**نام مقاله: بررسي رويكرد هاي موضوعي (زبان طبيعي و كنترل شده)‌ در بازيابي اطلاعات از پايگاههاي پيوسته كتابشناختي**

**نام نشريه: فصلنامه كتابداري و اطلاع رساني (اين نشريه در www.isc.gov.ir نمايه مي شود)**

**شماره نشريه: 36 \_ شماره چهارم، جلد 9**

**پديدآور: دكتر عبد الرسول جوكار، سعيده انوري**

**چكيده**

**در اين پژوهش كاربردي كه با استفاده از تحليل محتوا انجام شده است، كارايي دو رويكرد جستجوي موضوعي: زبان طبيعي و واژگان كنترل شده با استفاده از معيارهاي ربط، دقت، بازيافت و ريزش كاذب، در دو پايگاه كتابشناختي پيوسته (كتابخانه كنگره واريك) مقايسه شده است. سي عنوانمربوط به موضوعات كتابخانه، كتابداري و اطلاع‌رساني به صورت تصادفي انتخاب شدند. براي هر عنوان راهبردهاي جستجوي زبان طبيعي و كنترل شده به دست آمد. پس از انجام جستجوها و ارزيابي ربط پيشينه‌هاي حاصل، دقت، بازيافت و ريزش كاذب در دو رويكرد مذكور با هم مقايسه گرديد. نتايج نشان مي‌دهد بازيافت رويكرد جستجوي زبان طبيعي در هر دو پايگاه بيش از رويكرد موضوعي كنترل شده مي باشد. دقت رويكرد زبان طبيعي در پايگاه اريك بيش از كنترل شده مي‌باشد اما، در فهرست پيوسته كتابخانه كنگره تفاوتي معنادار مشاهده نشد. رويكرد زبان طبيعي در پايگاه اريك ريزش كاذب كمتري به همراه داشته است، در حالي كه در فهرست كتابخانه كنگره تفاوت معنادار نيست. پيشنهاد تحقيق حاضر اين است كه براي كسب نتايج جامع در جستجوهاي موضوعي در پايگاههاي كتابشناختي پيوسته، تركيبي از دو رويكرد زبان طبيعي و كنترل شده مورد استفاده قرار گيرد. براي انجام جستجوهايي كه جامعيت نتايج حاصل از آنها اهميت چنداني ندارد و هدف، به دست آوردن تعدادي پيشينه مربوط است، رويكرد زبان طبيعي مزيت بيشتري دارد.**

**كليدواژه‌ها: جستجوي موضوعي، زبان طبيعي، واژگان كنترل شده.**

**مقدمه**

پيشرفت و گسترش فناوري اطلاعات، به ايجاد نظامهاي اطلاعاتي پيچيده سرعت مي‌بخشد. در چنين نظامهايي بخش عمده‌اي از جستجوهاي پيوسته را جستجوهاي موضوعي تشكيل مي‌دهند، اما بسياري از كاربران اين نظام‌ها با وجود آشنايي پايه با فنون جستجو، از روشهاي مناسب براي جستجوي مؤثر اطلاعات مربوط مطلع نيستند. در واقع، اهميت دستيابي موضوعي، بيش از صد سال پيش وقتي كاتر[[3]](http://192.168.0.110/editor/main.htm" \l "_ftn3" \o ") اهداف فهرست را بيان كرد، روشن شد. كاتر معتقد بود فهرست بايد به افراد براي يافتن كتابي كه موضوع آن مشخص است كمك كند و همچنين به استفاده‌كنندگان نشان دهد كه كتابخانه در مورد يك موضوع خاص، چه منابعي را در اختيار دارد ( هوفمن[[4]](http://192.168.0.110/editor/main.htm" \l "_ftn4" \o ")، 2001).

   در عصر الكترونيك نيز استفاده از جستجوي موضوعي براي بازيابي اطلاعات، رويكردي عمده تلقي مي‌شود. نظامهاي جستجو كه به كاربران اجازه مي‌دهند زبان طبيعي را براي جستجو به كار ببرند، هنوز نمي‌توانند به طور موفقيت‌‌آميز نيازهاي كاربران را برآورده سازند، در يكي از مؤسسه‌هاي وابسته به OCLC، اعضاي هيئت علمي و مشاركت‌كنندگان به توافق جمعي رسيدند كه براي دسترسي به منابع الكترونيكي، به واژگان كنترل شده و رده‌بندي نياز است (تيلور[[5]](http://192.168.0.110/editor/main.htm" \l "_ftn5" \o ")،1381). از سوي ديگر، زمان آن رسيده است كه كتابخانه‌ها نظامهاي بازيابي موضوعي خود را با دنياي اطلاعات رقمي سازگار سازند (مور[[6]](http://192.168.0.110/editor/main.htm" \l "_ftn6" \o ") و كين[[7]](http://192.168.0.110/editor/main.htm" \l "_ftn7" \o ") ، 1996).

**مسئله پژوهش**

يكي از مهمترين عملكردهاي هر كتابخانه و مركز اطلاع رساني، بر طرف كردن نيازهاي جامعه استفاده‌كننده با استفاده از يك نظام بازيابي كار آمد است. بازيابي اطلاعات، فرايندي است كه طي آن مجموعه‌اي از مدارك براي مشخص كردن مداركي كه با يك موضوع خاص مرتبطند، مورد جستجو قرار مي‌گيرد. بر اين اساس، براي ارزيابي نظام و بررسي كارايي بازيابي آن، كافي است روشهايي براي بررسي پاسخهاي ارائه شده به استفاده‌كننده ايجاد شود.

كارايي[[8]](http://192.168.0.110/editor/main.htm" \l "_ftn8" \o ") ميزاني براي سنجش توانايي نظام با توجه به ربط مدارك بازيابي شده است براي سنجش كارايي نظام بازيابي، دو مفهوم دقت[[9]](http://192.168.0.110/editor/main.htm" \l "_ftn9" \o ") و بازيافت[[10]](http://192.168.0.110/editor/main.htm" \l "_ftn10" \o ") مورد توجه قرار مي‌گيرند. از سوي ديگر واژگان نقش مهمي در كارايي و توانايي يك نظام بازيابي اطلاعات بازي مي‌كند. اصطلاح جستجوي مورد استفاده در فرايند بازيابي، ممكن است از زبان طبيعي يا واژگان كنترل شده استخراج شده باشد. يكي از مهمترين مزاياي واژگان كنترل شده، توانايي آنها در ارائه اصطلاحات جستجوي متعارفي است كه فاقد ابهامند، اما اين روش نيازمند ايجاد و روزآمدسازي مداوم واژگان كنترل شده است. روزآمدسازي واژگان و انعطاف‌پذيري در برابر تغييرات بويژه در موضوعات جديد كه به سرعت توسعه مي‌يابند، از اهميت بيشتري برخوردار است.

استفاده از اصطلاحات زبان طبيعي كه به عنوان جايگزيني براي واژگان كنترل شده معرفي شده است، با نگرانيهاي عمده‌اي در مورد ابهام و كارايي آن در فرايند بازيابي روبرو شده است. رفع ابهام از ابزارهاي جستجوي موضوعي و به كارگيري آنها در نظامهاي بازيابي پيوسته، دغدغه خاطر كتابداران و طراحان اين نظامهاست.

**هدفها و سؤالهاي پژوهش**

اين پژوهش با استفاده از معيارهاي ربط، دقت، بازيافت و ريزش كاذب، به ارزيابي و مقايسه كارايي زبان طبيعي و واژگان كنترل شده در بازيابي اطلاعات از پايگاههاي كتاب‌شناختي پيوسته مي‌پردازد تا در پرتو اين بررسي كاربران و جستجوكنندگان اطلاعات و طراحان پايگاههاي اطلاعاتي، بتوانند روشهاي بهتري براي جستجوهاي موضوعي در پايگاهها به كار گيرند .براي نيل به اين هدف، پژوهش حاضر در صدد پاسخگويي به سؤالهاي زير است:

1. كدام رويكرد جستجوي موضوعي: كنترل شده ـ (سرعنوان موضوعي و اصطلاح نامه) يا زبان طبيعي ـ در جستجوي اطلاعات بازيافت بيشتري به همراه دارد؟

2. كدام رويكرد جستجوي موضوعي ـ كنترل شده (سرعنوان موضوعي و اصطلاح نامه) يا زبان طبيعي ـ در جستجوي اطلاعات دقت بيشتري به همراه دارد؟

3. كدام رويكرد جستجوي موضوعي ـ كنترل شده (سرعنوان موضوعي و اصطلاح نامه) يا زبان طبيعي ـ در جستجوي اطلاعات ريزش كاذب كمتري به همراه دارد؟

**پيشينه پژوهش**

تاريخچه زبان طبيعي در مقابل واژگان كنترل شده به سه دوره تقسيم مي‌شود. نخستين دوره در قرن نوزدهم با افزايش نمايه‌سازي با استفاده از اصطلاحات عناوين آغاز شد. دوره دوم، با اختراع رايانه‌ها و نويدي كه با ارائه جستجوي متن آزاد در مقابل واژگان كنترل شده به همراه داشت به وجود آمد. در اين دوره، پژوهشهاي زيادي در زمينه مورد بحث صورت گرفت كه عمده‌ترين آنها تحقيقاتي بود كه در دهه 1960 در كرانفليد[[11]](http://192.168.0.110/editor/main.htm" \l "_ftn11" \o ") انگلستان براي ارزيابي عملكردهاي زبانهاي نمايه‌سازي گوناگون در بازيابي اجرا شد.

در دوره سوم، از اواسط ده 1970 به بعد، روشهاي متفاوتي براي بررسي مسئله واژگان كنترل شده در برابر متن آزاد مورد توجه قرار گرفت (مودماله[[12]](http://192.168.0.110/editor/main.htm" \l "_ftn12" \o ")، 1998).

چشمه سهرابي ( 1378) پايان‌نامه خود را تحت عنوان تأثير استفاده از اصطلاح ‌نامه در بانكهاي اطلاعاتي كتاب شناختي بر ميزان جامعيت، مانعيت و مدت زمان جستجوي اطلاعات بازيابي شده انجام داده است. يافته‌هاي وي نشان مي دهد ميزان جامعيت در جستجوهاي اصطلاح نامه‌اي كاهش يافته اما زمان جستجو و ميزان مانعيت بيشتر شده است. بهمن‌آبادي (1380) در مقاله‌اي از طريق بررسي متون و تحقيقات صورت گرفته، به بحث در مورد كاربرد اصطلاح نامه در بازيابي اطلاعات پرداخته است. او در پايان نتيجه گرفته است كه، هرچند استفاده از اصطلاح نامه‌هاي موجود‌ ممكن است بر جامعيت جستجوها بيفزايد، اما در عين حال لازم است براي كمك به استفاده كنندگان در كشف مفاهيم مورد نياز و انجام كاوشهاي مؤثرتر، گونه ‌خاصي از اصطلاح نامه تحت عنوان اصطلاح نامه بازيابي فراهم شود.

 پيترز[[13]](http://192.168.0.110/editor/main.htm" \l "_ftn13" \o ") و كورت[[14]](http://192.168.0.110/editor/main.htm" \l "_ftn14" \o ") (1991) پژوهشي تحت عنوان «جستجوي موضوعي واژگان كنترل شده و كنترل نشده در فهرست پيوسته يك كتابخانه دانشگاهي» انجام داده‌اند. نتايج اين پژوهش نشان مي‌دهد كه اغلب كاربران براي بازيابي اطلاعات مورد نياز خود جستجوي موضوعي كليد واژه‌اي را با جستجوي كنترل شده تركيب كرده‌اند. شاو[[15]](http://192.168.0.110/editor/main.htm" \l "_ftn15" \o ") (1994) در پژوهش خود به توصيف روشهايي مي‌پردازد كه براي بررسي عملكرد بازيابي نمونه‌هاي موضوعي كنترل شده و كنترل نشده در يكي از پايگاههاي مدلاين[[16]](http://192.168.0.110/editor/main.htm" \l "_ftn16" \o ") به كار رفته است. او به اين نتيجه رسيد كه نمونه‌هاي كنترل‌شده و كنترل‌نشده در همه سطوح بازيابي مورد انتظار عملكرد يكساني داشته‌اند و تركيبي از نتايج بازيابي شده، از نتايج حاصل از هر كدام از نمونه‌ها به تنهايي مفيدتر است. مود ماله [[17]](http://192.168.0.110/editor/main.htm" \l "_ftn17" \o ")(1998) پژوهشي با عنوان «زبان طبيعي در مقابل واژگان كنترل شده در بازيابي اطلاعات» انجام داده است. در اين پژوهش، كارايي ابزارهاي بازيابي اطلاعات (اصطلاح نامه و زبان طبيعي) مورد بررسي قرار گرفت و در نهايت محقق با توجه به عملكردهاي متفاوت دو روش فوق، تركيبي از آنها را براي انجام جستجوها پيشنهاد مي‌كند. بالارد[[18]](http://192.168.0.110/editor/main.htm" \l "_ftn18" \o ") (1998) در مقاله‌اي گزارش مي‌كند كه اضافه كردن يك گزينه جستجوي كليدواژه‌اي در فيلد موضوع به فهرست كتابخانه‌اش، باعث كاهش 75درصدي ميزان عدم موفقيت در جستجوهاي موضوعي شده است. آرلانو[[19]](http://192.168.0.110/editor/main.htm" \l "_ftn19" \o ") (1999) پژوهشي با عنوان «جستجوي موضوعي در فهرستهاي پيوسته شامل مواد اسپانيايي و انگليسي»" انجام داده است. نتايج اين پژوهش نشان مي‌دهد استفاده از زبان كنترل شده در جستجوهاي موضوعي فهرست، سودمندي بيشتري به همراه داشته است، با وجود آن كه استفاده از كلمات عنوان، شيوه‌اي است كه به زبان كاربران نزديكتر است.

توجه به اين پژوهشها نشان مي‌دهد تأثير رويكردهاي موضوعي در بازيابي از نظامهاي اطلاعاتي، مقوله‌اي است كه از ديدگاههاي گوناگون مورد بررسي قرار گرفته است. استفاده از اصطلاح نامه‌هاي خاص،رفتارهاي اطلاع‌يابي،پايگاههاي دوزبانه و تلاش در جهت افزايش مزاياي يك رويكرد با استفاده از رويكردهاي ديگر، جنبه‌هاي مختلف اين پژوهشها را نشان مي‌دهند.

**روش‌شناسي پژوهش**

پژوهش حاضر كه تحقيقي كاربردي ـ توصيفي است، به گردآوري داده‌ها از طريق تحليل محتوا پرداخته است. براي استخراج موضوعاتي كه‌ بتوان‌ سه ‌رويكرد جستجوي موضوعي با استفاده از زبان طبيعي، سرعنوان موضوعي يا اصطلاح نامه را در مورد آنها بررسي كرد از عناوين موجود در راهنماي موضوعي Books In Print 2000-2001 (BIP) استفاده شد. تعداد 30 عنوان[[20]](http://192.168.0.110/editor/main.htm" \l "_ftn20" \o ") ذيل موضوعات مربوط به كتابخانه، كتابداري و اطلاع‌رساني به صورت تصادفي انتخاب شدند. سرعنوانهاي موضوعي مناسب اين عناوين از فهرست سرعنوانهاي موضوعي كتابخانه كنگره انتخاب شدند تا به عنوان راهبردهاي جستجوي كنترل شده در فهرست پيوسته كتابخانه كنگره به كار روند. اختصاص توصيفگرهاي اصطلاح نامه اريك به هر عنوان با استفاده از امكانات پايگاه پيوسته اريك و نظر متخصصان موضوعي صورت پذيرفت. براي مشخص كردن راهبردهاي جستجوي زبان طبيعي، از واژه‌هاي موجود در عناوين و عملگرهاي بولي استفاده شد.

در مرحله دوم تحقيق، سرعنوان‌ها، توصيفگرها و راهبردهاي زبان طبيعي به دست آمده، در پايگاه پيوسته كتاب شناختي اريك و فهرست پيوسته كتابخانه كنگره مورد جستجو قرار گرفتند. ده پيشينه اول حاصل از هرجستجوي موضوعي مورد بررسي قرار گرفت. اين بررسي شامل ارزيابي ربط موضوعي هر كدام از پيشينه‌هاي بازيابي شده به عنوان مربوط به خود بود. با توجه به اين كه عناوين منتخب براي انجام جستجوهاي موضوعي در اين پژوهش بر اساس نياز واقعي كاربران پايگاهها نيستند. بنابراين، ارزيابي ربط پيشينه‌هاي بازيابي شده بر اساس قضاوت متخصص موضوعي صورت پذيرفت. ربط تصوري نظري است كه از دو ديدگاه متفاوت عيني و ذهني مورد بررسي قرار مي‌گيرد.نگرش عيني، وضعيت خاصّ دانش استفاده‌كننده را مورد توجه قرار نمي‌دهد.در نتيجه، در اين نگرش تأكيد عمده در واقع بر نظامهاي بازيابي اطلاعات است. پارك[[21]](http://192.168.0.110/editor/main.htm" \l "_ftn21" \o ") عوامل مؤثر در قضاوتهاي ربط را در چند گروه خلاصه مي‌كند: ويژگيهاي فردي و دريافت و تصور استفاده‌كننده، شرايط خارجي و شرايط مسئله.از آنجا كه در اين پژوهش قضاوت ربط از ديدگاه عيني مورد توجه بوده و به وسيله متخصص موضوعي صورت گرفته است ،ويژگيهاي فردي و دريافت و تصور استفاده‌كنندهدر قضاوتهاي ربط مؤثر نبوده‌اند.براي ارزيابي شرايط خارجي، محصول نهايي براي هر موضوع منتخب يك پايان‌نامه تصور شده و هدف هر جستجو به دست آوردن اطلاعات جامع براي انجام كليه مراحل پژوهش تلقي شده است. همه نتايج حاصل از جستجوها دسترس‌پذير فرض شده‌اند. در مورد شرايط مسئله همه مداركي كه به روشن شدن گستره موضوع منتخب،تعريف و درك و پيشينه آن، روشهاي بررسي مسئله، چارچوب و جنبه‌هاي جديد و قابل توجه آن كمك كرده‌اند مربوط تلقي شده اند.

**تجزيه وتحليل يافته‌ها**

تجزيه و تحليل داده‌ها در اين پژوهش با استفاده از روشهاي آمار توصيفي و استنباطي با استفاده از نرم‌افزارSPSS صورت پذيرفت. براي مقايسه دقت و ريزش كاذب رويكردهاي موضوعي از آزمونt  در دو گروه مستقل و براي مقايسه بازيافت رويكردهاي موضوعي مختلف، از آزمون «خي دو» براي داده‌هاي ناپارامتريك استفاده شد. جدول1، تعداد جستجوها و پيشينه‌هاي بازيابي شده در دو پايگاه پيوسته كتاب شناختي اريك و فهرست پيوسته كتابخانه كنگره را نشان مي‌دهد.

**جدول 1. تعداد جستجوها و پيشينه هاي بازيابي شده**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **پايگاه پيوسته كتابشناختي** | **رويكرد جستجوي موضوعي** | **تعداد جستجوها** | **تعداد پيشينه بازيابي شده** | **تعداد پيشينه بازيابي شده تا ده پيشينه اول** | **تعداد پيشينه مربوط** |
| **فهرست كتابخانه كنگره** | **زبان طبيعي** | 81 | 55615 | 513 | 190 |
| **كنترل شده** | 77 | 20249 | 600 | 193 |
| **پايگاه اريك** | **زبان طبيعي** | 81 | 432005 | 715 | 374 |
| **كنترل شده** | 54 | 313193 | 503 | 81 |
| **جمع** |  | 293 | 821062 | 2331 | 838 |

براي بررسي سؤال اول تحقيق كه هدف آن مقايسه بازيافت رويكردهاي جستجوي موضوعي كنترل شده با زبان طبيعي بود، از آزمون غيرپارامتريك «خي دو» استفاده گرديده است.

**جدول 2.آزمون معناداري تفاوت بازيافت در دو رويكرد زبان طبيعي و كنترل‌شده**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **پايگاه پيوسته كتابشناختي** | **رويكرد جستجوي موضوعي** | **مشاهده شده** | **مورد انتظار** | **مجذور كاي** | **درجه آزادي** | **احتمال** |
| **فهرست كتابخانه كنگره** | **زبان طبيعي** | 55615 | 37932 | 79/16486 | 1 | 000/0 |
| **كنترل شده** | 20249 | 37932 |
| **پايگاه اريك** | **زبان طبيعي** | 432005 | 372599 | 5/9471 | 1 | 000/0 |
| **كنترل شده** | 313193 | 372599 |

با توجه به جدول2 و ميزان خي دو به دست آمده مي‌توان ادعا نمود كه در فهرست كتابخانه كنگره ميزان بازيافت در رويكرد موضوعي زبان طبيعي بيش از رويكرد موضوعي كنترل شده مي‌باشد و تقريباً 7/2 برابر آن مي‌باشد.

در پايگاه اريك نيز بازيافت پيشينه‌ها در رويكرد موضوعي زبان طبيعي بيش از رويكرد موضوعي كنترل شده مي‌باشد و اين تفاوت از نظر آماري معنادار است.

بنابراين، در پاسخ به سؤال اول پژوهش مي‌توان گفت كه در هر دو پايگاه مورد بررسي بازيافت رويكرد زبان طبيعي بيش از رويكرد واژگان كنترل شده بوده و اين تفاوت از نظر آماري معنادار است.

سؤال دوم تحقيق، بررسي ميزان دقت در دو رويكرد زبان طبيعي و كنترل شده مي‌باشد. ميزان دقت بر اساس فرمول زير محاسبه گرديد:

**100 × ) تعداد مدارك بازيابي شده تا 10 پيشينه اول/ تعداد مدارك مربوط) = دقت**

براي بررسي اين سؤال از آزمون t استفاده گرديده است .

**جدول 3. آزمون معناداري تفاوت ميانگين دقت در دو رويكرد كنترل شده و زبان طبيعي**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **پايگاه پيوسته كتابشناختي** | **رويكرد جستجوي موضوعي** | **ميانگين دقت (درصد)** | **SD** | **t-value** | **df** | **sig** |
| **فهرست كتابخانه كنگره** | **زبان طبيعي** | 9/41 | 1/28 | 22/0- | 58 | 829/0 |
| **كنترل شده** | 2/40 | 75/31 |
| **پايگاه اريك** | **زبان طبيعي** | 4/50 | 8/26 | 02/6- | 48 | 000/0 |
| **كنترل شده** | 7/15 | 45/16 |

دقت فهرست پيوسته كتابخانه كنگره در رويكرد زبان طبيعي و كنترل شده، تفاوتي معنادار با هم ندارند. بررسي ميزان دقت در دو رويكرد كنترل شده و زبان طبيعي در پايگاه اريك، نشان مي‌دهد كه دقت در رويكرد زبان طبيعي در اين پايگاه بيش از رويكرد كنترل شده مي‌باشد، به طوري كه ميانگين دقت در رويكرد زبان طبيعي در اين پايگاه 4/50% است در حالي كه اين ميانگين براي رويكرد كنترل شده 7/15% مي باشد كه تفاوتي در حدود 7/34% را نشان مي‌دهد. جدول3، اين نتايج را نشان مي‌دهد.

دقت در داده‌هاي حاصل از اين پژوهش، نتايج ديگري را نيز روشن مي‌كند. اگر سر عنوانهاي موضوعي موجود در پژوهش برحسب تعداد تقسيم فرعي آنها گروه‌بندي شده و تعداد پيشينه‌هاي حاصل از هر گروه و دقت بازيابي آنها محاسبه شود، جدول4 به دست خواهد آمد.

**جدول4. تعداد پيشينه‌ها و دقت حاصل از سر عنوانهاي موضوعي با تقسيمهاي فرعي متفاوت**

|  |  |  |  |
| --- | --- | --- | --- |
| **سرعنوانهاي موضوعي** | **تعداد سرعنوانها** | **تعداد پيشينه‌هاي بازيابي شده** | **دقت** |
| بدون تقسيم فرعي | 28 | 280 | 17% |
| داراي يك تقسيم فرعي | 24 | 222 | 34% |
| داراي دو تقسيم فرعي | 20 | 73 | 77% |
| داراي سه تقسيم فرعي | 5 | 18 | 78% |

داده‌هاي اين جدول نشان مي‌دهد هرچه تعداد تقسيم فرعي سرعنوان موضوعي افزايش يافته، تعداد پيشينه‌هاي بازيابي شده كاهش يافته و دقت بازيابي آنها افزايش پيدا كرده است.

سؤال سوم كه در حقيقت عكس سؤال دوم تحقيق است، بررسي ميزان ريزش كاذب در دو رويكرد زبان طبيعي و كنترل شده مي باشد . ريزش كاذب در فهرست پيوسته كتابخانه كنگره در رويكرد زبان طبيعي و كنترل شده تفاوتي معنادار با هم ندارد. بررسي ميزان ريزش كاذب در دو رويكرد كنترل شده و زبان طبيعي در پايگاه اريك، نشان مي‌دهد كه ريزش كاذب در رويكرد زبان طبيعي در اين پايگاه كمتر از رويكرد كنترل شده مي باشد.

**بحث، نتيجه‌گيري و پيشنهاد**

نتايج تحقيق و آزمونهاي انجام شده در پاسخ به سؤالهاي تحقيق نشان ‌مي‌دهد دقت و بازيافت جستجوهاي موضوعي انجام شده با استفاده از رويكرد زبان طبيعي در مقايسه با رويكرد كنترل، در مجموع بهتر بوده است.

به نظر مي‌رسد كارايي بهتر رويكرد زبان طبيعي در جستجوي اطلاعات پيوسته، ناشي از سنخيت بيشتري است كه اين رويكرد با محيط شبكه و وب دارد. موتورهاي جستجوي پايگاههاي اطلاعاتي اغلب براساس ويژگيهايي عمل مي‌كنند كه با رويكرد زبان طبيعي هماهنگ‌تر است. اين موتورها ارزيابيهاي ربط خود را براساس تحليلهاي كمّي از واژگان مدارك خود و راهبردهاي جستجوي كاربران انجام مي‌دهند و امكاني براي تحليل محتوي مدارك يا بررسي روابط معنايي موجود ميان واژگان جستجو ندارند.

از سوي ديگر شايد پيش از آن كه عملكرد بهتر رويكرد زبان طبيعي در اين پژوهش دليل مزيت اين رويكرد باشد ، بايد آن را ناشي از موانع موجود در استفاده از زبان كنترل شده تصور كرد. در رويكرد كنترل شده، وجود واژگان نمايه‌سازي كامل و روزآمد عنصري تعيين كننده در موفقيت نظام نمايه‌سازي است. اين نكته زماني روشن‌تر مي‌شود كه نتايج تحقيق در دو پايگاه اريك و فهرست پيوسته كتابخانه كنگره را مورد مقايسه قرار دهيم.

در فهرست پيوسته كتابخانه كنگره زبان نمايه‌سازي، فهرست سرعنوانهاي موضوعي كتابخانه كنگره است. اين فهرست به عنوان يك استاندارد ملي و تا حدي استاندارد بين‌المللي مورد استفاده قرار مي‌گيرد. نتايج پژوهش در فهرست پيوسته كتابخانه كنگره نشان مي‌دهد از نظر دقت تفاوتي معنا‌دار ميان دو رويكرد زبان طبيعي و كنترل شده در اين پايگاه وجود ندارد. در حالي كه در پايگاه اريك، اين تفاوت كاملاً معنا‌دار است. در پايگاه اريك توصيفگرها از اصطلاح نامه اريك انتخاب مي‌شوند. در مرحله گردآوري داده‌ها براي اين پژوهش، اختصاص توصيفگرها با موانعي همراه بود كه به نظر مي‌رسد ناشي از نقص اصطلاح نامه اريك و روزآمد نبودن آن دست كم در حوزه موضوعي كتابخانه، كتابداري و اطلاع‌رساني بوده است. با توجه به مطالب ذكرشده، شايد بتوان اين طور نتيجه‌گيري كرد كه درصدي از تفاوت بسيار زياد موجود ميان نتايج دو رويكرد در پايگاه اريك، ناشي از نقص اصطلاح نامه اريك باشد.

از سوي ديگر، اگر مقايسه اي ميان كلمات عناوين مورد نظر و كلمات اصطلاحات كنترل شده اختصاص يافته به همان موضوعات صورت پذيرد، روشن مي‌شود كه كلمات عناوين درصد كمي از مفاهيم موجود در موضوعات را تحت پوشش قرار مي‌دهند كه اين خود مي‌تواند نشانه‌اي از عدم جامعيت نتايج جستجوهاي موضوعي با استفاده از رويكرد در زبان طبيعي باشد.

با توجه به همه نكات ذكر شده در اين قسمت مي‌توان چنين نتيجه‌گيري كرد:

با وجود آن كه در جستجوهاي موضوعي با استفاده از رويكرد زبان طبيعي در مقايسه با رويكرد كنترل شده در اين پژوهش دقت و بازيافت بيشتر و ريزش كاذب كمتري حاصل شده است، نمي توان در مورد كارايي بهتر اين رويكرد نظر قطعي ارائه كرد. با توجه به اين كه تعيين تعداد پيشينه‌هاي مربوط كه در پايگاههاي مورد نظر وجود داشته ولي بازيابي نشده‌اند ممكن نيست، لذا تعيين داده‌هاي كمّي در مورد ميزان جامعيت پيشينه‌هاي حاصل از دو رويكرد و مقايسه آنها با هم نيز مقدور نمي‌باشد. بنابراين، نمي‌توان در مورد جامعيت نتايج حاصل از رويكرد زبان طبيعي بويژه با توجه به نكاتي كه در مورد نقص آنها نسبت به راهبردهاي كنترل شده ارائه شد مطمئن بود. از طرف ديگر، چنانكه گفته شد، احتمال دارد بخشي از اين برتري نسبي رويكرد زبان طبيعي نسبت به رويكرد كنترل شده ناشي از ايرادهاي وارد به زبان نمايه‌سازي پايگاه اريك باشد. بنابراين، پيشنهاد مي‌شود در انجام جستجوهاي موضوعي كامل در پايگاههاي پيوسته، تركيبي از دو رويكرد زبان طبيعي و كنترل‌شده براي به دست آوردن نتايج جامع استفاده شود. براي انجام جستجوهايي كه جامعيت نتايج حاصل از آنها اهميت چنداني ندارد و هدف به دست آوردن تعدادي پيشينه مربوط است، رويكرد زبان طبيعي از مزيت بيشتري برخوردار است. استفاده از واژگان كنترل شدة اخص تر در جستجو، سبب افزايش دقت بازيابي مي‌گردد.

بهتر است كاربران پايگاهها نيازهاي اطلاعاتي خود را در قالب عناوين روشن و كامل بيان كنند و مؤلفان، عناوين واضح و جامع براي مدارك خود بر گزينند. چنين عناويني، كارايي جستجوهاي موضوعي بويژه جستجوهايي را كه با استفاده از رويكرد زبان طبيعي صورت مي‌گيرد، افزايش مي‌دهند.

**منابع**

ـ بهمن‌آبادي، عليرضا (1380). «كاربرد اصطلاحنامه در بازيابي اطلاعات» **فصلنامه كتاب**، 12 (2 ): 77 .

ـ پائو، ميراندا لي(1378). **مفاهيم بازيابي اطلاعات**، ترجمه اسدالله آزاد و رحمت‌الله فتاحي. مشهد: دانشگاه فردوسي.

ـ تيلور، آرلين جي(1381). **سازماندهي اطلاعات**، ترجمه محمدحسين دياني. مشهد: كتابخانه رايانه‌اي.

ـ چشمه سهرابي، مظفر (1378). **تأثير استفاده از اصطلاحنامه در بانكهاي اطلاعاتي كتابشناختي بر ميزان جامعيت، مانعيت و مدت زمان جستجوي اطلاعات بازيابي‌شده**. پايان‌نامه كارشناسي ارشد، دانشگاه تربيت مدرس، دانشكده علوم.

ـ حريري، نجلا (1377). «مفهوم ربط در بازيابي از نظامهاي اطلاعاتي». **فصلنامه كتاب**، 9(2):7.

ـ زينس، كيم(1381). «راهبرد كاوش نظام‌مند بازيابي اطلاعات در عصر نظامهاي جهاني اطلاعات» ، ترجمه فاطمه مكي‌زاده. **فصلنامه كتاب**، 3 ( 2 ):155.

ـ شهبازي، رحيم (1378). «تجسس موضوعي در نظامهاي فهرست پيوسته»، **فصلنامه اطلاع‌رساني**، 16 (1و2): 38 .

ـ مور، كارول، جك‌ كين (1996). « تكنولوژي نوين و رقمي كردن سرعنوانهاي موضوعي»، ترجمه سعيد اكبري‌نژاد. **گزيدة مقالات ايفلا (چين:25ـ31 اوت 1996):** 66.

ـ ويسمان، هرمان (1378). **سيستمها، خدمات و مراكز اطلاع‌رساني**، ترجمه جعفر مهراد. شيراز: نويد.

-Arellano, F. (1999). “Subject searching in online catalogs including Spanish and English material” **Cataloging and-Classification Quarterly**, 28 (2):45.

-Ballard, T. (1998). “Keyword / subject: finding a middle path; this hybrid search approach dramatically improves patron success" **Information Today** at [http://www.findarticles.com](http://www.findarticles.com/) (2003/12/29).

-Connell, Tschera Harkness ; Tipple ,Jennifer E. (1999) "Testing the accuracy of information on the WWW using the AltaVista search engine" **Reference and User Services Quarterly,**38(4):360.

-Hoffman, H.H. (2001). “Subject access to works in online catalogs” **Technicalities**, 21(5):9, at [http://gateway](http://gateway/)proquest.com (2003/12/29).

-Hunter, R.N. (1991). “Success and failures of patrons searching the online catalog at a large academic library: a transaction log analysis” **RQ**, 30(3):395.

-Jansen, Bernard (2000)."The effect of query complexity on web searching results" [http://www.shef.ac.uk/is/publications/infres/paper 87.html](http://www.shef.ac.uk/is/publications/infres/paper%2087.html) (2005/5/10)

-Leighton, H. Vernon (1995) "Performance of four WWW index services: Infoseek, Lycos, Web Crawler and wwworm"http://www. winona.msus.edu/library/webind.htm/   (2005/10/5)

-Muddamalle, M.R. (1998). “Natural language versus controlled vocabulary in information retrieval: a case study in soil mechanics” **Journal of American Society for Information Science**, 49 (10):881.

-Peters, T.A., M. Kurth (1991). “Controlled and uncontrolled vocabulary subject searching in an academic library catalog” **Information Technology and Libraries**, 10(3): 201.

-Rudner, L.M. (2003). “How many people search ERIC database each day” **Dlibmagazine**, 9(4). [http://www.dlib.org/dlib/ april03/rudner/04rudner.html.2003/](http://www.dlib.org/dlib/%20april03/rudner/04rudner.html.2003/) (2003/12/5).

-Shaw, W.M. (1994). “Retrieval expectations, cluster-based effectiveness, and performance standards in the CF database” **Information Processing and Management**, 30 (3):23.

-**Subject Guide to Books In Print 2000-2001(2000)**. New York: Bowker .

-Williamson, N.J. (1996). “Standards and rules for subject access” **Cataloging and Classification Standards and Rules**, 21(314):155.

-Yee, Ingrid Hsieh (1998). “The retrieval power of selected search engines: How well do they address general reference questions and subject questions?” **Reference Librarian**, 60(27).

1. عضو هيأت علمي بخش علوم كتابداري و اطلاع‌رساني دانشگاه شيراز

**2. كارشناس ارشد علوم كتابداري و اطلاع‌رساني**

1. Cutter

2. Hoffman

3. Taylor

4. Moore

5. Cain

1. Effectiveness

2. Precision

3. Recall

1. Cranfield

2. Muddamalle

1. Peters

2. Kurth

3. Shaw

4. MEDLINE

5. Muddamalle

6. Ballard

1. Arellano

2. Topic

1. Park