

استفاده از ضریب وزنی در سرعنوانهای موضوعی در راستای بهبود بازیابی اطلاعات

دکتر مرتضی کوکی^۱

چکیده

در فهرستنویسی موضوعی، فهرستنویس برای بیان محتوای موضوعی کتاب در دست فهرستنویسی، اغلب از سرعنوان(های) موضوعی استفاده می‌کند. نحوه قرار گرفتن یک سرعنوان یکسان، در فهرستبرگه‌های متفاوت، ممکن است متفاوت باشد. یک سرعنوان ممکن است تنها سرعنوان تخصیص یافته به یک فهرستبرگه، یا با یک یا چند سرعنوان دیگر همراه باشد. در حالت دوم، بسته به رتبه سرعنوان در میان سرعنوانهای دیگر، آن سرعنوان به طور نظری، دارای وزن متفاوتی خواهد بود. اما در تمامی این موارد، این سرعنوانها به گونه‌ای مساوی در نظر گرفته می‌شوند؛ نکته‌ای که می‌تواند بر بازیابی اطلاعات تأثیری منفی بگذارد. نوشته حاضر تلاش می‌کند برای حل این مشکل در فهرستنویسی سنتی، نرم افزارهای کتابخانه‌ای، و مارک ایران، راه‌حلهایی پیشنهاد نماید.

کلیدواژه‌ها: فهرستنویسی موضوعی، ضریب وزنی، نرم افزارهای کتابخانه‌ای، مارک ایران.

مقدمه

«وزن‌دهی»^۲ که با عبارتهای دیگری مانند «ارزشگذاری وزنی اصطلاح» نیز در متون مربوط به بازیابی اطلاعات و از جمله متون کتابداری دیده می‌شود، عبارت است از

۱. دانشیار گروه کتابداری دانشگاه شهید چمران اهواز: kokabi80@Yahoo.com

2. Term weighting.

«ارزش‌گذاری اصطلاحات به کار رفته در عبارت جستجو، که طی آن بالاترین ارزش و وزن به اصطلاحاتی داده می‌شود که مربوط‌ترین یا مفیدترین اصطلاح برای درخواست جستجو باشند» (کینن^۱، ۱۳۷۸، ۸۴). «پولیت»^۲ نیز در کتاب «نظامهای ذخیره و بازیابی اطلاعات: خاستگاه، توسعه و کاربردها» در مبحث «وزن‌دهی اصطلاح»، چنین می‌نویسد: «اهمیت مفروض یک اصطلاح را می‌توان با فراوانی رویداد آن در پایگاه اطلاعاتی به هم پیوند زد: هر چه فراوانی اصطلاح بیشتر باشد، بین مدارک مربوط و سایر مدارک کمتر تفاوت می‌گذارد». (پولیت، ۱۳۸۰، ۱۱۱). در این کتاب، حدود چهار صفحه به مفهوم وزن‌دهی اصطلاح اختصاص داده شده است. «هارتلی»^۳ و دیگران نیز در کتاب «اصول و روشهای جستجوی پیوسته»، در هنگام بحث پیرامون «بازیابی رتبه‌ای» به وزن‌دهی واژه‌ها اشاره می‌کنند و هدف این وزن‌ها را «پدید آوردن امکان تطبیق خودکار پرسشها با رکوردهای مدارک از طریق محاسبه ارزش تطبیق ریاضی می‌دانند» (هارتلی و دیگران، ۱۳۸۰، ۲۸۶). «حسن‌زاده» در مقاله‌ای با عنوان «تأثیر مدل‌های بازیابی اطلاعات بر میزان ربط»، دربارهٔ ضریب وزنی نیز مطالبی را نوشته است (حسن‌زاده، ۱۳۸۳). «بیزا - بیتس و ریبرو - نتو»^۴ در صفحات ۱۸۴ تا ۱۹۲ و ۲۱۵ تا ۲۱۷ از کتاب خود با عنوان «قلمروهای نو در بازیابی اطلاعات»، در مورد وزن‌دهی و وزن‌دهی احتمالی سخن می‌گویند (بیزا - بیتس و ریبرو - نتو، ۱۳۸۴). همچنین، «گراسمن و فریدر»^۵ در فصل دوم کتاب خود با عنوان «بازیابی اطلاعات: الگوریتمها و روشهای اکتشافی» که به راهبردهای بازیابی اختصاص دارد، در مبحث «مدل فضای برداری»، محاسبه ضریب وزنی یک اصطلاح در یک مدرک را «ترکیبی از فراوانی اصطلاح (df) و فراوانی مدرک معکوس (idf)» تعریف می‌کنند (گراسمن و فریدر، ۱۳۸۴، ۲۱).

1. Keenan.
2. Pollitt.
3. Hartley.
4. Baeza-Yates and Ribeiro-Neto.
5. Grossman and Frieder.

اما نکته‌ای که باید مورد اشاره و تأکید قرار گیرد، اینکه وزندهی در نظامهای بازیابی که از روش نمایه‌سازی پس‌همارا^۱ استفاده می‌کنند، به کار گرفته می‌شود. نظامهای بازیابی مبتنی بر نمایه‌سازی پیش - همارا^۲ نمی‌توانند از وزندهی استفاده کنند. تا آن جا که نگارنده می‌داند، تاکنون دربارهٔ استفاده از ضریب وزنی یا وزن دهی در نظامهای بازیابی مبتنی بر نمایه‌سازی پیش‌همارا، مانند برگه دان سنتی کتابخانه و نظامهای بازیابی که بر ساختار برگه‌دان متکی هستند، حداقل در زبان فارسی مطلبی منتشر نشده است. آنچه در ادامه می‌آید، استفاده از ضریب وزنی را در این گونه نظامهای بازیابی اطلاعات مورد بررسی قرار می‌دهد.

استفاده از ضریب وزنی در نظامهای بازیابی مبتنی بر نمایه‌سازی پیش‌همارا

در فهرست‌نویسی موضوعی، فهرست‌نویس برای بیان محتوای موضوعی کتاب در دست فهرست‌نویسی، اغلب از سرعنوان(های) موضوعی استفاده می‌کند. این سرعنوان (ها) در مراحل بعدی، در بالای نسخه‌ای از همان فهرستبرگه تایپ می‌شوند و برگهٔ مادر را به برگهٔ موضوع تبدیل می‌کنند. در برابر هر سرعنوان موضوعی که به کتاب داده می‌شود یک برگهٔ موضوع تولید می‌شود و در برگه دان موضوع در ردیف الفبایی خود قرار می‌گیرد. در تمامی مراحل فهرست‌نویسی موضوعی، فهرست‌نویس تلاش می‌کند اصول متداول فهرست‌نویسی موضوعی، و از جمله «ویژگی» را رعایت کند. «ویژگی، یعنی به جای موضوع اعم و کلی تر باید سعی شود سرعنوان موضوعی اخص و مستقیم‌تر به کار برده شود؛ یعنی، موضوع مورد نظر در قالب واژه یا اصطلاح خاص مورد استفاده کاربران ارائه شود. مفهوم مستقیم یعنی سرعنوانها به طور مستقیم و بدون واسطهٔ واژه‌های هم‌خانواده در فهرست بیابند تا کاربر سریع تر و آسان تر به موضوع مورد نظر خود دست یابد». (فتّاحی و طاهری، ۱۳۸۴، ۶۴).

1. Post-coordinate indexing.
2. Pre-coordinate indexing.

نحوه قرار گرفتن یک سرعنوان یکسان در فهرستبرگه‌های متفاوت، ممکن است متفاوت باشد. برای نمونه، سرعنوان «آمار» ممکن است در فهرستبرگه‌ای به منزله تنها موضوع، به کتابی داده شود که درباره آمار به طور کلی است. همین سرعنوان ممکن است در فهرستبرگه‌ای دیگر، در کنار یک یا چند سرعنوان دیگر و در جایگاهی متفاوت قرار گیرد، مثلاً در کنار سرعنوانی دیگر مانند «احتمالات». در این حالت دوم، بسته به تعداد سرعنوانهای داده شده به کتاب و نیز، جایگاه سرعنوان «آمار» در میان این سرعنوانها، تقسیمهای متفاوتی پیش می‌آید. معمولاً فرض بر این است که اگر «آمار» سرعنوان اول است موضوع اصلی کتاب در دست فهرستنویسی، «آمار» است یا حداقل بخش عمده آن به «آمار» اختصاص دارد. اما وزن موضوعی این سرعنوان با وزن سرعنوان حالت اول متفاوت است زیرا اگر فرض را بر این بگذاریم که وزن سرعنوان «آمار» در حالت اول، عدد یک است در حالت دوم، بسته به تعداد سرعنوانهای داده شده، اگر سرعنوان «آمار»، سرعنوان اول باشد، وزن آن معادل کسری از یک، و اگر سرعنوان دوم یا سوم باشد، وزنی کمتر خواهد داشت. اما در تمامی این حالتها، سرعنوان «آمار» در بالای برگه موضوعی خاص خود قرار می‌گیرد و هنگامی که فهرستبرگه‌های مربوط به حالت‌های بالا در برگه‌دان موضوعی برگه‌آرایی می‌شوند، در عمل دارای وزنهایی یکسان به شمار می‌آیند؛ در حالی که عملاً این گونه نیست زیرا در حالت اول، یعنی زمانی که تنها یک سرعنوان «آمار» به کتاب داده شده است، کل کتاب به آمار اختصاص دارد؛ در حالی که در حالت‌های دوم، تنها بخشی از کتاب به آمار مربوط است. در این جا، شاید این نکته مطرح شود که کاربر می‌تواند با بررسی سایر سرعنوانهای موضوعی داده شده در بخش فهرست تحلیلی هر فهرستبرگه، رتبه هر سرعنوان و تعداد سرعنوانهای دیگر در آن فهرستبرگه را تشخیص دهد؛ اما در پاسخ می‌توان گفت که حتی اگر فهرستبرگه‌های معمولاً در هم فشرده موجود در برگه‌دانهای کتابخانه‌ها چنین فرصتی را برای کاربر ایجاد کند، بعید است او چنین بررسی‌ای را انجام دهد، ضمن آنکه در کنار هم قرار گرفتن سرعنوانهای موضوعی و

سایر شناسه افزوده‌ها، تشخیص تعداد سرعنوانها و رتبه هر سرعنوان را برای کاربر معمولاً ناآشنا به رموز فهرست‌نویسی، دشوار می‌سازد.

حالت سومی نیز معمولاً رخ می‌دهد و آن این است که سرعنوان «آمار»، به تنهایی موضوع قرار نمی‌گیرد و تقسیم فرعی بر سرعنوانی اصلی است؛ مانند «آموزش و پرورش - آمار». در این حالت نیز، اگر تنها یک سرعنوان به کتاب داده شود، تقسیم فرعی «آمار» وزنی کمتر از سرعنوان اصلی یعنی «آموزش و پرورش» دارد، اما اگر سرعنوانی مانند «آموزش و پرورش - ایران - آمار» به کتابی داده شود، وزن تقسیم فرعی «آمار» در این حالت، احتمالاً از حالت اول نیز کمتر خواهد بود. این حالت در بازیابی خودکار اطلاعات در نرم افزارهای کتابخانه‌ای، و در حالت جستجوی «موضوع کلیدواژه‌ای» پیش می‌آید که در آن، رایانه در هنگام جستجوی مدارک مربوط به «آمار»، ممکن است حتی تقسیمهای فرعی «آمار» را نیز بازیابی کند. در چنین حالتی نیز، مدارک مربوط به «آمار» با وزنهایی متفاوت - و نه لزوماً به ترتیب موضوعی وزنی - در کنار هم بازیابی خواهند شد.

شایان ذکر است، در تمامی این حالتها، سرعنوانهایی که با رعایت اصل «ویژگی» ارائه شده اند در عمل، بازیابی را به گونه‌ای «ویژه» انجام نمی‌دهند؛ زیرا مثلاً سرعنوان «آمار»ی که سومین سرعنوان از سه سرعنوان یک کتاب است، نمی‌تواند هموزن سرعنوان «آمار»ی باشد که تنها سرعنوان یک کتاب است. اما در عمل، به شرحی که در بالا آمد چنین پنداشته می‌شود و سرعنوان «آمار»ی که سومین سرعنوان از سه سرعنوان یک کتاب است، به همان اندازه تک سرعنوان «آمار»، «ویژه» در نظر گرفته می‌شود.

آنچه در این نوشته خواهد آمد، تلاش برای ارائه راهکارهایی برای حل این مشکل و بهبود بازیابی موضوعی اطلاعات در برگه‌دانهای کتابخانه، نرم افزارهای کتابخانه‌ای و مارک ایران است. برای ارائه راه حل برای سه حالت برگه‌دانهای کتابخانه، نرم افزارهای کتابخانه‌ای و مارک ایران، باید به این نکته توجه داشت که وزن هر سرعنوان، تابعی از سه متغیر زیر است:

۱. تعداد سرعنوانهای داده شده به یک کتاب

۲. مرتبه آن سرعنوان در بین سرعنوانهای داده شده

۳. حجم یا تعداد صفحات کتاب.

متغیر حجم یا تعداد صفحات کتاب، از این نظر می‌تواند مهم باشد که وزن یک سرعنوان «آمار»، در کتابی با تعداد صفحاتی با حدود ۳۰۰ صفحه، که یکی از سرعنوانهای آن «آمار» است و به طور نظری، حدود ۱۰۰ صفحه از آن به آمار اختصاص می‌یابد، بیشتر از وزن سرعنوان «آمار» در کتابی مثلاً ۸۰ صفحه‌ای است که تنها سرعنوان موضوعی آن «آمار» است. البته، از آنجا که معمولاً حجم کتابها در حدود معینی قرار دارد، متغیر حجم یا تعداد صفحات کتاب، چندان نمی‌تواند در این مورد تأثیر گذار باشد.

نکته دیگری که باید مورد توجه قرار گیرد، این است که در تمامی موارد مورد بحث، فرض بر این است که فهرست‌نویس در مواردی که چند سرعنوان به کتابی می‌دهد، سرعنوان اول را به منزله مرتب‌ترین یا سرعنوان عمده، و سرعنوان (های) بعدی را به منزله سرعنوانهایی که اهمیت آنها از سرعنوان اول به سرعنوانهای بعدی کاهش می‌یابد، در نظر می‌گیرد. در این مورد باید اذعان داشت، نمایه ساز انسانی همواره در معرض اشتباهات انسانی شامل پیشینه ذهنی و یا سایر موارد قرار دارد، اما در این جا می‌توان استدلال کرد چه از ضریب وزنی استفاده بشود چه نشود، این اشتباهات در فهرست‌نویسی موضوعی اجتناب‌ناپذیر است.

راه حل پیشنهادی برای فهرستبرگه‌ها

راه حل پیشنهادی برای فهرستبرگه‌ها عبارت است از افزودن دو عدد در پایان هر سرعنوان موضوعی، به گونه‌ای که عدد اول که در کمانک گذاشته می‌شود. رتبه سرعنوان و عدد دوم که بدون کمانک است، تعداد سرعنوانهای داده شده به کتاب را نشان دهد. می‌توان به طور قراردادی، برای تک سرعنوان از هیچ عددی استفاده نکرد. بدین ترتیب:

«آموزش و پرورش»، یک سرعنوان است.

«آموزش و پرورش (۱)»، اولین سرعنوان از دو سرعنوان است.

«آموزش و پرورش (۳۱)»، اولین سرعنوان از سه سرعنوان است.

«آموزش و پرورش (۳۲)»، دومین سرعنوان از سه سرعنوان است.

«آموزش و پرورش (۵۲)»، دومین سرعنوان از پنج سرعنوان است.

اگر این اعداد در هنگام برگه آرایی در نظر گرفته شود، کاربر در هنگام جستجوی مطالبی درباره آموزش و پرورش، نخست سرعنوانهایی را می‌یابد که تکی و در نتیجه، دارای بیشترین وزن موضوعی هستند. در مراحل بعدی، سرعنوانهایی را پیدا می‌کند که اولین هستند، اما تنها سرعنوان نیستند. اینها اگر چه از نظر موضوعی وزن سرعنوانهای گروه اول را ندارند، اما نسبت به سرعنوانی مانند «آموزش و پرورش (۵۲)» که دومین سرعنوان از پنج سرعنوان است، وزن بیشتری دارند. با انجام چنین کاری، کاربر مطالب مربوط را از مرتبط‌ترین تا کم ارتباط‌ترین پیدا می‌کند. در ادامه، چند نمونه از یک نسخه از کتابشناسی ملی ایران (کتابشناسی ملی ایران، ۱۳۷۵، ۱، ۲۹۷) ارائه خواهد شد. اگر چه همه نمونه‌ها از فهرستبرگه‌های کتابشناسی ملی ایران گرفته شده، اما به دلیل صرفه‌جویی در فضای نوشته، تنها عناوین و سرعنوانهای موضوعی هر فهرستبرگه ارائه خواهد شد:

آمار و کاربرد آن در مدیریت

۱. آمار ۲. مدیریت - روشهای آماری.

آمار کاربردی در اقتصاد و بازرگانی

۱. آمار ۲. آمار بازرگانی ۳. اقتصاد - روشهای آماری.

آمار و احتمال مقدماتی

۱. احتمالات ۲. آمار.

آمار کاربردی

۱. علوم اجتماعی - روشهای آماری ۲. مدیریت - روشهای آماری ۳. اقتصاد -

روشهای آماری ۴. آمار.

با بررسی چهار سرعنوان «آمار» در بالا، مشخص می‌شود وزنهای موضوعی یکسانی ندارند، زیرا رتبه هر سرعنوان و تعداد سرعنوانهای دیگر موجود در هر فهرستبرگه، متفاوت

است. اما در فهرست‌نویسی سنتی، تمام این سرعنوانها، هنگامی که در بالاترین سطر فهرستبرگه تایپ و فهرستبرگه‌های مربوط در جای الفبایی خود برگه‌آرایی می‌شوند، عملاً وزنی یکسان خواهند داشت. در حالی که اگر به روش بالا برای هر یک ضریب وزنی درخور تعیین شود و این ضریبها در برگه‌آرایی نیز منظور شوند، بازیابی از مرتبط‌ترین تا کم ارتباط‌ترین انجام و بدین ترتیب اصل ویژگی نیز رعایت می‌شود. بدین ترتیب، سرعنوانهای «آمار» به ترتیب زیر ضریب دار می‌شوند:

کتاب اول: آمار(۱) ۲

کتاب دوم: آمار(۱) ۳

کتاب سوم: آمار(۲) ۲

کتاب چهارم: آمار(۴) ۴

ترتیب برگه‌آرایی نیز طبیعتاً همان ترتیب بالاست؛ یعنی برای سرعنوانهای همانند، نخست شماره‌های مساوی درون کمانک در کنار هم و به ترتیب شماره‌های بعدی از کم به زیاد مرتب می‌شوند و آن‌گاه، شماره‌های درون کمانک که مقدار بیشتری دارند، می‌آیند. بدیهی است، در نمونه بالا، تنها سرعنوانهای «آمار» با ضریب وزنی نشان داده شده‌اند و سرعنوانهای دیگر نیز باید برای بازیابی‌های مربوط، ضریبهای وزنی مناسب داشته باشند.

راه‌حل پیشنهادی برای نرم افزارهای کتابخانه‌ای

در نرم افزارهای کتابخانه‌ای معمولاً دو نوع جستجوی موضوعی امکان پذیر است: «موضوع سرعنوانی» و «موضوع کلیدواژه‌ای». این عبارتها در نرم‌افزارهای متفاوت ممکن است به شکلهای متفاوتی به کار بروند. در جستجو از طریق موضوع سرعنوانی، فرض بر این است که کاربر عین سرعنوان موضوعی را در محل جستجو و گزینه «موضوع سرعنوانی» را در «فیلد جستجو» وارد می‌کند. در این حالت، شماره‌گذاری سرعنوانها به روش بالا می‌تواند راه حل مناسبی برای تعیین وزن سرعنوانها باشد. در این حالت، تنها

وجود الگوریتمی ساده، رایانه را قادر می‌سازد تا سرعنوانهای سنگین وزن تر را در آغاز و سبک وزن تر را در مراحل بعدی بازبایی کند. بدیهی است، به دلیل ناآشنایی اغلب کاربران با ساختار سرعنوانهای موضوعی، آنان ترجیح می‌دهند بیشتر از جستجوی «موضوع کلیدواژه‌ای» استفاده کنند.

در فرایند جستجوی «موضوع کلیدواژه‌ای»، فرض بر این است که رایانه، کلیدواژه‌های مندرج در سرعنوانهای موضوعی را بازبایی می‌کند. در این حالت، اگر کاربر کلیدواژه «آمار» را در محل جستجو، و گزینه «موضوع کلیدواژه‌ای» را در «فیلد جستجو» وارد نماید، رایانه باید کلید واژه «آمار» را، افزون بر سرعنوانهایی که دارای واژه «آمار» به منزله سرعنوان اصلی هستند، در سرعنوانی مانند «آموزش و پرورش - ایران - آمار» نیز پیدا کند. در چنین حالتی، این که واژه «آمار» در کجای سرعنوان قرار دارد نیز اهمیت پیدا می‌کند. نکته قابل بررسی این است که وقتی از مفهوم اهمیت موضوعی در بخشهای مختلف یک سرعنوان سخن به میان می‌آید، نمی‌توان با قاطعیت اظهار نظر کرد که در سرعنوانی مانند «آموزش و پرورش - ایران - آمار» کدام یک از دو مفهوم ایران و آمار اهمیت بیشتری دارند، زیرا چنین ترکیبی به نظر می‌رسد بیشتر تابع دستورات کتاب سرعنوانهای موضوعی باشد تا منطقی دیگر. در نتیجه، مهم این است که آیا واژگان «ایران» و «آمار» به منزله بخش اول یا مهم سرعنوان آمده اند یا نه. بدین ترتیب، آنچه در این حالت مهم است، اتخاذ تمهیدی است تا به وسیله آن رایانه بتواند بخش اول را از بخشهای بعدی در یک سرعنوان تمیز دهد. بدیهی است، این نکته نیز همچنان باقی است که این بخش اول در سرعنوانی با چه رتبه‌ای و در میان چند سرعنوان قرار دارد.

نکته دیگری که در جستجوی «موضوع کلیدواژه‌ای» باید مورد بررسی قرار گیرد، دو حالتی است که جستجو در آنها انجام می‌شود: جستجو بدون استفاده از عملگرهای بولی و جستجو با استفاده از این عملگرها. این دو حالت در ادامه بررسی می‌شوند.

در حالت اول، هنگامی که کاربر جستجوی «موضوع کلیدواژه‌ای» را با یک تک واژه یا عبارت انجام می‌دهد، اگر تک واژه در سرعنوانی با یک یا چند تقسیم فرعی وجود

داشته باشد، رایانه تنها کافی است پس از شناسایی اینکه سرعنوانی که کلید واژه مورد جستجو در آن قرار دارد، دارای چه رتبه‌ای و در میان چند سرعنوان دیگر است تشخیص دهد که آیا کلید واژه مورد نظر کاربر در بخش اول آمده است یا نه. مثال زیر این نکته را روشن تر می‌کند. فرض کنیم کاربری در هنگام جستجوی «موضوع کلیدواژه‌ای»، واژه «آمار» را در محل جستجو وارد کرده است. آن چه روشن است اینکه کاربر، قصد دارد کتابی درباره آمار بیابد اما این احتمال وجود دارد که رایانه، سرعنوانی مانند «آموزش و پرورش - ایران - آمار» را، به دلیل وجود واژه «آمار» در آن بازیابی کند که منظور نظر کاربر نیست. اما اگر سرعنوان بالا به روش زیر شماره گذاری شود: «آموزش و پرورش (۱) - ایران - آمار» به گونه‌ای که شماره یک نشان دهد بخش اول، «آموزش و پرورش» است و الگوریتم بازیابی به گونه‌ای نوشته شده باشد که در این حالت، تنها کلیدواژه‌هایی را پیدا کند که با شماره یک مشخص شده‌اند، این بازیابی نادرست یا ریزش کاذب رخ نخواهد داد.

در حالت دوم، الگوریتم بازیابی باید به گونه‌ای نوشته شود که رایانه بتواند سه کلید واژه مورد نظر کاربر را در سه حالت مختلف بررسی کند. دلیل این که از سه کلیدواژه نام برده شد این است که به نظر می‌رسد نرم افزارهای کتابخانه‌ای حداکثر تا سه محل برای سه کلیدواژه فراهم می‌کنند. اکنون فرض کنیم کاربری کتابی در مورد «آمار آموزش و پرورش ایران» می‌خواهد. او احتمالاً بنا به عادت زبان طبیعی خود، در محل اول، کلیدواژه «آمار» در محل دوم، کلیدواژه «آموزش و پرورش»، و در محل سوم، کلیدواژه «ایران» را تایپ خواهد کرد. در صورتی که سرعنوان «آموزش و پرورش (۱) - ایران - آمار» در پیشینه‌های مربوط وجود داشته باشد، رایانه هر سه کلیدواژه را از این نظر که کدام یک در بخش اول سرعنوانهای با تقسیمات فرعی قرار می‌گیرد آزمایش خواهد کرد و پس از گزینش سرعنوانهایی که بخش اول آنها «آموزش و پرورش» است، در تقسیمات فرعی این سرعنوانها به جستجوی دو کلیدواژه دیگر خواهد رفت. بدین ترتیب، اگر در هر سرعنوان، افزون بر مشخص شدن بخش اول، معین شده باشد که سرعنوان، دارای رتبه چندم و در

استفاده از ضرب و وزن در سرعنوانهای موضوعی ... / ۲۵۳

میان چند سرعنوان است، بازیابی مدارک مرتبط به صورتی بهینه انجام خواهد گرفت. شایان ذکر است، در این جا منظور از بخش اول هر سرعنوان، تنها بخش اول در سرعنوان با تقسیمهای فرعی نیست بلکه در سرعنوانهای با نقطه نیز می باشد.

نمونه‌های زیر از یک نسخه کتابشناسی ملی ایران (کتابشناسی ملی ایران، ۱۳۷۵، ۱، ۴۵۶-۴۵۷) احتمالاً به درک بهتر موضوع کمک می کند:

رویش و پرواز: چهار پاره

۱. شعر فارسی - قرن ۱۴ - مجموعه‌ها.

... کشتی طوفان کربلا ...

۱. شعر مذهبی - قرن ۱۴ - مجموعه‌ها. ۲. شعر فارسی - قرن ۱۴ - مجموعه‌ها.

مدایح و مراثی آل محمد (ص): مجموعه زیبا و جانسوز شعرای بنام و معاصر

ایران ... به ضمیمه‌ای از سروده‌های اسلامی

۱. چهارده معصوم - شعر. ۲. شعر مذهبی - قرن ۱۴ - مجموعه‌ها. ۳. شعر فارسی -

قرن ۱۴ - مجموعه‌ها.

عاشقانه‌ها: گزینه سروده‌های شاعران امروز ایران

۱. شعر عاشقانه فارسی - قرن ۱۴. ۲. شعر آزاد - مجموعه‌ها. ۳. شاعران ایرانی - قرن

۱۴. ۴. شعر فارسی - قرن ۱۴ - مجموعه‌ها.

در چهار کتاب بالا که تنها عناوین و سرعنوانهای موضوعی فهرستبرگه‌های آنها در

این جا ارائه شده است، سرعنوان «شعر فارسی - قرن ۱۴ - مجموعه‌ها». در چهار رتبه

متفاوت در چهار فهرستبرگه دیده می شود. بدیهی است، چنانچه این سرعنوانها به روش

بالا دارای ضرب و وزن شوند، در بازیابی، سرعنوان کتاب چهارم با ضرب و وزن (۴) در

آخر و سرعنوان کتاب اول بدون ضرب و وزن (به معنای یک از یک) در آغاز قرار

می گیرد. اما در این جا لازم می آید این سرعنوانها ضرب دیگری نیز داشته باشند، زیرا قرار

است مورد جستجوی «موضوع کلیدواژه‌ای» نیز قرار گیرند. بدین ترتیب و با توجه به

استدلال بالا که تنها بخش اول هر سرعنوان در این مرحله اهمیت دارد، بخشهای اول هر

سرعنوان (در این جا، «شعر فارسی») باید ضریبی داشته باشند تا بر مبنای آن به منزله کلیدواژه اصلی استخراج گردند. از آنجا که این ضریب تنها در سرعنوانهای با تقسیم فرعی کارآیی دارد، افزودن شماره یک در کمانک به سرعنوان اصلی، احتمالاً برای این منظور کفایتی کسب‌کننده سرعنوانهای موضوعی چهار کتب بلا بصوت زودری آینه

۱. شعر فارسی (۱) - قرن ۱۴ - مجموعه‌ها.

۱. شعر مذهبی (۱) - قرن ۱۴ - مجموعه‌ها (۱)۲. شعر فارسی (۱) - قرن ۱۴ -

مجموعه‌ها (۲)۲،

۱. چهارده معصوم (۱) - شعر (۱)۳. شعر مذهبی (۱) - قرن ۱۴ - مجموعه‌ها (۲)۳.۳.

شعر فارسی (۱) - قرن ۱۴ - مجموعه‌ها (۳)۳،

۱. شعر عاشقانه فارسی (۱) - قرن ۱۴ (۱)۴. شعر آزاد (۱) - مجموعه‌ها (۲)۴.۳.

شاعران ایرانی (۱) - قرن ۱۴ (۳)۴. شعر فارسی (۱) - قرن ۱۴ - مجموعه‌ها (۴)۴.

در این شرایط، رایانه در پاسخ به کاربری که خواهان مجموعه‌های شعر فارسی در

قرن ۱۴ است مدارک دارای سرعنوانهای موضوعی زیر را به ترتیب زیر بازیابی می‌کند:

۱. شعر فارسی (۱) - قرن ۱۴ - مجموعه‌ها.

۲. شعر فارسی (۱) - قرن ۱۴ - مجموعه‌ها (۲)۲،

۳. شعر فارسی (۱) - قرن ۱۴ - مجموعه‌ها (۳)۳،

۴. شعر فارسی (۱) - قرن ۱۴ - مجموعه‌ها (۴)۴.

در این حالت، مدارک مربوط با توجه به کلید واژه اصلی که با شماره (۱) مشخص

شده و بر حسب رتبه سرعنوانها بازیابی می‌شوند.

راه حل پیشنهادی برای مارک ایران

مارک ایران در بازیابی اطلاعات همانند سایر نرم افزارهای کتابخانه‌ای عمل

می‌کند؛ یعنی هم می‌تواند فهرستبرگه تولید کند و هم به منزله ابزاری برای بازیابی

اطلاعات مورد استفاده قرار گیرد، با این تفاوت که بر یک استاندارد جهانی مبتنی است.

در مارک ایران در بلوک - - 6 که بلوک تجزیه و تحلیل موضوعی نام دارد، هشت

فیلد به شرح زیر برای سرعنوانهای موضوعی وجود دارد (مارک ایران، ۱۳۸۱، ۲۰۵):

نام شخص به منزله موضوع	600
نام تالگان به منزله موضوع	601
نام خاندان به منزله موضوع	602
پدید آور و عنوان به منزله موضوع	604
عنوان به منزله موضوع	605
موضوع (اسم عام یا عبارت اسمی عام)	606
نام جغرافیایی به منزله موضوع	607
اصطلاحهای موضوعی کنترل نشده	610

همگی این فیلدها تکرار پذیرند و درون آنها نخستین فیلد فرعی یعنی \$a به عنصر شناسه‌ای اختصاص دارد که مثلاً در فیلد 606 «حاوی اصطلاح، به همان شکلی است که سرعنوان موضوعی مورد استفاده تعیین کرده است». (مارک ایران، ۱۳۸۱، ۲۳۰). فیلدهای فرعی \$j (تقسیم فرعی شکلی)، \$x (تقسیم فرعی موضوعی)، \$y (تقسیم فرعی جغرافیایی)، و \$z (تقسیم فرعی دوره‌ای) افزون بر سایر فیلدهای فرعی، کمابیش در تمامی این فیلدها وجود دارند.

اکنون برای نشان دادن راه حل پیشنهادی در مارک ایران، نمونه‌های بالا را در فیلدهای مارک ایران می‌گذاریم:

۱. آمار. ۲. مدیریت - روشهای آماری.

@\$2nli آمار\$آمار#6061

@\$2nli روشهای آماری\$x مدیریت#6062

۱. آمار. ۲. آمار بازرگانی. ۳. اقتصاد - روشهای آماری.

@\$2nli آمار\$آمار#6061

@\$2nli آمار بازرگانی#6062

@\$2nli روشهای آماری\$x اقتصاد#6062

۱. احتمالات. ۲. آمار.

@\$2nli احتمالات#6061

@2nli\$آمار\$6062#

۱. علوم اجتماعی - روشهای آماری. ۲. مدیریت - روشهای آماری. ۳. اقتصاد - روشهای آماری. ۴. آمار.

@2nli\$روشهای آماری\$علوم اجتماعی\$6061#

@2nli\$روشهای آماری\$مدیریت\$6062#

@2nli\$روشهای آماری\$اقتصاد\$6062#

@2nli\$آمار\$6062#

چنانکه مشاهده می‌شود، در تمام موارد بالا فیلد 606 بر حسب لزوم تکرار شده است و در همه فیلدها، فیلد فرعی a به منزله «عنصر شناسه‌ای» وجود دارد. اگر چه در صفحه ۲۱ از دستنامه مارک ایران در مورد «ترتیب فیلدها» نوشته شده که «نظم خاصی برای ضبط فیلدها در یک رکورد پیش بینی نشده است. نظم حاکم بر فهرست راهنمای فیلدها ممکن است با ترتیب ضبط فیزیکی فیلدها یکی نباشد». اما در مارک ایران، و مثالهای ارائه شده در انتهای «دستنامه مارک ایران» نیز نشان می‌دهد می‌توان نظم فیلدهای موضوعی را بر حسب نظم ارائه آنها در یک فهرستبرگه انتخاب نمود. این نکته در مارک ایران در مورد نظم فیلدهای فرعی نیز صدق می‌کند؛ بدین معنا که جز در مواردی خاص که فیلدهای فرعی از نظم خاصی پیروی می‌کنند، ترتیب خاصی برای فیلدهای فرعی در نظر گرفته نشده است. در مثالهای ارائه شده در انتهای کتاب «دستنامه مارک ایران» نیز نظم فیلدهای فرعی موضوعی، همان ترتیب سرعنوان اصلی و تقسیمهای فرعی در سرعنوانهای موضوعی است. بدین ترتیب، هم در پیشینه مارک ایران رتبه هر سرعنوان مشخص می‌شود و هم می‌توان با افزودن یک الگوریتم ساده، این امکان را فراهم ساخت که رایانه هنگام تولید فهرستبرگه، با احتساب فیلدهای موضوعی و رتبه هر سرعنوان، ضرایب وزنی لازم را در انتهای هر سرعنوان موضوعی اضافه کند. لازم به توضیح است، اگر چه در تمام نمونه‌های بالا از فیلد 606 استفاده شده است، اما استفاده از فیلدهای موضوعی دیگر نیز تفاوتی به وجود نمی‌آورد و راه حل پیشنهادی به قوت خود باقی است.

در نمونه‌های بالا، سرعنوانهای با تقسیم فرعی نیز وجود دارند و به اعتباری، مسأله تعیین بخش اول یا مهم هر سرعنوان با استفاده از فیلد فرعی «عنصر شناسه‌ای» تا حدودی

استفاده از ضریب وزنی در سرعنوانهای موضوعی ... / ۲۵۷

روشن شده است. با این حال، یکی از نمونه‌های مربوط به جستجوی «موضوع کلیدواژه‌ای» برای وضوح بیشتر، با استفاده از نشانه گذاری مارک ایران، در زیر ارائه می‌شود:

۱. شعر عاشقانه فارسی - قرن ۱۴. ۲. شعر آزاد - مجموعه‌ها. ۳. شاعران ایرانی - قرن ۱۴. ۴. شعر فارسی - قرن ۱۴ - مجموعه‌ها.

@2nli\$قرن\$۱۴\$شعر عاشقانه فارسی\$a#6061

@2nli\$مجموعه‌ها\$شعر آزاد\$a#6062

@2nli\$قرن\$۱۴\$شاعران ایرانی\$a#6062

@2nli\$مجموعه‌ها\$شعر فارسی\$a#6062

در نمونه بالا، باز هم مشاهده می‌شود که ترتیب فیلدهای موضوعی را می‌توان همان ترتیب سرعنوانهای موضوعی در نظر گرفت و فیلد فرعی \$a (عنصر شناسه‌ای) نیز مشخص می‌کند که جستجوی موضوع کلیدواژه‌ای، چگونه در سرعنوان اصلی می‌تواند انجام شود.

نتیجه‌گیری و پیشنهادها

در صورت پذیرش پیشنهادهای بالا بهتر است:

۱. این ضرایب وزنی در شناسه‌های کتابشناسی ملی ایران - چه به صورت چاپی و چه به صورت لوح فشرده - اعمال شود تا از این پس در کتابخانه‌های گوناگونی که از این کتابشناسی در راستای فهرستنویسی بهره می‌گیرند، به کار رود.
۲. در نرم افزارهای کتابخانه‌ای نیز این ضرایب - خواه به صورت رویکردی آینده نگر و خواه به صورت رویکردی گذشته‌نگر - انجام شود.
۳. در نرم افزارهای آتی فرمت مارک ایران، الگوریتمهای لازم برای تولید این ضریبها در فهرستبرگه‌های تولیدی از طریق جستجوی «موضوع کلیدواژه‌ای» در مارک ایران در نظر گرفته شود.

منابع

- بیضا - بیتس، برتیه ریکاردو و نتوریرو (۱۳۸۴). قلمروهای نو در بازیابی اطلاعات، ترجمه علی حسین قاسمی با همکاری سیروس آزادی، علی جوامع، تهران: چاپار.

۲۵۸ / کتابداری و اطلاع‌رسانی - جلد ۱۱ شماره ۱

- پولیت، ا. استون (۱۳۸۰). نظامهای ذخیره و بازیابی اطلاعات: خاستگاه، توسعه و کاربردها/ ترجمه محمدحسین دیانی، جعفر مهراد. - شیراز: کتابخانه منطقه‌ای علوم و تکنولوژی شیراز.

- حسن‌زاده، محمد (۱۳۸۳). تأثیر مدل‌های بازیابی اطلاعات بر میزان ربط. اطلاع‌شناسی، ۲ (۱)، ۶۴-۸۹.

- فتاحی، رحمت‌الله و مهدی طاهری (۱۳۷۵). فهرستنویسی: اصول و روشها - ویراست ۴. - تهران: کتابدار.

- کتابشناسی ملی ایران، فروردین ۱۳۴۹. - تهران: کتابخانه ملی ایران، ۱۳۴۹ - (شماره ۸۳، نیمه اول ۱۳۷۵)

- کینن، استلا (۱۳۷۸). فرهنگ فشرده کتابداری و اطلاع‌رسانی، ترجمه و تدوین فاطمه اسدی کرگانی، مقابله و ویرایش عبدالحسین آذرننگ. - تهران: کتابدار.

- گراسمن، دیوید ا. و فریدر افیر (۱۳۸۴). بازیابی اطلاعات: الگوریتمها و روشهای اکتشافی؛ ترجمه جعفر مهراد و سارا کلینی، مشهد: کتابخانه رایانه‌ای؛ شیراز: کتابخانه منطقه‌ای علوم و تکنولوژی.

- مارک ایران/ کمیته ملی مارک ایران. - تهران: کتابخانه ملی جمهوری اسلامی ایران، ۱۳۸۱.

- هارتلی، آر. ج. ... [و دیگران] (۱۳۸۰). اصول و روشهای جستجوی پیوسته/ ترجمه زاهد بیگدلی. - مشهد: کتابخانه رایانه‌ای.