

# پیوندهای کور، چالشی در ایده داده‌های پیوندی: واکاوی سرعنوان‌های موضوعی فارسی

عاطفه شریف

دکتری علم اطلاعات و دانش‌شناسی؛ استادیار؛ گروه علم اطلاعات و دانش‌شناسی؛ دانشگاه  
فردوسی مشهد  
atefehsharif@gmail.com

مقاله پژوهشی

دریافت: ۱۳۹۲/۱۱/۰۵

پذیرش: ۱۳۹۳/۰۱/۲۰

دوره ۳۰ شماره ۱

صص. ۲۲۳-۲۴۴

دانشگاه  
فردوسی مشهد

پژوهشنامه پردازش و مدیریت اطلاعات

فصلنامه علمی پژوهشی

شاپا (چاپی) ۸۲۲۳-۲۲۵۱

شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱

نمایه در LISA، ISC و Scopus

http://jipm.irandoc.ac.ir

پژوهشگاه علوم و فناوری اطلاعات ایران

**چکیده:** در این نوشتار ضمن طرح ایده داده‌های پیوندی، رویکرد انتشار، اشتراک، و برقراری پیوند میان پاره‌های داده، اطلاعات و دانش در وب معنایی، بخشی از اشکالات احتمالی در تبدیل رکوردهای مستند سرعنوان‌های موضوعی فارسی به داده‌های پیوندی با روش پیمایشی شناسایی شد. ۱۱۲۳۳ رکورد مستند موضوع فارسی در سه نظام - فهرست پیوسته کتابخانه ملی ایران، فهرست پیوسته کتابخانه کنگره، و گنجینه کتاب‌های نوسا مورد کاوش قرار گرفت و چگونگی برقراری پیوند میان موضوعات فارسی و انگلیسی در ۹۵۱۹ رکورد مشترک در فارسی و کنگره ثبت گردید. نتایج، نشانگر آن است که برقراری پیوند میان موضوعات فارسی و انگلیسی در ۲۰ درصد رکوردها با خطا مواجه می‌شود. بیشترین بروز خطا مربوط به نامشخص بودن نظام موضوعی (۷/۶ درصد) در پایگاه مستند موضوعی کتابخانه ملی ایران است. پیشنهاد می‌شود جهت تبدیل صحیح و فراهم‌ساختن امکان پیوند میان موضوعات فارسی و انگلیسی، قبل از تبدیل رکوردهای مستند موضوع به داده‌های پیوندی، خطاهای موجود در جریان پیش‌پردازش‌ها شناسایی و برطرف گردد.

**کلیدواژه‌ها:** داده‌های پیوندی؛ وب معنایی؛ سرعنوان‌های موضوعی فارسی؛ سرعنوان‌های موضوعی کنگره؛ گنجینه کتاب‌های نوسا

\* این مقاله در نخستین کنفرانس کتابخانه دیجیتال (بهمن ۱۳۹۲) به‌عنوان مقاله برتر برگزیده شده است.

## ۱. مقدمه

نظام مبتنی بر واژگان کنترل شده، به سیاهه‌ای اصلی از عبارات وابسته است که می‌توان آنها را به مدارک تخصیص داد و در مورد اغلب فهرست‌های کتابخانه‌ای دستی یا پیوسته، این عبارات به نام سرعنوان موضوعی خوانده می‌شود. سیاهه سرعنوان‌های موضوعی یا اصطلاحنامه، حاوی عباراتی برای دستیابی موضوعی است که می‌بایست در عملیات فهرست‌نویسی یا نمایه‌سازی به کار روند (چان<sup>۱</sup>، ۱۹۸۱، ۱۵۸).

فهرست سرعنوان‌های موضوعی به منزله یکی از نظام‌های سازماندهی دانش و با توجه به پشتوانه انتشاراتی در زبان‌های مختلف تدوین می‌شود و توسعه می‌یابد. این نظام، همچون سایر نظام‌های مستند-به‌طور مثال اصطلاحنامه‌ها- نیاز به نگهداری و نظارت مداوم دارد. چان، این نظام کنترلی را نظام مستند موضوعی می‌نامد. نظامی که در آن تصمیمات اتخاذ شده پیرامون گزینش عبارت‌ها و پیوندهایی که میان آنها وجود دارد، با استناد به مدارک بیان می‌شود. هر سرعنوان موضوعی به محض ساخته شدن، جهت تضمین یکپارچگی و یکدستی در واژگان سرعنوان‌های موضوعی و ارجاعات متقابل، در نظام مستند موضوعی ثبت می‌گردد (چان، ۱۹۸۱، ۱۵۹).

از زمان ظهور فهرست سرعنوان‌های موضوعی کنگره در سال‌های پایانی قرن ۱۹، این نظام به منزله ابزاری جهت فراهم‌آوری نقاط دسترسی موضوعی برای مجموعه‌های سنتی کتابخانه، از سوی جامعه کتابداری و اطلاع‌رسانی مورد استفاده گسترده‌ای قرار گرفته است. فهرست سرعنوان‌های موضوعی کنگره، واژگانی موضوعی است که تمامی رشته‌ها را تحت پوشش قرار می‌دهد و بزرگترین زبان نمایه‌سازی عمومی در زبان انگلیسی محسوب می‌شود (Yi and Chan 2010).

در زبان فارسی و در چهار دهه گذشته، فهرست سرعنوان‌های موضوعی فارسی با مبنای قرارداد فهرست سرعنوان‌های موضوعی کنگره توسعه یافته است و هر روز بر حجم پیشینه‌های آن افزوده می‌شود. این فهرست، به منزله ابزاری استاندارد و ملی جهت بازنمایی موضوعی کتاب‌های فارسی از سوی کتابخانه ملی ایران و سایر کتابخانه‌ها در ایران مورد پذیرش قرار گرفته است و استفاده می‌شود. جستجوی موضوعی در فهرست‌های پیوسته و

1. Chan

نرم‌افزارهای کتابخانه‌های دیجیتال بر پایه همین موضوعات مستندسازی شده انجام می‌گیرد.

به‌منظور کاهش مشکلاتی چون محدودیت دامنه معنایی و انعطاف‌پذیری اندک نظام‌های ذخیره و بازیابی، همچنین افزایش قابلیت‌های استنتاج در نظام‌ها، حرکت‌هایی به سمت نظام‌های جدید بازنمون دانش در جریان است. این حرکت‌ها که بیشتر در حوزه هوش مصنوعی مشاهده می‌شود، با ظهور کتابخانه‌های دیجیتال و اندیشه وب معنایی رنگ دیگری به خود گرفته است.

در حالی که وب کنونی برای استفاده مستقیم انسان طراحی شده، وب معنایی در صدد خلق فضایی است که در آن رایانه‌ها قادر به پردازش و درک محتوای اطلاعات صفحات وب باشند. برنرزلی و همکارانش وب معنایی را گسترش وب کنونی می‌دانند، به طوری که در آن، معنای اطلاعات تعریف می‌شود و تعامل میان انسان و رایانه فراهم می‌آید (Berners-Lee, Hendler and Lassila 2001). وب معنایی برای رسیدن به این هدف، نیازمند فراهم ساختن دو مجموعه است: اول مجموعه‌ای از داده‌های ساختاریافته<sup>۱</sup> [دانش بازنمون شده در قالب هستی‌شناسی‌ها] و دوم مجموعه‌ای از قواعد استنتاج<sup>۲</sup> که جهت پردازش و استنتاج خودکار به کار رود (شریف ۱۳۸۶). در وب معنایی لازم است شبکه‌ای از اسناد (که در وب سنتی وجود داشت) به شبکه‌ای از داده‌ها تبدیل شود. ضرورت ایجاد این ارتباط معنادار بین منابع موجود در وب باعث به وجود آمدن مفهومی جدید به نام داده‌های پیوندی<sup>۳</sup> شده است (Berners-Lee 2006) نقل در بهکمال، کاهانی، دادخواه، زرین کلام، پایدار (۱۳۹۰).

پژوهش‌های صورت گرفته در زمینه هستی‌شناسی<sup>۴</sup> -ها- که از ابزارهای جدید بازنمون دانش به‌شمار می‌روند- از جمله این حرکت‌هاست (شریف ۱۳۸۵، ۶۷). هستی‌شناسی‌ها به منزله ابزار بازنمون دانش در نظام‌های ذخیره و بازیابی مورد استفاده قرار می‌گیرند (Weng, Tsai, Liu and Hsu 2006) و آن را مجموعه‌ای از مفاهیم، خصیصه‌ها<sup>۵</sup>، و روابط

- 
1. structured data
  2. inference rules
  3. linked data
  4. ontology
  5. properties

میان آن مفاهیم تعریف کرده‌اند (Hess and Schlieder 2006, 545).

ساخت و توسعه ابزارهای بازنمون دانش از جمله هستی‌شناسی‌ها، که یکی از لایه‌های اساسی وب معنایی به‌شمار می‌روند، فرایندی زمان‌بر، پرهزینه، و پیچیده است<sup>۱</sup>. علاوه بر موضوع مهندسی و انتشار هستی‌شناسی‌ها، اهمیت توجه به لایه اعتماد<sup>۲</sup> در وب معنایی و استفاده مجدد از داده‌های موجود (Hannemann and Kett 2010) شکل‌گیری پژوهش‌های فراوانی را در فضای پژوهشی وب معنایی رقم زده است. کدگذاری سرعنوان‌های موضوعی کنگره با استاندارد نظام ساده سازماندهی دانش (اسکاس)<sup>۳</sup> و سایر نظام‌های مستند کتابخانه‌ای از جمله اصطلاحنامه‌ها و انتشار آنها به‌گونه‌ای که به‌عنوان داده‌های پیوندی<sup>۴</sup> در فضای ابر داده‌های پیوندی<sup>۵</sup> وب<sup>۶</sup> کاربرد یابد، از آن جمله است:

"Harper 2006; Ed, Antoine, Clay and Dan 2008; Tillett and Harper 2007; Summers, Isaac, Redding and Krech 2008; Cantara 2006; Malmsten 2009; Hannemann and Kett 2010; Neubert 2009; Fripp 2010; Styles, Ayers and Shabir 2008."

کاوش و بازیابی داده‌های ساخت‌یافته در فهرست‌های کتابخانه‌ای به پروتکل‌هایی چون Z39.50 محدود است و در فضای نرم‌افزارهای کتابخانه‌ای باقی می‌ماند؛ حال آنکه، همین اطلاعات موجود می‌تواند بخشی از جنبشی بزرگ‌تر باشد و به دنیای خارج از کتابخانه‌ها گسترش یابد. اشتراک داده‌ها و استفاده مجدد از آنها درهای کتابخانه‌ها را، که با جهان کتابشناختی مأنوسند، به روی سایر متخصصان باز خواهد کرد و این جهان را به جهان داده‌های پیوندی مرتبط خواهد ساخت (Malmsten 2009).

۱. برای آشنایی با روش‌های ساخت هستی‌شناسی‌ها و فرایند مهندسی و توسعه آن رجوع شود به شریف، عاطفه (۱۳۸۸). مهندسی خودکار هستی‌شناسی: امکان‌سنجی استخراج روابط معنایی از متون فارسی و تعیین میزان پیدایی آنها. کتابداری و اطلاع‌رسانی ۱۲ (۲): ۲۴۳-۲۶۲.

2. trust

۳. SKOS: Simple Knowledge Organization System یکی از استانداردهای مطرح‌شده از سوی کنسرسیوم وب جهانی می‌باشد و در کدگذاری نظام‌های دانش مانند طرح‌های رده‌بندی، سرعنوان‌های موضوعی و اصطلاحنامه‌ها کاربرد دارد. این استاندارد زبانی به منزله یکی از استانداردهای وب معنایی در کدگذاری هستی‌شناسی‌ها نیز شناخته می‌شود (Miles and Perez-Aguera 2007, 70).

۴. Linked data حاصل این طرح در سایت رسمی کتابخانه کنگره (<http://id.loc.gov>) منتشر شده است.

5. LOD (Linked Open Data) cloud

6. Web 3.0

قدرت داده‌های پیوندی را در انتشار و پیوند داده‌ها به یکدیگر می‌دانند ( Fripp 2010)، به گونه‌ای که علاوه بر پیوند میان داده‌ها در درون یک پایگاه مشخص، برای مثال پیوند میان رکوردها در یک پایگاه اطلاعات کتابشناختی، می‌توان پیوندهایی متنوع میان داده‌های منتشر شده در فضای ابر داده‌های پیوندی برقرار ساخت. به‌طور مثال، چنانچه نویسنده‌ای دارای اطلاعاتی در پرتال دانشگاه، صفحه‌ای در ویکی‌پدیا، و آثاری در فهرست‌های پیوسته کتابخانه باشد، چنانچه این اطلاعات به‌صورت داده‌های پیوندی به انتشار رسیده باشد، می‌توان از طریق نام وی - یعنی مؤلفه مشترک میان این سه منبع - نتایج کاوش را گسترش داد.

به این ترتیب، چنانچه کتابخانه قصد حضور در فضای ابر داده‌های پیوندی را داشته باشد، لازم است به بازنشر داده‌های معتبر و مستند خود بپردازد. اتفاقی که در این فضای پژوهشی رخ می‌دهد، تبدیل و نگاشت<sup>۱</sup> میان دو نظام کدگذاری و بازنشر داده‌های موجود و معتبر کتابخانه‌ها در فضای وب است؛ به گونه‌ای که استفاده مجدد و امکان برقراری پیوندهای داخلی<sup>۲</sup> و خارجی<sup>۳</sup> میان انواع داده‌ها در مجموعه‌های داده‌ای<sup>۴</sup> مختلف فراهم آید. چنین رویکردی در پروژه<sup>۵</sup> VIAF در حال اجراست و بر مستندسازی نام‌ها متمرکز شده است. ضمن آنکه رکوردهای مستند موضوعی کتابخانه کنگره و چند کتابخانه دیگر در این فضا به انتشار رسیده است.<sup>۶</sup>

این پیوند نه تنها به سود متخصصان سایر حوزه‌هاست و به بازنشر و استفاده مجدد داده‌های معتبر و ساخت یافته می‌انجامد، بلکه سرانجام، بهبود نظام ذخیره و بازیابی و رضایت کاربران نظام‌های کتابخانه‌ای را در پی دارد. چنانچه در نرم‌افزارهای کتابخانه‌ای حرکت به سمت وسوسه کاربرد داده‌های پیوندی باشد، کاربرانی که کاوش را از موتورهای کاوش و وب آغاز می‌کنند، به نظام کتابخانه‌ای (مانند اپک، کتابخانه دیجیتال

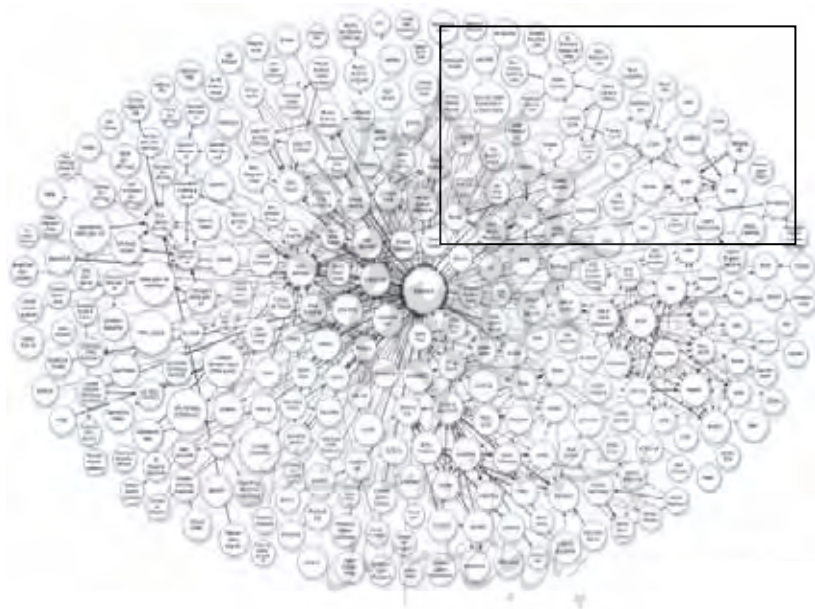
- 
1. mapping
  2. internal links
  3. external links
  4. data sets
  5. Virtual International Authority File (<http://viaf.org>)

۶. اطلاعات بیشتر در این زمینه در

<http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets> موجود

است.

و ... هدایت خواهند شد و اگر نقطه آغاز کاوش در نظام کتابخانه‌ای باشد، به دانش وسیع‌تری که در فضای ابر داده‌های پیوندی منتشر شده، دسترسی خواهند یافت. تبدیل رکوردهای مارک به قالب‌های هم‌خوان با استانداردهای وب معنایی اولین گام‌ها در این زمینه است (Harper 2006). در حال حاضر فهرست سرعنوان‌های موضوعی کنگره در قالب اسکاس در فضای ابر داده‌ها منتشر شده و در سایت کتابخانه کنگره آمریکا نیز (<http://id.loc.gov>) قابل جستجو است. تصویر شماره ۱ و ۲ شمایی از ابر داده‌های پیوندی است. فهرست سرعنوان‌های موضوعی کنگره به‌عنوان یکی از مجموعه داده‌های منتشر شده قابل مشاهده است.



تصویر ۱. شمایی کلی از ابر داده‌های پیوندی در سال ۲۰۱۳ از سایت رسمی <http://linkeddata.org>



قابلیت‌های کاوش، به‌ویژه در جستجوهای بین‌زبانی<sup>۱</sup> بیفزاید. افزودن بر قابلیت‌های کاوش و برقراری پیوند میان انواع داده‌های مرتبط، در فضای ابر داده‌ها - چه پیوندهای داخلی در درون کتابخانه و چه میان داده‌های کتابخانه با سایر مجموعه داده‌های منتشرشده در این ابر- از اهداف داده‌های پیوندی است. یکی از امکانات نرم‌افزاری ایران یعنی گنجینه کتاب‌های نوسا<sup>۲</sup> با برقراری پیوند میان دو مجموعه داده‌های مستند فهرست سرعنوان‌های موضوعی فارسی و کنگره<sup>۳</sup> امکان کاوش در دو مجموعه داده را فراهم آورده است. سازوکار کاوش در گنجینه کتاب‌های نوسا با ایده داده‌های پیوندی فاصله دارد<sup>۴</sup> و به نظر می‌رسد<sup>۵</sup> تنها بر پایه تطابق رشته‌های حرفی<sup>۶</sup> انجام پذیرفته است.

#### 1. cross-language

۲. در وبسایت رسمی شرکت در توضیح این خدمت نرم‌افزاری، که آن را نسل سوم نرم‌افزارهای شرکت نامیده است، آمده: هر کتابخانه یا مرکز اسناد مجهز به سیمرغ در واقع یک میزبان در این شبکه است که کاربران با اتصال به هر کدام از این میزبان‌ها می‌توانند در کل شبکه سیمرغ به‌طور هم‌زمان به دنبال کتاب یا سایر منابع کتابخانه‌ای خود بگردند. این شبکه که بیش از ۱۰ سال از عمر آن می‌گذرد، در حال حاضر بزرگترین و مهم‌ترین شبکه جستجوی منابع کتابخانه‌ای در ایران است و دارای بیش از ۸۰ میزبان با موجودی بیش از ۱۰ میلیون جلد انواع کتاب و سایر منابع کتابخانه‌ای است. <http://www.nosa.com/Nosaweb/Products/Simorgh/InnerPages.aspx?PageId=375> گنجینه نوسا ایده‌ای مشابه با Google Books است و با هدف سازماندهی و ارائه متن کامل کتاب‌های فارسی شکل گرفته است.

۳. شرکت نوسا برای این محصول نرم‌افزاری خود از اطلاعات پایگاه‌های مستند موضوع سازمان اسناد و کتابخانه ملی ایران و داده‌های مستند موضوع کتابخانه کنگره آمریکا استفاده می‌کند. در مورد موضوعات فارسی پیوندی به وبسایت کتابخانه ملی ارائه نمی‌شود، اما در مورد موضوعات معادل لاتین پیوندی به بخش مستند موضوع وبسایت کنگره (<http://id.loc.gov/authorities/subjects.html>) که به انتشار مستندا در قالب داده‌های پیوندی پرداخته است، برقرار می‌باشد.

۴. دلیل این ادعا نبود شواهدی از نگاشت سرعنوان‌های موضوعی فارسی در قالب زبان‌های وب معنایی از جمله اسکاس و عدم انتشار رسمی آنها در ابر داده‌های پیوندی در <http://linkeddata.org> است. ۵. اطلاعاتی پیرامون چگونگی عملکرد گنجینه کتاب‌های نوسا در سایت رسمی شرکت منتشر نشده است.

#### 6. String matching



در جریان کاوش‌ها مشاهده می‌شود که در برخی موارد، اگرچه سرعنوان موضوعی فارسی ما به‌ازایی در فهرست سرعنوان‌های کنگره دارد، پیوند میان این دو ابزار زبانی در گنجینه کتاب‌های نوسا برقرار نشده است. چنانچه کتابخانه ملی قصد هم‌سویی با سایر کتابخانه‌های ملی جهان و انتشار داده‌های خود را داشته باشد، لازم است در باب چالش‌های تبدیل و انتشار بیندیشد. انتشار مجموعه داده‌ها زمانی به هدف خود می‌رسد که پیوند میان داده‌ها به‌خوبی برقرار شود. به‌طور مثال، اگر دو مجموعه موضوعات فارسی و انگلیسی در ابر داده‌ها منتشر شده باشد، آن هنگام که پیوند میان این دو مجموعه در سطح موضوعات - پیوند موضوع به زبان فارسی با موضوع به زبان انگلیسی - برقرار شود، نظام به هدف خود دست یافته است. روش‌های مختلف و مکملی جهت برقراری پیوند میان URI‌ها وجود دارد. بخش قابل توجهی از این فعالیت از طریق تطابق زنجیره‌های حرفی انجام می‌پذیرد. به این ترتیب، به نظر می‌رسد اشکالاتی که در هنگام کاوش در گنجینه نوسا در برقراری پیوند میان موضوعات فارسی و انگلیسی رخ می‌دهد، می‌تواند فرایند تطابق رشته‌های حرفی را در مورد داده‌های مستند موضوع کتابخانه ملی ایران نیز - در صورت انتشار - در پی داشته باشد. در این نوشتار از گنجینه کتاب‌های نوسا به‌عنوان ابزاری جهت شناسایی اشکالات احتمالی در برقراری پیوند میان موضوعات فارسی و انگلیسی استفاده می‌شود. پژوهش در پی پاسخ‌گویی به دو پرسش است:

- این داده می‌تواند نویسنده‌ای مشخص، محل جغرافیایی، موضوع و یا منبعی در چند مجموعه داده متفاوت باشد که با اختصاص URI منحصر به فرد منتشر شده است. در مرحله شناسایی پیوندها، میان دو URI مشابه، رابطه SameAs شکل می‌گیرد و از این طریق می‌توان به ایده داده‌های پیوندی تحقق بخشید.
- هر موجودیتی (داده‌ای)، مانند نام فرد، نام کتاب، محل جغرافیایی و ...، در فضای داده‌های پیوندی دارای نشانگری منحصر به فرد است که URI نام دارد. برای مثال، مقاله‌ای با عنوان «سیستم تشخیص نفوذ مبتنی بر پروتکل‌های فزایی از طریق تکنیک داده‌کاوی» با یوآرآی <http://wtlab.um.ac.ir/linkedata/papers/1005161>، از مجموعه انتشارات یکی از اساتید دانشگاه فردوسی مشهد یعنی «دکتر محسن کاهانی» با یوآرآی <http://wtlab.um.ac.ir/linkedata/profs/kahani> است که در مجموعه داده‌های پیوندی دانشگاه فردوسی مشهد منتشر شده است. برای کسب اطلاعات بیشتر پیرامون داده‌های پیوندی دانشگاه فردوسی مشهد (FUM-LD) بهتر است به مقاله به‌کمال و همکاران (۱۳۹۰) مراجعه شود.

۱. درصد برقراری پیوندهای صحیح<sup>۱</sup> میان دو ابزار سرعنوان‌های موضوعی فارسی و کنگره در گنجینه کتاب‌های نوسا چه قدر است؟
۲. چه عواملی باعث عدم برقراری پیوند میان این دو ابزار بازنمون در گنجینه کتاب‌های نوسا می‌شود؟

### ۳. ضرورت پژوهش

پاسخ‌گویی به دو پرسش پژوهش از آن جهت اهمیت دارد که علاوه بر آشکارشدن بخشی از اشکالات موجود در رکوردهای مستند موضوع فارسی، چالش‌های احتمالی پیش رو در راه انتشار داده‌های پیوندی را تا اندازه‌ای آشکار می‌سازد. یکی از بزرگترین چالش‌های وب معنایی تأمین لایه اعتماد است. با توجه به آغاز حرکت‌های معنایی در ایران، که نشانه‌هایی از آن قابل مشاهده است (مانند Eslami and Vaghefzadeh 2013)، و پذیرفتن این پیش‌فرض که رکوردهای موجود در کتابخانه‌ها و مراکز اطلاع‌رسانی در طی سالیان دراز توسط تعداد قابل توجهی از کتابداران و موضوع‌سازان متخصص توسعه یافته، مورد بازنگری قرار گرفته، و مجموعه‌ای ارزشمند فراهم آورده است، این نتیجه حاصل می‌شود که می‌توان در تأمین لایه اعتماد وب معنایی از این داده‌ها بهره برد. به این ترتیب، شناسایی اشکالات احتمالی در تبدیل قالب‌ها، انجام پیش‌پردازش‌ها و رفع اشکالات، پیش از اقدام به تبدیل و انتشار داده‌های پیوندی ضرورت می‌یابد.

پژوهش حاضر اگرچه به انتشار داده‌های پیوندی نپرداخته است، بر پایه داده‌های پیوندی موجود در کتابخانه کنگره (مستند موضوع) و پیوندی که گنجینه کتاب‌های نوسا میان این داده‌ها با رکوردهای مستند موضوعی فارسی برقرار ساخته، بخشی از اشکالات احتمالی در زمان تبدیل داده‌های کتابشناختی به داده‌های پیوندی و برقراری پیوندها را آشکار خواهد ساخت.

### ۴. مروری بر پژوهش‌های پیشین

مقالات مروری و همچنین پژوهش‌هایی که به ضرورت انتشار داده‌های پیوندی در

۱. منظور از پیوند صحیح، برقراری پیوند میان موضوع فارسی و انگلیسی به شرط اشتراک دو سرعنوان موضوعی در آن موضوع است.

کتابخانه‌ها پرداخته‌اند، اندک نیست. در این پژوهش‌ها مدل‌های مفهومی، چالش‌های حقوقی، چالش‌های فنی، نوع داده‌های انتخابی و هستی‌شناسی‌های مناسب مورد بحث قرار گرفته و گاهی پیش‌فرض‌ها و ضرورت‌های این انتقال بیان شده است (مانند Hannemann and Kett 2010; Tillett and Harper 2007; Yee 2009; Heath and Bizer 2011; Bizer, Heath and Berners-Lee 2009). البته باید اشاره نمود که مقاله‌های فارسی در این زمینه بسیار نادر هستند و هنوز پژوهشگران ایرانی به‌طور جدی وارد این قلمرو نشده‌اند.

اما انتشار و دسترس‌پذیر ساختن داده‌های کتابشناختی به‌صورت داده‌های پیوندی، بیش از هر چیز، موضوعی فنی است و نوعی انتقال داده محسوب می‌شود. به همین جهت است که بیشتر پژوهش‌های صورت گرفته در این حوزه به تبدیل داده‌ها، بیان استانداردها، رویکردها و گاه ابزارهای تبدیل پرداخته‌اند (Bowen 2010). در برخی از آنها تنها تبدیل بخشی از داده‌های کتابشناختی مورد توجه است؛ برای مثال، تبدیل سرعنوان‌های موضوعی از قالب مارک ایکس‌ام‌ال<sup>۱</sup> (MARC XML) به اسکاس (SKOS) (Summers, Isaac, ) (Redding and Krech 2008; Harper 2006; Liang and Sini 2006; Morshed, Caracciolo, ) (Johannsen and Keizer 2011). در برخی دیگر از پژوهش‌ها، گستره وسیع‌تری در نظر گرفته شده است که طرح‌هایی چون پیاده‌سازی داده‌های پیوندی در کتابخانه سلطنتی سوئد با نام LIBRIS (Malmsten 2009) و طرح تبدیل مارک ۲۱ به آردی‌اف با استفاده از هستی‌شناسی‌های FOAF، MarcOnt، و SKOS (Styles, Ayers and Shabir 2008) از آن جمله‌اند. حاصل این پژوهش‌ها ترسیم چارچوبی تقریباً مشابه و پیاده‌سازی آن در قالب سه نگاهت است؛ ۱. نگاهت رکوردهای کتابشناختی از انواع قالب‌های مارک از جمله مارک ۲۱ به آردی‌اف (RDF)، ۲. نگاهت مستند نام‌ها به هستی‌شناسی FOAF، و ۳. نگاهت مستند موضوعات به SKOS.

در داخل، پژوهش‌های حوزه وب معنایی در کتابخانه‌ها و مراکز اطلاع‌رسانی بیش از آنکه پژوهشی باشد، مروری است و تنها پژوهشی که به گزارش مراحل انتشار رکوردهای کتابخانه ملی ایران به شکل داده‌های پیوندی می‌پردازد، به‌تازگی منتشر شده است (Eslami

۱. انواع قالب‌های مارک با اندک تفاوتی در مقالات مورد اشاره قرار گرفته است.

and Vaghefzadeh 2013). در حال حاضر کتابخانه‌های ملی سوئد<sup>۱</sup>، فرانسه<sup>۲</sup>، کتابخانه سلطنتی انگلستان<sup>۳</sup>، کتابخانه ملی ژاپن<sup>۴</sup>، کتابخانه کنگره آمریکا<sup>۵</sup> از جمله کتابخانه‌هایی هستند که به انتشار داده‌های پیوندی پرداخته‌اند و کتابخانه ملی ایران نیز این حرکت را آغاز کرده و در ابتدای راه تبدیل است.

مرور پژوهش‌هایی که ذکر شد، آشکار می‌سازد که حرکت‌هایی در جهت هم‌نوایی با تغییرات و هم‌سویی با تحولات نظام‌های ذخیره و بازیابی اطلاعات در حال آغاز است. حوزه وب معنایی در راستای علایق متخصصان حوزه کتابداری و نویدبخش نگاه معنایی به مقوله بازیابی اطلاعات است. مجموع پژوهش‌های صورت گرفته در حوزه وب معنایی و حوزه کتابداری مشخص می‌سازد که این حوزه پژوهشی و عملیاتی، مورد توجه متخصصان کتابداری و اطلاع‌رسانی قرار گرفته است و آنچه بیش از هر چیزی در جریان است، فرایندهای تبدیل داده‌های موجود به داده‌هایی است که در طرح عظیم وب معنایی قابلیت استفاده خواهد داشت.

## ۵. روش‌شناسی پژوهش

پاسخ‌گویی به پرسش‌های پژوهش به روش پیمایشی و از طریق کاوش در سه نظام فهرست پیوسته کتابخانه ملی ایران<sup>۶</sup>، فهرست پیوسته کتابخانه ملی کنگره<sup>۷</sup> و گنجینه کتاب‌های نوسا<sup>۸</sup> انجام پذیرفت.

جامعه آماری پژوهش تمامی رکوردهای موجود در فهرست پیوسته کتابخانه ملی ایران است. از آن میان مجموعه‌ای ۱۱۲۳۳ رکوردی از سرعنوان‌های موضوعی فارسی که از سرعنوان‌های موضوعی فارسی چاپی (۱۳۸۱) و پیوست‌های آن (۱۳۸۲ و ۱۳۸۵) استخراج شده به‌عنوان نمونه مورد بررسی قرار گرفت<sup>۹</sup>. درستی برقراری پیوند میان

1. <http://libris.kb.se/?language=en>

2. <http://data.bnf.fr>

3. <http://bnb.data.bl.uk>

4. <http://datahub.io/dataset/national-diet-library-authorities/> <http://id.ndl.go.jp/auth/ndla>

5. <http://id.loc.gov>

6. <http://opac.nlai.ir/opac-prod/index.jsp>

7. <http://id.loc.gov>

8. <http://www.nosabooks.com/WebUI/Home.aspx>

۹. این تعداد سرعنوان موضوعی پس از حذف ۵۳۲۹ سرعنوان غیرموضوعی (مانند نام شخص، نام خاندان،

سرعنوان‌هایی مد نظر است که در هر دو فهرست مستند فارسی و کنگره موجود باشد. لذا در بررسی پیوند میان دو ابزار فارسی و کنگره، آن دسته از موضوع‌هایی که در مستند فارسی ساخته شده و فاقد مابه‌ازایی در کنگره است، از شمارش‌ها حذف شد.

داده‌ها در کتابچه طراحی شده در نرم‌افزار اکسل وارد شد. به‌منظور جلوگیری از بروز خطا در گردآوری داده‌ها، هر سرعنوان موضوعی فارسی ابتدا در فهرست کتابخانه ملی مورد کاوش قرار گرفت و معادل انگلیسی آن در کتابچه اکسل ثبت شد. سپس همان سرعنوان در گنجینه کتاب‌های نوسا مورد کاوش قرار گرفت و وجود یا نبود پیوند میان سرعنوان‌های موضوعی فارسی و کنگره مشخص گردید. بروز خطا دقیقاً زمانی اتفاق می‌افتد که معادل انگلیسی سرعنوان فارسی در کنگره وجود داشته باشد و به‌علت خطاهای سیستمی بازیابی نشود و پیوند میان این دو ابزار زبانی برقرار نشده باشد. جهت شناسایی این شکل از خطاها، که پرسش دوم پژوهش است، در هر مورد از سرعنوان‌ها، معادل انگلیسی ثبت شده برای هر یک از سرعنوان‌ها در فهرست فارسی، در فهرست پیوسته کنگره مورد کاوش قرار گرفت. عدم بازیابی عین عبارت استخراج شده از سرعنوان‌های فارسی، دلیل بر نبود آن قلمداد نشد و در صورت عدم بازیابی، مجدداً بخشی از سرعنوان نیز در فهرست پیوسته کنگره کاوش و با مرور سرعنوان‌های کنگره بازیابی شده نسبت به شناسایی آن اقدام گردید. در برخی موارد برای یافتن املائی صحیح عبارت‌های انگلیسی ثبت شده در فهرست پیوسته کتابخانه ملی ایران، کاوشی در موتور کاوش گوگل انجام شد و از امکان تصحیح املائی آن استفاده گردید. پس از حدس شکل صحیح عبارت‌ها، بار دیگر کاوشی در فهرست پیوسته کنگره انجام پذیرفت تا شکل صحیح سرعنوان‌ها شناسایی و ثبت شود. با این روند، امکان بروز اشکال و ندیده‌شدن سرعنوان‌ها در دو ابزار زبانی فارسی و انگلیسی تا حد ممکن کاهش می‌یابد و اشکالات مورد شناسایی قرار می‌گیرد. در مواردی

---

نام جغرافیایی، تالگان، عنوان قراردادی و ...) و برخی سرعنوان‌های دارای تقسیم فرعی شامل تمام مدخل‌های فهرست‌های مستند چاپی است. تعداد رکوردهای مستند موضوعی فارسی در فهرست مستند بنا به اطلاعات دریافتی از کتابخانه ملی ایران در ابتدای سال ۱۳۹۰ حدود ۸۴۶۰۰ رکورد بوده است. به این ترتیب، نمونه ۱۱۲۳۳ رکوردی، نمونه‌ای قابل قبول در پاسخ‌گویی به پرسش‌های پژوهش به نظر می‌رسد و حدود ۱۲ درصد کل مستندهای موضوعی موجود در پایگاه مستند موضوع کتابخانه ملی ایران می‌باشد.

که پیوند میان دو سرعنوان فارسی و کنگره برقرار است، سیستم به‌خوبی ایفای نقش کرده است.

## ۶. یافته‌ها و بحث

پرسش نخست: درصد برقراری پیوندهای صحیح میان دو ابزار سرعنوان‌های موضوعی فارسی و کنگره در گنجینه کتاب‌های نوسا چه قدر است؟

از میان ۱۱۲۳۳ رکورد مورد کاوش تعداد ۹۵۱۹ یعنی ۸۴/۷ درصد آنها ما به‌ازایی در فهرست پیوسته کنگره دارند. به این معنا که ۸۴/۷ درصد از موضوعات ساخته‌شده در فهرست سرعنوان‌های موضوعی فارسی عیناً در فهرست پیوسته کنگره نیز وجود دارد. طبق روش‌شناسی پژوهش و انجام کاوش‌های مکرر جهت شناسایی دقیق سرعنوان‌های مشترک در دو فهرست کنگره و فارسی مشخص شد که در این ۹۵۱۹ رکورد مشترک شناسایی شده، تنها ۷۶۱۲ رکورد یعنی ۸۰ درصد بدون اشکال بازیابی شده و پیوند میان آنها در گنجینه کتاب‌های نوسا برقرار بوده است. در ۲۰ درصد دیگر کاوش‌ها، علی‌رغم وجود ما به‌ازای انگلیسی در فهرست کنگره، خطایی در کاوش رخ داده است.

به نظر می‌رسد، خطای ۲۰ درصدی در شناسایی پیوند میان رکوردهای مستند موضوع فارسی و کنگره در گنجینه کتاب‌های نوسا، خطای اندکی نباشد. لازم به ذکر است که این خطا در مورد نمونه ۱۱۲۳۳ رکوردی به‌دست آمده است و چنین خطایی در حجم بالای رکوردهای سرعنوانی، اشکالات قابل توجهی را رقم خواهد زد. بدیهی است واکاوی دلایل این اشکالات می‌تواند برای بالابردن کارآمدی مفهوم داده‌های پیوندی در جستجوهای کاربران در سطح وب بسیار مهم باشد، مسئله‌ای که پرسش دوم پژوهش به آن پرداخته است.

پرسش دوم: چه عواملی باعث عدم برقراری پیوند میان این دو ابزار بازنمون در گنجینه کتاب‌های نوسا می‌شود؟

شناسایی انواع اشکالات موجود در برقراری پیوندها، اشکالات داده‌های ثبت‌شده در فهرست پیوسته کتابخانه ملی ایران را آشکار می‌سازد. در هنگام کاوش، برای هر یک از رکوردهایی که پیوند میان دو فهرست فارسی و کنگره در گنجینه کتاب‌های نوسا برقرار نشده، به شرط مشترک بودن سرعنوان در هر دو فهرست (در ۹۵۱۹ رکورد)، علت عدم

برقراری پیوند، مورد کنکاش قرار گرفت و به تدریج ثبت و دسته‌بندی شد. جدول ۱ نوع و فراوانی اشکالات به دست آمده در برقراری پیوندها را در رکوردهای مشترک میان دو فهرست سرعنوان‌های موضوعی فارسی و کنگره در گنجینه کتاب‌های نوسا نشان می‌دهد.

جدول ۱. وضعیت رکوردهای سرعنوانی از نظر انواع اشکالات موجود در نظام

وضعیت رکوردهای سرعنوانی	فراوانی	درصد
<b>بدون اشکال</b>	۷۶۱۲	۸۰/۰
<b>دارای اشکال</b>		
اشکال در ورود اطلاعات (اشکال املائی و ...)	۵۶۶	۵/۹
نامشخص بودن نظام موضوعی <sup>۱</sup>	۷۲۱	۷/۶
تغییر موضوع و عدم ارجاع به موضوع جدید	۹	۰/۱
بدون اشکال در ورود اطلاعات ولی عدم بازیابی در گنجینه نوسا	۵۹۰	۶/۲
ارجاع یک سرعنوان فارسی به دو سرعنوان موضوعی در فهرست کنگره <sup>۲</sup>	۲۱	۰/۲
<b>مجموع</b>	<b>۹۵۱۹</b>	<b>۱۰۰</b>

در ۲۰ درصد از سرعنوان‌های مشترک، پیوند میان سرعنوان‌های فارسی و انگلیسی برقرار نشده و نامشخص بودن نظام موضوعی در کتابخانه ملی ایران بیشترین درصد علت بروز اشکال را به خود اختصاص داده است. خطا از آنجا رخ می‌دهد که در تعداد نه‌چندان

۱. در فهرست پیوسته کتابخان ملی ایران امکان محدودسازی کاوش به مستند موضوع فارسی نیست. هنگامی که جستجوی موضوعی انجام می‌شود، این جستجو در مستند موضوع اصفا نیز به انجام می‌رسد و بخشی از یافته‌ها مربوط به آن نظام موضوعی است. یکی از فیلهای پایگاه که نام نظام مستند موضوعی است، مشخص می‌سازد که این موضوع مربوط به نظام موضوعی اصفا و یا سرعنوان موضوعی فارسی است. متأسفانه پُر نشدن این فیلد و نامشخص بودن نظام موضوعی مانع از برقراری پیوند میان موضوعات فارسی و کنگره شده است.

۲. با توجه به تفاوت‌های زبانی، این مورد را نمی‌توان اشکال و نقص تلقی کرد. دلیل ذکر این مورد آن است که این شکل از ارجاع مانع از تحقق پیوند میان دو موضوع در فهرست فارسی و کنگره در فهرست کتاب‌های نوسا شده است.

اندکی از رکوردهای مستند، فیلد مربوط به نوع نظام مستند که می‌تواند مستند موضوع فارسی، اصطلاحنامه‌ها اصفا و یا سایر نظام‌های مستند از جمله مستند نام و ... باشد، پر نشده است. به این ترتیب، گنجینه کتاب‌های نوسا امکان برقراری پیوند میان دو نظام را، که لازم است هر دو مستند موضوع باشند (یکی موضوع فارسی و دیگری کنگره) نخواهد یافت. این اشکال قابل حل است و می‌تواند در هنگام تبدیل رکوردهای مستند موضوع به قالب اسکاس اصلاح شود.

تعداد ۵۹۰ مورد از سرعنوان‌های مشترک (۶/۲ درصد) اگر چه اشکالی در ورود اطلاعات نداشته‌اند، در گنجینه کتاب نوسا بازیابی نشده‌اند. برای مثال، موضوعی مانند «خیابان‌ها» که در فهرست سرعنوان‌های موضوعی فارسی به همین شکل موجود است و در فهرست پیوسته ملی نیز بازیابی می‌شود، در گنجینه نوسا بازیابی نمی‌شود و موجود نیست. در مورد این موضوع و موضوعات مشابه، نام پایگاه مستند موضوع وارد شده، نیم‌فاصله رعایت، و معادل انگلیسی آن نیز به شکل صحیح وارد شده است. حال، پرسش آن است که اگر گنجینه نوسا بر پایه پایگاه مستند موضوعی کتابخانه ملی به کاوش می‌پردازد و به آن مرتبط است، چگونه می‌توان بازیابی نشدن این دست از موضوعات را که درصد قابل توجهی را نیز به خود اختصاص می‌دهند، توضیح داد؟ یکی از پاسخ‌ها آن است که در گنجینه نوسا تنها موضوعاتی بازیابی می‌شوند که در مجموعه کتاب‌های نوسا به‌عنوان موضوع به منبعی اختصاص یافته باشند و کتابی در آن موضوع در مجموعه گنجینه نوسا موجود باشد. در این صورت، تنها بخشی از داده‌های مستند موضوع فارسی به مستند موضوع کنگره پیوند خورده است؛ موضوعاتی که کتابی در آن موضوع در گنجینه نوسا موجود است.

وجود اشکالات تاییبی در معادل انگلیسی سرعنوان‌های موضوعی فارسی از آن جهت که ارتباط میان سرعنوان‌های کتابخانه ملی و کتابخانه کنگره آمریکا در گنجینه کتاب‌های نوسا فراهم آمده است، باعث بروز اشکال و عدم بازیابی می‌شود. گاه تفاوت معادل انگلیسی ثبت شده در رکورد فارسی و سرعنوان کنگره ثبت شده در آمریکا، از اطلاعاتی است که مستندساز کتابخانه ملی ایران در پراپرتز به معادل‌های انگلیسی خود افزوده است. برای نمونه، محتوای مندرج در پراپرتز در موضوعی با معادل انگلیسی (Prejudices in literature (N. S. G توسط مستندساز کتابخانه ملی ایران افزوده شده است و در فهرست



سرعنوان‌های کنگره موجود نمی‌باشد. هر گونه تغییر در شکل نگارشی در معادل‌های انگلیسی باعث قطع پیوند میان دو ابزار فارسی و انگلیسی شده است. به نظر می‌رسد تحقق ایده داده‌های پیوندی و برقراری پیوند میان موضوعات فارسی و انگلیسی نیازمند بازبینی و اصلاح معادل‌های انگلیسی موضوعات در رکوردهای مستند موضوع فارسی است. با این بازنگری بخش قابل توجهی از اشکال عدم برقراری پیوند قابل مدیریت خواهد بود.

ارجاع از دو سرعنوان موضوعی فارسی به یک معادل انگلیسی (مانند دو موضوع «نانوذره‌های بنیادی» و «نانوذرات» که هر دو معادل انگلیسی Nanoparticles را پذیرفته‌اند) ممکن است به واسطه تعیین دو معادل قابل قبول برای یک سرعنوان موضوعی کنگره و یا نشانی از مستندسازی دوباره باشد. این شکل از ارجاعات در ایجاد پیوند میان سرعنوان‌های موضوعی فارسی و کنگره اختلال ایجاد کرده است (۰/۲ درصد). همچنین تغییر موضوع و عدم ارجاع به موضوع جدید (۰/۱ درصد) نیز از دیگر اشکالات ثبت شده در فرایند گردآوری داده‌هاست.

## ۷. برخی یافته‌های دیگر

از دیگر یافته‌های این بررسی می‌توان به بخشی از اشکالات موجود در گنجینه کتاب‌های نوسا و فهرست پیوسته کتابخانه ملی ایران اشاره داشت. این اشکالات گاه به موضوع داده‌های پیوندی باز نمی‌گردد و صرفاً اشکالاتی است که در جریان گردآوری داده‌ها مشاهده شده و در این بخش گزارش می‌شود:

برخی اشکالات در گنجینه کتاب‌های نوسا:

◇ عدم نمایش روابط معنایی در گنجینه کتاب‌های نوسا به گونه‌ای که بخشی از اطلاعات از کتابخانه ملی ایران فراخوانی شده و نمایش داده می‌شود و بخشی دیگر نمایش داده نمی‌شود. نبود یکدستی در نمایش روابط، آگاهی از شبکه روابط محلی سرعنوان‌ها را از طریق این وبسایت با دشواری رو به رو می‌سازد. به همین دلیل، برای هر یک از سرعنوان‌های موضوعی چنانچه ارتباط سرعنوان با سایر سرعنوان‌های هم‌ارز، مرتبط، اعم یا اخص مورد توجه کاربر باشد، لازم است با مراجعه به فهرست پیوسته کتابخانه ملی ایران، کاوش دیگری به انجام رسد.

برخی اشکالات موجود در فهرست پیوسته سازمان اسناد و کتابخانه ملی ایران:

◇ نبود یکدستی در شکل ارجاعات و وجود سرعنوان‌هایی با ارجاعات «نیز نگاه کنید»؛ مانند سرعنوان موضوعی «استعمار» و یا «حفاری‌ها (باستان‌شناسی)»؛ این در حالی است که مدت‌هاست نشانه‌های مربوط به انواع روابط معنایی در فهرست سرعنوان‌ها تغییر کرده است و به این ترتیب، استفاده از برچسب‌های قدیمی برای اشاره به روابط میان موضوعات غیرمنطقی به نظر می‌رسد.

◇ نامشخص بودن نوع رابطه حاکم بر سرعنوان‌ها؛ برای نمونه دو سرعنوان «فلسفه ایرانی» و «فلسفه و اسلام» سرعنوان‌های اخص سرعنوان «فلسفه اسلامی» می‌باشند، اما در رکورد نمایش داده شده در فهرست پیوسته کتابخانه ملی ایران، برچسب نوع رابطه خالی است.

#### ۸. سخن پایانی

گام‌های عملیاتی در جهت اجرای ایده داده‌های پیوندی بخش قابل توجهی از اطلاعات ارزشمند و ساخت یافته کتابخانه‌ها و مراکز اطلاع‌رسانی را در جهت تحقق وب معنایی و بازیابی معنایی فراهم می‌سازد. اطلاعاتی که در پایگاه‌های کتابشناختی ذخیره شده و می‌تواند در خدمت لایه اعتماد، یکی از لایه‌های مهم وب معنایی باشد و بزرگ‌ترین چالش وب معنایی را برطرف سازد. به این ترتیب، اطلاعات ارزشمندی که در طول سال‌ها به شکل «طرح‌های مفهومی» بازنمون شده‌اند (harper 2006) نه تنها در انحصار کتابخانه‌ها و مراکز اطلاع‌رسانی باقی نخواهد ماند، بلکه به صورت مشترک در اختیار تمامی کاربران وب معنایی قرار خواهد گرفت.

در جریان نگاشت و انتشار مجموعه داده‌ها در ابر داده‌های پیوندی، عملیات شناسایی و برقراری پیوند میان داده‌ها انجام می‌پذیرد. قدرت داده‌های پیوندی در برقراری پیوندهای خارجی است. برای نمونه برقراری پیوند از فهرست پیوسته کتابخانه به سایر داده‌های مرتبط با رکوردها در هستی‌شناسی‌هایی چون DBpedia - نسخه وب معنایی و یکی پدیا - و سایر داده‌های مرتبط خارج از فضای کتابخانه، امکان معنایی‌تری را برای کاربران فراهم می‌آورد و نظام‌های معنایی را تحقق می‌بخشد. به این ترتیب، مجموعه‌ای از داده‌ها در خدمت پاسخ‌های معنایی قرار می‌گیرد، کاربر با پیوندهای گسترده‌تری روبه‌رو می‌شود و بر قابلیت‌های نرم‌افزارهای کتابخانه‌ای از جمله کتابخانه‌های دیجیتال خواهد افزود.

در این پژوهش تلاش شد ضمن معرفی اجمالی ایده‌داده‌های پیوندی در فضای پایگاه‌های کتابشناختی، با توجه به حرکت آغازشده در تبدیل و انتشار داده‌ها، سه نظام فهرست پیوسته کتابخانه کنگره آمریکا، فهرست پیوسته کتابخانه ملی ایران و گنجینه کتاب‌های نوسا در نظر گرفته شود و چگونگی برقراری پیوند میان موضوعات فارسی و انگلیسی در گنجینه کتاب‌های نوسا، به‌عنوان ابزار، ردگیری شود. در جریان پژوهش، بخشی از اشکالات موجود در برقراری پیوند میان موضوعات فارسی و انگلیسی شناسایی و دسته‌بندی شد. شناسایی سرعنوان‌های مشابه در دو فهرست موضوعی از طریق گنجینه کتاب‌های نوسا انجام پذیرفت. برقراری پیوند میان سرعنوان‌های فارسی و کنگره تا اندازه‌ای در گنجینه کتاب‌های نوسا تحقق یافته است؛ به‌گونه‌ای که با کاوش موضوعی که در فهرست کتابخانه ملی ایران موجود است، در صورت وجود معادل آن رکورد در کتابخانه کنگره آمریکا، امکان مشاهده موضوع در کتابخانه کنگره آمریکا وجود دارد و پیوند میان این دو فهرست پیوسته برقرار گشته است.

در عمل مشاهده می‌شود که پیوند میان حدود ۲۰ درصد از رکوردهای مشترک در فهرست مستند موضوع فارسی و کنگره برقرار نشده است. بیشترین دلیل بروز اشکال در برقراری این پیوند، نامشخص بودن نظام موضوعی (۷/۶ درصد) و وجود اشکال در ورود اطلاعات به‌ویژه بی‌توجهی به املاء سرعنوان‌ها و معادل انگلیسی آنهاست (۵/۹ درصد). این دو اشکال که درصد بالایی از خطاها مربوط به آنهاست، قابل مدیریت و اصلاح می‌باشد.

#### ۹. پیشنهاد

با توجه به وجود اشکال در ورود داده‌های کتابشناختی در پایگاه‌های مستند موضوع کتابخانه ملی ایران، پیشنهاد می‌شود شیوه‌نامه‌های مدونی جهت ورود اطلاعات در پایگاه کتابخانه ملی تدوین شود و به‌صورت دوره‌ای، درستی ورود اطلاعات مورد بررسی قرار گیرد. جهت تسهیل فرایند بازنگری دوره‌ای، می‌توان به‌گزینه تشکیل کمیته دائمی بازنگری و توسعه سرعنوان‌های موضوعی فارسی نیز اندیشید. از آنجا که امکان انتقال رکوردهای مستند از پایگاه مستند کتابخانه ملی ایران به سایر پایگاه‌ها در کتابخانه‌های ایران امکان‌پذیر است، تصحیح خطا و جلوگیری از بروز آن، به یکدستی و درستی قابل

قبول تری در سطح کتابخانه‌های کشور خواهد انجامید، ضمن آنکه ثبت صحیح معادل‌های انگلیسی سرعنوان‌های فارسی و به صفر رساندن اشکالاتی از این دست، ایده داده‌های پیوندی را به واقعیت نزدیک‌تر خواهد ساخت.

با توجه به نبودن حوزه داده‌های پیوندی و پژوهش‌های اندکی که در این حوزه به انجام رسیده، موضوعات متنوعی قابل پژوهش است. نکته اصلی آن است که با توجه به میان‌رشته‌ای بودن کار، بهتر آن است که پژوهش‌ها در قالب گروه‌های پژوهشی متشکل از متخصصان حوزه‌هایی چون کامپیوتر و کتابداری و اطلاع‌رسانی تعریف شود. از منظر مدیریتی بخشی از چالش‌های مدیریتی نگاشت و بازنشر داده‌های کتابخانه‌ای قابل بررسی است. روش‌ها و استانداردهای مناسب جهت نگاشت و بازنشر از دیگر موضوعات قابل بررسی است.

انتشار داده، یک بُعد و استفاده از آن، بُعدی دیگر است. پرسش‌هایی از این دست که رابط کاربر نظام‌های معنایی چه ویژگی‌هایی خواهد داشت؟ مطالعه رفتار اطلاع‌یابی، معیارهای قضاوت ربط و رضایت کاربران از نظام‌های معنایی چگونه است؟ در مرحله بعد اهمیت خواهد یافت. در حال حاضر نخستین گام‌ها در حال برداشته شدن است و شاید بیش از هر چیزی بُعد مدیریتی و انتشار داده‌ها در قالبی مناسب و استاندارد از اولویت برخوردار باشد.

#### ۱۰. فهرست منابع

اخوتی، مریم. ۱۳۸۸. تحلیل و مقایسه شبکه مفهومی حاکم بر سرعنوان‌های موضوعی فارسی با متون کتابداری و اطلاع‌رسانی و ابزارهای بین‌المللی ذخیره و بازیابی اطلاعات بر اساس رویکرد تحلیل حوزه. پایان‌نامه دکترای کتابداری و اطلاع‌رسانی. دانشکده علوم تربیتی و روان‌شناسی دانشگاه فردوسی مشهد. به‌کمالت، بهشید، محسن کاهانی، محبوبه دادخواه، فتنه زرین کلام، و صمد پایدار. ۱۳۹۰. ارائه چارچوبی برای انتشار مجموعه داده‌های فارسی به‌صورت داده‌های پیوندی روی وب. مجله فنی مهندسی دانشگاه آزاد مشهد ۴(۱): ۱-۱۹.

چان، لوئیس می. ۱۹۸۱. فهرست‌نویسی و رده‌بندی. ترجمه زهیر حیاتی و زهرا ستوده. ۱۳۷۹. تهران: نشر کتابدار.

- سلطانی، پوری، کامران فانی، و مهناز رهبری اصل. ۱۳۷۳. سرعنوان‌های موضوعی فارسی (ویرایش ۲). تهران: سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران.
- شریف، عاطفه. ۱۳۸۵. شناختی از روابط معنایی در هستی‌شناسی وب. *اطلاع‌شناسی* ۴ (۱ و ۲): ۶۵-۸۴.
- \_\_\_\_\_. ۱۳۸۶. *بازنمایی معنا در زبان هستی‌شناسی وب*. مقاله ارائه‌شده در همایش ملی اصطلاحنامه‌ها و کاربردهای آن در محیط الکترونیکی، قم.
- \_\_\_\_\_. ۱۳۹۱. شناسایی و تحلیل ابعاد ساختاری و مفهومی فهرست سرعنوان‌های موضوعی فارسی در مقایسه با فهرست سرعنوان‌های موضوعی کنگره. پایان‌نامه دکترای کتابداری و اطلاع‌رسانی. دانشکده علوم تربیتی و روان‌شناسی دانشگاه فردوسی مشهد.
- Berners-Lee, T., J. Hendler & O. Lassila. 2001 The Semantic Web: a new form of web content that is meaningful to computers will unleash a revolution of new Possibilities, <http://www.w3c.org/2001/sw.html> (accessed Oct 5, 2006)
- Bizer, C., T Heath, and T. Berners-Lee. 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)* 5 (3): 1-22. <http://eolo.cps.unizar.es/Docencia/doctorado/Articulos/LinkedData/bizer-heath-berners-lee-ijswis-linked-data.pdf> (accessed Jan. 14, 2012).
- Bowen, J. 2010. *Moving Library Metadata Toward Linked Data: Opportunities Provided by the eXtensible Catalog*. In: Proceedings of the International Conference on Dublin Core and Metadata Applications, 44-59. <http://dcpapers.dublincore.org/pubs/article/view/1010/979>, Pittsburgh, Pennsylvania, USA (accessed May 20, 2011).
- Cantara, Linda. 2006. Encoding controlled vocabularies for the Semantic Web using SKOS Core. *OCLC Systems & Services* 22 (2): 111-114.
- Ed, S., I. Antoine, R. Clay, and K. Dan. 2008. *LCSH, SKOS and linked data*. In: Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications, 25-33. Berlin, Germany: Dublin Core Metadata Initiative.
- Eslami, Saeedeh and Mohammad Hossein Vaghefzadeh. 2013. *Publishing Persian linked data of National library and Archive of Iran*. Paper presented at; IFLA World Library and Information Congress 17-23 August 2013, Singapore.
- Fripp, Dominic. 2010. *Using linked data to classify web documents*. In: Aslib Proceedings, 62 (6): 585- 595.
- Hannemann, Jan and Jurgen Kett. 2010, August 17. *Linked Data for Libraries*. In: Presented at the World Library and Information Congress: 76th IFLA General Conference and Assembly, Gothenburg, Sweden. <http://conference.ifla.org/past/2010/149-hannemann-en.pdf> (accessed Feb. 20, 2011).
- Harper, C. A. 2006. *Encoding library of congress subject headings in SKOS: authority control for the semantic web*. Proceedings of the 2006 international conference on Dublin Core and Metadata Applications: metadata for knowledge and learning, 89-94. Manzanillo, Colima, Mexico: Dublin Core Metadata Initiative.
- Heath, T., and C. Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology, 1(1): 1-136. [Abstract] <http://www.morganclaypool.com/doi/abs/10.2200/s00334ed1v01y201102wbe001> (accessed Jan. 14, 2012).
- Hess, C., C. Schlieder. 2006. Ontology-based verification of core model conformity in

- conceptual modeling. *Computers, Environment and Urban Systems* 30: 543-561. Elsevier Database (May 5, 2007).
- Liang, A. C., and M. Sini. 2006. Mapping AGROVOC and the Chinese Agricultural Thesaurus: Definitions, tools, procedures. *New Review of Hypermedia and Multimedia* 12(1): 51-62. [Abstract] <http://www.tandfonline.com/doi/abs/10.1080/13614560600774396> (accessed Jan. 14, 2012).
- Malmsten, M. 2009. *Exposing Library Data as Linked Data*. Presented at the IFLA 2009 Satellite Meetings in Florence, Italy. <http://www.ifla2009satelliteflorence.it/meeting3/program/assets/MartinMalmsten.pdf> (accessed Jan. 20, 2010).
- Miles, A., J. R. Perez-Aguera. 2007. SKOS: Simple Knowledge Organisation for the web. *Cataloging & Classification Quarterly* 43 (3/4): 69-83.
- Morshed, A., C. Caracciolo, G. Johannsen, and J. Keizer. 2011. *Thesaurus Alignment for Linked Data Publishing*. In: Proceedings of the International Conference on Dublin Core and Metadata Applications, 37-46. Hague, Netherlands <http://dcpapers.dublincore.org/pubs/article/view/3623/1849> (accessed Jan. 14, 2012).
- Neubert, Joachim. 2009. *Bringing the "Thesaurus for Economics" on to the Web of Linked Data*. Proceedings of the Linked Data on the Web Workshop LDOW2009, Madrid, Spain
- Styles, R., D. Ayers, and N. Shabir. 2008. *Semantic Marc, MARC21 and The Semantic Web*. In C. Bizer, T. Heath, K. Idehen & T. Berners-Lee (eds.), LDOW.; CEUR-WS.org. <http://events.linkedata.org/ldow2008/papers/02-styles-ayers-semantic-marc.pdf> (accessed Jan. 20, 2010).
- Summers, E., A. Isaac, C. Redding, and D. Krech. 2008. *LCSH, SKOS and linked data*. In: Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications (pp. 25-33). Berlin, Germany: Dublin Core Metadata Initiative.
- Tillett, B., and C. Harper. 2007. Library of Congress Controlled Vocabularies and Their Application to the Semantic Web. *Cataloging & Classification Quarterly* 43 (3-4): 47-68.
- Weng, S.-S., H.-J. Tsai, S.-C Liu, and C.-H Hsu. 2006. Ontology construction for information classification. *Expert Systems with Applications* 31(1): 1-12. Elsevier Database (May 5, 2007).
- Yee, Martha M. 2009. Can Bibliographic Data be Put Directly onto the Semantic Web? *Information Technology and Libraries* 28 (2). <https://ejournals.bc.edu/ojs/index.php/ital/article/view/3175/2788> (accessed May 20, 2011).
- Yi, K., and L. M. Chan. 2010. Revisiting the syntactical and structural analysis of Library of Congress Subject Headings for the digital environment. *Journal of the American Society for Information Science and Technology* 61: 677-687.

# Blind Links, a Big challenge in the Linked Data Idea: Analysis of Persian Subject Headings

Atefeh Sharif

Assistant Professor; Ferdowsi University of Mashhad; Mashhad  
Iran  
atefehsharif@gmail.com

Iranian Journal of  
**Information  
Processing &  
Management**

**Abstract:** In this survey, Linked data concept as exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web and some potential problems in converting Persian subject headings (PSHs) records into linked data were discussed. A data set (11233 records) of PSHs was searched in three information retrieval systems including National Library of Iran (NLI) online catalog, Library of Congress (LC) online catalog and NOSA books. Correct links between Persian and English subject headings in the 9519 common records of two catalogs were recorded. The results indicate that the links between Persian and English subjects in %20 of records were failed. The maximum error was associated with the anonymous databases (%6/7) in NLI online catalog. It is recommended to preprocess the PSHs records before any conversion projects. It seems that, during the preprocessing, the potential errors could be identified and corrected.

**Keywords:** Linked Data, Semantic Web, Persian Subject Headings (PSHs); Library of Congress Subject Headings (LCSHs); Nosa Books

Iranian Research Institute  
for Science and Technology

ISSN 2251-8223

eISSN 2251-8231

Indexed in SCOPUS, ISC & LISA

Vol.30 | No.1 | pp: 223-244

Autumn 2014

پرتال جامع علوم انسانی