

بررسی ویژگی‌های معنایی و هستی‌شناسانه نظام‌های بازیابی اطلاعات مبتنی بر اصطلاحنامه و هستی‌شناسی

المیرا کریمی *

دانشجوی دکتری علم اطلاعات و دانش‌شناسی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

محمود بابایی

دکتری علوم ارتباطات
استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

ملوک السادات حسینی بهشتی

دکتری زبان‌شناسی
استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

دریافت: ۹۵/۱۰/۰۷

پذیرش: ۹۶/۰۳/۲۴

چکیده: نظام‌های بازیابی اطلاعات برای افزایش نتایج مرتبط بازیابی شده و کسب رضایت کاربران به دنبال تغییر جستجوی مبتنی بر کلیدواژه به جستجوی مبتنی بر مفاهیم هستند. این نظام‌ها برای انجام این کار از ابزارهایی استفاده می‌کنند که بتوانند واژگان را به مفاهیم تبدیل کنند. این امر طی سالیان متمادی زمینه تولید ابزارهایی همچون طرح‌های رده‌بندی، سرعنوان‌های موضوعی، واژگان و اصطلاحنامه‌ها را فراهم آورده است. به تازگی نیز هستی‌شناسی‌ها در تکامل این روند به کار گرفته شده‌اند.

این مسئله که نظام‌های بازیابی اطلاعات مطرح در دنیا که مدعی استفاده از هستی‌شناسی و اصطلاحنامه هستند چگونه از ویژگی‌های معنایی این ابزارها در روند بازیابی اطلاعات خود بهره می‌گیرند، مشکل اصلی است که در این پژوهش به آن پرداخته شده است. در این مقاله ۷ نظام بازیابی اطلاعات با توجه به استفاده از این ابزارها انتخاب شدند، ویژگی‌های معنایی نظام‌ها با بررسی متون در قالب سیاهه واری محقق ساخته‌ای ارائه شد. این ویژگی‌ها ۵ مقوله کلی جستجوی مفهومی، جستجوی مترادف‌ها، پیشنهاد برای جستجو، استفاده از هستی‌شناسی‌ها و نمایش هستند. پژوهشگر با مشاهده ساختاریافته در قالب سیاهه، این ویژگی‌ها را مورد بررسی قرار داد. با وجود این که این نظام‌ها از این ابزارها برای جستجو و بازیابی اطلاعات بهره گرفته‌اند، اما ویژگی‌های

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۸۲۲۳-۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS و LISA، ISC و
<http://jlist.irandoc.ac.ir>
دوره XX | شماره X | صص XX-XX
۱۳XX X

نوع مقاله: پژوهشی

به این مقاله به شکل زیر استناد کنید:

دورن متن:

(کریمی، المیرا، بابایی و حسینی بهشتی، زودآیند)

در فهرست منابع:

کریمی، المیرا، بابایی، محمود و حسینی بهشتی، ملوک السادات. زودآیند. بررسی ویژگی‌های معنایی هستی‌شناسانه نظام‌های بازیابی اطلاعات مبتنی بر اصطلاحنامه و هستی‌شناسی. پژوهشنامه پردازش و مدیریت اطلاعات.

<http://jipm.irandoc.ac.ir> (دسترسی در

۹۵ / ماه / سال)

معنایی و هستی‌شناسانه مورد انتظار را نشان ندادند. پایگاه اطلاعاتی اریک و کتابخانه ملی استرالیا بیشترین امتیاز (۳۲ و ۲۴) را در معیارهای بررسی شده به دست آوردند و آرشیو ملی استرالیا و موزه بریتانیا کمترین امتیاز (۱۲ و ۱۳) را داشتند.

کلیدواژه‌ها: جستجوی معنایی، ویژگی‌های معنایی بازیابی اطلاعات، نظام‌های بازیابی اطلاعات، جستجوی مفهومی، جستجوی مترادف‌ها، هستی‌شناسی

*المیرا کریمی karimi@students.irandoc.ac.ir

۱. مقدمه و مسئله

جستجوی اطلاعات، فعالیت انسانی پیچیده‌ای است که با یک نیاز اطلاعاتی آغاز می‌شود و در بعضی موارد شامل چندین شکل استراتژی جستجو و مرور منابع اطلاعاتی گوناگونی است. هر مدلی از جستجوی اطلاعات شرحی از مفهوم نیاز اطلاعاتی، که فرایند جستجو با آن شروع می‌شود، فراهم می‌کند. از دیدگاه شناختی، این نیازها، که در بسیاری موارد مبهم هستند، ممکن است از طریق مفاهیم سطح بالا بیان شوند (Sutcliffe and Ennis 1998) که به نوبه خود به مفاهیمی دیگر مرتبط هستند که بخشی از دامنه خاص یا درون تعداد کمی از آنهاست. (García and Sicilia 2003). یافتن پاسخی برای این نیاز اطلاعاتی از سوی نظام‌ها، نیازمند دریافت دقیق معنای پرس و جوی کاربر است.

نظام‌های بازیابی اطلاعات تا کنون اغلب به کلمات و کلیدواژه‌ها بیش از معنای آنها می‌پرداختند به نوعی که در اکثر موارد معنای کلیدواژه مورد نظر کاربر نادیده گرفته می‌شد. این نظام‌ها از فهرست‌های کلیدواژه‌ای برای توصیف محتوای اطلاعاتی استفاده می‌کنند، اما مسئله این است که چنین فهرستی چیزی در مورد ارتباطات معنایی بین کلیدواژه‌ها نمی‌گوید و معنای کلمات و عبارات را مورد توجه قرار نمی‌دهد. (Styltsvig 2006). همان‌طور که آلن والکر رید اشاره می‌کند:

« خطر بزرگ در پرداختن به کلمات این است که کلمه را به جای معنا در نظر می‌گیرند. این غلط است. یک کلمه، در چارچوب اجتماعی ایجاد می‌شود که می‌تواند با چند جنبه یا بعد که همگی

¹ . Allen Walker Read

در یک زمان ارائه می‌شوند، در نظر گرفته شود. توصیف معنایی کامل یک کلمه شامل حداقل چهار جنبه یا بعد است، «واجگان»، «نحو»، «ساختواژه» و «کاربرد».

اغلب برای کاربران عادی استفاده از سیستم‌های اطلاعاتی بر پایه فنون کلیدواژه محور مشکل است. کاربران معمولاً در بیان نیازهای اطلاعاتی خود و ترجمه نیازها به سؤال با مشکل مواجهند.

نوع رفتار جستجوی کاربران مستلزم عرضه دانش جدید و فنون بازیابی و راهبردهای بازیابی جدید است. بازیابی اسناد از طریق شیوه‌های معنایی¹، مبتنی بر تحلیل ساختارهای معنایی محتوای اسناد صورت می‌گیرد. (Hynek 2002). فنون مختلفی براساس هوش مصنوعی و پردازش زبان طبیعی برای پردازش معنایی به کار می‌روند و بیشتر آنها به استفاده از ساختارهای کمکی همچون واژگان کنترل شده و هستی‌شناسی‌ها متکی هستند. با استفاده از واژگان کنترل شده (واژه‌نامه‌ها و اصطلاحنامه‌ها) اصطلاحات اخص، اعم و مرتبط در پرس‌وجوها گنجانده می‌شوند. واژگان کنترل شده یکی از روش‌های غلبه بر بعضی از شدیدترین موانع پرس‌وجوهای کلیدواژه‌ای هستند. در طی سال‌ها، ساختارهای کمکی دیگری نیز، مثل مجموعه‌های بزرگ مترادف وردنت، برای این هدف ایجاد شد. مشخص شده است که جستجوی مفهومی براساس ساختارهای کمکی مثل وردنت می‌تواند به طور کارآمدتری با استفاده مجدد مدل‌های بازیابی اطلاعات و ساختارهای داده‌ای از بازیابی اطلاعات کلاسیک به کار روند. مدل‌های داده‌ای که مجموعه‌هایی از مفاهیم در دامنه خاص (هستی‌شناسی‌های دامنه‌ای) را ارائه می‌دهند، و آنهایی که می‌توانند ارتباطات میان اصطلاحات را دربرگیرند، در سال‌های اخیر به کار گرفته شده‌اند.

بنابراین طبق آنچه گفته شد مشخص است که مشکلات جستجوی کلیدواژه‌ای و معنایی می‌تواند منجر به بازیابی منابع غیرمرتبط با نیاز اطلاعاتی کاربر شود. به این دلیل، لزوم توجه به معنا کاملاً احساس می‌شود. نظام‌های بازیابی بزرگ دنیا به دنبال راهی برای حل این مشکل هستند و هر یک از ابزارهای برای رفع مشکل استفاده کرده‌اند. نظام‌های بازیابی اطلاعات موجود در ایران نیز در پی رسیدن به جستجوی معنایی و حل مشکل جستجوی کلیدواژه‌ای هستند. تقریباً در ایران نیز مانند دنیا، نظام جامعی که بتواند به ویژگی‌های معنایی پردازد، وجود ندارد. بنابراین، این نظام‌ها در مرحله اول می‌توانند از رویکردهای مورد استفاده در نظام‌های بازیابی اطلاعات بزرگ دنیا ایده

¹ . Semantic Methods

گرفته و سپس با توجه به رویکردها و ویژگی‌های مورد استفاده آنها، نظام بازیابی معنایی خود را پایه گذاری کنند.

این مقاله به دنبال پاسخ به این پرسش است که کدام ویژگی‌های معنایی در نظام‌های بازیابی اطلاعات که مدعی استفاده از ابزارهای معنایی هستند، مورد استفاده قرار گرفته‌اند و آیا با استفاده از این ابزارها به اهداف معنایی اشاره شده در متون رسیده‌اند یا خیر؟ برای بررسی این مسئله محقق با بررسی متون به سیاهه‌ای دست یافت که براساس آن به مشاهده این نظام‌ها پرداخت.

۲. هدف‌های پژوهش

هدف اصلی این پژوهش بررسی عملکرد نظام‌های بازیابی اطلاعات موردنظر پژوهش از نظر استفاده از ویژگی‌های معنایی است. برای رسیدن به این هدف، اهداف فرعی زیر دنبال خواهد شد:

۱. ارزیابی ویژگی‌های جستجوی مفهومی در نظام‌های بازیابی اطلاعات مورد نظر پژوهش
۲. ارزیابی ویژگی‌های جستجوی مترادف‌ها در نظام‌های بازیابی اطلاعات موردنظر پژوهش
۳. ارزیابی ویژگی‌های پیشنهاد برای جستجو در نظام‌های بازیابی اطلاعات مورد نظر پژوهش
۴. ارزیابی ویژگی‌های مربوط به استفاده از هستی‌شناسی‌ها در نظام‌های بازیابی اطلاعات مورد نظر پژوهش
۵. ارزیابی ویژگی‌های نمایش در نظام‌های بازیابی اطلاعات مورد نظر پژوهش

۳. سؤال‌های پژوهش

سؤال اصلی این پژوهش این است که عملکرد نظام‌های بازیابی اطلاعات مورد نظر پژوهش از نظر استفاده از ویژگی‌های معنایی چگونه است؟ برای پاسخ به این سؤال اصلی، سؤالات فرعی زیر دنبال خواهد شد:

۱. جستجوی مفهومی در نظام‌های بازیابی اطلاعات مورد نظر پژوهش چگونه است؟
۲. ویژگی جستجوی مترادف‌ها در نظام‌های بازیابی اطلاعات موردنظر پژوهش چگونه است؟

۳. ویژگی پیشنهاد برای جستجو در نظام‌های بازیابی اطلاعات مورد نظر پژوهش چگونه است

۴. ویژگی مربوط به استفاده از هستی‌شناسی‌ها در نظام‌های بازیابی اطلاعات مورد نظر پژوهش چگونه است؟

۵. ویژگی نمایش در نظام‌های بازیابی اطلاعات مورد نظر پژوهش چگونه است؟

۴. پیشینه پژوهش

مروری بر کارهایی که تا کنون در حوزه ارزیابی نظام‌های بازیابی اطلاعات انجام شده است نشان داد که به طور کلی تقریباً به هیچ یک از این نظام‌ها از دیدگاه معنایی نگاه نشده است و محققان اغلب برای ارزیابی این نظام‌های بازیابی اطلاعات رویکردهای کلی و تخصصی ارزیابی عملکرد به جز ویژگی‌های معنایی را در نظر گرفته‌اند. علاوه بر این، مرور متون نشان داد که بیشتر بررسی‌های انجام شده بر موتورهای جستجوی معنایی متمرکز هستند. از میان بررسی‌های انجام شده بر روی موتورهای جستجوی معنایی نیز تعداد کمی از آنها به بررسی ساختاریافته پرداخته‌اند و اغلب مطالعات به مطالعه اجمالی و بررسی کلی این موتورهای جستجو اکتفا نموده‌اند. در واقع اغلب مطالعات با انجام جستجوهای مختلف در این موتورهای جستجو ویژگی‌های معنایی آنها را به تصویر کشیده‌اند. «ترنر»، «شاه» و «بیتیریم»^۱ (۲۰۰۹)، «آنداگو»، «فوبی»، و «تانون»^۲ (۲۰۱۰)، «اولری»^۳ (۲۰۱۰)، «سودپتی»، «انورادها» و «بابو»^۴ (۲۰۱۲)، «عزیزان»^۵ و همکاران (۲۰۱۳)، «نگی» و «کومار»^۶ (۲۰۱۴) تماماً با انجام جستجوهای معنایی به بررسی ویژگی‌های معنایی در موتورهای جستجو پرداختند. اغلب آنها فهرستی از امکانات و ویژگی‌های مورد استفاده در موتورهای جستجو را ارائه کردند. در بعضی از آنها از جمله «ترنر»، «شاه» و «بیتیریم» (۲۰۰۹) مقایسه‌ای بین موتورهای جستجوی معنایی و موتورهای جستجوی عادی انجام شده است و در نهایت به این نتیجه رسیدند که عملکرد جستجوی معنایی در هر دو گروه موتور جستجوی کلیدواژه‌ای و موتور جستجوی

¹ . Turner, Shah, and Bitirim

² . Andago, Phoebe, and Thanoun

³ . O, leary

⁴ Sudeepthi, Anuradha and Babu

⁵ .Azizan et al

⁶ . Negi and Kumar

معنایی در سطح پایینی قرار دارد. «آنداگو»، «فوبی»، و «تانون» (۲۰۱۰) نیز این مقایسه را بین گوگل و هاکیا انجام دادند. نتایج این پژوهش نشان داد که گوگل عملکرد بهتری نسبت به هاکیا دارد. «اولری» (۲۰۱۰) نیز مانند «آنداگو»، «فوبی» و «تانون» به مقایسه گوگل و هاکیا پرداخت و نتایج پژوهش او نیز عملکرد بهتر گوگل نسبت به هاکیا را نشان داد. «نگی» و «کومار» (۲۰۱۴) این مقایسه را بین دو موتور جستجوی معمولی گوگل و یاهو و سه موتور جستجوی معنایی هاکیا، داک داک گو و بینگ انجام دادند و در نهایت به این نتیجه رسیدند عملکرد جستجوی معنایی موتورهای جستجوی معنایی بالاتر از موتورهای جستجوی عادی است.

پژوهش‌های «سودپتی»، «انورادها» و «بابو» (۲۰۱۲) و «عزیزان» و همکاران (۲۰۱۳) بر روی موتورهای جستجوی معنایی انجام شد. هدف آنها ارزیابی این موتورهای جستجو نبود بلکه آنها به دنبال یافتن ویژگی‌ها و امکانات معنایی در این موتورهای جستجو بودند و در نهایت فهرستی از این ویژگی‌ها را ارائه کردند.

در ایران نیز «دری» در مقاله‌ای عملکرد ۵ موتور جستجوی معنایی را که به صورت رایگان در وب قابل دسترس هستند با استفاده از ۴۵ معیار مشخص در قالب یک سیاهه واری محقق ساخته مورد ارزیابی و مقایسه قرار داد. این سیاهه شامل دو بخش ویژگی‌های معنایی و عادی موتورهای جستجو بود. نتایج پژوهش نشان داد که موتورهای جستجوی مورد بررسی در شاخص ویژگی‌های عادی و معنایی، عملکرد مطلوب و کارایی مورد انتظار را نداشتند. نتایج مقایسه ویژگی‌های معنایی در این موتورهای جستجو نشان داد که DuckDuckGo و Hakia بیشترین امتیاز را در این مسئله دارا هستند. (دری ۱۳۹۳).

همانطور که در بالا اشاره شد، پژوهش‌های بسیار کمی به بررسی ویژگی‌های معنایی در نظام‌های بازیابی شناخته شده جهان پرداخته‌اند و از طرفی از دیدگاه هستی‌شناسانه این ویژگی‌ها مورد توجه قرار نگرفته است. این پژوهش با بررسی ویژگی‌های معنایی در نظام‌های بزرگ که از ابزارهای معنایی استفاده می‌کنند به استفاده از هستی‌شناسی‌ها و ویژگی‌های مربوط به آنها نگاه ویژه خواهد داشت.

روش پژوهش

این پژوهش از نوع پژوهش‌های توصیفی کاربردی است که به پیمایش نظام‌های بازیابی اطلاعات بزرگ دنیا می‌پردازد. در این پژوهش از مشاهده ساختاریافته با استفاده از سیاهه واری محقق

ساخته‌ای برای بررسی معیارها و دست‌یابی به داده‌های پژوهش استفاده شده است. برای تدوین این سیاهه با مطالعه متون، ویژگی‌های معنایی مورد نظر در پژوهش دری (۱۳۹۳) به عنوان اساس اولیه مورد توجه قرار گرفت که شامل ۴۵ معیار است. این سیاهه با مطالعه سایر منابع تکمیل گردید. روایی این سیاهه با نظر متخصصان این حوزه تأیید شده است. جامعه این پژوهش نظام‌های بازیابی اطلاعات شناخته شده دنیا هستند که از ابزارهای معنایی (شامل فرهنگ‌ها، سرعنوان‌های موضوعی، اصطلاحنامه‌ها و هستی‌شناسی‌ها) در بازیابی اطلاعات استفاده می‌کنند. به این دلیل که این بررسی مقایسه‌ای انجام نگرفته و هدف مطالعه، نمایش ویژگی‌های معنایی متفاوت موجود در نظام‌های بازیابی اطلاعات مختلف اعم از فهرست‌های پیوسته کتابخانه‌ها، پایگاه‌های اطلاعاتی و مراکز اسناد و آرشیوی و ... است از هر یک از موارد موجود در جهان و ایران مواردی انتخاب شد و مورد بررسی قرار گرفت.

جدول ۱. لیست نظام‌ها برای بررسی ویژگی‌های معنایی

ردیف	نام نظام بازیابی اطلاعات	ویژگی خاص برای انتخاب
۱	وبگاه کتابخانه کنگره آمریکا	استفاده از سرعنوان‌های موضوعی و اصطلاحنامه‌های مختلف در بازیابی اطلاعات
۲	وبگاه کتابخانه ملی استرالیا	امکان جستجو براساس اصطلاحنامه
۳	وبگاه آرشیو ملی استرالیا	استفاده از مدل‌های مفهومی در بازیابی اطلاعات
۴	وبگاه موزه بریتانیا	استفاده از اصطلاحنامه مواد در سازماندهی و بازیابی اطلاعات
۵	پایگاه اطلاعاتی اریک	استفاده از اصطلاحنامه در جستجو
۶	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)	استفاده از اصطلاحنامه در سازماندهی اطلاعات
۷	پایگاه پایان‌نامه‌های پژوهشگاه علوم و فناوری	استفاده از اصطلاحنامه در نمایه‌سازی

ردیف	نام نظام بازیابی اطلاعات	ویژگی خاص برای انتخاب
	اطلاعات ایران (گنج)	

در میان نظام‌های اشاره شده در جدول ۱ پایگاه گنج ایراندک از اصطلاحنامه تنها در نمایه‌سازی به عنوان مرحله‌ای از سازماندهی اطلاعات استفاده می‌کند و شواهد آن در بازیابی اطلاعات مشهود نیست، بنابراین از لیست بررسی حذف گردید.

با مطالعه متون و منابع موجود در این زمینه مشخص شد که محققان مختلف ویژگی‌های مختلفی را برای یک جستجوی معنایی مورد توجه قرار داده‌اند. این ویژگی‌ها شامل پنج مقوله کلی و مقوله‌های فرعی مربوط به آنها هستند. در ویژگی‌های مربوط به جستجوی مفهومی ۲۹ مقوله فرعی شناسایی شد، در گزینه‌های مربوط به پیشنهاد برای جستجو ۳ مقوله، در گزینه‌های مربوط به جستجوی مترادف‌ها ۲ مقوله و در گزینه‌های مربوط به نمایش ۹ مقوله ارائه و مورد بررسی قرار گرفت.

در گروه ویژگی‌های مربوط به هستی‌شناسی‌ها چهار گزینه برای بررسی براساس مطالعات تعریف شد. مقوله خاصی که در این پژوهش براساس مطالعه بیتس (۱۹۹۰) مورد توجه قرار گرفت مربوط به تاکتیک‌هایی است که می‌تواند به طور مؤثری با هستی‌شناسی‌ها با حرکت درون هستی‌شناسی، حمایت شود. این تاکتیک‌ها شامل ۱۰ تاکتیکی هستند که می‌توانند نشان دهنده استفاده از هستی‌شناسی و ویژگی‌های هستی‌شناسانه در نظام بازیابی اطلاعات باشند.

تعدادی از معیارها با بررسی نظام قابل مشاهده بود اما برای بررسی سایر معیارها با توجه به ماهیت آنها پرس‌وجوهایی براساس نمونه‌های موجود تعریف شد و در نظام بازیابی مورد نظر مشاهده گردید. نتایج این مشاهدات به صورت صفر و یک در جدولی مشخص گردید.

۵. نظام‌های بازیابی اطلاعات

کتابخانه کنگره آمریکا

فهرست آن‌لاین کتابخانه کنگره آمریکا یکی از شناخته شده‌ترین نظام‌های بازیابی اطلاعات جهان است که در بازیابی اطلاعات خود از اصطلاحنامه‌ها و سرعنوان‌های موضوعی استفاده می‌کند. فهرست‌نویسی موضوعی به لیستی از موضوعات کنترل شده در دسترس اصطلاحنامه‌ها و واژگان

وابسته است. رکوردهای فهرست کتابخانه کنگره عموماً نقاط دسترسی موضوعی را از یک یا چند اصطلاحنامه و سرعنوان‌های موضوعی زیر می‌گیرد.

- سرعنوان‌های موضوعی کتابخانه کنگره
- اصطلاحات سبک هنری/فرم کتابخانه کنگره برای مواد آرشیوی و کتابخانه‌ای
- اصطلاحنامه متوسط عملکرد موسیقی کتابخانه کنگره
- اصطلاحنامه‌ای برای مواد گرافیکی: اصطلاحات موضوعی و ژانری
- واژگان کنترل شده RBMS: واژگان کنترل شده برای استفاده در کتب نایاب و فهرست‌نویسی مجموعه‌های خاص
- راهنمایی در دسترسی موضوعی به آثار شخصی داستان، درام و ...،
- راهنمای ژانر/فرم تصاویر متحرک، راهنمای اصطلاحات سبک هنری/فرم رادیو، اصطلاحات موضوعی قانون‌گذاری (Thesauri & Controlled Vocabularies (2016)

در سیستم جستجوی کاتالوگ‌های کتابخانه کنگره از این ساختارهای اصطلاحنامه‌ای و همچنین سرعنوان‌های موضوعی برای بهبود نتایج جستجو بهره گرفته می‌شود. به عنوان مثال اگر در جستجوی پیشرفته عبارت Iranian Revolution و Revolution را به عنوان موضوع جستجو وارد کنیم، این درخواست به مجموعه سرعنوان موضوعی خواهد رفت و عبارت مناسب با آن با استفاده از ساختار درختی موجود استخراج می‌شود و به عبارات درخواستی کاربر افزوده می‌شود و جستجو با همه این عبارات انجام خواهد شد. در این مثال یکی از نتایج کتابی است از خاطرات جبهه مربوط به لشکر امام حسین اصفهان که در تقسیمات فرعی مجموعه سرعنوان‌ها قرار گرفته است. در قسمت محتویات موضوعی، محل قرار گرفتن موضوع این مدرک را در مجموعه سرعنوان‌های موضوعی مشخص می‌کند که می‌توان با کلیک کردن روی آن، مجموعه رکوردهای جدیدی را با اصطلاح جستجوی جدید در یک لیست مشاهده کرد. (استفاده از اصطلاحنامه در پایگاه‌های اطلاعاتی مرکز اطلاعات و مدارک علمی ایران ۱۳۸۳)

کتابخانه ملی استرالیا

بزرگ‌ترین کتابخانه و یکی از مهم‌ترین مراکز پژوهشی استرالیا، کتابخانه ملی آن است. سابقه فعالیت این کتابخانه به ۱۹۰۲ باز می‌گردد. وب‌گاه کتابخانه استرالیا شامل جستجوی ساده، پیشرفته و مرور در فهرست کتابخانه ملی استرالیا و سایر کتابخانه‌های استرالیا از جمله کتابخانه‌های عمومی است. همچنین می‌توان در منابع دیجیتالی کتابخانه و نیز منابع الکترونیکی آن جستجو انجام داده و به اطلاعات مورد نیاز خود دست یافت.

کتابخانه ملی استرالیا از اصطلاحنامه APAIS استفاده می‌کند. این اصطلاحنامه اصطلاحات موضوعی مورد استفاده برای نمایه‌سازی مقالات برای APAIS را لیست می‌کند: خدمات اطلاعاتی روابط عمومی استرالیا، راهنمای موضوعی برای متون در علوم اجتماعی و انسانی. این اصطلاحنامه در ۱۹۷۸ وقتی که APAIS به عنوان پایگاه اطلاعاتی آن‌لاین در دسترس شد آغاز به کار کرد و اولین بار توسط کتابخانه ملی استرالیا در ۱۹۸۰ برای تسهیل جستجو ایجاد شد.

این اصطلاحنامه هنوز به عنوان ابزار نمایه‌سازی استفاده می‌شود، همچنین ابزار کمکی برای جستجوی آن‌لاین APAIS و پایگاه اطلاعاتی تمام متن روابط عمومی استرالیا است. علاوه بر خدمت به مشتریان APAIS اصطلاحنامه منتشر شده در تعدادی پروژه که نیاز به فهرست ساختاربندی شده اصطلاحات موضوعی استرالیا دارند نیز مورد استفاده قرار گرفت. (National library of Australia 2016)

آرشیو ملی استرالیا

آرشیو ملی استرالیا از اصطلاحنامه عملکرد تعاملی دولت استرالیا (AGIFT) در بازیابی اطلاعات خود بهره می‌گیرد. این اصطلاحنامه چارچوبی استاندارد برای توصیف عملکردهای دولت (AGIFT) است. این اصطلاحنامه، اصطلاحنامه‌ای با سلسله مراتب سه سطحی است که عملکردهای تجاری انجام شده در نواحی مشترک المنافع، ایالات و دولت‌های محلی در استرالیا را توصیف می‌کند. اهداف این اصطلاحنامه شامل موارد زیر است:

- فراهم آوری اصطلاحات استاندارد برای ادارات دولتی برای استفاده در عنصر «عملکرد» مجموعه عناصر فراداده‌ای AGLS

- کمک به کاربران در جستجوی نقاط مدخل دولتی که مطمئن نیستند کدام اصطلاح مورد استفاده قرار گرفته یا کدام سطح دولتی برای اطلاعات یا خدمات مورد نیاز مسئول است.
- فراهم‌آوری چارچوبی برای ادارات دولتی برای توسعه اصطلاحنامه‌های عملکردی مبتنی بر اداره برای نیازهای طبقه‌بندی آنها

کتابخانه موزه بریتانیا

مجموعه پایگاه اطلاعاتی از یک سیستم از اصطلاحات کنترل شده برای نمایه‌سازی و طبقه‌بندی مجموعه موزه بریتانیا استفاده می‌کند. اینها شامل انواع اشیاء مثل اشیاء چاپی، مجسمه‌سازی یا کاسه تا مکان‌های مرتبط به این اشیاء می‌شوند. این اصطلاحات کنترل شده در اصطلاحنامه و فایل‌های مستند آمده‌اند که به موزه داران و سایر کاربران پایگاه اطلاعاتی اجازه دسته‌بندی یکدست اشیاء را می‌دهد.

یکی از اصطلاحنامه‌ها، اصطلاحنامه‌ای برای نام‌های اشیاء و مواد است. اینها هم با اعتماد به مجموعه‌ها منتشر شده (سابقاً MDA/ انجمن دکوماناسیون موزه) و با دیگر موزه‌ها در سراسر جهان پذیرفته شده است.

پایگاه اطلاعاتی اریک

«مرکز اطلاعاتی منابع تعلیم و تربیت» (اریک) دسترسی نامحدود به بیش از ۱,۲ میلیون سند کتابشناختی از مقالات نشریات و سایر مطالب مربوط به تعلیم و تربیت را فراهم می‌کند و هر دو هفته یکبار صدها سند جدید به آن اضافه می‌شود. در صورت دسترسی، لینک‌ها به متن کامل مقالات هم اضافه می‌شود.

اریک مرکز منابع اطلاعات علمی و آموزش و پرورش است و نیز یک کتابخانه دیجیتالی آنلاین با اطلاعات گسترده علمی و پژوهشی این پایگاه امکان دستیابی به اطلاعات را جهت پشتیبانی و استفاده از پژوهش‌ها و در آموزش، تدریس و پژوهش را فراهم می‌کند.

در این پایگاه علاوه بر امکان جستجوی ساده و پیشرفته، امکان جستجو در اصطلاحنامه این پایگاه نیز وجود دارد. کاربرانی که مهارت استفاده از جستجوی پیشرفته را ندارند یا می‌خواهند مناسب‌ترین عبارت را برای جستجوی مورد نظرشان بیابند یا می‌خواهند با یافتن عبارات مرتبط با موضوع، حجم اطلاعات بازیابی شده را افزایش دهند، می‌توانند از طریق جستجو در اصطلاحنامه موضوعات مورد علاقه خود را یافته و اطلاعاتی در مورد آن بازیابی کنند، بنابراین بسط درخواست کاربر و افزودن عبارات هم معنا و متناسب به صورت اتوماتیک، منتفی است و این وظیفه به خود کاربر واگذار شده است. (استفاده از اصطلاحنامه در پایگاه‌های اطلاعاتی مرکز اطلاعات و مدارک علمی ایران ۱۳۸۳)

پایگاه اطلاعاتی مجلات نورمگز

مرکز تحقیقات کامپیوتری علوم اسلامی، پیشگام در رایانه‌ای کردن منابع دینی و پژوهش‌های اسلامی است. داده‌پردازی علوم اسلامی و به کارگیری رایانه در پژوهش‌های دینی و اسلامی، از همان ابتدا در رأس فعالیت‌های این مرکز بوده است. در حال حاضر، این مرکز در همه موضوعات علوم اسلامی نرم‌افزارهایی را با رویکردهای پژوهشی، تبلیغی و اطلاع‌رسانی برای سطوح تخصصی و عمومی تولید کرده است. پایگاه‌های مرکز هر یک برای اهداف خاص یا در شرایط معینی طراحی و تولید شده‌اند. پایگاه اطلاعاتی مجلات نورمگز نیز جزو این پایگاه‌ها است. شبکه جهانی نور از سال ۱۳۷۷ اقدام به عرضه مجلات مربوط به علوم اسلامی و علوم انسانی بر روی پایگاه حوزه نمود. این پایگاه از مهرماه سال ۱۳۸۴ کار خود را آغاز کرده که ایجاد بایگانی مجلات و مقالات علوم انسانی و اسلامی در اینترنت، یکی از مهم‌ترین اهداف راه‌اندازی پایگاه مجلات تخصصی بوده است. درج بیش از سی دو هزار شماره از هفتصد عنوان مجله تخصصی در قالب موضوعات مختلفی چون «قرآن و حدیث»، «فلسفه، کلام و عرفان»، «فقه، اصول و حقوق»، «کتابداری و اطلاع‌رسانی»، «روان‌شناسی»، «اقتصاد»، «هنر»، «مدیریت»، «حقوق»، «علوم انسانی» و «تاریخ»، این پایگاه را به یکی از مهم‌ترین بانک‌های اطلاعات مجلات در زمینه علوم اسلامی و انسانی تبدیل کرده است. (اکبری ۱۳۸۹).

۶. تجزیه و تحلیل یافته‌های پژوهش

جستجوی مفهومی

جستجوی مفهومی با جستجوی فراداده‌ای یا کلیدواژه‌ای تفاوت بسیاری دارد. جستجوی مفهومی در تحلیل نسبی، مطابقت مفهومی بین پرس‌وجو و سند را نشان می‌دهد. در واقع این نوع جستجو، فنی است که کلماتی را که مشابه به مفهوم یک کلمه پرس‌وجو هستند فراهم می‌کند. این نوع جستجو اسنادی را بازمی‌گرداند که مرتبط به همان مفهوم کلمه جستجو بدون توجه به این است که کلمه پرس‌وجو در اسناد نتایج جستجو وجود دارد یا خیر.

در این بخش و با توجه به تعاریف و ویژگی‌هایی که متون برای این نوع جستجو ارائه کرده‌اند، ۲۹ گزینه مربوط به ویژگی‌های آن شناسایی شد و در نظام‌های بازیابی اطلاعات مورد بررسی قرار گرفت.

جدول ۲. معیارهای جستجوی مفهومی در نظام‌های بازیابی اطلاعات

ردیف	معیارهای جستجوی مفهومی	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)
۱	استفاده از واژگان کنترل شده	۱	۱	۱	۱	۱	۱
۲	توسعه اصطلاحنامه‌ای	۰	۰	۰	۱	۱	۱
۳	جستجوی هستی‌شناسی	۰	۰	۰	۰	۰	۰
۴	استفاده دستی از اصطلاحنامه	۱	۰	۰	۰	۱	۰
۵	جستجوی مترادف‌ها با اصطلاحنامه	۱	۰	۰	۰	۱	۰

ردیف	معیارهای جستجوی مفهومی	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)
۶	ارائه مجموعه‌ای از مفاهیم در حوزه خاص	۰	۱	۱	۱	۱	۰
۷	انتخاب دامنه	۰	۰	۱	۰	۱	۱
۸	انتخاب مفاهیم سطح بالا در هر دامنه	۰	۰	۱	۰	۱	۰
۹	تنوع ریخت شناختی نتایج در واژه جست و جوشده مانند حالت‌های مختلف زمان فعل، جمع و مفرد بودن	۱	۱	۰	۰	۱	۱
۱۰	امکان عمومیت بخشی	۰	۱	۰	۱	۱	۰
۱۱	توجه به تغییرات مورفولوژیکی	۱	۱	۰	۰	۰	۰
۱۲	در نظر گرفتن آیتم‌های شناخته شده	۱	۰	۰	۰	۰	۰
۱۳	ارائه ارتباطات میان اصطلاحات	۰	۰	۰	۰	۱	۰
۱۴	ارائه نتایج مرتبط با نیاز اطلاعاتی بیان شده	۱	۱	۱	۱	۱	۱
۱۵	رتبه‌بندی و دخیره براساس ارتباط	۱	۱	۱	۱	۱	۱

ردیف	معیارهای جستجوی مفهومی	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	بایگاه اطلاعاتی اریک	بایگاه مرکز تحقیقات کامپیوتری علوم اسلامی(نور)
۱۶	تعیین محل و نمایش سریع نتایج	۱	۱	۱	۱	۱	۱
۱۷	ثابت نبودن طول پرس و جو	۱	۰	۰	۰	۱	۰
۱۸	امکان مرور سلسله مراتبی	۱	۱	۱	۰	۱	۰
۱۹	ایجاد سریع نمایه‌های پرس و جوی آماده	۰	۰	۰	۰	۱	۱
۲۰	عدم نیاز به نحو پیچیده و خاص	۱	۱	۱	۱	۱	۱
۲۱	اجازه پرس و جوی ترکیبی با استفاده از مفاهیم، کلیدواژه‌ها و فرا داده‌ها	۱	۱	۱	۱	۱	۰
۲۲	عدم تأثیرپذیری از املای نادرست کلمات، اشتباهات تایپی یا اشتباهات اسکن OCR در متن پرس و جو	۰	۰	۰	۰	۱	۰
۲۳	بازیابی نتایج مرتبط بدون توجه به وجود یا عدم وجود مفهوم در پرس و جو	۱	۰	۰	۰	۱	۰
۲۴	استفاده از ریشه لغت	۰	۰	۱	۰	۱	۱

ردیف	معیارهای جستجوی مفهومی	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)
۲۵	استخراج معنای بالقوه پرس و جو	۱	۰	۰	۰	۱	۰
۲۶	همخوانی با تغییرات معنایی اصطلاحات	۱	۱	۰	۰	۰	۰
۲۷	مشخص کردن تعداد نتایج	۱	۰	۰	۰	۰	۰
۲۸	استخراج موجودیت، نام های خاص و دیگر اطلاعات خاص برای اهداف جست و جو	۰	۱	۰	۰	۰	۰
۲۹	امکان انتخاب معنای مورد نظر در مورد چند معنایی ها	۰	۰	۰	۰	۰	۰
	کل	۱۷	۱۳	۱۱	۹	۲۲	۱۰
	درصد	۵۸/۶۲	۴۴/۸۲	۳۷/۹۳	۳۱/۰۳	۷۵/۸۶	۳۴/۴۸
	میانگین	۱۳/۶۶					

بررسی ۲۹ ویژگی مربوط به جستجوی مفهومی، همانطور که در جدول ۲ مشخص است، نشان داد پایگاه اطلاعاتی اریک با داشتن ۲۲ امتیاز و ۷۵/۸۶ درصد از امتیاز معیارهای مشخص شده بیشتر از سایر نظامها از جستجوی مفهومی پشتیبانی می کند. پس از اریک، کتابخانه کنگره آمریکا با ۱۷ امتیاز و ۵۸/۶۲ درصد از امتیاز معیارهای جستجوی مفهومی در رتبه دوم قرار گرفت. موزه بریتانیا با ۹ امتیاز و ۳۱/۰۳ درصد از این معیارها در رتبه آخر قرار گرفت. امتیاز میانگین کل نظامهای بازیابی اطلاعات ۱۳/۶۶ است.

همانطور که در جدول مشخص است، ویژگی‌های استفاده از واژگان کنترل شده، اجازه پرس و جوی ترکیبی با استفاده از مفاهیم، کلیدواژه‌ها و فراداده‌ها، ارائه نتایج مرتبط با نیاز اطلاعاتی بیان شده، رتبه‌بندی و ذخیره براساس ارتباط و تعیین محل و نمایش سریع نتایج در تمامی نظام‌ها به صورت مشترک وجود داشته است.

ارائه مجموعه‌ای از مفاهیم در حوزه خاص و توسعه اصطلاحنامه‌ای نیز در میان ویژگی‌های جستجوی مفهومی در بیشتر نظام‌ها (۴ نظام از ۶ نظام موردنظر) به کار رفته است.

پیشنهاد برای جستجو

زمانی که جستجوگر، جستجوی اولیه خود را انجام می‌دهد از وسعت واژگانی که می‌تواند برای جستجوی خود استفاده کند آگاه نیست و اغلب جستجو را محدود به کلیدواژه‌های خاص و در زبان طبیعی ارائه می‌دهد. پیشنهاد برای جستجو به کاربر اجازه می‌دهد که مفاهیم مرتبط به کلیدواژه‌های موردجستجوی خود را ببیند و در صورت لزوم جستجوی خود را براساس آن انجام دهد.

در این معیار، ۳ مقوله مدنظر قرار گرفت و در نظام‌های بازیابی اطلاعات مورد نظر مورد بررسی قرار گرفت.

جدول ۳. معیارهای پیشنهاد برای جستجو در نظام‌های بازیابی اطلاعات

معیارهای پیشنهاد برای جستجو	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)
ارائه پیشنهاد براساس اصطلاحات نتایج مرتبط	۰	۱	۰	۰	۰	۰

۱	۰	۰	۰	۰	۰	ارائه پیشنهاد براساس اصطلاحنامه
۰	۰	۰	۰	۰	۰	پیشنهاد رفع ابهام اصطلاحات پرس وجو
۱	۰	۰	۰	۱	۰	کل
۳۳/۳۳	۰	۰	۰	۳۳/۳۳	۰	درصد
۰/۳۳						میانگین

همانطور که در جدول ۳ مشخص است، اغلب نظام‌های مورد بررسی این پژوهش معیارهای پیشنهاد برای جستجو را ندارند و در به کار بردن این ویژگی‌ها ضعیف هستند. کتابخانه ملی استرالیا نتایج مشابه به نتایج بازیابی شده را در فهرستی ارائه می‌دهد.

جستجوی مترادف‌ها

جستجوی مترادف‌ها اسنادی را که شامل اصطلاحات مشابه در معنا به کلمات پرس‌وجو هستند معمولاً با استفاده از اصطلاحنامه برای تعیین این که کدام اصطلاحات با کلمات پرس‌وجو مطابقت دارند، باز می‌گرداند.

در این معیار، ۲ مقوله مدنظر قرار گرفت و در نظام‌های بازیابی اطلاعات مورد نظر مورد بررسی قرار گرفت.

جدول ۴. معیارهای جستجوی مترادف‌ها در نظام‌های بازیابی اطلاعات

بایگاه مرکزی تحقیقات کامپیوتری علوم اسلامی (نور)	بایگاه اطلاعاتی اریک	موزه بریتانیا	آرشیو ملی استرالیا	کتابخانه ملی استرالیا	کتابخانه کنگره آمریکا	معیارهای جستجوی مترادفها
۱	۱	۰	۰	۰	۱	جستجوی مترادفها با اصطلاحنامه
۰	۰	۰	۰	۰	۰	جستجوی مترادفها با هستی‌شناسی
۱	۱	۰	۰	۰	۱	کل
۵۰	۵۰	۰	۰	۰	۵۰	درصد
۰/۵						میانگین

به این دلیل که اغلب نظام‌ها از هستی‌شناسی در باز یابی اطلاعات خود استفاده نمی‌کردند، معیار جستجوی مترادفها بر اساس اصطلاحنامه مدنظر قرار می‌گیرد که آن طور که در جدول ۴ آمده است، تنها سه مورد از نظام‌ها آن را مورد توجه قرار داده بودند. امتیاز میانگین ۰/۵ از ۶ امتیاز کل، امتیاز بسیار پایینی است. ویژگی استفاده از اصطلاحنامه‌ها در جستجوی مترادفها، ویژگی است که در نیمی از نظام‌ها به کار رفته است و حضور آن در پرداختن به ویژگی‌های معنایی اهمیت زیادی دارد.

استفاده از هستی‌شناسی

در نظام‌های باز یابی اطلاعات استفاده از هستی‌شناسی به عنوان راه‌حلی برای مشکلات جستجوی کلیدواژه‌ای و تبدیل جستجوی سنتی به جستجوی معنایی پیشنهاد شده است. بنابراین استفاده از

آنها می‌تواند در بازیابی معنایی مفید باشد. به طور کلی ۴ مقوله کلی برای این معیار شناسایی شد. معیار استفاده از هستی‌شناسی در تاکتیک‌های جستجو خود شامل ۱۰ مقوله فرعی بوده است.

بیتس (۱۹۹۰) اشاره می‌کند، گاهی کاربر ترجیح می‌دهد که فرایند جستجو را تاحدی، بر طبق استراتژی‌ها و تاکتیک‌های مشخصی که بتواند به طور مؤثر در رابط کاربر حمایت شود، کنترل کند. بیتس تاکتیک‌ها را به عنوان « یک یا تعداد انگشت‌شماری از حرکاتی که جستجو را پیش می‌برند» تعریف می‌کند و به عنوان مثال، آنها را شامل تغییر یک اصطلاح به اصطلاحات خاص تر یا اضافه کردن اصطلاحات مشابه می‌داند. در واقع هستی‌شناسی‌ها می‌توانند برای **حمایت مستقیم تاکتیک‌ها** استفاده شوند. (García and Sicilia 2003).

جدول خلاصه‌ای از تاکتیک‌هایی اشاره شده در بیتس (۱۹۹۰) را که می‌تواند به طور مؤثری از طریق هستی‌شناسی‌ها حمایت شود، توصیف می‌کند. در جدول حروف بزرگ برای مشخص کردن اصطلاحات (A,B,C) نشانه زیرمجموعه برای نشان دادن توابع استفاده شده است. (García and Sicilia 2003)

جدول ۵. تاکتیک‌های جستجوی حمایت‌شونده از سوی هستی‌شناسی (García and Sicilia 2003)

تاکتیک	توصیف بیتس	پیاده‌سازی مبتنی بر هستی‌شناسی	مثال
SUPER	حرکت سلسله‌مراتبی به اصطلاح اعم (بالتر)	تغییر B با A، B زیرمجموعه A	حرکت از گربه چشم آبی به گربه
SUB	حرکت سلسله‌مراتبی به اصطلاح اخص (فرعی)	تغییر A با B ، B زیرمجموعه A است	حرکت از گربه به گربه چشم آبی
RELATE	حرکت یک طرفه سلسله‌مراتبی به هم پایه	تغییر B با C جایی که مفهوم A وجود دارد به طوری که B زیرمجموعه A، C، A زیرمجموعه A	حرکت از گربه چشم آبی به گربه براق از طریق گربه

مثال	پیاده‌سازی مبتنی بر هستی‌شناسی	توصیف پیتس	تاکتیک
حرکت از محصول نهایی به محصول اولیه (به عنوان دسته‌بندی مشروبات)	تغییر B با C جایی که اشتراک B و C تهی است (حداقل)	جستجوی اصطلاحی که صورت منطقی مخالف توصیف موردنظر است	CONTRARY
مسیر حیوان خانگی به گربه به گربه چشم آبی ثبت شده، بنابراین سیستم ممکن است بعداً به دیگر مشاهده‌ها پیشنهاد دهد	ذخیره هستی‌شناسی مسیرهای مشاهده شده در حال مرور	پیگیری مسیرهای آزمایشی که دنبال شده و آزمایش‌های مطلوبی که دنبال یا تکمیل نشده	RECORD
وقتی که جستجو برای اطلاعاتی در مورد ابزارهای آزمایش نرم افزار انجام می‌شود، تجزیه فنون مختلف (واحد، یکپارچه و ...)	تجزیه جستجوی مداوم، ویرایش بعضی از شاخه‌هایی که می‌توانند بعد اضافه شوند	تجزیه پرس‌وجوی جستجوی پیچیده به مشکلات فرعی و کاربر بر روی یک مشکل در یک زمان	SELECT
وقتی جستجو برای اطلاعات در مورد بازرسی نرم‌افزار شامل نقد نرم افزار نیز می‌شود	افزودن اصطلاح C وقتی که C مشابه به هر اصطلاحی که بخشی از جهت‌یابی جاری را تشکیل می‌دهد	گسترش فرمول‌بندی جستجو (یا گسترده‌تر) با شامل کردن مترادف‌ها یا دیگر اصطلاحات برابر مفهومی	PARALLEL
-	مخالف PARALLEL	ایجاد فرمول‌بندی جستجوی دقیق با حداقل‌سازی (یا کاهش) تعداد اصطلاحات برابر،	PINPOINT

تاکتیک	توصیف بیس	پیاده سازی مبتنی بر هستی شناسی	مثال
	حفظ اصطلاحات توصیفی تر		
SPECIFY	جستجوی اصطلاحاتی که به اندازه اطلاعات مطلوب خاص هستند	اجازه اصلاح قدم به قدم با پیش بردن به سوی خاص تر شدن در سلسله مراتب تابعی، ترکیب چند اصطلاح با «و» شبیه معناها	بعد به عنوان اصول طراحی ONTOIR راهنمای اصلی توضیح داده خواهد شد.
EXHAUST	شامل بیشتر یا تمام عناصر پرس و جو در فرمول بندی جستجوی اولیه: برای افزودن یک یا چند عنصر پرس و جو به فرمول جستجوی آماده	اجازه به انتخاب چند اصطلاح و «یا»- شبیه معناها	(بعدا به عنوان ویژگی OntoIR توضیح داده خواهد شد)

تاکتیک های جستجوی مبتنی بر هستی شناسی طبق جدول ۵ مورد بررسی قرار گرفت. نتایج بررسی این مقوله ها در نظام ها در جدول ۶ آمده است.

جدول ۶. معیارهای استفاده از هستی شناسی در نظام های بازبای اطلاعات

معیارهای استفاده از هستی شناسی	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)
استفاده از رابط کاربر برای تایپ اصطلاحات در هستی شناسی	۰	۰	۰	۰	۰	۰

پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)	پایگاه اطلاعاتی اریک	موزه بریتانیا	آرشیو ملی استرالیا	کتابخانه ملی استرالیا	کتابخانه کنگروه آمریکا	معیارهای استفاده از هستی‌شناسی
۰	۰	۰	۰	۰	۰	استفاده از هستی‌شناسی برای توسعه پرس و جو
۰	۰	۰	۰	۰	۰	ارائه ساختار فنی هستی شناسی در رابط کاربر
۰	۱	۱	۰	۱	۰	حمایت از تاکتیک‌های جستجو
۰	۱	۱	۰	۱	۰	SUPER
۰	۱	۱	۰	۱	۰	SUB
۰	۱	۰	۰	۱	۰	RELATE
۰	۱	۰	۰	۰	۰	CONTRARY
۰	۰	۰	۰	۰	۰	RECORD
۰	۰	۰	۰	۰	۰	SELECT
۰	۰	۰	۰	۰	۰	PARALLEL

معیارهای استفاده از هستی‌شناسی	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)
PINPOINT	۰	۰	۰	۰	۰	۰
SPECIFY	۰	۱	۰	۰	۰	۰
EXHAUST	۰	۱	۰	۰	۱	۰
کل	۰	۶	۰	۳	۶	۰
درصد	۰	۴۲/۸۵	۰	۲۱/۴۲	۴۲/۸۵	۰
میانگین	۱/۸۷۵					

همانطور که در جدول ۶ آمده است نظام‌های بازیابی اطلاعات ویژگی‌های هستی‌شناسانه را خیلی کم در بازیابی خود به کار برده‌اند. کتابخانه ملی استرالیا و پایگاه اطلاعاتی اریک که با ۴۲/۸۵ درصد رتبه اول را دارند اما این میزان استفاده آنها کمتر از نیمی از ویژگی‌ها است. در میان تاکتیک‌های جستجوی مبتنی بر هستی‌شناسی، تاکتیک SUPER و SUB (در ۳ نظام از ۶ نظام) که به نوعی اصطلاحات اعم و اخص را نشان می‌دهد بیشتر از سایر تاکتیک‌ها در نظام‌ها به کار گرفته شده است و این نشان دهنده محدود بودن این نظام‌ها به روابط اصطلاح‌نامه‌ای است.

نمایش

استفاده از ویژگی‌های معنایی در بازیابی اطلاعات، نمودهایی در نمایش دارد. این معیارهای استخراج شده از متون شامل ۹ معیار است. این معیارها در هر یک از نظام‌های بازیابی اطلاعات بررسی شد.

جدول ۷. معیارهای نمایش در نظام‌های بازیابی اطلاعات

معیارهای نمایش	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)
ارائه بخش مرتبط به نتایج به شکل ساده	۱	۱	۱	۱	۱	۱
نمایش موضوعات مرتبط	۰	۱	۰	۰	۰	۱
نمایش برجسته کلیدواژه‌های مرتبط به با پرس و جو	۱	۰	۰	۰	۱	۱
تعمیم یک مفهوم کلی به اجزاء آن	۰	۱	۰	۰	۰	۰
ارائه مدلی تصویری از نتایج کاوش	۰	۰	۰	۰	۰	۰
نمایش رکورد به همراه خلاصه	۱	۱	۱	۱	۱	۱
ارائه نتایج یکسان برای پرس و جوهای حاوی نحو متفاوت	۰	۰	۰	۰	۰	۰

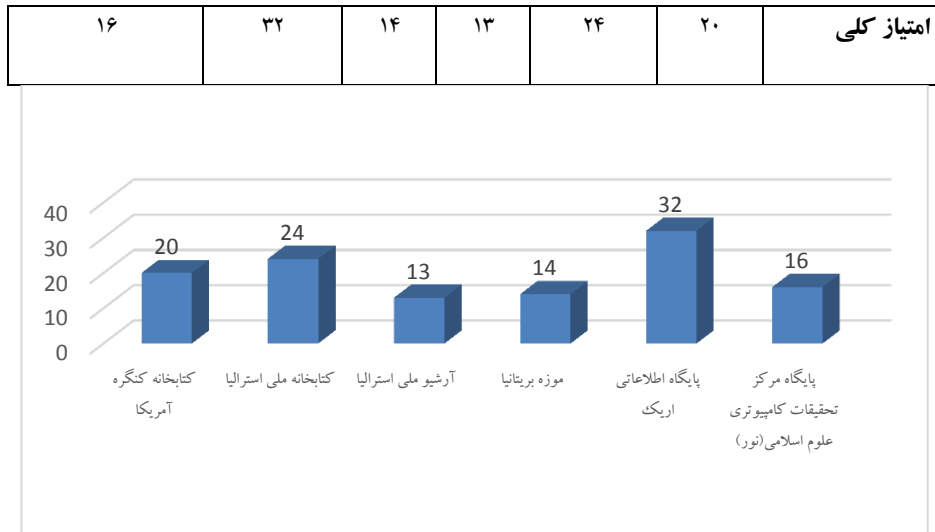
معیارهای نمایش	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)
و معنای مشابه						
نمایش اشکال مختلف دستوری پرس و جو	۰	۰	۰	۰	۰	۰
ارائه تعریف	۰	۰	۰	۰	۰	۰
کل	۲	۴	۲	۲	۳	۴
درصد	۲۲/۲۲	۴۴/۴۴	۲۲/۲۲	۲۲/۲۲	۳۳/۳۳	۴۴/۴۴
میانگین	۱/۳					

همانطور که در جدول ۷ مشخص است، نظام‌های بازیابی اطلاعات مورد بررسی این پژوهش، به ندرت در نمایش خود ویژگی‌های معنایی را نشان دادند. در میان آنها، پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی و کتابخانه ملی استرالیا تعداد بیشتری از این معیارها (۴ معیار از ۹ معیار، ۴۴/۴۴ درصد) را دارا بودند. در این میان آرشیو ملی استرالیا و موزه بریتانیا خیلی کم (۲ معیار از ۹ معیار، ۲۲/۲۲ درصد) از این معیارها استفاده کردند.

ارائه بخش مرتبط به نتایج به شکل ساده و نمایش رکورد به همراه خلاصه در تمامی نظام‌ها وجود داشت. نمایش برجسته کلیدواژه‌های مرتبط به با پرس و جو که نمونه ساده‌ای از نمایش ویژگی‌های معنایی است تنها در نیمی از نظام‌ها مورد توجه قرار گرفته است.

جدول ۸. ویژگی‌های معنایی کلی در نظام‌های بازیابی اطلاعات

معیارهای معنایی	کتابخانه کنگره آمریکا	کتابخانه ملی استرالیا	آرشیو ملی استرالیا	موزه بریتانیا	پایگاه اطلاعاتی اریک	پایگاه مرکز تحقیقات کامپیوتری علوم اسلامی (نور)



نمودار ۱. ویژگی‌های معنایی در نظام‌های بازیابی اطلاعات

جدول ۸ و نمودار ۱ وضعیت نظام‌های مورد بررسی را از لحاظ ویژگی‌های معنایی و هستی‌شناسانه نشان می‌دهد. همانطور که مشخص است پایگاه اطلاعاتی اریک و کتابخانه ملی استرالیا با ۳۲ و ۲۴ امتیاز از نظر ویژگی‌های معنایی در صدر نظام‌های بررسی شده قرار گرفته‌اند و آرشیو ملی استرالیا و موزه بریتانیا نیز از این لحاظ با ۱۴ و ۱۳ امتیاز کمترین امتیاز را داشتند.

۷. نتیجه‌گیری

در این پژوهش ویژگی‌های معنایی نظام‌های بازیابی مطرح دنیا که از اصطلاحنامه و هستی‌شناسی در بازیابی اطلاعات خود استفاده می‌کردند مورد بررسی قرار گرفت. هدف از این پژوهش ارزیابی این نظام‌ها نبود بلکه تلاش شد تا ویژگی‌هایی که بیشتر در نظام‌ها به کار رفته است و ویژگی‌هایی که متون برای این نظام‌ها در نظر گرفته‌اند شناسایی شود. ویژگی‌هایی که به طور کلی از متون به عنوان ویژگی‌های معنایی شناسایی شده بود شامل این موارد است: جستجوی مفهومی که خود شامل استفاده از واژگان کنترل شده (اصطلاحنامه، هستی‌شناسی‌ها، وردنت)، توسعه اصطلاحنامه‌ای، جستجوی هستی‌شناسی، استفاده دستی از اصطلاحنامه، جستجوی مترادف‌ها با اصطلاحنامه، ارائه مجموعه‌ای از مفاهیم در حوزه خاص، انتخاب دامنه، انتخاب مفاهیم سطح بالا در هر دامنه، تنوع ریخت‌شناختی نتایج در واژه جست و جوشده مانند حالت‌های مختلف زمان فعل، جمع و مفرد بودن، امکان عمومیت بخشی، توجه به تغییرات مورفولوژیکی، در نظر گرفتن

آیتم‌های شناخته شده، ارائه ارتباطات میان اصطلاحات، ارائه نتایج مرتبط با نیاز اطلاعاتی بیان شده، رتبه‌بندی و دخیره براساس ارتباط، تعیین محل و نمایش سریع نتایج، ثابت نبودن طول پرس و جو، امکان مرور سلسله مراتبی، ایجاد سریع نمایه‌های پرس و جوی آماده، عدم نیاز به نحو پیچیده و خاص، اجازه پرس و جوی ترکیبی با استفاده از مفاهیم، کلیدواژه‌ها و فراداده‌ها، عدم تأثیرپذیری از املائی نادرست کلمات، اشتباهات تایپی یا اشتباهات اسکن OCR در متن پرس و جو، بازیابی نتایج مرتبط بدون توجه به وجود یا عدم وجود مفهوم در پرس و جو، استفاده از ریشه لغت، استخراج معنای بالقوه پرس و جو، همخوانی با تغییرات معنایی اصطلاحات، مشخص کردن تعداد نتایج، استخراج موجودیت، نام‌های خاص و دیگر اطلاعات خاص برای اهداف جست و جو، امکان انتخاب معنای مورد نظر در مورد چند معنایی‌ها است، نتایج این بررسی نشان داد که بزرگترین نظام‌های بازیابی اطلاعات دنیا دارای ویژگی‌های معنایی شامل ویژگی‌های استفاده از واژگان کنترل شده، اجازه پرس و جوی ترکیبی با استفاده از مفاهیم، کلیدواژه‌ها و فراداده‌ها، ارائه نتایج مرتبط با نیاز اطلاعاتی بیان شده، رتبه‌بندی و دخیره براساس ارتباط و تعیین محل و نمایش سریع نتایج بود که در تمامی نظام‌ها به صورت مشترک وجود داشته است. ارائه مجموعه‌ای از مفاهیم در حوزه خاص و توسعه اصطلاحنامه‌ای نیز در میان ویژگی‌های جستجوی مفهومی در بیشتر نظام‌ها به کار رفته بود. در جستجوی مفهومی پایگاه اطلاعاتی اریک با ۲۲ امتیاز در بالاترین سطح امتیازات قرار گرفت و این پایگاه از نظر جستجوی مفهومی ویژگی‌های بیشتری را در خود نشان داد.

پیشنهاد برای جستجو که شامل ارائه پیشنهاد براساس اصطلاحات نتایج مرتبط، ارائه پیشنهاد براساس اصطلاحنامه، پیشنهاد رفع ابهام اصطلاحات پرس و جو بود و در بررسی نظام‌ها این ویژگی‌ها دیده نشد. تنها مورد موجود ارائه پیشنهاد بر اساس اصطلاحات نتایج مرتبط بود که در کتابخانه ملی استرالیا و در بخشی با عنوان Similar Items اسناد مشابه با نتایج را نشان می‌داد و در پایگاه اطلاعاتی نورمگز نیز قسمتی با عنوان پیشنهاد برای جستجو وجود دارد که براساس اصطلاحات موجود در اصطلاحنامه کلماتی برای جستجو پیشنهاد می‌کند. پیشنهاد رفع ابهام اصطلاحات پرس و جو به عنوان ویژگی مهمی که در نظام‌های معنایی به کار می‌رود در هیچ یک از این نظام‌ها مشاهده نشد. کتابخانه ملی استرالیا و پایگاه مجلات نورمگز تنها مواردی بودند که پیشنهاد برای جستجو را ارائه می‌کردند.

جستجوی مترادف‌ها که خود شامل جستجوی مترادف‌ها با اصطلاحنامه، جستجوی مترادف‌ها با هستی‌شناسی است. ویژگی استفاده از اصطلاحنامه‌ها در جستجوی مترادف‌ها، در نیمی از نظام‌ها به کار رفته است و این موضوع اهمیت حضور آن را در ویژگی‌های معنایی نشان می‌دهد. جستجوی مترادف‌ها نیز در نورمگز، کتابخانه کنگره آمریکا و اریک انجام می‌شد.

استفاده از هستی‌شناسی به عنوان یکی دیگر از ویژگی‌های معنایی مورد توجه شامل استفاده از رابط کاربر برای تایپ اصطلاحات در هستی‌شناسی، استفاده از هستی‌شناسی برای توسعه پرس و جو، ارائه ساختار فنی هستی‌شناسی در رابط کاربر، حمایت از تاکتیک‌های جستجو شامل تاکتیک‌های SUB، SUPER، RELATE، CONTRARY، RECORD، SELECT، PARALLEL، PINPOINT، SPECIFY، XHAUST است، این ویژگی‌ها در نظام‌های مورد بررسی مورد استفاده قرار نگرفته است و از میان تاکتیک‌هایی که ویژگی‌های معنایی را حمایت می‌کنند تنها تاکتیک‌های SUB و SUPER که به ترتیب حرکت سلسله‌مراتبی به اصطلاح اعم (بالا تر) و حرکت سلسله‌مراتبی به اصطلاح اخص (فرعی) است، مورد توجه این نظام‌ها قرار گرفتند. این تاکتیک‌ها اشاره به روابط اصطلاحنامه‌ای اعم و اخص دارد و نشان می‌دهد که این نظام‌ها تاکنون توانسته‌اند که روابط معنایی موجود در اصطلاحنامه‌ها را در بازیابی اطلاعات به کار گیرند. کتابخانه ملی استرالیا و پایگاه اطلاعاتی اریک از نظر توجه به ویژگی‌های هستی‌شناسی در سطح بالاتری نسبت به سایر نظام‌ها قرار داشتند.

ویژگی‌های مربوط به نمایش ویژگی‌های معنایی شامل ارائه بخش مرتبط به نتایج به شکل ساده، نمایش موضوعات مرتبط، نمایش برجسته کلیدواژه‌های مرتبط به با پرس و جو، تعمیم یک مفهوم کلی به اجزاء آن، ارائه مدلی تصویری از نتایج کاوش، نمایش رکورد به همراه خلاصه، ارائه نتایج یکسان برای پرس و جوهای حاوی نحو متفاوت و معنای مشابه، نمایش اشکال مختلف دستوری پرس و جو، ارائه تعریف بود که ارائه بخش مرتبط به نتایج به شکل ساده و نمایش رکورد به همراه خلاصه در تمامی نظام‌ها وجود داشت. نمایش برجسته کلیدواژه‌های مرتبط به با پرس و جو که نمونه ساده‌ای از نمایش ویژگی‌های معنایی است تنها در نیمی از نظام‌ها مورد توجه قرار گرفته است. نورمگز و کتابخانه ملی استرالیا از لحاظ ویژگی‌های مربوط به نمایش در سطح بالاتری قرار گرفتند.

علاوه بر این، نتایج نشان داد که در بین نظام‌های بازیابی اطلاعات مورد بررسی پایگاه اطلاعاتی اریک با ۳۲ امتیاز از مجموع ۵۶ امتیازی که امکان اکتساب آن بود بهتر از سایر نظام‌ها در زمینه معیارهای معنایی بوده است. پس از آن نیز کتابخانه ملی استرالیا با ۲۴ امتیاز رتبه دوم را دارد. این نشان می‌دهد می‌توان تا حد زیادی از ویژگی‌های این نظام‌ها در طراحی نظام بازیابی معنایی فارسی الگو گرفت. آرشیو ملی استرالیا، موزه بریتانیا و پایگاه اطلاعاتی نورمگز نیز به ترتیب با ۱۳، ۱۴ و ۱۶ امتیاز از ۵۶ امتیاز را کسب کرده‌اند و این نشان از ضعف این نظام‌ها در ویژگی‌های معنایی است.

فهرست منابع

استفاده از اصطلاحنامه در پایگاه‌های اطلاعاتی مرکز اطلاعات و مدارک علمی ایران. شرکت خدمات کامپیوتری پیک داده گستر، تهران: پژوهشگاه علوم و فناوری اطلاعات ایران، ۱۳۸۳.

اکبری، احمد. ۱۳۸۹. پایگاه‌های مرکز تحقیقات کامپیوتری علوم اسلامی در یک نگاه. فصلنامه ره آورد نور. ۵۰ (۳۳).
<http://www.rahavardnoor.com/index.php/authors/item/155-tahghighat>
(دسترسی در ۱۳۹۵/۷/۱۵).

دری، راحله. ۱۳۹۳. مقایسه و ارزیابی موتورهای جستجوی معنایی. پژوهشنامه پردازش و مدیریت اطلاعات ۳۰ (۲): ۴۶۷-۴۸۷.

علیمردای، مصطفی. ۱۳۹۴. نگرشی به گونه‌های جست و جو و بازیابی اطلاعات. فصلنامه ره آورد نور ۵۹ (۶۲): ۲۶.

Andago, Martin o, Teh P.L Phoebe, and Bassam A.M Thanoun. 2010. "Evaluation of a Semantic Search Engine against a Keyword Search Engine Using First 20 Precision." *INTERNATIONAL JOURNAL FOR THE ADVANCEMENT OF SCIENCE & ARTS* 1 (2): 55-63.

n.d. APAIS Thesaurus. Accessed 10 24, 2016. <https://www.nla.gov.au/apais/thesaurus/about.html>.

García , Elena, and Miguel-Ángel Sicilia Sicilia. 2003. "User Interface Tactics in Ontology-Based Information Seeking." *PsychNology Journal* 1 (3): 242 – 255.

Hynek, Jiř. 2002. Document Classification in a Digital Library. Pilsen: *University of West Bohemia*.

Jones , Karen Sparck, and Martin Kay. 1973. *Linguistics and information science*. New York and London: Academic press.

Linckels, Serge, and Christoph Meinel. 2011. *E-Librarian Service; User-Friendly Semantic Search in Digital Libraries*. BerlinHeidelberg: Springer-Verlag.

2016. *National library of Australia*. Accessed 10 24, 2016. <https://www.nla.gov.au/australian-public-affairs-information-service>.
- Read, Allen Walker . 1942. "the lexicographer and general semantics." *general semantics monographs* (Institute of general semantics, Lakeville, Connecticut) (3).
- Styltsvig, Henrik Bulskov. 2006. *ontology-based information retrieval*. Denmark: PhD thesis, computer science section Roskilde University.
- Sudeepthi, G, G Anuradha, and M Surendra Pra. 2012. "A Survey on Semantic Web Search Engine." *IJCSI International Journal of Computer Science Issues* 9 (2): 241-245.
- Sutcliffe, A, and M Ennis. 1998. "Towards a cognitive theory of information retrieval." *Interacting with Computers* 10 (3): 321-351.
2016. *Thesauri & Controlled Vocabularies*. Accessed 10 23, 2016. <http://www.loc.gov/library/libarch-thesauri.html>.
- Turner, Duygu, M. A Shah, and Yiltan Bitirim. 2009. "An empirical evaluation on semanticsearch performance of keyword-based and semantic search engines: google, yahoo, msn." *In proceedings of Fourth international conference on internet monitoring and. venice, mestre*.

The study of semantic and ontological features of thesaurus and ontology based information retrieval systems

Elmira Karimi, PhD Candidate in library and information science,
Iranian Research Institute for Information Science and Technology
(IranDoc)

Dr Mahmoud Babaei, Assistant Professor in Iranian Research Institute
for Information Science and Technology (IranDoc)

Dr Moluko Sadat Hosseini Beheshti. Assistant Professor in Iranian
Research Institute for Information Science and Technology (IranDoc)

Abstract

In order to increase in relevant retrieved results and users satisfaction achievement, Information retrieval systems, are seeking a way to change keyword-based search to concept-based search. These systems use the tools that can convert terms to concepts. Over the years, this provided the basis of tools production such as classification schema, vocabularies, subject headings and thesauri. Recently, ontologies use in Complementing of this process.

The main problem of this research is that The Great information retrieval systems of the world that claim to use of ontologies and thesauri how use the semantic features of these tools in their information retrieval process. In this paper 7 information retrieval systems due use of these tools, were selected. The semantic features presented in form of research checklist with literature review. These features include 5 general categories concept search, synonyms search, suggestion for search, use of ontologies and display. The researcher examined these features with structured observation in form of checklist. Though these systems use these tools for searching and retrieving, but don't show the expected features. In studied factors, ERIC database and national library of Australia achieved the maximum scores (32 and 24) and national archive of Australia and Britain museum had minimum scores (12 and 13).

Keywords: semantic search, information retrieval semantic features, information retrieval systems, concept search, synonyms search, ontology



المیرا کریمی، دارای مدرک کارشناسی ارشد رشته علم اطلاعات و دانش‌شناسی از دانشگاه تربیت مدرس تهران است. ایشان هم‌اکنون دانشجوی دکتری رشته علم اطلاعات و دانش‌شناسی پژوهشگاه علوم و فناوری اطلاعات ایران است. هستی‌شناسی و بازیابی اطلاعات است از جمله علایق پژوهشی وی است.



محمود بابایی، دارای مدرک تحصیلی دکتری در رشته علوم ارتباطات از دانشگاه علامه طباطبائی است. ایشان استادیار پژوهشگاه علوم و فناوری اطلاعات ایران بوده است. وی هم اکنون مدیرعامل شرکت دانش بنیان «سبوی دانش پارسیان» است. سایبرنتیک از جمله علایق پژوهشی وی است.





بررسی ویژگی‌های معنایی و هستی‌شناسانه نظام‌های بازیابی اطلاعات مبتنی بر اصطلاحنامه و هستی‌شناسی | کریمی و دیگران

ملوک‌السادات حسینی بهشتی، دارای مدرک تحصیلی دکتری زبان‌شناسی همگانی از دانشگاه تهران است.
ایشان هم‌اکنون استادیار و رئیس پژوهشکده مدیریت دانش پژوهشگاه علوم و فناوری اطلاعات ایران است.