

شناسایی نظام مفهومی حاکم بر فهرست سرعنوان‌ها و تطابق آن با نظام رده‌بندی کنگره و نظام دانشی در ایران

عاطفه شریف*

دانشجوی دکتری،

اطلاعات و دانش‌شناسی دانشگاه فردوسی مشهد

دکتر محمدرضا داورپناه^۱

استاد،

گروه کتابداری و اطلاع‌رسانی دانشگاه فردوسی مشهد

دکتر اعظم صنعت‌جو^۲

استادیار،

گروه کتابداری و اطلاع‌رسانی دانشگاه فردوسی مشهد

دریافت: ۱۳۹۱/۰۳/۲۸ | پذیرش: ۱۳۹۱/۰۶/۲۷

فصلنامه علمی پژوهشی
پژوهشنامه پردازش و مدیریت اطلاعات
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۲۲۵۱-۸۲۲۳
شاپا(الکترونیکی) ۲۲۵۱-۸۲۳۱
نمایه در SCOPUS، LISA و ISC
http://jipm.irandoc.ac.ir
دوره ۲۸ | شماره ۴ | ص ص ۸۵۷-۸۳۱
تابستان ۱۳۹۲
نوع مقاله: پژوهشی

*atefehsharif@gmail.com
1.mrdavarpanah@yahoo.com
2. sanatjoo@gmail.com

چکیده: پژوهش حاضر از نوع کاربردی است و با هدف کشف نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی و ویژگی‌های آن، به روش تحلیل محتوا انجام پذیرفته است. در شناسایی نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی، ۸۴۵۴۵ رکورد سرعنوانی مورد بررسی قرار گرفت و ۲۴۱ درخت سلسله مراتبی شناسایی شد. نظام مفهومی شناسایی شده، با نظام دانشی موجود در ایران و نظام رده‌بندی کنگره مطابقت داده شد. بیشترین ارتباط میان نظام موضوعی و رده‌بندی کنگره در رده‌های علوم (۱۹/۵٪)، فناوری (۱۴/۴٪)، علوم اجتماعی (۱۰/۶٪)، پزشکی (۹/۵٪) و فلسفه و روان‌شناسی (۶/۶٪) است. نیمی از سرعنوان‌های رأس با عنوان رشته‌های تحصیلی تطابق واژگانی ندارد و فقط ۲۱/۰۵ درصد از سرعنوان‌ها در تطابق کامل‌اند. در بررسی تطابق واژگانی سرعنوان‌های رأس با رده‌بندی کنگره مشخص گردید که ۷۱/۰۵ درصد از سرعنوان‌های رأس به‌طور دقیق با همان واژگان در دو ابزار وجود دارند. در مجموع، نظام مفهومی فهرست سرعنوان‌های موضوعی فارسی فقط در تعداد معدودی از حوزه‌ها نظم یافته‌اند و در سایر حوزه‌ها پیوند سلسله مراتبی میان موضوعات، آن‌گونه که انتظار می‌رود، برقرار نشده است. پیشنهاد می‌شود این فهرست، در جهت غنی‌سازی شبکه معنایی میان سرعنوان‌ها و پوشش تمامی حوزه‌های دانش در ایران مورد بازنگری قرار گیرد.

کلیدواژه‌ها: فهرست سرعنوان‌های موضوعی فارسی، نظام مفهومی، نظام دانشی ایران، رده‌بندی کنگره

۱. سخن نخست

ساختارگرایی در مقابل تفکر اتمیسم^۱ راسل^۲ و شرط اتمی بودن^۳ آن قرار می‌گیرد. در شرط اتمی بودن، معنی هر تعبیر بنیادین باید از نظر منطقی و روان‌شناسی، مستقل از معنی دیگر تعبیرهای بنیادین باشد. ساختارگرایی بر وابستگی متقابل امور، به جای وجود مجزا و مشخص امور، تأکید می‌کند (لاینز^۴ ۱۹۸۱، ۱۳۱). در این رویکرد، هیچ واژه‌ای را نمی‌توان مستقل از واژگان دیگری که از جهت ساختار معنی با آن مرتبط هستند و به روشن شدن معنی آن کمک می‌کند، تعریف کرد. در این دیدگاه، ساختار لغتنامه‌ای یک زبان (یعنی ساختار ذخیره واژگان آن زبان) را از منظر معنی‌شناسی می‌توان به منزله شبکه‌ای از روابط مفهومی^۵ [روابط معنایی] در نظر آورد که هر گره آن یک واژه و هر قطعه نخ آن یک رابطه مفهومی [رابطه معنایی] است (لاینز ۱۹۸۱، ۱۴۵). سوسور^۶ معتقد است که در خارج از چارچوب گفتار، واژه‌هایی که وجه مشترکی دارند در حافظه با یکدیگر ارتباط می‌یابند و به این ترتیب، گروه‌هایی را تشکیل می‌دهند که روابط بسیار گوناگونی در آنها حکمرواست. برای مثال، واژه آموزش ناخودآگاه انبوهی از واژه‌های دیگر را در ذهن برمی‌انگیزد؛ آموختن، اطلاع دادن، تجهیز تسلیحاتی، تغییر و غیره یا آموزش و پرورش و یادگیری. هر یک از این واژه‌ها به گونه‌ای، وجه اشتراکی با یکدیگر دارند (سوسور ۱۹۸۶، ۱۷۷).

در حالی که ساختارگرایی به دهه‌های آغازین قرن ۲۰ باز می‌گردد، چنین به نظر می‌رسد که عبارت شبکه‌های معنایی نخستین بار، توسط کویلین مطرح شد. اما ایده شبکه معنایی که شبکه‌ای از مفاهیم به هم مرتبط است به خیلی پیش‌تر برمی‌گردد و حتی می‌توان پیشینه آن را به ارسطو نیز نسبت داد (Quillian 1966 در پهلوان‌نژاد و نامور فرگی ۱۳۸۸، ۳). در هر شبکه، خوشه‌هایی را می‌توان تشخیص داد که نشان‌دهنده حوزه‌های معنایی^۷ است و مهم‌ترین رابطه معنایی در این شبکه، رابطه سلسله مراتبی است به گونه‌ای که شبکه‌ای سلسله مراتبی از واژگان به هم مرتبط را تشکیل می‌دهد.

از ابتدای شکل‌گیری فهرست سرعنوان‌های موضوعی کنگره و به ویژه بعد از سال ۱۹۷۵ به واسطه عدم رعایت منطقی روابط اصطلاحنامه‌ای و تبدیل ماشینی روابط اولیه به روابط

1. Atomism

2. Russell

3. condition of atomicity

4. Lyonse

۵. sense-relations در معنی‌شناسی زبانی sense را دلالت درون زبانی و یا دلالت مفهومی ترجمه می‌کنند و به رابطه‌ی معنایی واحدهای زبان با یکدیگر گفته می‌شود (صفوی ۱۳۸۴، ذیل واژه) به همین دلیل و با توجه به پژوهش حاضر، عبارت رابطه معنایی مناسب‌تر به نظر می‌رسد.

6. Ferdinand de Saussure

7. semantic filed

اصطلاحنامه‌ای - هم‌ارز، سلسله‌مراتبی و هم‌بسته - این ابزار مورد انتقاد قرار گرفت (مانند Dykstra 1998, 43; Rolland-Thomas 1993). با وجود این، به طور تلویحی سه نوع رابطه معنایی هم‌ارزی، سلسله‌مراتبی و هم‌بسته میان سرعنوان‌ها حاکم است و این ساختار از فهرست سرعنوان‌های موضوعی کنگره به فهرست سرعنوان‌های موضوعی فارسی منتقل شده است. پژوهشگرانی چون یی و چان با پذیرفتن این شکل از روابط معنایی در فهرست سرعنوان‌ها به بیان ایده‌ای در باب رده‌بندی پنهان در فهرست سرعنوان‌ها پرداخته‌اند (Yi and Chan 2008; 2010). این رده‌بندی پنهان با پیگیری روابط سلسله‌مراتبی موجود در فهرست سرعنوان‌های موضوعی قابل شناسایی است. با پیگیری رابطه سلسله‌مراتبی، درخت‌هایی تشکیل می‌شود که می‌تواند نشانی از نظام مفهومی موجود در فهرست سرعنوان‌ها باشد. چنانچه تمام سرعنوان‌های موضوعی مورد توجه باشد و درخت‌های سلسله‌مراتبی کلی^۱ - که هم اکنون پنهان از دید در فهرست موجودند - ترسیم گردد، تعدادی درخت مجزا تشکیل خواهد شد. استقلال این درخت‌ها از یکدیگر، پژوهشگران را به این فرضیه رهنمون ساخته است که هر یک از این درخت‌ها، به بازنمایی یک مفهوم می‌پردازند و هیچ دو مفهومی وجود ندارد که در دو درخت مستقل ظاهر شود و در سلسله‌مراتبی‌اش هم‌پوشانی داشته باشد. به این ترتیب، تمام فهرست سرعنوان‌های موضوعی را می‌توان در شکل n تعداد درخت مستقل ترسیم نمود (Yi and Chan 2010, 681).

فلبر تفاوتی را میان نظام مفهومی و نظام موضوعی طرح می‌کند؛ در نظام‌های موضوعی، روابط نظام‌ساز با وقوع مفاهیم به شکل موضوع در اسناد ارتباط دارد و نتیجه می‌گیرد که بر همین مبنا، امکان حذف اعضا در ساختار یک نظام موضوعی بیش از ساختار نظامی مفهومی است (۱۹۸۴، ۱۸۰). در مقابل نظر فلبر، در نظریه حوزه‌های معنایی^۲ عضوهای یک حوزه معنایی، واحدهای یک نظام مفهومی را تشکیل می‌دهند. این عضوها [صرف‌نظر از نوع رابطه معنایی] در ارتباط متقابل با یکدیگرند. به طور مثال، مفهوم‌هایی [واژه‌هایی] چون کتاب، مقاله، نقشه به واسطه ویژگی " رسانه بودن" با مفهوم‌های [واژه‌های] دیگری از این دست در یک حوزه قرار می‌گیرند (صفوی ۱۳۸۳، ۴ و ۷-۸). به این اعتبار، با در نظر گرفتن رابطه معنایی میان سرعنوان‌های موضوعی فارسی، مجموعه‌ای از مفاهیم در حوزه‌های معنایی تشکیل می‌شوند و این حوزه‌ها را می‌توان نشانی از طبقات مفهومی و مجموعه آنها را نظام مفهومی حاکم بر

۱. Global Hierarchical Structure (GHS) نمایشی سلسله‌وار از سرعنوان‌هایی است که به واسطه رابطه سلسله‌مراتبی به یکدیگر پیوند خورده‌اند. ساختاری درختی دارد و فقط به یک سطح از رابطه سلسله‌مراتبی بسنده نمی‌شود، بلکه رابطه سلسله‌مراتبی تا جایی که ادامه داشته باشد، دنبال می‌شود.

2. filed theory

سرعنوان‌های موضوعی به شمار آورد. ایده یی و چان (Yi and Chan 2010; 2008) در باب درخت‌های سلسله مراتبی و رده‌بندی پنهان، بی‌شبهت با نظریه حوزه‌های معنایی نیست.

۲. بیان مسأله و پرسش‌های پژوهش

از آن جایی که واژگان موضوعی در فهرست سرعنوان‌های موضوعی به مرور زمان و به واسطه پشتوانه انتشاراتی به زبان فارسی تولید می‌شوند و بازنمونی از مفاهیم حوزه‌های موضوعی به زبان فارسی تلقی می‌گردند منطقی به نظر می‌رسد که سرعنوان‌های موضوعی در پیوند با دانش بومی به زبان فارسی و نظام دانشی حاکم بر ایران گسترش یابد. همچنین، با توجه به فرایند سازماندهی منابع و اختصاص شماره رده بر مبنا و پس از اختصاص موضوع، نوعی از انطباق و ارتباط میان ابزار موضوع‌دهی - فهرست سرعنوان‌های موضوعی - و رده‌بندی‌ها - رده‌بندی کنگره - وجود دارد. اگر چه به لحاظ نظری، صحبت اندکی از آن به میان آمده باشد، این پیش فرض به صورت تلویحی در پژوهش‌هایی چون خوشخویی (Khosh-Khui 1985)، لارسن (Larson 1992)، دالین (Dolin 1998) و فرانک و پینتر (Frank and Paynter 2004) حضور دارد. از آن جایی که فهرست سرعنوان‌های موضوعی فارسی با الگوگیری از سرعنوان‌های کنگره شکل گرفته است، می‌توان این ارتباط را میان فهرست سرعنوان‌های موضوعی فارسی و رده‌بندی کنگره نیز تصور نمود.

مطالعه مقدماتی پژوهشگران نشان از آن دارد که تاکنون پژوهش‌های اندکی فهرست سرعنوان‌های موضوعی فارسی را مورد بررسی قرار داده‌اند. به این ترتیب مشخص نیست که نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی چه ویژگی‌هایی دارد، از چه تطابقی با نظام رده‌بندی کنگره برخوردار است، و چگونه پیوندی با نظام دانشی ایران دارد. پژوهش حاضر بر آن است تا پاسخی بر این پرسش‌ها بیابد:

۱. نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی چه ویژگی‌هایی دارد؟

۱-۱. با توجه به رده‌بندی کنگره، تنوع، حجم و نسبت سرعنوان‌های موضوعی در هر یک از رده‌های موضوعی چگونه است؟

۱-۲. با نظر به درخت‌های سلسله مراتبی، طبقات مفهومی^۱ موجود در سرعنوان‌های موضوعی فارسی کدام‌اند؟

۱. در ساختار کلی رابطه سلسله مراتبی، تعدادی درخت سلسله مراتبی به دست می‌آید. هر یک از این درخت‌ها نشان‌گر یک طبقه مفهومی هستند و سرعنوانی که در هر یک از این درخت، در رأس قرار گرفته است و سایر سرعنوان‌ها اخص آن به شمار می‌روند، به منزله نماینده آن طبقه مفهومی شناسایی می‌شود و مورد تحلیل قرار می‌گیرد. در پژوهش یی و چان (Yi, Chan, 2010) از این سرعنوان‌ها با عنوان سرعنوان رأس (top subject heading) یاد شده است.

- ۱-۳. نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی، تا چه میزان بازتابی واقع‌گرایانه از نظام دانشی موجود در ایران^۱ است؟
۲. میزان گم‌شدگی اطلاعات بر اساس تعیین سطح هم‌خوانی میان نظام مفهومی حاصل از درخت‌های سلسله‌مراتبی با نظام مفهومی رده‌بندی کنگره^۲ تا چه میزان است؟

۳. مروری بر پژوهش‌های پیشین

در داخل کشور، پژوهش‌های اندکی در باب فهرست سرعنوان‌های موضوعی فارسی انجام شده و بیشتر این پژوهش‌ها در حیطه تطابق واژگانی صورت پذیرفته است (مانند بزرگی ۱۳۷۰؛ فتاحی و ارسطوپور ۱۳۸۶). فقط پایان‌نامه دکترای اخوتی (۱۳۸۸) است که ورودی به حیطه مفهوم در فهرست سرعنوان‌ها داشته و به شناسایی طبقات مفهومی پرداخته است. یافته‌های اخوتی (۱۳۸۸) نشانگر آن است که طبقه مفهومی "منابع اطلاعاتی" در متون مورد بررسی، بیش از سایر طبقات مفهومی مورد توجه قرار گرفته است. شبکه مفهومی حاکم بر کتاب‌ها با شبکه مفهومی مقالات تفاوت معنی‌دار داشته و فقط ۱۰/۶۵ درصد از مفهوم-واژگان متون مورد بررسی در بین مفاهیم مرتبط با کتابداری و اطلاع‌رسانی سرعنوان‌های موضوعی فارسی و ۲۹/۳۲ درصد آنها در کل سرعنوان‌های موضوعی فارسی وجود داشته‌اند. بر این اساس تفاوت معنی‌دار میان شبکه‌های مفهومی سرعنوان‌های موضوعی فارسی و متون مورد بررسی مشاهده شده است. وی همچنین، گزارش کرده است که شبکه مفهومی سرعنوان‌های موضوعی فارسی به طور معنی‌دار با شبکه مفهومی سرعنوان موضوعی کنگره، سرعنوان موضوعی سیرز و لیزا متفاوت بوده است.

از سویی دیگر، فرشچی (۱۳۸۴) در پژوهشی به بررسی تطبیقی رده تاریخ در اصطلاحنامه فرهنگی فارسی (اصفا) و رده DSR (تاریخ ایران) و رده‌بندی دهدهی دیویی پرداخته است. مقایسه کمی اصطلاحات و موضوعات نشانگر آن بوده است که این تعداد در اصطلاحنامه اصفا (با ۴۹۷ مورد) به نسبت رده DSR تاریخ ایران (با ۲۶۵ مورد) و تاریخ ایران در رده‌بندی دیویی (با ۲۶۷ مورد) به طور تقریبی دو برابر است. سرشاخه‌ها (سرشناسه و موضوع‌ها)ی

۱. این نظام دانشی ناظر بر رشته‌های دانشگاهی موجود در نظام آموزش عالی کشور است که از طریق دفترچه راهنمای ثبت‌نام و شرکت در آزمون سراسری دوره کارشناسی سال ۱۳۹۰ شناسایی می‌شود. این شیوه، نمونه‌ای واقع‌گرایانه از نظام دانشی کشور به‌دست خواهد داد.

۲. بر اساس تطابق سرعنوان‌های رأس به منزله نماینده‌های نظام مفهومی فهرست سرعنوان‌های موضوعی با نشانه‌های طرح کلی رده‌بندی کنگره آمریکا

تاریخی در اصفاء، رده DSR و رده دیویی با یکدیگر یکسان بوده است، اما زیررده و موضوعات وابسته یکسان نبوده‌اند، به طوری که اصفاء از دو منبع دیگر مفصل‌تر بوده است و بنابراین، مقایسه این سه منبع با یکدیگر، از نظر موضوع امکان‌پذیر نیست.

پیشینه پژوهش در بخش خارج از کشور نشان از تعداد بیشتر پژوهش‌ها پیرامون ویژگی‌های سرعنوان‌های موضوعی کنگره و همچنین اصطلاحنامه‌ها دارد. در پژوهش‌هایی، اشکالات پیش آمده در روابط سلسله مراتبی در شبکه معنایی نظام یکپارچه زبان پزشکی (یوام ال اس)^۱ با کمک الگوریتم‌های نوشته شده و به صورت خودکار تعیین شده است (Chen et al. 2009a). همین پژوهشگران در مطالعه‌ای دیگر با استفاده از روش‌های خودکار و دستی به گروه‌بندی مفاهیم و شناسایی مفاهیمی که به اشتباه در یک حوزه وارد شده‌اند، پرداخته‌اند (Chen et al 2009b). کای، هینر و مگنس نیز به ارزیابی ساختار شبکه‌ای اصطلاحنامه‌ها پرداخته‌اند و در این ارزیابی از فنون‌های بصری‌سازی در ترسیم درخت‌های سلسله مراتبی بهره جسته‌اند (Kai, Heiner and Magnus 2007).

یی و چان بر وجود نوعی از رده‌بندی پنهان در سرعنوان‌های موضوعی کنگره باور دارند. پژوهشگران با بیان این مطلب که تاکنون^۲ ابزاری جهت نمایش این رده‌بندی پنهان طراحی نشده است، به شرح طراحی نرم‌افزار VisaLCSH پرداخته‌اند (Yi and Chan 2008). همین پژوهشگران در پژوهشی دیگر که با هدف امکان‌سنجی کاربرد سرعنوان‌های موضوعی کنگره در فضای شبکه‌های دیجیتال انجام پذیرفته است، به بازمینی و بررسی ویژگی‌های نحوی^۳ و ساختاری^۴ سرعنوان‌های موضوعی پرداخته‌اند. یکی از پرسش‌های اساسی این پژوهش پیرامون ویژگی‌های ساختار سلسله مراتبی کلی فهرست سرعنوان‌های موضوعی کنگره است. با بهره‌گیری از برنامه‌هایی که به زبان پرل^۵ نوشته شد، تعداد ۱۵۵۹۰۵ درخت کلی سلسله مراتبی (جی اچ اس) ترسیم شد. سرعنوان موضوعی علم^۶ با ۱۴۵۵۰۲۶ سرعنوان در ساختار سلسله مراتبی‌اش

1. Unified Medical Language System (UMLS)

۲. همان‌گونه که در متن مقاله نیز تصریح شده است دو منبع معتبر کاوش در سرعنوان‌های موضوعی کنگره تا سال ۲۰۰۸ که این مقاله به چاپ رسیده است، یعنی وب سایت مستندات کتابخانه کنگره آمریکا (<http://authorities.loc.gov>) و OCLC (<http://www.oclc.org/connexion>)، به گونه‌ای طراحی شده‌اند که فقط یک رکورد مستند مورد بازبایی قرار می‌گیرد و ارتباط آن با سایر رکوردهای مستند موضوع مشخص نیست. در حال حاضر، نوعی از بصری‌سازی در نمایش سرعنوان‌های موضوعی کنگره انجام پذیرفته است که در وب سایت کتابخانه کنگره (<http://id.loc.gov>) قابل مشاهده و جستجو است، اما در این نمایش نیز، فقط یک رکورد مستند به همراه سرعنوان‌های اعم، اخص و مرتبط آن نمایش داده می‌شود.

3. syntactical

4. structural

5. Perl

6. science

بزرگ‌ترین درخت را تشکیل داده است. پس از آن تاریخ^۱ (۷۶۹۵۲۶ سرعنوان)، تمدن^۲ (۳۶۷۹۶۴ سرعنوان)، مذهب^۳ (۲۹۲۶۳۹ سرعنوان)، زندگی^۴ (۱۴۵۰۵۰ سرعنوان) و توانبخشی^۵ (۱۰۳۴۹۵ سرعنوان) قرار گرفته‌اند. متوسط تعداد سرعنوانها در هر یک از درخت‌های جی اچ اس، ۲/۱۹ است. پژوهشگران بر این باورند که ساختار درخت‌ها، ساختاری نامتعادل از روابط سلسله مراتبی را نشان می‌دهد و توصیه می‌کنند که لازم است تلاش‌هایی در زمینه پیش‌بینی‌پذیری و یکدستی^۶ سرعنوانها انجام پذیرد (Yi and Chan 2010).

در پیوند دو ابزار سازماندهی، یعنی فهرست سرعنوانها و رده‌بندی، پژوهش‌های عمیقی انجام شده است، به طوری که امکان برقراری ارتباط و تعامل میان سرعنوان‌های موضوعی کنگره با نظام رده‌بندی به منظور ایجاد نظامی که فرایند تخصیص شماره رده‌بندی را خودکار نماید، مورد مطالعه قرار گرفته است (Frank and Khosh-Khui 1985; Larson 1992; Dolin 1998; Paynter 2004).

خوشخویی در مجموعه‌ای از پیشینه‌های کتابشناختی به بررسی ارتباط احتمالی میان سرعنوان‌های موضوعی کنگره با نشانه‌های^۷ رده‌بندی دیویی و رده‌بندی کنگره پرداخته است. در مورد ارتباط میان سرعنوانهای کنگره و نشانه‌های رده‌بندی کنگره با سرعنوان‌های کنگره و نشانه‌های رده‌بندی دیویی اختلاف معنی‌دار گزارش شده است و درجه مشابهت سرعنوانهای کنگره و نشانه‌های رده‌بندی کنگره بیش از سرعنوانهای کنگره و نشانه‌های رده‌بندی دیویی است. وی همچنین نشان داده است که نشانه‌های رده اختصاص یافته به سرعنوانهای کنگره فاقد یکدستی قابل قبول است و هر سه ابزار رده‌بندی کنگره، رده‌بندی دیویی و سرعنوان‌های موضوعی کنگره نیازمند تقویت و بهبود ارتباط میان سرعنوانها و نشانه‌های رده‌بندی هستند (Khosh-Khui 1985).

فرانک و پینتر نیز در پژوهشی به بررسی پیش‌بینی‌پذیری شماره رده‌بندی کنگره از روی سرعنوان‌های موضوعی کنگره می‌پردازند و به طور خاص، بر مشکلات ناشی از انتساب خودکار شماره کنگره به آثار از روی سرعنوان‌های موضوعی تمرکز کرده‌اند (Frank and Paynter 2004). پژوهش‌های دیگری نیز چون لارسن (Larson 1992) و دالین (Dolin 1998) همچون فرانک و پینتر (Frank and Paynter 2004) به حوزه یادگیری ماشینی پرداخته‌اند. در این پژوهش‌ها، فرایند خودکارسازی اختصاص شماره رده‌بندی کنگره به منابع، با بهره‌گیری از سرعنوان‌های موضوعی

1. auxiliary science of history
4. Life

2. culture
5. rehabilitation

3. religion
6. consistent

7. notations

انجام شده است. پژوهش لارسن نشان داد که با در نظر گرفتن اولین سرعنوان موضوعی در اختصاص شماره رده‌بندی، ۴۶/۶ درصد کتاب‌های مورد بررسی شماره‌ای صحیح دریافت می‌کنند (Larson 1992). دالین نیز در پایان‌نامه دکترای خود نظامی به نام فراس^۱ طراحی کرد. در این پژوهش، ۱/۵ میلیون پیشینه برای یادگیری مورد استفاده قرار گرفته است و پژوهشگر، درستی ۱۴ درصدی نظام را گزارش کرده است. در این نظام نیز یادگیری ماشینی از فیلدهای عنوان و سرعنوان‌های موضوعی پیشینه‌ها انجام گرفته است (Dolin 1998).

در مقایسه‌ای کوتاه میان طرح‌های پژوهشی خارج از کشور و پژوهش‌های داخلی این واقعیت آشکار می‌گردد که تنوع موضوعی در پژوهش‌های خارج از کشور بیش از پژوهش‌های داخلی است، به طوری که به عنوان نمونه، امکان برقراری پیوند میان رده‌بندی کنگره و سرعنوان‌های موضوعی و تسهیل فرایند اختصاص خود کار شماره رده به منابع نیز مورد مطالعه قرار گرفته است. همچنین، تنوع در روش‌های اجرا در پژوهش‌های خارج از کشور بیش از پژوهش‌های داخلی است. در حالی که پژوهش‌های انجام گرفته در داخل، بیشتر بر شیوه‌های دستی استوار بوده‌اند، در پژوهش‌های انجام شده در خارج از کشور هر جا که لازم بوده از برنامه‌نویسی، ابزارهای موجود و روش‌های یادگیری ماشینی استفاده شده است. در برخی موارد، الگوریتم‌ها یا برنامه‌هایی نیز جهت عملیاتی شدن پژوهش‌ها نوشته شده است؛ طراحی نرم‌افزارهای ویژه‌ای چون VisaLCSH نمونه‌هایی از این دست است.

در این پژوهش، نظام مفهومی سرعنوان‌های موضوعی فارسی با تکیه بر انگاره‌های ساختارگرایی، ویژگی‌های شبکه‌های معنایی در نظریه شبکه‌های معنایی و خوشه‌های معنایی در نظریه حوزه‌های معنایی - که پیش‌تر به آنها اشاره شد - مورد بررسی قرار می‌گیرد و به منظور عملیاتی شدن کار، از برنامه‌نویسی، نگاشت، و برخی از ابزارهای وب‌معنایی استفاده می‌شود. با توجه به پیشینه موجود می‌توان گفت که تاکنون پژوهشی با این گستره (از نظر پوشش کل رکوردهای مستند موضوعی) و با تمرکز بر نظام مفهومی سرعنوان‌های موضوعی فارسی اجرا نشده است و پژوهشگران در هم‌سویی با پژوهش‌های صورت پذیرفته پیرامون فهرست سرعنوان‌های موضوعی کنگره تلاش داشته‌اند.

۴. روش پژوهش

پژوهش حاضر از نوع کاربردی است و به روش تحلیل محتوا انجام پذیرفته است.

1. Pharos

شناسایی نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی از طریق بررسی تمامی (سرشماری) سرعنوان‌های موجود در مستند موضوعی کتابخانه ملی ایران انجام پذیرفت (۸۴۵۴۵ رکورد با فرمت ایکس ام ال^۱، دریافت شده از سازمان اسناد و کتابخانه ملی ایران در تاریخ ۸۹/۱۰/۲۵). این در حالی است که پاسخگویی به پرسش ۱-۱ پژوهش با در نظر گرفتن شماره‌های رده اختصاص یافته به سرعنوان‌های موجود در فهرست سرعنوان‌های موضوعی فارسی، ویرایش سوم (سلطانی، فانی، و رهبری اصل ۱۳۸۱) و دو پیوست آن (سلطانی، فانی، و رهبری اصل ۱۳۸۲؛ ۱۳۸۵) انجام پذیرفته است (۱۱۲۳۳ رکورد). در تعیین انطباق میان نظام مفهومی سرعنوان‌ها با نظام دانشی ایران (پرسش ۱-۳)، مجموعه رشته‌های تحصیلی دانشگاه‌های کشور مندرج در راهنمای ثبت نام و شرکت در آزمون سراسری سال ۱۳۹۰ سازمان سنجش آموزش کشور و در تعیین میزان انطباق این نظام با نظام مفهومی رده‌بندی کنگره (پرسش دوم) طرح کلی رده‌بندی کتابخانه کنگره آمریکا^۲، موجود در وب‌سایت رسمی کتابخانه کنگره مورد توجه قرار گرفته است.

۵. ملاحظات گردآوری داده‌ها

به منظور گردآوری داده‌های مربوط به ساختار کلی سلسله مراتبی، بنا به نظر مشورتی متخصصان حوزه رایانه دانشگاه فردوسی مشهد، نگاهی^۳ از رکوردهای کتابخانه ملی ایران به قالب اسکاس (SKOS)^۴ انجام پذیرفت. به این منظور، جدول تطبیق عناصر مارک (MARC) با عناصر اسکاس (SKOS) با توجه به منابع (Harper 2006; Ed et al. 2008; International Federation of Library Associations and Institutions, 2001) و بررسی اجمالی رکوردهای سازمان اسناد و کتابخانه ملی ایران تهیه شد. طی جلسات مشترک، نرم افزارهای مناسب و چگونگی انجام نگاشت مورد توافق و در هر مرحله، خروجی کار از نظر درستی مورد بازنگری

1. XML 2. Library of Congress Classification Outline 3. Mapping
 4. Simple Knowledge Organization System: این استاندارد یکی از استانداردهای مطرح شده از سوی کنسرسیوم وب جهانی است و در کدگذاری نظام‌های دانش مانند اصطلاحنامه‌ها، سرعنوان‌های موضوعی و طرح‌های رده‌بندی مورد استفاده قرار می‌گیرد. RDF یکی از استانداردهای زبانی کنسرسیوم وب جهانی است و امکان تعریف خصیصه‌های منطقی را فراهم می‌آورد. این قالب از انتشار توزیع شده داده‌ها پشتیبانی می‌کند، به این ترتیب واژگان ساخت یافته و کنترل شده‌ای که با اسکاس منتشر شده باشد از قابلیت پیوند (link) و یکی شدن (merge) با سایر منابع داده‌ای از جمله نمایه‌های موضوعی یا سایر واژگان کنترل شده برخوردار است. به این ترتیب، نیازی به پایگاه‌های پیچیده و یکپارچه وجود نخواهد داشت (Miles and Perez- Aguera 2007, 70). همچنین، تبدیل رکوردهای مستند از استاندارد مارک به استاندارد اسکاس این امکان را فراهم می‌آورد تا بتوان در شناسایی و نمایش گرافیکی روابط معنایی از نرم‌افزارهای وب معنایی چون RDF-Gravity استفاده کرد.

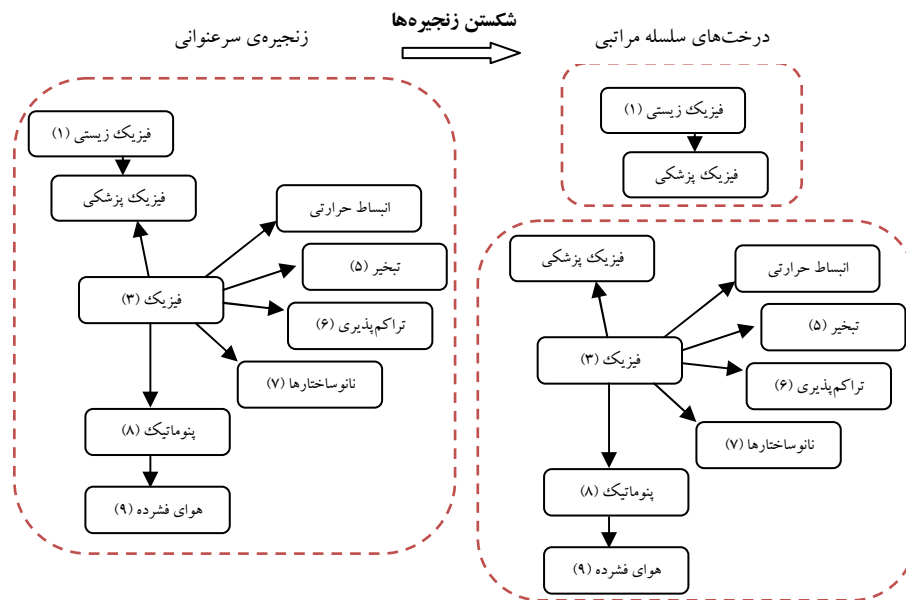
قرار گرفت. به منظور انجام این نگاشت از زبان برنامه‌نویسی جاوا (محیط NetBeans 6.9)^۱ و برای XQuery از API 3-9 Saxon^۲ استفاده شد. فایل XML رکوردهای کتابخانه ملی ایران به وسیله زبان XQuery ایجاد شد و درستی فایل RDF تولید شده در سایت رسمی کنسرسیوم وب جهانی^۳ مورد ارزیابی قرار گرفت. استفاده از رکوردهای XML سرعنوان‌های موضوعی فارسی سازمان اسناد و کتابخانه ملی ایران از یک سو و بهره‌گیری از زبان‌های برنامه‌نویسی استاندارد، نرم‌افزارهای مورد تأیید کنسرسیوم وب جهانی^۴، و ارزیابی درستی فایل RDF حاصل آمده، از سویی دیگر نشان از اعتبار داده‌های حاصل شده در این بخش از پژوهش است.

۶. یافته‌های پژوهش

۶-۱. ویژگی‌های نظام مفهومی حاکم بر سرعنوان‌های موضوعی

شبکه کلی رابطه سلسله مراتبی در این پژوهش، شبکه‌ای بزرگ از سرعنوان‌های به هم پیوسته است. اگر چه گسست‌ها و زنجیره‌های کوچکی نیز در آن به وجود می‌آید. به منظور تحلیل بهتر شبکه کلی حاصل شده، ابتدا زنجیره‌های پیوند یافته و سپس درخت‌های سلسله مراتبی تشکیل شده است. تصویر ۱ نحوه تشکیل زنجیره‌ها و سپس درخت‌های سلسله مراتبی را نمایش می‌دهد. با دقت در تصویر مشخص می‌گردد که برخی از سرعنوان‌ها (در این جا سرعنوان ۲) در بیش از یک درخت ظاهر شده‌اند. در این تصویر، یک زنجیره سرعنوانی شامل دو سرعنوان موضوعی (۱ و ۳) به همراه سرعنوان‌های اخص هر کدام با یک حلقه پیوندی به یکدیگر پیوند یافته‌اند. در واقع سرعنوان اخص ۲ برای سرعنوان ۱ و سرعنوان‌های اخص ۴ تا ۸ برای سرعنوان شماره ۳ شکل گرفته است و خود سرعنوان شماره ۸، سرعنوان اخص دیگری به شماره ۹ دارد. این زنجیره پس از شکسته شدن، دو درخت سلسله مراتبی را تشکیل داده است. شبکه کلی رابطه سلسله مراتبی مجموعه‌ای از همین زنجیره‌هاست. در مجموع، تعداد ۱۸۴۲ زنجیره در شبکه کلی رابطه سلسله مراتبی شکل گرفته است.

1. <http://netbeans.org/>
 2. <http://sourceforge.net/project/saxon>
 3. W3C RDF validation Service at <http://www.w3.org/RDF/Validator/>
 4. RDF-Gravity (Version 1.0) & IsaViz (version: 2.1, October 2004)



تصویر ۱. نمونه و نمای ساده شده‌ای از تشکیل زنجیره‌ها و درخت‌های سلسله مراتبی

آن گونه که از شمارش زنجیره‌ها برمی‌آید، درصد بالایی از سرعنوان‌های موضوعی فقط دارای یک مدخل سرعنوانی به همراه سرعنوان‌های اخص آن هستند. پس از آن، ۲۷۴ زنجیره با دو مدخل سرعنوانی شکل گرفته است. هر چه بر بزرگی زنجیره و تعداد سرعنوان‌هایی که در زنجیره مدخل قرار گرفته‌اند، افزوده می‌شود، فراوانی زنجیره‌ها کاهش می‌یابد. به طوری که بزرگ‌ترین زنجیره شکل گرفته در این شبکه، دارای ۳۹۵ مدخل سرعنوانی است^۱.

۱-۱-۶. تنوع، حجم و نسبت سرعنوان‌های موضوعی در هر یک از رده‌های موضوعی

یکی از ویژگی‌های فهرست سرعنوان‌های موضوعی، بیان شماره رده برای موضوعات است. به این ترتیب ارتباط منطقی میان این دو ابزار سازماندهی و رده‌بندی فراهم می‌آید. بیان شماره رده در تمامی مدخل‌های سرعنوانی تحقق نیافته و در مقابل گاه در برخی از مدخل‌ها

۱. در این بزرگ‌ترین زنجیره به وجود آمده در مجموع تعداد ۱۳۰۷ سرعنوان حضور دارند که ۳۹۵ سرعنوان از آن‌ها مدخل بوده‌اند. هر یک از این ۳۹۵ سرعنوان دارای سرعنوان‌های اخصی می‌باشند. از این روست که مجموع تعداد سرعنوان‌ها - سرعنوان‌هایی که مدخل هستند و سرعنوان‌های اخص آن‌ها - عدد ۱۳۰۷ شده است. لازم به ذکر است که هر یک از این ۱۳۰۷ سرعنوان موضوعی پذیرفته شده‌اند.

بیش از یک شماره رده ثبت شده است. چگونگی برقراری پیوند ۱۱۲۳۳ رکورد مورد توجه بوده و از طریق کاوش در فهرست پیوسته^۱ سرعنوان‌های موضوعی فارسی در وب‌گاه سازمان اسناد و کتابخانه ملی ایران^۲ بازیابی و ثبت شده است. درصد‌های محاسبه شده برای تعداد شماره رده‌های مندرج در هر سرعنوان موضوعی بیانگر این مطلب است که پیوند میان موضوع و شماره رده در ۴۶/۹ درصد از سرعنوان‌ها برقرار نشده است. در عین حال، برای ۳۹/۵ درصد از سرعنوان‌های مورد بررسی دست کم یک شماره رده کنگره ثبت شده است. هر چه بر تعداد شماره‌های رده ثبت شده در سرعنوان افزوده می‌شود، از تعداد سرعنوان‌ها کاسته می‌شود. به طوری که ۹/۴۵ درصد از سرعنوان‌ها، دو شماره رده، ۳/۸۸ درصد، سه شماره رده، ۰/۲۰ درصد، چهار شماره رده، و ۰/۰۴ درصد، پنج شماره رده داشته‌اند. برای هر یک از طبقات شش، هفت و ده شماره رده‌ای، فقط یک سرعنوان موضوعی (معادل ۰/۰۱٪ برای هر یک از طبقات) در فهرست سرعنوان‌ها مشاهده شد. در مجموع، ۸۰۳۷ شماره رده برای ۱۱۲۳۳ رکورد مورد بررسی ثبت شده است که متوسط شماره رده‌های ارائه شده در این فهرست به ازای هر سرعنوان موضوعی، معادل ۰/۷۱ و کمتر از یک است.

در جدول ۱ چگونگی پراکندگی ۸۰۳۷ شماره رده ثبت شده برای ۱۱۲۳۳ رکورد سرعنوانی به تفکیک رده‌های رده‌بندی کنگره آمده است. با توجه به این جدول، تنوع و نسبت سرعنوان‌های موضوعی در هر یک از رده‌های رده‌بندی کنگره آشکار می‌گردد.

جدول ۱. تنوع و نسبت شماره‌های رده ثبت شده به تفکیک رده‌های فرعی رده‌بندی کنگره

رده	ترجمه رده	فراانما	ترجمه فراانما	فراوانی رده	درصد
A	کلیات	A	کلیات: آثار عمومی	۲۲	۰/۲۷
B	فلسفه، روان‌شناسی، مذهب	B-BJ	فلسفه، روان‌شناسی	۵۲۸	۶/۵۷
		BL, BM, BP ^۳ , BQ	دین: ادیان هندوئیسم، یهودیت، اسلام، بودیسم	۲۰۴	۲/۵۴
		BR-BV	دین: مسیحیت، کتاب مقدس	۱۱۲	۱/۳۹
		BX	دین: فرقه‌های مسیحی	۲۳	۰/۲۹

←

1. Online Catalog

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

۳. تعداد سرعنوان‌های موجود در نمونه مورد بررسی که شماره رده دین اسلام را گرفته‌اند، ۷۸ مورد از ۲۰۴ مورد است.

→

درصد	فراوانی رده	ترجمه فرانما	فرانما	ترجمه رده	رده
۰/۹۵	۷۶	علوم کمکی تاریخ	C	تاریخ	C-F
۰/۸۸	۷۱	آثار کلی (عمومی) تاریخ، تاریخ اروپا (قسمت اول)	D-DJ		
۰/۱۹	۱۵	تاریخ اروپای شرقی: عمومی، اتحاد جماهیر شوروی، لهستان	DJK-DK		
۰/۱۰	۸	تاریخ اروپا (قسمت دوم)	DL-DR		
۱/۰۰	۸۰	تاریخ آسیا، آفریقا، استرالیا، زلاندنو و غیره	DS-DX		
۰/۳۲	۲۶	تاریخ: آمریکا	E-F		
۵/۳۶	۴۳۱	جغرافیا، نقشه‌ها، مردم‌شناسی، سرگرمی	G	جغرافیا، مردم‌شناسی، اوقات فراغت	G
۱۰/۶	۸۵۰	علوم اجتماعی	H	علوم اجتماعی	H
۱/۷۲	۱۳۸	علوم سیاسی	J	علوم سیاسی	J
۰/۱۲	۱۰	حقوق عمومی	K	حقوق	K
-	-	جداول رده K: جداول تقسیمات شکلی حقوقی	K		
۰/۰۰	۰	حقوق انگلستان و ایرلند	KD		
۰/۰۱	۱	حقوق آمریکا، آمریکای لاتین، جزایر هند غربی	KDZ, KG-KH		
۰/۰۰	۰	حقوق کانادا	KE		
۰/۰۰	۰	حقوق ایالات متحده	KF		
۰/۰۰	۰	حقوق اروپا	KJ-KKZ		
۰/۰۰	۰	حقوق فرانسه	KJV-KJW		
۰/۰۰	۰	حقوق آلمان	KK-KKC		
۰/۰۴	۳	حقوق اروپا - آسیا، آفریقا، اقیانوس آرام و قطب جنوب	KL-KWZ		
۰/۰۱	۱	حقوق ملل	KZ		

←

→

درصد	فراوانی رده	ترجمه فرانما	فرانما	ترجمه رده	رده
۲/۴۰	۱۹۳	آموزش و پرورش	L	آموزش	L
۱/۵۴	۱۲۴	موسیقی و کتب موسیقی	M	موسیقی	M
۴/۶۷	۳۷۵	هنرهای زیبا	N	هنرهای زیبا	N
۰/۸۷	۷۰	لغت‌شناسی و زبان‌شناسی (آثار کلی)، زبان و ادبیات یونانی، زبان و ادبیات لاتین	P-PA	زبان و ادبیات	P
-	-	جداول زبان و ادبیات	P-PZ		
۰/۵۶	۴۵	زبان‌های جدید اروپا	PB-PH		
۰/۰۰	۰	ادبیات روسی	PG		
۰/۰۲	۲	ادبیات فارسی	PIR		
۱/۲۷	۱۰۲	لغت‌شناسی و ادبیات شرقی، لغت‌شناسی و ادبیات هندوایرانی	PJ-PK		
۰/۶۳	۵۱	زبانهای آسیای شرقی، آفریقا، اقیانوسه؛ هایپریوری، هندی و زبان‌های ساختگی	PL-PM		
-	-	ضمیمه‌ها: نمایه زبان‌ها و لهجه‌ها	P-PM		
۲/۳۳	۱۸۷	ادبیات عمومی	PN		
۰/۶۶	۵۳	ادبیات انگلیسی و آمریکایی، ادبیات جوانان	PR, PS, PZ		
۰/۵۲	۴۲	ادبیات فرانسوی، ایتالیایی، اسپانیایی و پرتغالی	PQ		
۰/۳۹	۳۱	بخش اول: ادبیات آلمانی	PT (Part1)		
		بخش دوم: ادبیات هلندی و اسکاندیناوی	PT (Part2)		
۱۹/۵۳	۱۵۷۰	علوم	Q	علوم	Q
۹/۴۹	۷۶۳	پزشکی	R	پزشکی	R
۴/۷۸	۳۸۴	کشاورزی	S	کشاورزی، معماری	S
۱۴/۴۳	۱۱۶۰	فناوری	T	فناوری	T
۱/۱۱	۸۹	علوم نظامی، علوم دریایی	U-V	علوم نظامی	U

←

→

رده	ترجمه رده	فرانما	ترجمه فرانما	فراوانی رده	درصد
V	علوم دریایی			۳۸	۰/۴۷
Z	کتابشناسی، نشر، کتابداری	Z	کتابشناسی و علم کتابداری	۱۵۹	۱/۹۸
مجموع				۸۰۳۷	۱۰۰

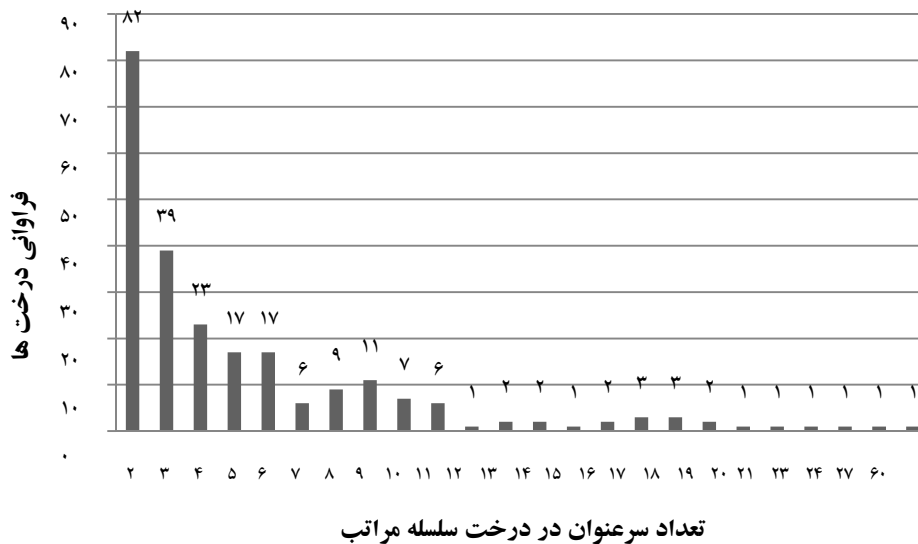
با استناد به درصد‌های به دست آمده از جدول ۱ چنین به نظر می‌رسد که بیشترین تعداد سرعنوان‌های موضوعی به ترتیب در حوزه علوم (۱۹/۵٪)، فناوری (۱۴/۴٪)، علوم اجتماعی (۱۰/۶٪)، پزشکی (۹/۵٪) و فلسفه و روان‌شناسی (۶/۶٪) ساخته شده است. اما می‌توان چنین نیز بیان داشت که بیشترین ارتباط میان نظام موضوعی و رده‌بندی کنگره در این رده‌ها برقرار شده است. همچنین، با توجه به درصد بسیار پایین شماره رده‌هایی مانند PIR که مربوط به گسترش‌های فارسی است، می‌توان چنین نتیجه‌گیری کرد که شماره‌های رده برای موضوعاتی که به طور مستقیم از فهرست سرعنوان‌های کنگره ترجمه شده، در فهرست سرعنوان‌های فارسی منعکس گشته است و به نظر می‌رسد تلاشی جدی در جهت تعیین شماره رده برای موضوعات ساخته شده در کتابخانه ملی ایران از سوی مستندسازان کتابخانه ملی ایران انجام پذیرفته است. تأیید این گمان، در مقدمه سرعنوان‌های موضوعی فارسی (سلطانی، فانی، رهبری اصل ۱۳۸۱، نوزده) آمده است.

آزمون آماری مربع کای^۱ (خی دو)، از نظر اختصاص شماره رده، تفاوت معنی‌داری را میان موضوعاتی که دارای معادل در فهرست سرعنوان‌های کنگره هستند و موضوعات فاقد معادل در آن فهرست نشان می‌دهد ($\chi^2=1987/743$ ، $P\text{-value}=0/00$ ، $df=1$). به این ترتیب، بین داشتن معادل در فهرست سرعنوان‌های موضوعی کنگره و داشتن شماره رده‌بندی، رابطه‌ای قوی برقرار است. به بیانی، آن دسته از موضوعاتی که بر مبنای فهرست سرعنوان‌های موضوعی کنگره ساخته شده‌اند - در کنگره معادل دارند - از نظر داشتن شماره رده‌بندی، متفاوت با آن دسته از سرعنوان‌هایی هستند که توسط کتابخانه ملی ایران ساخته شده‌اند. از میان ۱۷۱۴ سرعنوان ساخته شده توسط کتابخانه ملی ایران، فقط ۷/۵ درصد آنها دارای شماره رده هستند و تلاش اندکی برای اختصاص شماره رده به موضوعات از سوی مستندسازان کتابخانه ملی ایران انجام پذیرفته است. این در حالی است که به ۶۱/۳ درصد موضوعات ساخته شده در کتابخانه کنگره آمریکا، شماره رده اختصاص یافته است.

1. Chi-Square

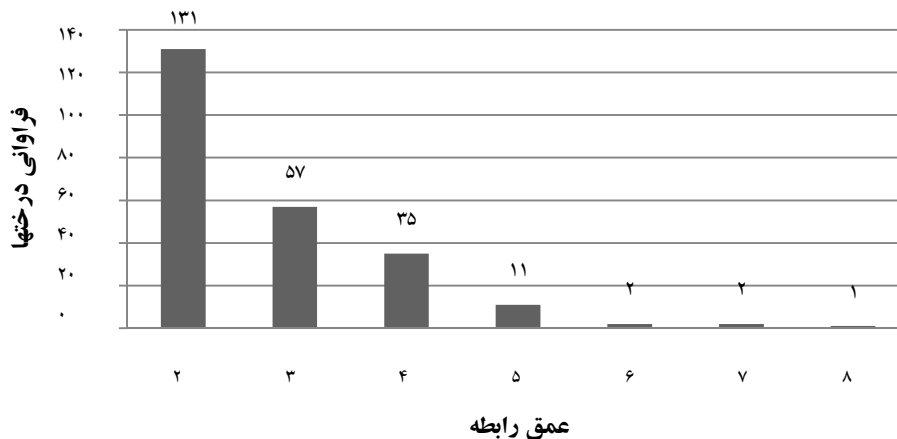
۶-۱-۲. شناسایی طبقات مفهومی موجود در فهرست با توجه به درخت‌های سلسله مراتبی

در پاسخ به پرسش شماره ۱، توضیحاتی پیرامون چگونگی تشکیل زنجیره‌های سرعنوانی و همچنین فرایند شکستن زنجیره‌ها و تشکیل درخت‌های مستقل ارائه شد (تصویر ۱). در همان بخش مشخص شد که بزرگ‌ترین زنجیره سرعنوانی شامل ۱۳۰۷ سرعنوان موضوعی است که ۳۹۵ مورد از آنها مدخل سرعنوانی هستند. پس از شکستن زنجیره‌ها در این بزرگ‌ترین زنجیره، ۲۴۱ درخت سلسله مراتبی تشکیل شد. نمودار ۱ تعداد سرعنوان‌ها را در هر یک از ۲۴۱ درخت سلسله مراتبی به دست آمده نشان می‌دهد.



نمودار ۱. تعداد سرعنوان‌های موضوعی در درخت‌های سلسله مراتبی

تعداد ۸۲ درخت (۳/۳۴٪) از درخت‌های حاصل شده، فقط شامل دو سرعنوان موضوعی هستند. بزرگ‌ترین درخت سلسله مراتبی نیز شامل ۶۰ سرعنوان موضوعی است. هر یک از این درخت‌ها عمق متفاوتی از رابطه سلسله مراتبی را دارا هستند. نمودار ۲ چگونگی عمق رابطه سلسله مراتبی را در ۲۴۱ درخت شناسایی شده نشان می‌دهد.



نمودار ۲. چگونگی عمق رابطه در درخت‌های سلسله مراتبی

۱۳۱ درخت سلسله مراتبی یعنی ۵۴/۸ درصد فقط یک عمق از رابطه سلسله مراتبی را دارا هستند و در مجموع ۹۸/۲ درصد از درخت‌ها، ۲ تا ۵ مرتبه اخص شده‌اند. بیشترین عمق رابطه سلسله مراتبی، ۸ و مربوط به یک درخت است. میانگین تعداد سرعنوان در هر درخت، ۵/۷۰ و میانگین عمق رابطه نیز ۲/۸۱ است.

از میان ۲۴۱ درخت سلسله مراتبی به دست آمده، بزرگ‌ترین درخت‌ها^۱ گزینش و نخستین سرعنوان آن درخت که از نظر سلسله مراتبی بالاتر از سایر سرعنوان‌ها قرار گرفته است، به منزله نماینده آن درخت و با نام سرعنوان رأس نام‌گذاری شد. ۳۸ سرعنوان رأس به دست آمده به این شرح است: پزشکی و روان‌شناسی (۶۰-۸)، فلسفه (۲۷-۵)، قراردادها (۲۴-۴)، جامعه‌شناسی (۲۴-۶)، ریخته‌گری (۲۳-۷)، ریاضیات (۲۱-۵)، مدیریت (۲۰-۵)، حقوق بین‌الملل (۲۰-۴)، حقوق تجارت (۱۹-۴)، دفاع از غیرنظامیان (۱۸-۵)، کشاورزی (۱۸-۴)، ادراک (۱۷-۵)، حقوق جزا (۱۷-۴)، کودکان (۱۷-۴)، نفس (۱۶-۶)، شیمی (۱۶-۴)، روش‌شناسی (۱۵-۶)، مهندسی (۱۴-۴)، دادگاه‌ها (۱۴-۴)، خون - گردش (۱۳-۶)، پزشکی قانونی (۱۳-۳)، قانون دریایی (۱۳-۴)، روابط بین‌الملل (۱۳-۴)، مدیریت صنعتی (۱۲-۳)، مواد

۱. ملاک بزرگ‌ترین درخت‌ها، تعداد سرعنوان‌ها در درخت است. در این پژوهش، درخت‌هایی که بیش از ۱۰ سرعنوان دارند در گروه درخت‌های بزرگ قرار گرفته‌اند.

۲. نخستین عدد مربوط به تعداد سرعنوان‌ها در درخت و دومین عدد مربوط به عمق رابطه است. در این درخت، ۶۰ سرعنوان موجودند و عمق رابطه معنایی، ۸ است.

چندسازه (۴-۱۱)، ژنتیک (۴-۱۱)، جرم‌یابی (۲-۱۱)، جوانان (۴-۱۱)، روان‌درمانی (۲-۱۱)، منطق (۴-۱۰)، کم‌خونی (۴-۱۰)، بیماری‌ها (۴-۱۰)، جنین‌شناسی (۳-۱۰)، دریانوردی (۱۰-۳)، آیین دادرسی (۴-۱۰)، پزشکی قرون وسطا (۴-۱۰).

در مقایسه، پژوهش‌های و چنان به شناسایی ۱۵۵۹۰۵ درخت سلسله مراتبی کلی در فهرست سرعنوان‌های موضوعی کنگره انجامید. سرعنوان موضوعی علم با ۱۴۵۵۰۲۶ سرعنوان در ساختار سلسله مراتبی‌اش بزرگترین درخت را تشکیل داده است. پس از آن تاریخ (۷۶۹۵۲۶ سرعنوان)، تمدن (۳۶۷۹۶۴ سرعنوان)، مذهب (۲۹۲۶۳۹ سرعنوان)، زندگی (۱۴۵۰۵۰ سرعنوان) و توانبخشی (۱۰۳۴۹۵ سرعنوان) قرار گرفته‌اند (Yi and Chan 2010). هیچ یک از درخت‌های تشکیل شده در فهرست سرعنوان‌های موضوعی فارسی مشابه با درخت‌های کنگره نیست. با دقت در درخت‌های حاصل شده در فهرست سرعنوان‌های فارسی به نظر می‌رسد تمام این درخت‌ها، می‌تواند زیرمجموعه درخت بزرگ‌تری یعنی "علم" قرار گیرد. همچنین، تفاوتی فراوان میان تعداد سرعنوان‌ها در هر درخت سلسله مراتبی در هر یک از این دو ابزار سرعنوانی مشاهده می‌شود. بزرگ‌ترین درخت در فهرست فارسی، پزشکی و روان‌شناسی است که فقط شامل ۶۰ سرعنوان است، در حالی که بزرگ‌ترین درخت فهرست کنگره شامل ۱۴۵۵۰۲۶ سرعنوان موضوعی - حدود ۲۴۲۵۰ برابر - است.

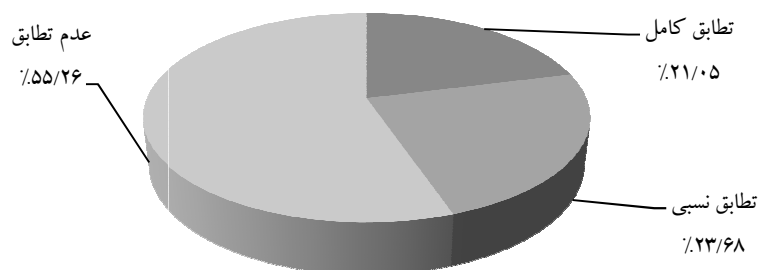
این در حالی است که فرناندز در مقایسه عبارات‌های عنوانی استخراج شده از عنوان منابع با سرعنوان‌های موضوعی کنگره به این یافته رسید که عمق رابطه معنایی سلسله مراتبی در فهرست سرعنوان‌های کنگره اندک است و عنوان منابع اطلاعاتی اخص‌تر از سرعنوان‌های موضوعی است (Fernandez 1991). با توجه به این یافته، آن جایی که عمق رابطه معنایی در فهرست کنگره اندک ارزیابی شده است، نمی‌توان عمق رابطه در فهرست سرعنوان‌های فارسی را در وضعیت مطلوبی ارزیابی کرد. اگر چه تعیین درستی رابطه سلسله مراتبی درخت‌های حاصل شده خارج از دایره تحلیل‌های این پژوهش است، اما با اندک دقتی در درخت‌ها اشکالاتی ساختاری قابل مشاهده است. تشکیل حلقه‌های بسته به واسطه تعریف یک سرعنوان در دو سطح از سلسله مراتب از آن جمله است^۱. در فهرست سرعنوان‌های کنگره نیز ساختاری

۱. به عنوان نمونه در درخت شماره ۷۸ از مجموعه درخت‌های شناسایی شده در پژوهش، سرعنوان موضوعی "بیماری‌های شغلی" با دو سرعنوان اخص یعنی "ریه‌ها -- بیماری‌های ناشی از استنشاق غبار" و "سندرم ارتعاش" همراه شده است. موضوع "سندرم ارتعاش"، خود سرعنوان اخصی دارد که آن "بیماری‌های شغلی" است. به این ترتیب موضوع "بیماری‌های شغلی" در دو سطح از سلسله مراتب در این درخت تکرار شده و حلقه‌ای بسته تشکیل داده است.

نامتعادل از روابط سلسله مراتبی گزارش شده است. پژوهشگران توصیه کرده‌اند تا به منظور بهره‌گیری از اطلاعات ارزشمند معنایی، تعریف دقیق روابط سلسله مراتبی در دستور کار کتابخانه کنگره قرار گیرد، چرا که در شرایط فعلی در فهرست کنگره، محتوای سلسله مراتبی مبهم و نامشخصی از سرعنوان‌های مرتبط با یکدیگر شکل گرفته است (Yi and Chan 2010).

۳-۱-۶. مقایسه نظام مفهومی شناسایی شده با نظام دانشی موجود در ایران

در مجموعه ۲۴۱ درخت سلسله مراتبی، علاوه بر سرعنوان‌های رأسی که معرفی شدند، سرعنوان‌هایی چون فیزیک، مخابرات، مهندسی سیستم‌ها، علوم زیستی، آمار، گیاه‌شناسی، جغرافیای انسانی نیز مشاهده می‌شود که با رشته‌های دانشگاهی در ایران همخوان است. در عین حال ۳۸ سرعنوان رأس بزرگ‌ترین درخت‌های سلسله مراتبی که مشخصات آنها بیان شد، به منزله بازنمونی از نظام مفهومی مورد نظر قرار گرفت. تطابق این نظام مفهومی با نظام دانشی موجود در ایران در دو لایه تطابق واژگانی عنوان رشته تحصیلی و تطابق دامنه موضوعی یا محتوایی رشته‌های تحصیلی دوره کارشناسی مندرج در دفترچه راهنمای آزمون سراسری ۱۳۹۰ مورد سنجش قرار گرفت.

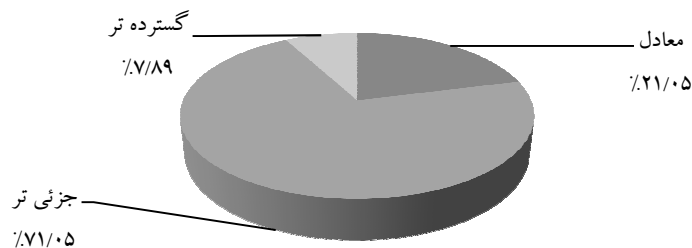


نمودار ۳. تطابق واژگانی سرعنوان‌های نشانه‌گر نظام مفهومی در سرعنوان‌های موضوعی فارسی با عنوان رشته‌های تحصیلی دانشگاهی در ایران (n=۳۸)

نمودار ۳ نشان می‌دهد که ۵۵/۲۶ درصد یعنی بیش از نیمی از سرعنوان‌های رأس با عنوان رشته‌های تحصیلی تطابق واژگانی ندارد و فقط ۲۱/۰۵ درصد از سرعنوان‌های رأس در تطابق کامل با عنوان رشته‌های تحصیلی در ایران است.

در تعیین تطابق محتوایی و دامنه موضوعی سرعنوان‌های رأس با رشته‌های تحصیلی در ایران، کدگذاری علاوه بر پژوهشگر، توسط دو کدگذار دیگر از دانشجویان دوره دکترای

کتابداری و اطلاع‌رسانی، انجام پذیرفت^۱. نمودار ۴ حاصل توافق کدگذاران است. نسبت توافق اسکات^۲ میان کدگذاران به صورت دو به دو ۰/۷۷، ۰/۸۱ و ۰/۷۱ و میانگین توافق‌ها ۰/۷۶ است. با توجه به نمودار ۴، مشخص است که مجموعه سرعنوان‌های رأس شناسایی شده جزئی‌تر از محتوای رشته‌های تحصیلی در ایران است. در عین حال، ۲۱/۰۵ درصد از سرعنوان‌های رأس در تطابق کامل با محتوای رشته‌های تحصیلی در ایران هستند. به این معنی که سرعنوان رأس شناسایی شده به طور دقیق موضوع اصلی در آن رشته تحصیلی محسوب می‌شود.



نمودار ۴. تطابق دامنه موضوعی سرعنوان‌های نشانگر نظام مفهومی در سرعنوان‌های موضوعی فارسی با محتوای رشته‌های تحصیلی دانشگاهی در ایران (n=۳۸)

در سنجش تطابق، فقط ۳۸ سرعنوان رأس به منزله نماینده نظام مفهومی با رشته‌های تحصیلی دانشگاهی ایران مورد مقایسه قرار گرفت. با توجه به این واقعیت که رشته‌های تحصیلی در ایران با استناد به دفترچه راهنمای آزمون سراسری ۱۳۹۰، ۱۴۶ عنوان رشته در گروه علوم ریاضی و فنی، ۱۱۸ عنوان رشته در گروه علوم تجربی، ۱۱۷ عنوان رشته در گروه علوم انسانی، ۴۴ عنوان رشته در گروه هنر، و ۱۸ عنوان رشته در گروه زبان‌های خارجی معرفی شده است^۳. چنانچه سرعنوان‌های رأس گسترده‌تر از رشته‌های تحصیلی بودند این احتمال وجود داشت که چند رشته تحصیلی زیر یک سرعنوان موضوعی رأس و در یک درخت سلسله مراتبی مورد توجه باشند. این در حالی است که ۷۱/۰۵ درصد از سرعنوان‌های رأس، محتوایی

۱. در سنجش تطابق محتوایی سرعنوان‌های رأس با رشته‌های تحصیلی، دفترچه کنکور سراسری سال ۱۳۹۰ در اختیار کدگذاران قرار گرفت تا با توجه به رشته‌های موجود در دفترچه و بدون در نظر گرفتن تطابق واژگانی میان سرعنوان‌های رأس و عنوان رشته‌ها، نظرشان را پیرامون گسترده‌تر، جزئی‌تر و یا معادل بودن سرعنوان رأس با محتوای کلی رشته‌های تحصیلی موجود در دفترچه مشخص سازند.

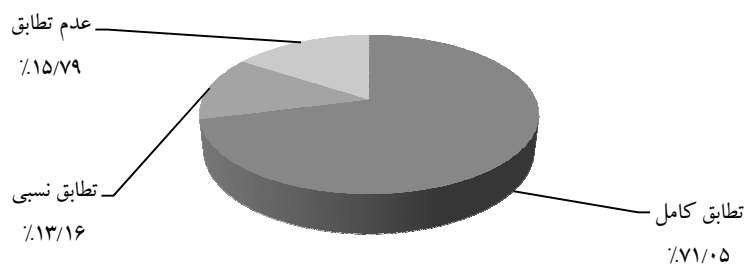
2. Scott's pi

۳. تعدادی از عنوان‌های شمارش شده در این گروه‌ها، در چند گروه تکرار شده‌اند. در این جا موارد تکراری حذف نشده است.

جزئی‌تر از رشته‌های تحصیلی در ایران دارند. به این ترتیب و با احتیاط می‌توان چنین استنباط کرد که در مجموع، نظام مفهومی شناسایی شده در فهرست سرعنوان‌های موضوعی فارسی از تطابق اندکی با محتوای رشته‌های تحصیلی و در نتیجه، نظام دانشی در ایران برخوردار است و بخش کوچکی از محتوای رشته‌های دانشگاهی ایران به شکل درخت‌های سلسله‌مراتبی در فهرست سرعنوان‌های موضوعی فارسی بازنمایی شده است.

اما اظهار نظر قطعی در این مورد که آیا موضوعات ساخته شده، متناسب با رشته‌های تحصیلی و نظام دانشی در ایران گسترش یافته‌اند یا خیر، زمانی دشوار می‌شود که بدانیم اصل ساخت موضوعات اخص و نه عمومی و کلی در فهرست سرعنوان‌های موضوعی، اصلی پذیرفته شده است. مشخص است که مجموعه موضوعات مطرح در یک رشته تحصیلی، محتوای آن رشته را می‌سازد و الزامی قطعی در تطابق کامل میان موضوعات ساخته شده و عنوان رشته تحصیلی و محتوای آن با سرعنوان‌های رأس وجود ندارد. بنابراین این امکان وجود دارد که محتوای موضوعی رشته‌های مختلف به صورت پراکنده و بدون برقراری پیوندهای سلسله‌مراتبی قوی ساخته شده و در قالب درخت‌های سلسله‌مراتبی ظاهر نشده باشند.

۲-۶. تعیین سطح هم‌خوانی میان نظام مفهومی حاصل از درخت‌های GHS با نظام مفهومی رده‌بندی کنگره در پاسخ به این پرسش، با روالی مشابه با پرسش فرعی ۱-۳، معادل انگلیسی ۳۸ سرعنوان رأس شناسایی شده، نخست از نظر تطابق واژگانی با عبارت‌های توضیحگر رده‌ها در رده‌بندی کنگره مورد بررسی قرار گرفت. سپس، گسترده‌گی دامنه شماره‌های اختصاص یافته به آن موضوع در طرح رده‌بندی تعیین گردید. نمودار ۵ میزان تطابق واژگانی نظام مفهومی حاکم بر سرعنوان‌های موضوعی را با نظام مفهومی رده‌بندی کنگره نشان می‌دهد.



نمودار ۵. تطابق واژگانی سرعنوان‌های نشانگر نظام مفهومی در سرعنوان‌های موضوعی فارسی با رده‌های اصلی و فرعی متناظر در طرح کلی رده‌بندی کنگره (n=۳۸)

همان گونه که در نمودار ۵ مشاهده می شود، ۷۱/۰۵ درصد (۲۷ مورد از ۳۸ مورد) از سرعنوان های رأس با تطابق کامل در طرح رده بندی کنگره حضور دارند. میزان تطابق نسبی ۱۳/۱۶ درصد (۵ مورد از ۳۸ مورد) است و ۱۵/۷۹ درصد (۶ مورد از ۳۸ مورد) از سرعنوان های رأس، با عبارتی که در فهرست سرعنوان های موضوعی کنگره آمده است، در طرح کلی رده بندی کنگره بازیابی نشده اند. جدول ۲ در بررسی گستره دامنه رده های متناظر با سرعنوان های رأس شناسایی شده، رسم شد.

جدول ۲. گستره دامنه رده های اصلی و فرعی متناظر با سرعنوان های رأس (n=۳۸)

گستره دامنه	فراوانی	درصد فراوانی
گسترده تر از یک رده اصلی	۱	۲/۶
به گستردگی یک رده اصلی	۱	۲/۶
گسترده تر از یک رده فرعی	۷	۱۸/۴
به گستردگی یک رده فرعی	۳	۷/۹
کوچکتر از یک رده فرعی	۱۶	۴۲/۱
پراکندگی در چندین رده اصلی	۴	۱۰/۵
عدم وجود	۶	۱۵/۸
مجموع	۳۸	۱۰۰

بیشترین درصد تعلق یافته در جدول ۳، یعنی ۴۲/۱ درصد، متعلق به سرعنوان های رأسی است که در طرح کلی رده بندی کنگره حضوری کوچک تر از یک رده فرعی داشته اند. پس از آن، ۷ سرعنوان رأس (۱۸/۴٪) وجود دارند که متناظر با وضعیتی گسترده تر از یک رده فرعی هستند. به این معنی که علاوه بر تمام یک رده فرعی، در سایر بخش های طرح کلی رده بندی مشاهده شده اند. در این صورت، در بیشتر موارد، حضور سرعنوان رأس در رده های دیگر به شکل ترکیبی و با تطابق نسبی بوده است. به طور مثال، موضوع ریاضیات^۱ رده فرعی QA را به خود اختصاص داده است و با ترکیب هایی چون ریاضیات کاربردی^۲ در رده فرعی T و ریاضیات مهندسی^۳ در رده TA ظاهر شده است. از سویی دیگر، ۱۵/۸ درصد از سرعنوان های رأس از طریق تطابق واژگانی در طرح کلی رده بندی کنگره بازیابی نشده اند.

سرعنوان های رأس در نظام مفهومی سرعنوان های موضوعی فارسی، تمام رده های اصلی R

پزشکی) و S (کشاورزی) و رده‌های فرعی B (فلسفه)، BC (منطق)، BF (روان‌شناسی)، HM (جامعه‌شناسی)، JX (حقوق بین‌الملل)، QA (ریاضیات)، QD (شیمی)، TA (مهندسی)، و VK (دریانوردی) را به خود اختصاص داده‌اند. در بخش تحلیل شماره‌های رده ثبت شده برای سرعنوان‌های موضوعی فارسی نیز مشخص شد که بیشترین تعداد سرعنوان‌های موضوعی به ترتیب در رده Q متعلق به حوزه علوم (۱۹/۵٪)، رده T مربوط به حوزه فناوری (۱۴/۴٪)، رده H مربوط به علوم اجتماعی (۱۰/۶٪)، رده R مربوط به پزشکی (۹/۵٪)، و رده B حوزه فلسفه و روان‌شناسی (۶/۶٪) ساخته شده است و یا می‌توان چنین نیز بیان داشت که بیشترین ارتباط میان نظام موضوعی و رده‌بندی کنگره در این رده‌ها برقرار شده است. این دو یافته، یکدیگر را تقویت می‌کنند.

در مجموع می‌توان چنین استنباط کرد که نظام مفهومی در فهرست سرعنوان‌های موضوعی فارسی فقط در تعداد معدودی از حوزه‌ها نظم یافته‌اند و در سایر حوزه‌ها پیوند میان موضوعات به واسطه رابطه سلسله مراتبی، آن گونه که انتظار می‌رود، برقرار نشده است. با نگاهی دوباره به سرعنوان‌های رأس شناسایی شده، این واقعیت آشکار می‌شود که هیچ‌گونه ساخت سلسله مراتبی از موضوعات بومی و ملی مانند ادبیات فارسی و یا اسلام به چشم نمی‌خورد.^۱ این مطلب در مورد تمام ۲۴۱ درخت سلسله مراتبی شناسایی شده نیز به چشم می‌خورد.

پیش از این، در مورد فهرست سرعنوان‌های موضوعی کنگره خوشخویی با بررسی فیلدهایی از رکوردهای کتابشناختی پیوند میان سرعنوان‌های موضوعی کنگره و شماره رده‌بندی کنگره را مورد بررسی قرار داد. یافته‌های وی نشانگر آن است که درجه مشابهت سرعنوان‌های موضوعی کنگره و نشانه‌های رده‌بندی کنگره بیش از سرعنوان‌های کنگره و نشانه‌های رده‌بندی دیویی است. وی همچنین نشان داد که نشانه‌های رده اختصاص یافته به سرعنوان‌های کنگره فاقد یکدستی قابل قبول است و هر سه ابزار رده‌بندی کنگره، رده‌بندی دیویی و سرعنوان‌های موضوعی کنگره نیازمند تقویت و بهبود ارتباط میان سرعنوان‌ها و نشانه‌های رده‌بندی هستند (Khosh-Khui 1985). یافته‌های پژوهش حاضر نشان داد که مشابه با فهرست سرعنوان‌های موضوعی کنگره، فهرست سرعنوان‌های موضوعی فارسی نیز نیازمند تقویت ارتباط با رده‌بندی کنگره است.

۱. در این جا تأکید بر ساخت سلسله مراتبی است و جایگاه این موضوعات در ساخت کلی نظام مفهومی سرعنوان‌های موضوعی فارسی مورد توجه قرار گرفته است. بنابراین امکان وجود پراکنده و ساخت نیافته این دست از موضوعات در فهرست سرعنوان‌ها نفی نمی‌شود.

پیوند میان دو ابزار فهرست سرعنوان‌ها و رده‌بندی کنگره از دو مسیر امکان‌پذیر است: نخست اختصاص شماره رده به سرعنوان‌های موضوعی و دوم تطابق نشانه‌های زبانی مربوط به توضیح و توصیف رده‌ها با سرعنوان‌های موضوعی. در مورد اختصاص شماره رده به سرعنوان‌ها مشخص گردید که درصد قابل توجهی از سرعنوان‌های مورد بررسی (۴۶/۸۹٪) فاقد شماره رده هستند. در تحلیل تطابق سرعنوان‌های رأس با رده‌بندی کنگره نیز مشخص شد که ۷۱/۰۵ درصد از سرعنوان‌ها با همان عبارت در رده‌بندی کنگره قابل بازیابی هستند، اما گستره رده اختصاص یافته به سرعنوان‌های رأس در ۴۲/۱ درصد از موارد، کوچک‌تر از یک رده فرعی است. حال آنکه انتظار می‌رود سرعنوان‌های رأس گستره وسیعی از شماره‌های رده را به خود اختصاص داده باشند. با این وصف، به نظر می‌رسد نوعی از گم‌شدگی اطلاعات در فرایند اختصاص شماره رده از روی موضوعات اختصاص یافته به منابع رخ می‌دهد. با توجه به یافته‌ها این میزان از گم‌شدگی قابل توجه به نظر می‌رسد.

۷. سخن پایانی

پژوهش حاضر با هدف شناسایی نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی و تعیین انطباق آن با نظام مفهومی رده‌بندی کنگره از یک سو و نظام دانشی ایران از سویی دیگر، انجام پذیرفت. پس از تبدیل فایل مستند موضوعی شامل ۸۴۵۴۵ رکورد مستند موضوع از قالب مارک/ایکس ام ال به قالب اسکاس / آر دی اف، زنجیره‌های سرعنوانی و درخت‌های سلسله مراتبی شناسایی شد و سرعنوان‌های رأس به منزله نشانی از مفاهیم موجود در فهرست سرعنوان‌های موضوعی فارسی، با عنوان رشته‌های دانشگاهی و نشانه‌های رده‌بندی کنگره، تطبیق داده شد.

در مجموع، تعداد ۱۸۴۲ زنجیره در شبکه کلی رابطه سلسله مراتبی شکل گرفت که بزرگ‌ترین زنجیره شکل گرفته در این شبکه، زنجیره‌ای با ۳۹۵ مدخل سرعنوانی است. بزرگ‌ترین زنجیره پس از شکستن، به ۲۴۱ درخت سلسله مراتبی تبدیل شد. تطبیق ۳۸ سرعنوان رأس از این مجموعه با نظام رده‌بندی کنگره نشان از تطابق کامل ۷۱/۰۵ درصدی در سطح تطابق واژگانی دارد. از نظر گستره دامنه رده‌های تخصیص یافته به سرعنوان‌های رأس، مشخص است که بیشترین درصد، یعنی ۴۲/۱ درصد، متعلق به سرعنوان‌های رأسی است که در طرح کلی رده‌بندی کنگره، حضوری کوچک‌تر از یک رده فرعی داشته و ۳۱/۵ درصد از سرعنوان‌ها، حضوری معادل یا بیش از یک رده فرعی داشته‌اند. تحلیل رده‌ها، پژوهشگران را

به این نتیجه می‌رساند که نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی، بخشی از نظام رده‌بندی کنگره را تحت پوشش دارد. به بیانی دیگر به نظر می‌رسد ساختار سرعنوان‌های موضوعی فقط در برخی از حوزه‌های مفهومی بسط یافته و شکل گرفته است.

در تطبیق سرعنوان‌های رأس با نظام دانشی کشور، یافته‌ها نشان داد که ۵۵/۲۶ درصد از سرعنوان‌های رأس با عنوان رشته‌های تحصیلی تطابق وازگانی ندارد و ۲۱/۰۵ درصد از سرعنوان‌های رأس در تطابق کامل با عنوان رشته‌های تحصیلی در ایران است. همچنین، ۲۳/۶۸ درصد از سرعنوان‌های رأس در تطابق نسبی با عنوان رشته‌های تحصیلی در ایران است. از نظر تطابق محتوایی و دامنه موضوعی سرعنوان‌های رأس با رشته‌های تحصیلی در ایران نیز مشخص است که نظام مفهومی حاکم بر سرعنوان‌های موضوعی جزئی‌تر از محتوای رشته‌های تحصیلی در ایران است. در عین حال، ۲۱/۰۵ درصد از سرعنوان‌های رأس در تطابق کامل با محتوای رشته‌های تحصیلی در ایران هستند و به نظر می‌رسد بخش کوچکی از نظام دانشی ایران و نه تمامی آن به صورت درخت‌های سلسله‌مراتبی در فهرست سرعنوان‌های موضوعی فارسی بازنمایی شده است. اخوتی (۱۳۸۸، ۹۵) نیز در یافته‌های پژوهش خویش تفاوت معنی‌داری را میان شبکه مفهومی سرعنوان‌های موضوعی فارسی و متون حوزه کتابداری و اطلاع‌رسانی گزارش کرده است.

نظام مفهومی حاکم بر سرعنوان‌های موضوعی جزئی‌تر از نظام دانشی در ایران است و از سویی نظام مفهومی حاکم بر سرعنوان‌های موضوعی فارسی در مقایسه با نظام رده‌بندی کنگره نشان داد که پیوند میان موضوعات فقط در چند حوزه محدود انجام پذیرفته است و بسیاری از رده‌ها مورد غفلت قرار گرفته است. با توجه به رشد علم در ایران و حرکت علمی ایران در عرصه بین‌الملل بعید به نظر می‌رسد که این ضعف ناشی از عدم وجود پشتوانه انتشاراتی باشد. بنابراین پیشنهاد می‌شود در مستندسازی موضوعات توجه یکسانی به تمامی حوزه‌های موضوعی مبذول گردد، حوزه‌های کمتر پوشش یافته شناسایی و تقویت گردد و یا این که از اصل، گسترش سرعنوان‌های موضوعی با حرکت به سمت نظام اصطلاحنامه‌ای و به صورت مستقل در حوزه‌های مختلف دنبال شود.

همچنین، این واقعیت آشکار شد که تدوین و گسترش ساخت سلسله‌مراتبی در موضوعات ملی و بومی کشور ایران مانند ادبیات فارسی و اسلام، چندان مورد توجه نبوده است. به گونه‌ای که هیچ یک از سرعنوان‌های رأس شناسایی شده به این موضوعات اختصاص نیافته است. بدیهی است که در فهرست سرعنوان‌های موضوعی کنگره، این موضوعات بسط نیافته‌اند

و وظیفه گسترش و برقراری پیوند میان موضوعات در این بخش‌ها بر عهده مستندساز ایرانی است. بنابراین، پیشنهاد می‌شود در این مورد رویکرد مناسبی دنبال شود.

۸. قدردانی

با سپاس ویژه از دکتر محسن کاهانی ریاست محترم مرکز اطلاعات، آمار و امور رایانه‌ای دانشگاه فردوسی مشهد، سرکار خانم سودابه غفوریان، دانشجوی کارشناسی ارشد مهندسی کامپیوتر و مهندس رضا شهبازی، مدیرکل فناوری اطلاعات و سرپرست منابع دیجیتالی سازمان اسناد و کتابخانه ملی ایران.

۹. 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.
- بزرگی، اشرف السادات. ۱۳۷۰. بررسی میزان سازگاری موضوع‌ها و عنوان‌ها در کتاب‌های فارسی. فصلنامه کتاب ۲ (۴-۱): ۱۱۴-۱۴۵.
- پهلوان‌نژاد، محمدرضا و مجتبی نامور فرگی. ۱۳۸۸. کره معنایی، رویکردی جدید به نظریه شبکه‌های معنایی. زبان پژوهی ۱ (۱): ۱-۳۴. <http://www.sid.ir> (دسترسی در ۸۹/۱/۲).
- سلطانی، پوری، کامران فانی و مهناز رهبری اصل. ۱۳۸۱. سرعنوان‌های موضوعی فارسی (ویرایش ۳). تهران: سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران.
- سلطانی، پوری، کامران فانی و فیروزان زهادی. ۱۳۸۲. سرعنوان‌های موضوعی فارسی ویراست سوم - پیوست یک. تهران: سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران.
- سلطانی، پوری، کامران فانی و فیروزان زهادی. ۱۳۸۵. سرعنوان‌های موضوعی فارسی ویراست سوم - پیوست دو. تهران: سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران.
- سوسور، فردینان دو. ۱۹۸۶. دوره زبان‌شناسی عمومی. ترجمه کورش صفوی. تهران: هرمس. (یادداشت نویسنده برای ویراستار: اثر انگلیسی مربوط به ۱۹۸۶ است که در این صورت شماره صفحه دیگر منطقی نیست چون در ترجمه آن که در سال ۱۳۷۸ صورت گرفته است در صفحه ۱۷۷ این مطلب ذکر شده است).
- صفوی، کورش. ۱۳۸۴. فرهنگ توصیفی معنی‌شناسی. تهران: فرهنگ معاصر.
- فتاحی، رحمت‌الله و شعله ارسطوپور. ۱۳۸۶. بررسی میزان همخوانی سرعنوان‌های موضوعی فارسی با کلید واژه‌های عنوان و فهرست مندرجات کتابهای فارسی در حوزه‌های علوم انسانی، علوم اجتماعی، علوم کاربردی و علوم محض. کتابداری و اطلاع‌رسانی، ۱۰ (۳۹): ۲۷-۳۹.
- http://www.aqlibrary.org/index.php?module=TWArticles&file=index&func=view_publications&did=153&pid=10 (دسترسی در ۸۸/۱/۲)

- فرشچی، معصومه. ۱۳۸۴. بررسی تطبیقی رده تاریخ در اصطلاحنامه فرهنگی فارسی (اصفا) و رده DSR (در تاریخ ایران). فصلنامه کتاب. ۱۶ (۴): ۴۳ - ۶۰
- فلبر، هلموت. ۱۹۸۴. *مبانی اصطلاح‌شناسی*. ترجمه محسن عزیزی. ۱۳۸۱ تهران: مرکز اطلاعات و مدارک علمی ایران.
- لاینز، جان. ۱۹۸۱. *مقدمه‌ای بر معناشناسی زبان‌شناسی*. ترجمه حسین واله. ۱۳۸۳ تهران: گام نو.
- Chen, Y., H. Gu, Y. Perl, and J. Geller. 2009a. Structural group-based auditing of missing hierarchical relationships in UMLS. *Journal of Biomedical Informatics* 42(3): 452-467.
- Chen, Y., H. Gu, Y. Perl, J. Geller, and M. Halper. 2009b. Structural group auditing of a UMLS semantic type's extent. *Journal of Biomedical Informatics* 42(1): 41-52.
- Dolin, R.A. 1998. Pharos: A scalable distributed architecture for locating heterogeneous information sources. Ph.D. thesis, University of California, Santa Barbara.
- Dykstra, M. 1988. can subject headings be saved? *Library Journal* 113(15): 55-67.
- 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.
- Fernandez, C.W. 1991. Semantic relationships between title phrases and LCSH. *Cataloging & Classification Quarterly* 13 (1): 51 - 77.
- Frank, E., and G., W. Paynter., 2004. Predicting Library of Congress Classifications from Library of Congress Subject Headings. *Journal of the American Society for Information Science & Technology* 55(3): 214-227.
- Harper, C. A. 2006. Encoding Library of Congress Subject Headings in SKOS: authority control for the semantic web. In *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.
- International Federation of Library Associations and Institutions .2001. *UNIMARC manual - Authorities format* (2 ed.). Munchen: Saur. <http://www.ifla.org/files/uca/unimarc-authorities-format.pdf> (accessed Jan. 2010).
- Kai, E., S. Heiner, and P. Magnus. 2007. Interactive thesaurus assessment for automatic document annotation, In *Proceedings of the 4th international conference on Knowledge capture*, 103-110. Whistler, BC, Canada: ACM.
- Khosh-Khui, A. 1985. *Statistical analysis of the association between library of congress subject headings and their corresponding class notations in main classes of LCC and DDC*. Unpublished PhD, Indiana University, United States - Indiana.
- Larson, R.R. 1992. Experiments in automatic Library of Congress Classification. *Journal of the American Society for Information Science* 43 (2): 130-148.
- Miles, A., and J.R. Perez-Aguera. 2007. SKOS: simple knowledge organisation for the web. *Cataloging & Classification Quarterly* 43 (3/4): 69-83.
- Rolland-Thomas, P. 1993. Thesaural codes: an appraisal of their use in the Library of Congress subject headings. *Cataloging & Classification Quarterly* 16(2): 71-91.
- Yi, K., and L. M. Chan. 2008. A visualization software tool for Library of Congress subject headings. In *Proceedings of the 10th International Conference of International Society for Knowledge Organization*. In C. Arsenault & J.T. Tennis (Eds.), *Advances in knowledge organization* (Vol. 11, pp. 170-176). Würzburg, Germany: Ergon.
- 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(4): 677-687.

Identification and Analysis of Conceptual Structure of Persian Subject Headings in Comparison with the Library of Congress Classification System and Academic System of Iran

Atefeh Sharif*

PhD Candidate, Ferdowsi University of Mashhad

Mohammad Reza Davarpanah¹

Professor of LIS, Ferdowsi University of Mashhad

Azam Sanatjoo²

Assistant professor of LIS, Ferdowsi University of Mashhad

Iranian Journal of
**Information
Processing &
Management**

Abstract: This paper aims to discover the conceptual system of Persian Subject Headings (PSH) through the content analysis method. In order to discover the PSH's global hierarchical structure (GHS), all PSH records (84545 records) were investigated. Conceptual system of PSH was compared to the both Academic system of Iran and Library of Congress Classification system (LCC). Findings show strong relationships between PSH and LCC in Science (%19.5), Technology (%14.4), Social Sciences (%10.6) Medicine (%9.5) and Philosophy and Psychology (%6.6) classes. Identifying the hierarchical structures (tree structures) embedded in the entire PSH, 241 trees were found. Large trees and top headings (e.g. Medicine and Psychology, Sociology, Philosophy, Contracts, Founding, Mathematics, Management, law and ...) were selected as the representation of the conceptual system of PSH to compare with the two other conceptual systems in this study. It shows that only %21.05 of top headings match the academic system of Iran and it is narrower than it. It means that a limited part of the academic system has been represented in PSH. The matching rate of top headings with the LCC is %71.05. In conclusion, the conceptual system of PSH is narrower than it is expected. It is recommended to correct the errors and enrich both its relational structure and conceptual system.

Keywords: Persian Subject Headings (PSH), conceptual system, Academic system of Iran, Library of Congress Classification System (LCC)

Iranian Research Institute
For Science and Technology
ISSN 2251-8223
eISSN 2251-8231
Indexed in LISA, SCOPUS & ISC
Vol.28 | No.4 | pp: 831-857
summer 2013

*Corresponding author: atefehsharif@gmail.com

1. mrdavarpanah@yahoo.com

2. sanatjoo@gmail.com