی.اف در مخزن آر.دی.اف جینا فوسکی بارگذاری و در نهایت در رابط گرافیکی اسکاسموس نمایش داده شدند. مراحل پیاده‌سازی به شکل یک روندنا در شکل ۵ ترسیم شده‌است.

شکل ۱. روندنمای کلی برای پیاده‌سازی داده‌های پیوندی در سرعنوان‌های موضوعی



بحث و نتیجه‌گیری


با وارد شدن به فاز سوم وب یعنی همان وب ۳، کتابخانه‌ها نیز از این مرحله عقب نماندند و اقداماتی برای گام نهادن در این عرصه انجام شد. انتشار داده‌های کتابشناختی در قالب داده‌های پیوندی مصداقی از حضور وب ۳ در نظام کتابخانه‌ها است. چندین کتابخانه در سطح جهان به‌منظور پیوندی کردن داده‌ها اقدام کرده‌اند. از جمله آن می‌توان به پیوندی کردن سرعنوان‌های موضوعی اشاره کرد. با این حال در انتشار داده‌های فارسی و سرعنوان‌های موضوعی فارسی خلأ پژوهشی وجود دارد و پژوهشی تاکنون به جنبه‌های عملیاتی و روش پیاده‌سازی آن اشاره نداشته است. پژوهش حاضر با شناسایی چارچوب مناسب و پیاده‌سازی هم‌زمان انتشار سرعنوان‌های موضوعی فارسی در قالب داده‌های پیوندی تلاشی در راستای برطرف کردن این خلأ بود.

با توجه به روش‌شناسی به‌کار برده‌شده در این پژوهش می‌توان نتیجه گرفت سرعنوان‌های موضوعی فارسی قابلیت تبدیل به داده‌های پیوندی را برای انتشار در محیط وب دارند. مراحل پاک‌سازی در این پژوهش مشابه با پژوهش هانسون (۲۰۱۴) بود. در مرحله پیوند نیز با پژوهش هانسون (۲۰۱۴) شباهت داشت ولی با پژوهش رایان و همکاران (۲۰۱۵) تفاوت داشت. همچنین از نظر استفاده از ابزار این‌ریفاین با پژوهش کار و داس (۲۰۲۱) شباهت داشت ولی با پژوهش آپنیت و بوجار (۲۰۲۱) متفاوت بود. به دیگر بیان، پژوهش آن‌ها برای تبدیل سرعنوان‌های موضوعی از یک سلسله کد پایتون به نام Mc2skos استفاده کرد، اما استفاده از این کد در پژوهش حاضر امکان‌پذیر نبود، زیرا سرعنوان‌های پژوهش آن‌ها از مارک ۲۱ برای ماشینی‌سازی پیشینه‌ها استفاده کرده بود در حالی که مارک ایران بر پایه یونی‌مارک است. شباهت پژوهش حاضر با آپنیت و بوجار (۲۰۲۱) در استفاده از رابط کاربری اسکاسموس بود. با این حال محدودیت‌هایی نیز برای پژوهش حاضر می‌توان در نظر گرفت. در این پژوهش تنها روش‌شناسی برای انتشار سرعنوان‌های موضوعی فارسی مطرح شد و مشکلات و موانع احتمالی ناشی از اشکالات زبانی و فنی در فرایند انتشار سرعنوان‌ها مورد بررسی قرار نگرفت؛ بنابراین در پژوهش‌های


آتی می‌توان به بررسی این مشکلات در فرایند انتشار پرداخت. با توجه به دستنامه نرم‌افزار اپن‌ریفاین این نرم‌افزار قابلیت ورود اطلاعات تا دو گیگابایت را دارد. در نتیجه داده‌های سرعنوان‌های موضوعی فارسی در صورتی که بیشتر از این مقدار شوند ممکن است نیاز به تکرار مراحل تبدیل باشد. در واقع به نظر می‌رسد در صورت حجیم‌شدن داده‌های سرعنوان‌های موضوعی فارسی لازم است تا فرایند تبدیل به صورت چندمرحله‌ای شکل بگیرد. به‌طور کلی می‌توان نتیجه گرفت از آنجایی که این چارچوب، بازیابی داده‌های مستند را تسهیل می‌کند، می‌تواند برای انتشار اصطلاحات موضوعی فارسی مفید واقع شود.

ORCID


Zeinab Sabbaghi

 <https://orcid.org/0000-0001-8910-0013>

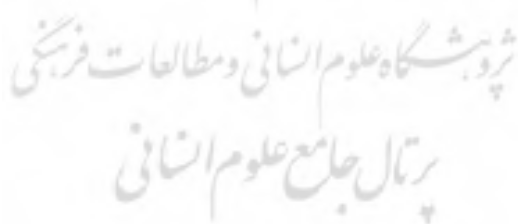
Bidgoli

 <https://orcid.org/0000-0003-4761-6761>

Atefeh Sharif

 <https://orcid.org/0000-0003-4542-7457>

Fatemeh Zandian



منابع

- بهکمال، بهشید، کاهانی، محسن، دادخواه، محبوبه، زرین کلام، فتانه و پایدار، صمد. (۱۳۹۰). ارائه چارچوبی برای انتشار مجموعه داده‌های فارسی به صورت داده‌های پیوندی روی وب. فنی مهندسی-دانشگاه آزاد مشهد، ۴(۱)، ۱-۱۹.
- حسینی، الهه. (۱۳۹۹). چهارچوب معنایی برای یکپارچه‌سازی و بازیابی معنایی اشیای محتوایی وب: رویکرد داده پیوندی در بافت سرطان. پایان‌نامه دکتری، دانشگاه الزهرا، تهران.
- سختاوتی، الهه. (۱۳۹۰). ارائه چارچوبی جهت انتشار اطلاعات کتابخانه‌ای بر پایه اصول داده‌های پیوندی. پایان‌نامه کارشناسی ارشد، دانشگاه پیام نور.
- سختاوتی، الهه، فراهی، احمد و جلالی، مهرداد. (۱۳۹۰). ارائه چارچوبی جهت انتشار اطلاعات دارو بر پایه *Linked data*. دومین همایش فناوری اطلاعات، حال، آینده. بازیابی از <https://civilica.com/doc/130594>
- شریف، عاطفه. (۱۳۹۳). پیوندهای کور، چالشی در ایده داده‌های پیوندی: واکاوی سرعنوان‌های موضوعی فارسی. *پردازش و مدیریت اطلاعات*، ۳۰(۱)، ۲۲۳-۲۴۴.
- فتحیان دستگردی، اکرم، طاهری، سید مهدی، صنعت جو، اعظم و کاهانی، محسن. (۱۳۹۹). پیاده‌سازی روش داده‌های پیوندی در نظام‌های کتابخانه‌ای: بررسی مؤلفه‌های موردنیاز و ارائه یک الگو. *کتابخانه‌های دیجیتال: پردازش و سازماندهی اطلاعات و دانش*، ۷(۲۵)، ۶۷-۹۴.

References

- Akbaridaryan, S., Khosravi, F., Ebrahimi, M. & Mahabadi, H. B. (۲۰۱۷). *SKOSification of Trilingual Cultural Thesaurus (TCH) of National Library of Iran (NLI): A step in line with NLI's Linked Data strategy*. Presented at the IFLA WLIC ۲۰۱۶ – Columbus, OH – Connections. Collaboration. Community –Retrieved from <https://library.ifla.org/id/eprint/۲۰۹۱>
- Apenīte, M. & Bojārs, U. (۲۰۲۱). National Library of Latvia Subject Headings as Linked Open Data. In R. Verborgh, A. Dimou, A. Hogan, C. d'Amato, I. Tididi, A. Bröring, ... M. Alam (Eds.). *The Semantic Web: ESWC 2021 Satellite Events* (pp. ۳۳-۳۷). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-80418-3_6
- Bushman, B., Anderson, D. & Fu, G. (۲۰۱۵). Transforming the Medical Subject Headings into Linked Data: Creating the Authorized Version of MeSH in RDF. *Journal of Library Metadata*, 15(۳-۴), ۱۵۷-۱۷۶. <https://doi.org/10.1080/19387389.2015.1099967>
- Candela, G., Escobar, P., Carrasco, R. C. & Marco-Such, M. (۲۰۲۰).

- Evaluating the quality of linked open data in digital libraries. *Journal of Information Science*, ۰۱۶۵۵۵۱۵۲۰۹۳۰۹۵. <https://doi.org/10.1177/0165551520930951>
- Eslami, S. & Vaghefzadeh, M. H. (۲۰۱۳). *Publishing Persian linked data of National library and Archive of Iran*. Retrieved ۹ November ۲۰۲۲ from <http://library.ifla.org/id/eprint/۱۹۳/>
- Goswami, S. & Biswas, P. (۲۰۱۱). THE CONCEPT OF SEMANTIC WEB IN LIBRARY SERVICES. *International Journal of Information Dissemination and Technology*, ۱(۳), ۱۶۵-۱۷۰.
- Graube, M., Pfeffer, J., Ziegler, J. & Urbas, L. (۲۰۱۲). Linked Data as Integrating Technology for Industrial Data: *International Journal of Distributed Systems and Technologies*, 3(۳), ۴۰-۵۲. <https://doi.org/10.4018/jdst.2012070104>
- Halla, M. (۲۰۱۳). Linked Data in Libraries: Library of Congress' Bibliographic Framework Transition Initiative. *Library Philosophy and Practice (e-Journal)*, ۱۰۱۵.
- Hallo, M., Luján-Mora, S. & Trujillo, J. (۲۰۱۴). Transforming Library Catalogs into Linked Data. In *ICERI2014* (pp. ۱۸۴۵-۱۸۵۳). Seville, Spain.
- Hannemann, J. & Kett, J. (۲۰۱۰). Linked Data for Libraries. Presented at the World Library and Information Congress: ۷۶th Ifla General Conference and Assembly, Gothenburg, Sweden: Ifla.
- Hanson, E. M. (۲۰۱۴). A Beginner's Guide to Creating Library Linked Data: Lessons from NCSU's Organization Name Linked Data Project. *Serials Review*, ۴۰(۴), ۲۵۱-۲۵۸. <https://doi.org/10.1080/00987913.2014.975887>
- Kamila, K. (۲۰۰۸). Application of Semantic Web in Modern Library and Information Services. *GNIMS - International E-Journal on Library Science*, ۲. Retrieved ۹ November ۲۰۲۲ from <https://gnimswebsite.sr.ap-south-1.amazonaws.com/files/2022/02/20122439/December-2014.pdf#page=7>
- Kar, S. & Das, R. (۲۰۲۰). Publishing E-resources of Digital Institutional Repository as Linked Open Data: an experimental study. *Library Philosophy and Practice (e-Journal)*, (۴۶۹۹). Retrieved ۱ December ۲۰۲۲ from https://digitalcommons.unl.edu/libphilprac/4699?utm_source=digitalcommons.unl.edu%2Flibphilprac%2F4699&utm_medium=PDF&utm_campaign=PDFCoverPages
- Kar, S. & Das, R. (۲۰۲۱). A Methodology for Transforming MARC ۲۱ Personal Name Authority Metadata into Linked Open Data with Integration of VIAF and LCNAF Datasets: An experimental study. *Library Philosophy and Practice (e-journal)*, (۵۴۵۸).
- Lampert, C. K. & Southwick, S. B. (۲۰۱۳). Leading to Linking: Introducing Linked Data to Academic Library Digital Collections. *Journal of Library Metadata*, 13(۲-۳), ۲۳۰-۲۵۳. <https://doi.org/10.1080/>

۱۹۳۸۶۳۸۹,۲۰۱۳,۸۲۶۰۹۵.

- Mulwad, V., Finin, T., Syed, Z. & Joshi, A. (۲۰۱۰). Using linked data to interpret tables. *Proceedings of the the First International Workshop on Consuming Linked Data*. co-located at ISWC ۲۰۱۰.
- Park, H. & Kipp, M. (۲۰۱۹). Library Linked Data Models: Library Data in the Semantic Web. *Cataloging & Classification Quarterly*, ۵۷(5), ۲۶۱-۲۷۷. <https://doi.org/10.1080/01639374.2019.1641171>
- Ryan, C., Grant, R., Carragáin, E. Ó., Collins, S., Decker, S. & Lopes, N. (۲۰۱۵). Linked data authority records for Irish place names. *International Journal on Digital Libraries*, 15(۲-۴), ۷۳-۸۵. <https://doi.org/10.1007/s00799-014-0129-8>
- Southwick, S. B. (۲۰۱۵). A Guide for Transforming Digital Collections Metadata into Linked Data Using Open Source Technologies. *Journal of LibraryMetadata*, ۱۵(۱), ۱-۳۵. <https://doi.org/10.1080/19386389.2015.1007009>
- Summers, E., Isaac, A., Redding, C. & Krech, D. (۲۰۰۸, January). *LCSH, SKOS and Linked Data*.
- Suominen, O., Ylikotila, H., Sini, P., Mikko, L. & Frosterus, M. (۲۰۱۵). Publishing SKOS vocabularies with Skosmos. *Manuscript Submitted for Review*. Retrieved from <https://skosmos.org/publishing-skos-vocabularies-with-skosmos.pdf>
- Tian, C. T., Cole, T. W. & Yu, K. (۲۰۲۱). Name and Subject Heading Reconciliation to Linked Open Data Authorities using Virtual International Authority File and Library of Congress Linked Data Service APIs: A Case Study featuring Emblematica Online. *Library Resources & Technical Services (ALA LRTS)*. Retrieved from https://scholarship.law.nd.edu/law_faculty_scholarship/1467
- Zengenene, D., Casarosa, V. & Meghini, C. (۲۰۱۴). Towards a Methodology for Publishing Library Linked Data. In T. Catarci, N. Ferro, & A. Poggi (Eds.). *Bridging Between Cultural Heritage Institutions*. IRCDL ۲۰۱۳. Communications in Computer and Information Science, vol ۳۸۵. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04347-0_10

References [In Persian]

- Behkamal, B., Kahani, M., Dadkhah, M., Zarin Kalam, F. & Paydar, S. (۲۰۱۱). Providing a framework for publishing Persian datasets as linked data on the web. *Technical and Engineering Mashhad Azad University*, ۴(۱), ۱-۱۹. [In Persian]
- Fathian Dastgerdi, A., Taheri, S. M., Sanatjo, A. & Kahani, M. (۲۰۱۹). *implementing the linked data method in library systems: examining the required components and presenting a model*. Digital libraries: processing and organizing information and knowledge. [In Persian]
- Hosseini, E. (۲۰۱۹). A Semantic Framework for Semantic Integration and Retrieval of Web Content Objects: A Linked Data Approach in Cancer Tissue. Ph.D Dissertation, Al-Zahra University, Tehran. [In

Persian]

Sekhavati, E. (۲۰۱۱). *Providing a framework for publishing library information based on the principles of linked data* (Master's thesis, Payam-e Noor university). [In Persian]

Sekhavati, E., Farahi, A. & Jalali, M. (۲۰۱۱). *Implementing a Framework for Publishing Medicine Information Based on Linked Data*. ۲nd Information Technology, Present, Future. Retrieved from <https://civilica.com/doc/۱۳۰۵۹۴> [In Persian]

Sharif, A. (۲۰۱۳). Blind links, a challenge in the idea of linked data: analysis of Persian subject headings. *Information processing and management*, ۳۰(۱), ۲۲۳-۲۴۴. [In Persian]



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

استناد به این مقاله: صباغی بیدگلی، زینب، شریف، عاطفه و زندیان، فاطمه. (۱۴۰۲). چارچوبی برای تبدیل و انتشار سرعنوان‌های موضوعی فارسی به داده‌های پیوندی. *فصلنامه علمی بازیابی دانش و نظام‌های معنایی*، ۱۰(۳۷)، ۱-۳۰. DOI: ۱۰,۲۲۰۵۴/JKS.۲۰۲۳,۷۲۵۳۸,۱۵۶۵



Name of Journal is licensed under a Creative Commons Attribution-NonCommercial ۴,۰ International License.