

کاربرد هستی‌شناسی در فرایند بازیابی اطلاعات: مروری بر پژوهش‌های جاری و ارائه الگویی مفهومی

مهديه ميرزايي*

دانشجوی دکترای کتابداری و اطلاع‌رسانی،
دانشگاه فردوسی مشهد

دریافت: ۱۳۸۸/۰۸/۳۰ | پذیرش: ۱۳۸۸/۱۱/۰۷

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۸۲۲۳-۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS، LISA و ISC
<http://jst.irandoc.ac.ir>
ویژه‌نامه ذخیره، بازیابی و مدیریت اطلاعات
ص-ص ۲۲۷-۲۵۳ زمستان ۱۳۹۰
نوع مقاله: مروری

*mmirzabei@gmail.com

چکیده: ناکارآمدی اصطلاحنامه‌ها و دیگر ابزارهای بازنمون اطلاعات، از لحاظ عدم تطابق با محیط‌های الکترونیکی، اطلاع‌رسانان را به تفکر مجدد و بازنگری در ساختار این ابزار واداشته است. یکی از نتایج این بازنگری‌ها، ظهور دیگر ابزارهای سازماندهی اطلاعات همچون هستی‌شناسی‌هاست. این مقاله، ضمن بررسی عملکرد هستی‌شناسی‌ها در بازیابی اطلاعات، دو مورد از مهمترین الگوهای بازیابی اطلاعات مبتنی بر هستی‌شناسی - الگوی لینک‌پنگ و الگوی دان- را که بر عناصری مانند نیاز اطلاعاتی، استخراج متن، نمایه‌سازی، هستی‌شناسی، و رتبه‌بندی در فرایند بازیابی تأکید دارند، بررسی کرده است و درنهایت، یک الگوی مفهومی جدید ارائه نموده است.

کلیدواژه‌ها: هستی‌شناسی، بازیابی اطلاعات، الگوی مفهومی، الگوی بازیابی اطلاعات

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

هستی‌شناسی اصطلاحی است که خاستگاه اولیه آن را باید در فلسفه جستجو کرد (Wikipedia 2009). از این دیدگاه، کل هستی متشکل از موجودیت‌هایی است که به‌نحوی در ارتباط مفهومی با یکدیگرند؛ این ارتباط‌ها شبکه‌ای درهم‌تنیده از مفاهیم را تشکیل می‌دهند که کل آن تصویری از هستی را متبادر می‌سازد (صفری ۱۳۸۳، ۷۵).

در علم کتابداری و اطلاع‌رسانی، در مورد پیدایش هستی‌شناسی اختلاف نظرهایی وجود دارد و تعاریف گوناگونی نیز از آن ارائه شده است. به عقیده سورگل، هستی‌شناسی شکل گسترش یافته رده‌بندی است. به نظر وی، رده‌بندی به‌مدت طولانی در نظام‌های کتابداری و اطلاع‌رسانی برای ارائه ساختاری در مرور نتایج جستجو به کار گرفته شده است و اکنون سایر حوزه‌ها مانند هوش مصنوعی و زبان‌شناسی برای پاسخ به رده‌بندی مفاهیم در وب از آن استفاده می‌کنند. گروهی دیگر آن را نوع تکامل یافته اصطلاحنامه معرفی می‌کنند که قادر است مفاهیم را در قالب زبان طبیعی ارائه کند. عده‌ای دیگر آن را ابزاری جدید می‌دانند که عاریت گرفته از رشته علوم رایانه و پدیده وب معنایی است (Soergel 1999).

چنانچه با رویکرد فلسفی به هستی‌شناسی نگریسته شود، تمام ابزارهای سازماندهی اطلاعات را می‌توان نوعی هستی‌شناسی دانست و بر این اساس با هستی‌شناسی اصطلاحنامه، هستی‌شناسی رده‌بندی و جز آن مواجه خواهیم بود، اما نگاه این مقاله به هستی‌شناسی یک نگاه ابزاری است. ابزاری که می‌تواند ضعف‌های موجود در اصطلاحنامه‌های کنونی را برطرف کند و تفاوت‌های اساسی با دیگر ابزارهای بازیابی دارد (مانند ارائه روابط معنایی دقیق و بدون ابهام میان مفاهیم و قابلیت فهم و پردازش برای ماشین و انسان)^۱. به‌همین دلیل، در اینجا تعریف آنکسفلد^۲ برای هستی‌شناسی پذیرفته می‌شود که معتقد است: "هستی‌شناسی مجموعه‌ای از مفاهیم (موجودیت‌ها و نگرش‌ها و فرایندها)، تعاریف آنها، و روابط داخلی میان آنهاست" (شریف ۱۳۸۷، ۹۸).

این مقاله سعی دارد ابتدا، به‌طور مختصر عملکردهای بالقوه هستی‌شناسی در بازیابی اطلاعات را مرور کند، سپس برخی از الگوهای بازیابی اطلاعات که به نقش هستی‌شناسی در فرایند بازیابی توجه کرده‌اند، مورد بررسی و نقد قرار دهد و در نهایت، الگوی جدیدی ارائه نماید. به نظر می‌رسد ارائه این الگو بتواند به فهم و درک بهتر نقش هستی‌شناسی در نظام‌های بازیابی اطلاعات کمک کند.

۲. عملکرد هستی‌شناسی در بازیابی اطلاعات

کاربران نظام‌های بازیابی اطلاعات همواره از پیدا کردن منابع مرتبط مورد نیاز از میان

۱. در مورد تفاوت بین اصطلاحنامه و هستی‌شناسی به‌طور مفصل در مقاله صنعت‌جو (۱۳۸۴) بحث شده است.

2. Anksfeld

هزاران منبع موجود در وب گلابه دارند (Andreou 2005; Ding, Ding, and Foo 2001; Vallet, Fernández, and Castells 2005). یکی از راه‌حل‌های حل این مسأله و کسب رضایت کاربران، گسترش نظامی است که بتواند جستجویی را انجام دهد که در آن بیشترین مطالب مرتبط و کمترین مطالب نامرتب با بازیابی شود. راه‌حل سنتی که جستجویی مبتنی بر انطباق کلیدواژه‌های کاربر با کلیدواژه‌های مدارک انجام می‌داد، به منابع زیادی که مفاهیم مدنظر کاربران را دربرداشتند، توجه نمی‌کرد. حل این مشکل را در استفاده از ابزارهایی دانستند که بتوانند واژگان را به مفاهیم تبدیل کنند. این امر طی سالیان متمادی زمینه تولید ابزارهایی همچون طرح‌های رده‌بندی و سرعنوان‌های موضوعی و اصطلاحنامه‌ها را فراهم آورد. هریک از این ابزارها توانستند تاحدودی به سازماندهی منابع و بازیابی بهتر کمک کنند، اما همان‌گونه که می‌لستند نیز معتقد است حتی در مورد اصطلاحنامه‌های الکترونیکی که می‌توان آنها را نسل پیشرفته‌تر این ابزارها دانست، مشکلات بنیادی وجود دارد که سودمند بودن آنها را برای نظام‌های بازیابی به چالش می‌کشد (Milstead 1998).

گستره موضوعی بیشتر اصطلاحنامه‌ها حتی اصطلاحنامه‌های تخصصی بسیار عام است و روابط موجود در طراحی آنها فقط به سه نوع رابطه محدود می‌شوند. علاوه بر این، پژوهش‌ها نشان می‌دهند این نوع روابط خاص برای کاربران، مبهم و یا حتی فاقد اهمیت هستند و آنها بیشتر نشانه‌های روابط سه‌گانه اصطلاحنامه را به راحتی درک نمی‌کنند (Hernandez 2006). نکته مهم دیگر اینکه ارائه مفاهیم به صورت مبهم، اصطلاحنامه‌ها را برای پردازش خودکار در محیط وب نامناسب می‌کند. این شواهد مبنی بر ناکارآمدی اصطلاحنامه‌ها، اطلاع‌رسانان را به تفکر مجدد و بازنگری در ساختار این ابزار به منظور تطابق با محیط‌های الکترونیکی وادار نمود که از نتایج آن ظهور دیگر ابزارهای سازماندهی اطلاعات، همچون هستی‌شناسی هاست.

برای درک عملکرد هستی‌شناسی در فرایند بازیابی اطلاعات می‌توان به آنچه که در این فرایند رخ می‌دهد و نیز عوامل و عناصر دخیل در هر مرحله از فرایند توجه کرد. مرور متون منتشر شده در زمینه هستی‌شناسی‌ها نشان می‌دهد که طی سالیان اخیر، این حوزه مورد توجه متخصصان علوم مختلف از جمله علوم کامپیوتر، زبان‌شناسی، و علوم کتابداری و اطلاع‌رسانی قرار گرفته است. هریک از این متخصصان سعی کرده‌اند از جنبه‌های مختلفی به بررسی عملکرد هستی‌شناسی‌ها در فرایند بازیابی اطلاعات بپردازند. گرچه بسیاری از عملکردهای مطرح شده در متون به صورت بالقوه و آزمایشی بوده است، به نظر می‌رسد که با به فعل در آمدن آنها بسیاری از چالش‌های نظام‌های بازیابی کنونی برطرف شود. در ادامه، به اجمال به برخی از مهمترین عملکردهای هستی‌شناسی در بازیابی اطلاعات پرداخته می‌شود.

۱-۲. فهم دقیق نیاز اطلاعاتی کاربر

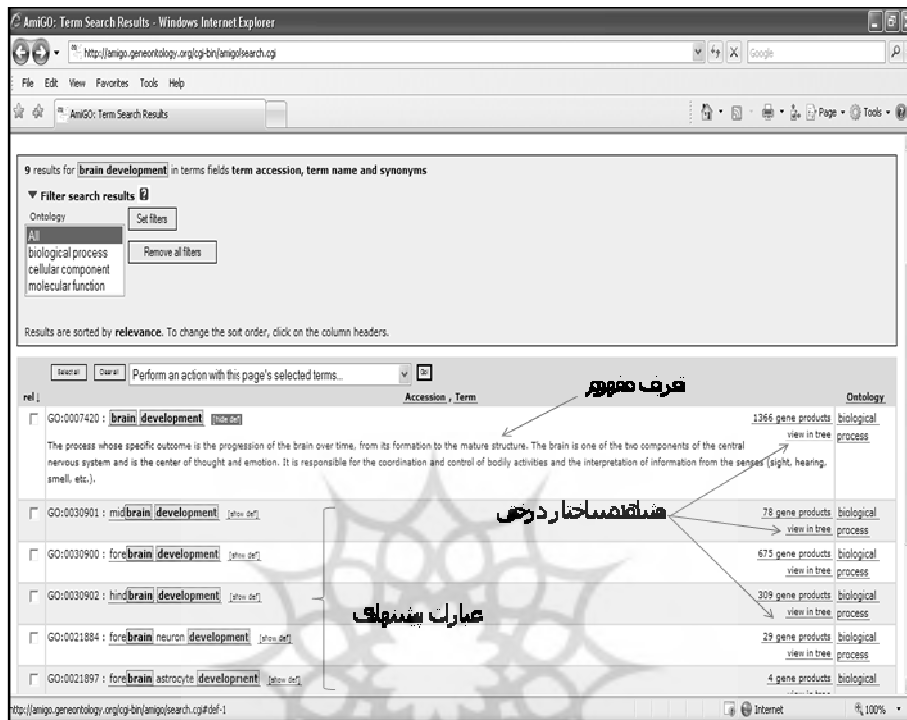
بیشتر اوقات، عبارت جستجوی کاربران را به‌عنوان نقطه شروع کشف نیاز اطلاعاتی وی در نظر می‌گیرند، اما بسیاری اوقات کاربران بدون توجه به روابط خاص میان اصطلاحات، آنها را به‌صورت جداگانه در یک عبارت جستجو قرار می‌دهند. با استفاده از هستی‌شناسی، نظام می‌تواند اطلاعات دقیقی را که مورد نیاز کاربر است درک و با رفع ابهام از پرسش کاربر، آن را به‌صورت دقیق بیان کند.

هنگامی که کاربر پرسش خود را به‌رابطه نظام بازیابی ارائه داد، تحلیل پرسش توسط بخش منطق هستی‌شناسی و با کمک فنون پردازش زبان طبیعی انجام می‌شود (Chen 2009)؛ برای مثال، کاربر ممکن است عبارت "غذا، سرطان" را وارد کند. دو مفهوم "غذا" و "سرطان" در این عبارت جستجو وجود دارد، اما رابطه بین این دو اصطلاح گم شده است. به‌عبارت دیگر، معلوم نیست که کاربر به‌دنبال چه رابطه‌ای بین این دو اصطلاح است. با استفاده از هستی‌شناسی، عبارت جستجو می‌تواند به‌صورت روابطی مثل "غذا می‌تواند باعث سرطان شود" و یا "غذا می‌تواند بر سرطان تأثیر بگذارد" به کاربر پیشنهاد شود. از طرفی، رابطه "تأثیر" یک معنای وابسته به‌نام "درمان" دارد: "درمان یکی از انواع تأثیر است". بنابراین، منطق هستی‌شناسی استنباط می‌کند که "غذا می‌تواند سرطان را درمان کند". این عبارتهای معنایی از طریق رابط به کاربر برگردانده می‌شوند و این عمل با گرفتن بازخوردهای مکرر از کاربر تا فهم دقیق نیاز وی ادامه پیدا می‌کند.

در برخی هستی‌شناسی‌ها نیز زمانی که کاربر کلیدواژه(های) خود را وارد می‌کند، هستی‌شناسی مجموعه‌ای از عبارات مرتبط با آن را به وی ارائه می‌دهد تا مناسب‌ترین عبارت را برای ادامه جستجو انتخاب کند. برای مثال، در "هستی‌شناسی ژن"^۱ پس از اینکه کاربر عبارت "brain development" را وارد کرد، هستی‌شناسی مجموعه‌ای از عبارات را که شامل عبارت گفته شده است، نمایش می‌دهد و کاربر می‌تواند با مشاهده تعاریف و ساختار درختی درنهایت، عبارت جستجوی مورد نظر خود را انتخاب کند.^۲

1. The Gene Ontology

۲. در حال حاضر، برخی از موتورهای جستجو مانند گوگل نیز در واکنش به ورود نخستین واژه از عبارت جستجوی کاربران، مجموعه‌ای از عبارتهای پیشنهادی را به آنها ارائه می‌دهند تا آنها مناسب‌ترین عبارت(ها) را برای ادامه جستجو انتخاب کنند، اما نوع ارتباط میان واژگان مشخص نشده است.



شکل ۱. عبارات پیشنهادی هستی‌شناسی ژن برای عبارت Brain Development

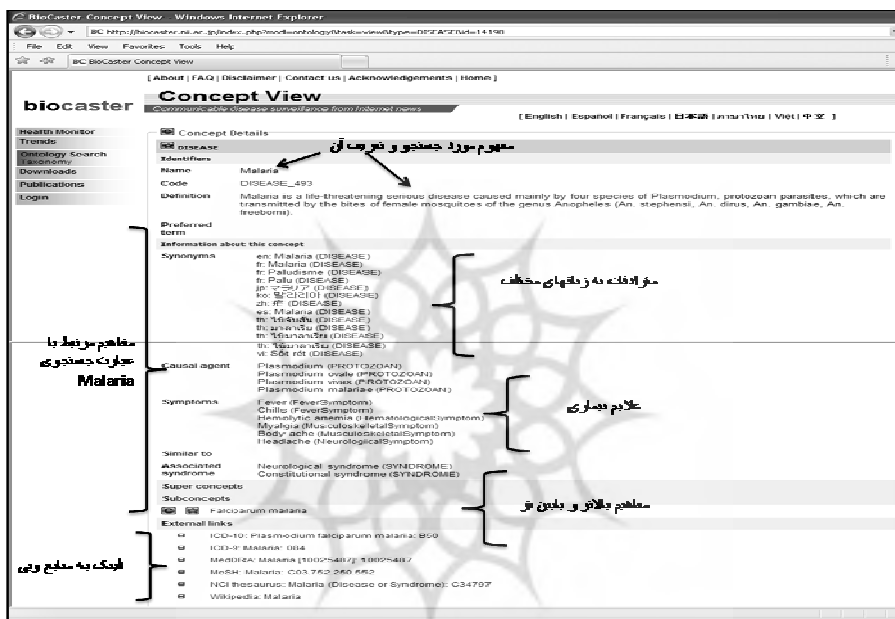
مسئله مهم دیگر این است که با استفاده از هستی‌شناسی‌ها کاربران می‌توانند به فهم و شناخت بیشتری از حوزه موضوعی برسند. برای مثال در شکل ۲، کاربر می‌تواند از طریق هستی‌شناسی علائم بیماری مالاریا را نیز مشاهده و با کلیک بر روی هر یک از علائم، شمای مفهومی آن را مرور کند. این اطلاعات به‌خصوص در مورد کاربرانی که دانش موضوعی کافی در مورد نیاز اطلاعاتی خود ندارند، بسیار مهم است.

۲-۲. بسط جستجو

بسط جستجو که از فرایندهای مهم و گاهی مفقوده نظام‌های بازیابی کنونی است، از طریق هستی‌شناسی‌ها می‌تواند به‌خوبی صورت پذیرد. در این رویکرد، اصطلاحاتی که به‌عنوان پرسش وارد نظام می‌شوند با مفاهیم مرتبط جابجا و به کاربر پیشنهاد می‌شوند (Andreou 2005)؛ برای مثال، عبارت جستجوی Malaria می‌تواند از طریق جایگزینی عبارات موجود در شکل ۲

1. Query Expansion

بسط پیدا کند. همان گونه که در این شکل مشاهده می شود، کاربر می تواند مترادف های عبارت Malaria را که به زبان های مختلف ارائه شده است، انتخاب کند و براساس آن به منابع موجود در وب که در ارتباط با آن مفهوم هستند، دست یابد (شکل ۲). علاوه بر این در صورت نیاز، کاربر می تواند از مفاهیم بالاتر^۱ و پایین تر^۲ عبارت جستجو نیز برای عام یا خاص کردن جستجوی خود استفاده کند.



شکل ۲. مفاهیم مرتبط با عبارت جستجوی Malaria برای بسط فرایند جستجو در هستی شناسی Biocaster

۲-۳. نمایه سازی مدارک براساس مفهوم (نمایه سازی مفهومی)

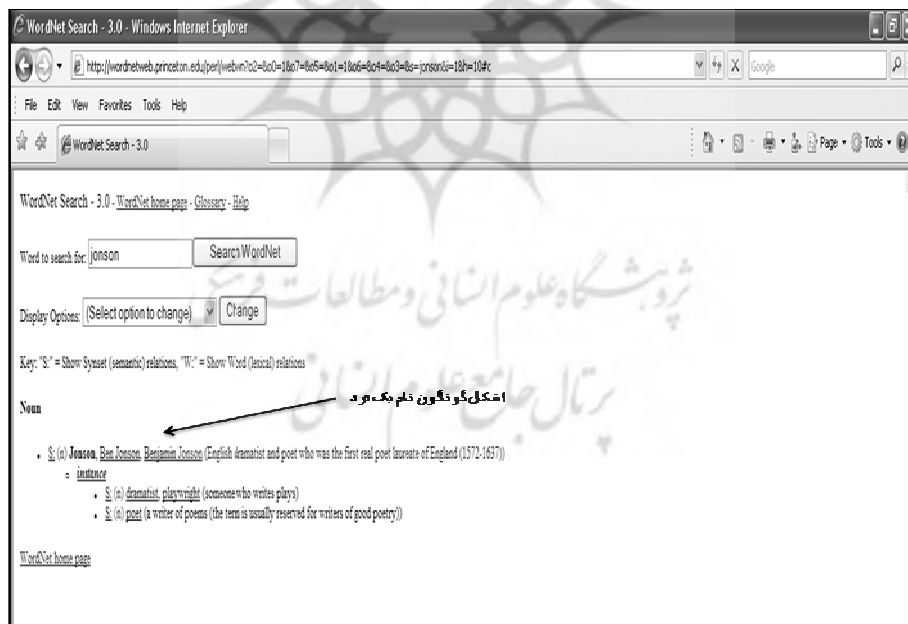
با به کارگیری هستی شناسی در مرحله نمایه سازی می توان مفاهیم مدارک را به طور خودکار و نیمه خودکار بیرون کشید و در نتیجه، فهم دقیقی از یک مدرک به دست آورد. این امر به واسطه ویژگی های زیر امکان پذیر است:

- در هستی شناسی، هر مفهوم می تواند صفات خاصه را از چندین مفهوم بالاتر از خود در نظام سلسله مراتبی دریافت کند؛ برای مثال، "پرستو" می تواند صفت خاصه "بال" را که از صفات خاصه "پرندگان" است، از چندین مفهوم بالاتر از خود در سلسله مراتب دریافت کند.

1. Super Concept
2. Subconcept

- همه مؤلفه‌های اطلاعاتی تعیین شده برای مفاهیم - اعم از تعاریف، سلسله مراتب، صفات خاصه، و ارث‌بری، به‌طور واضح و صریح - به‌گونه‌ای رمزبندی شده‌اند که هم برای انسان و هم برای ماشین قابل فهم باشد. برای مثال، اگر مفهوم "ب" یکی از شرط‌های بودن مفهوم "الف" باشد، هستی‌شناسی می‌تواند استدلال بیاورد که مفهوم "ب" "شاهدی" از مفهوم "الف" است و سپس، به‌طور خودکار مفهوم "ب" را زیر مفهوم "الف" قرار دهد (Dan et al. 2006).

از سوی دیگر، امروزه یکی از مشکلات نمایه‌سازی، بازنمایی نامتجانس اطلاعات شامل اشکال مختلف نام‌ها، مترادف‌ها، شبه مترادف‌ها و غیره است. با کمک هستی‌شناسی می‌توان نام دقیق یک شخص یا سازمان، یا نام دقیق یک اختصار را شناسایی کرد. به‌عبارت دیگر، هستی‌شناسی می‌تواند مجموعه ساختار بازنمایی اشتراکی از واژگان و مفاهیم را فراهم آورد که در اشکال مختلف نام‌های ناهمگن به‌عنوان یک کل درک شوند. برای مثال، با استفاده از هستی‌شناسی وردنت^۲ می‌توان نام‌های گوناگون شاعر انگلیسی بن جانسون^۳ را پیدا کرد و به‌نام کامل آن دست یافت (شکل ۳).

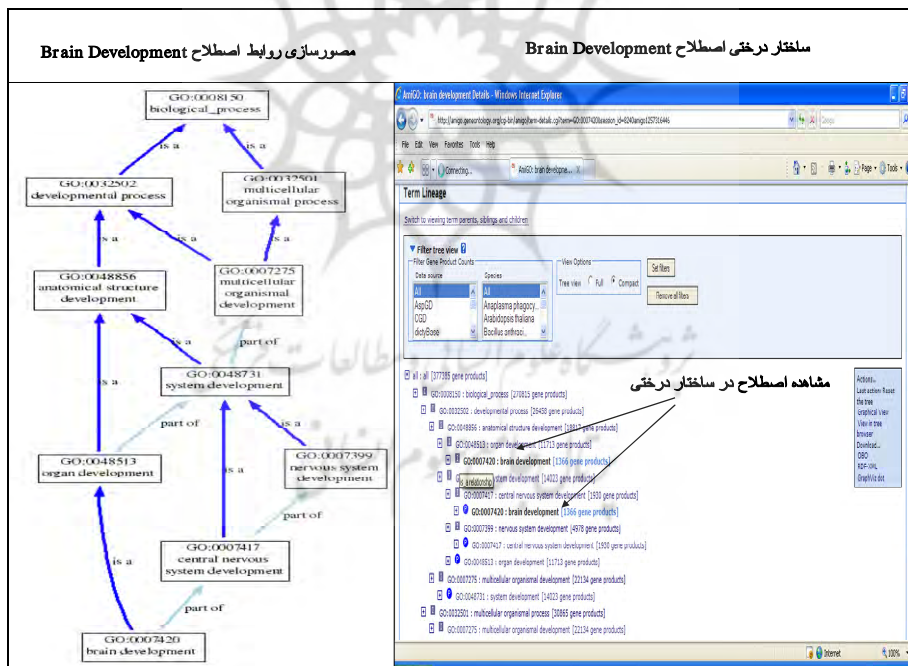


شکل ۳. بازنمایی نام کامل بن جانسون شاعر انگلیسی از طریق هستی‌شناسی وردنت

1. Instance 2. WordNet 3. Ben Jonson

۴-۲. مصورسازی معنایی^۱

در هستی‌شناسی، قابلیت بازنمایی منطقی و منسجم از ساختار مفهومی روابط دو یا چند مفهوم از طریق ابزارهای مصورسازی فراهم می‌شود. در این شیوه، با توجه به قابلیت بالای تبادل و انتقال معنی، تلاش می‌شود تا با نمایش تصویری مفاهیم و روابط میان آنها، قدرت فهم و یادگیری کاربر افزایش یابد و میزان قابل توجه اطلاعات به صورت فشرده، با حجم کمتر و به شیوه مصور ارائه شود (Albertoni, Riccardo, Alessio 2005). در واقع، ارائه بصری روابط میان مفاهیم، کاربران را قادر می‌سازد که روابط بین مفاهیم - به خصوص مقادیر بزرگ اطلاعاتی - را مشاهده، مرور، دریافت، و درک نمایند. برای مثال در هستی‌شناسی ژن، روابط مفهومی عبارت "Brain Development" را می‌توان در قالب‌های مختلف و از جمله، در ساختار درختی و نیز روابط مصورسازی شده دید. همان‌گونه که مشاهده می‌شود، با وجود اینکه هدف هر دو، نمایش افزایش سرعت انتقال و فهم روابط است، به نظر می‌رسد در مصورسازی، قابلیت فهم و درک روابط بیشتر از ساختار درختی است (شکل ۴).



شکل ۴. روابط مفهومی عبارت Brain Development در قالب ساختار درختی و روابط مصورسازی شده در هستی‌شناسی ژن

1. Semantic Visualization

لازم به اشاره است که هستی‌شناسی‌های مختلف از ابزارهای متعددی برای مصورسازی و تعامل استفاده می‌کنند که برخی از آنها عبارتند از آنتوویز^۱، تی‌جی‌ویژ^۲، جامبالیا^۳، و داپ برازر^۴.

۲-۵. شخصی کردن نظام بازیابی اطلاعات

با افزایش تعداد کاربران اینترنت از یک سو و تعداد صفحات وب از سوی دیگر، پیدا کردن منابعی که برطرف‌کننده نیازهای خاص کاربران باشد، به دشواری انجام می‌پذیرد. مشکل این است که خواه کاربران مرور کنند یا جستجوی کلیدواژه‌ای انجام دهند، خواه آنها دانش آموز دبیرستانی باشند یا برنده جایزه نوبل، اطلاعات نمایش داده شده یکسان است. به بیان ساده‌تر، مشکل این است که الگوریتم‌های جستجو و مرور - به عنوان دو روش بنیادی پیدا کردن اطلاعات - برای همه کاربران یکسان است. مسأله دیگر اینکه مفاهیم مشابه می‌توانند با اصطلاحات متفاوتی ظاهر شوند یا اصطلاحات مشابه، مفاهیم متفاوتی را به ذهن متبادر سازند. برای مثال، دو کاربر عبارت "Wildcats" را با دو هدف متفاوت (تیم‌های ورزشی و حیوانات وحشی) جستجو می‌کنند، اما هر دو آنها به نتایج مشابهی می‌رسند. به نظر می‌رسد کلید حل این چالش‌ها می‌تواند استفاده از "نظام بازیابی شخصی مبتنی بر هستی‌شناسی"^۵ باشد (Gauch et al. 2003).

بیشتر نظام‌های شخصی مبتنی بر پروفایل کاربران هستند. برای ساختن پروفایل کاربران لازم است تا برخی اطلاعات مربوط به آنها گردآوری شود. نظام‌های تجاری همچون "مای یاهو"^۶ به طور صریح از کاربران می‌خواهند تا اطلاعاتی را در مورد خودشان ارائه دهند، سپس با کمک این اطلاعات پروفایل هر کاربر به سادگی تشکیل می‌شود. مشکل تهیه پروفایل‌های عینی^۷ این است که علاوه بر اینکه مسئولیتی را بر دوش کاربران می‌گذارد، ممکن است کاربران به دقت علایقشان را گزارش ندهند. از طرفی، با گذشت زمان این علایق می‌تواند تغییر پیدا کنند. بدین منظور، بیشتر اوقات ایجاد پروفایل ضمنی^۸ مبتنی بر مشاهده و ثبت رفتارهای کاربران در حین مرور وب پیشنهاد می‌شود (Pretschner and Gauch 1999).

در نظام‌های مبتنی بر هستی‌شناسی نیز با کمک ماشین یادگیری^۹ هستی‌شناسی می‌تواند پروفایل کاربران را در هنگام مرور و جستجوی آنها در وب استخراج کند. در واقع، پروفایل هر کاربر از طریق تحلیل رفتارهای کاربران (مانند مرور، جستجو، زمان صرف شده برای هر صفحه،

1. OntoVis

4. Dope browser

6. MyYahoo

9. Learning machine

2. TGViz

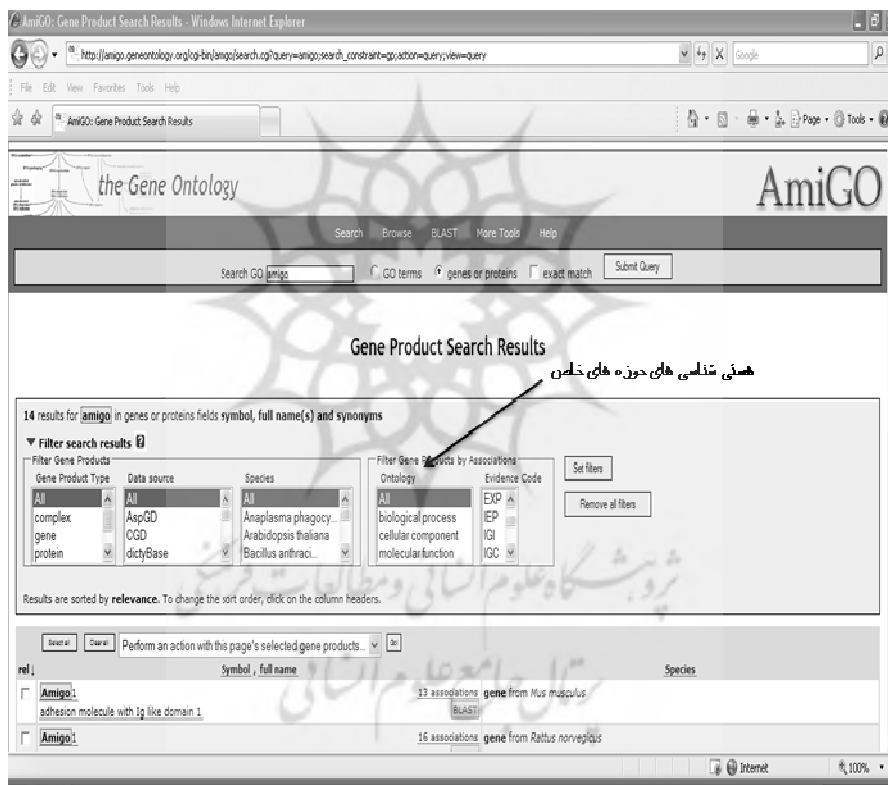
5. Ontology-based personalized retrieval system

7. Explicit profile

3. Jambalaya

8. Implicit profile

تحلیل محتوای هر صفحه جهت استخراج کلمات کلیدی آن، و وزن‌دهی و ثبت کلیدواژه‌ها در هستی‌شناسی کاربر) انجام می‌گیرد (Gauch et al. 2003). همچنین، با کمک هستی‌شناسی می‌توان موضوعات مختلف را جدا کرد و واژگان خاص هر حوزه را به‌طور دقیق شرح داد. به‌عبارت دیگر، می‌توان هستی‌شناسی‌های مختلفی داشت که هر یک، حوزه یا زیرحوزه موضوعی خاصی را پوشش می‌دهند. این امر به کاربران امکان می‌دهد تا هستی‌شناسی موضوعی خاص خود را انتخاب کنند و آن را مبنای جستجوهای آتی خود قرار دهند (شکل ۵).



شکل ۵. امکان انتخاب هستی‌شناسی حوزه خاص در مجموعه هستی‌شناسی ژن

۲-۶. خوشه‌بندی مدارک بازیابی شده^۱

پیشتر اشاره شد که ممکن است بین اصطلاحات عبارت جستجو، روابط چندگانه‌ای وجود داشته باشد. مدارک بازیابی شده می‌توانند برپایه این روابط چندگانه سازماندهی شوند. برای مثال، در عبارت جستجوی "غذا، سرطان" در مرحله بازیابی، هستی‌شناسی می‌تواند مدارک را در قالب چند گروه مجزا بازیابی کند: مدارکی در ارتباط با عبارت "غذا می‌تواند باعث سرطان شود"، مدارکی مرتبط با عبارت "غذا می‌تواند بر سرطان تأثیر بگذارد"، و مدارکی در ارتباط با عبارت "غذا می‌تواند سرطان را درمان کند" (Chen 2009).

۳. الگوهای بازیابی مبتنی بر هستی‌شناسی

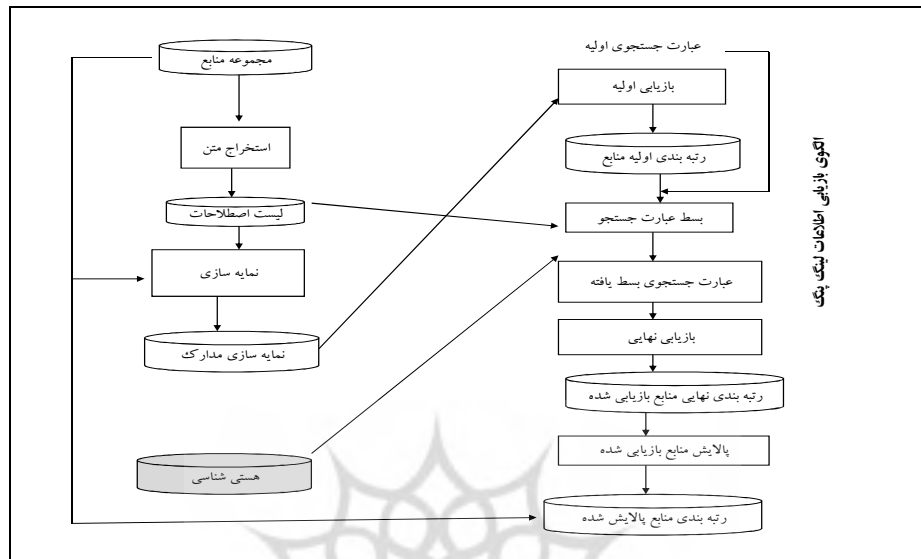
پس از مطرح شدن هستی‌شناسی‌ها در نظام بازیابی اطلاعات، الگوهایی از سوی متخصصان ارائه شده است که نقش و جایگاه هستی‌شناسی را در مراحل مختلف ذخیره و بازیابی اطلاعات نشان می‌دهند. بررسی این الگوها که به روشی ساده روابط بین قضایای نظری و فرایندهای مربوط به ذخیره و بازیابی اطلاعات را نشان می‌دهند، می‌تواند دانش ما را نسبت به نقش هستی‌شناسی افزایش دهد. در ادامه، به دو مورد از مدل‌های مفهومی بازیابی اطلاعات مبتنی بر هستی‌شناسی پرداخته می‌شود:

۳-۱. الگوی بازیابی اطلاعات لینگ‌پنگ و همکاران^۲

در این مدل، ابتدا به‌طور خودکار، اصطلاحات از مجموعه منابع استخراج می‌شوند و از آنها برای ساختن نمایه‌ها استفاده می‌شوند. در مرحله دوم، از این اصطلاحات در عبارت جستجو به کار گرفته می‌شوند و براساس آن، بازیابی اولیه‌ای انجام می‌گیرد. در مرحله سوم، برای هر عبارت جستجو، بسط جستجو از طریق هستی‌شناسی انجام می‌گیرد و سپس، بازیابی نهایی انجام می‌شود. در نهایت براساس اصطلاحات اولیه، منابع رتبه‌بندی و پالایش می‌شوند (شکل ۶).

1. Retrieved document organization

2. Lingpeng et al.



شکل ۶. فرایند بازیابی اطلاعات مبتنی بر الگوی لیکنگ پیکنگ و همکاران (Lingpeng et al. 2004)

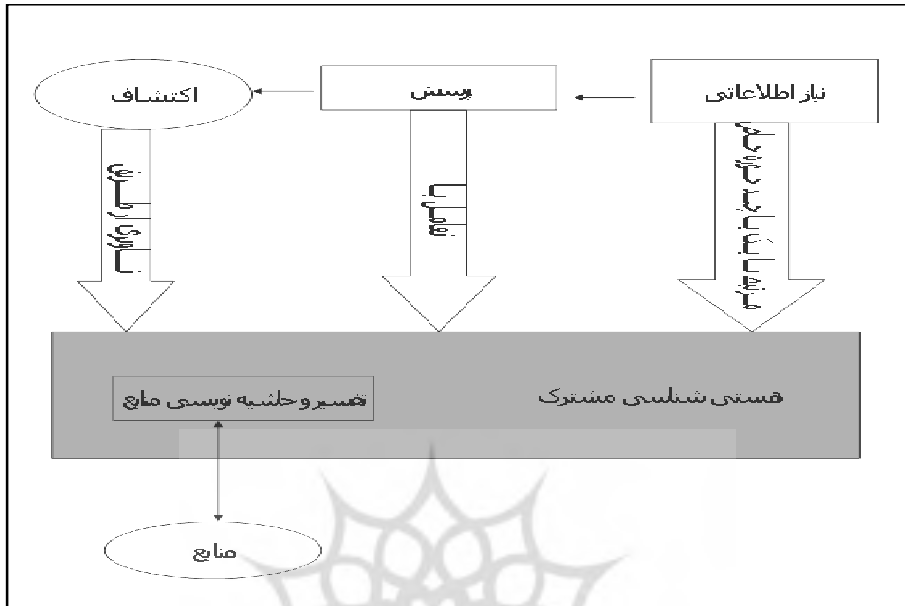
از ویژگی‌های اصلی این الگو این است که براساس مطالعات تجربی شکل گرفته است. در واقع، پژوهشگر شرایطی آزمایشگاهی را فراهم آورده است و تمامی عناصر موجود در الگوی مطرح شده را به‌طور عملی آزمایش کرده است (Lingpeng et al. 2004). به‌نظر می‌رسد ضعف اصلی این مدل محدود بودن کاربرد هستی‌شناسی به بسط جستجو است و در سایر مراحل بازیابی از آن استفاده نمی‌شود.

۲-۳. الگوی بازیابی اطلاعات دان و دیگران^۱

در این مدل، نیاز اطلاعاتی کاربر از طریق "هستی‌شناسی مشترک"^۲، به‌عبارت جستجویی که برای ماشین قابل فهم باشد تبدیل می‌شود. از طرفی، فرمول‌بندی عبارت جستجو می‌تواند از طریق تعامل با یک یا چند هستی‌شناسی - که در ابتدا براساس نیاز اطلاعاتی حوزه‌ای خاص انتخاب شده‌اند - انجام شود (شکل ۷). همچنین، مدارک موجود در نظام بازیابی باید از طریق هستی‌شناسی‌های مشترک تفسیر (نمایه‌سازی) شوند. در مرحله نهایی، فرایند شناسایی و انطباق عبارت جستجو با منابع نمایه‌شده به کمک منطق و استدلال هستی‌شناسی انجام می‌گیرد (Dan et al. 2006).

1. Dan et al.

2. Shared Ontology



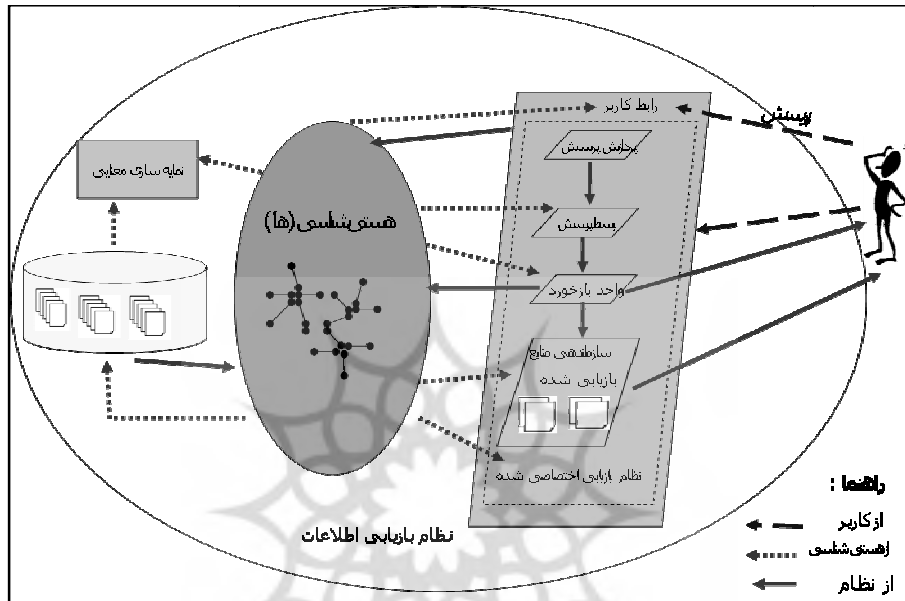
شکل ۷. فرایند بازیابی اطلاعات مبتنی بر الگوی دان و دیگران (Dan et al. 2006)

این الگو نسبت به الگوی لینک‌پنگ، هستی‌شناسی را در مراحل بیشتری مورد استفاده قرار می‌دهد. در مدل لینک‌پنگ - بیشتر اشاره شد - کاربرد هستی‌شناسی فقط در مرحله بسط عبارت جستجو مورد توجه قرار می‌گیرد، اما در این الگو از هستی‌شناسی برای درک نیاز اطلاعاتی کاربر توسط ماشین، فرمول‌بندی عبارت پرسش، نمایه‌سازی مدارک موجود در نظام بازیابی اطلاعات، و در نهایت، شناسایی و انطباق عبارت جستجو با منابع نمایه‌شده استفاده می‌شود. با وجود مزیت‌های اشاره‌شده، از چالش‌های مطرح در مدل دان، عدم توجه وی به رابط کاربر، شخصی کردن نظام بازیابی اطلاعات، و خوشه‌بندی مدارک بازیابی‌شده است که هر یک می‌تواند نقش مهمی در تعاملات کاربر با نظام داشته باشد.

۳-۳. الگوی پیشنهادی نظام بازیابی اطلاعات مبتنی بر هستی‌شناسی

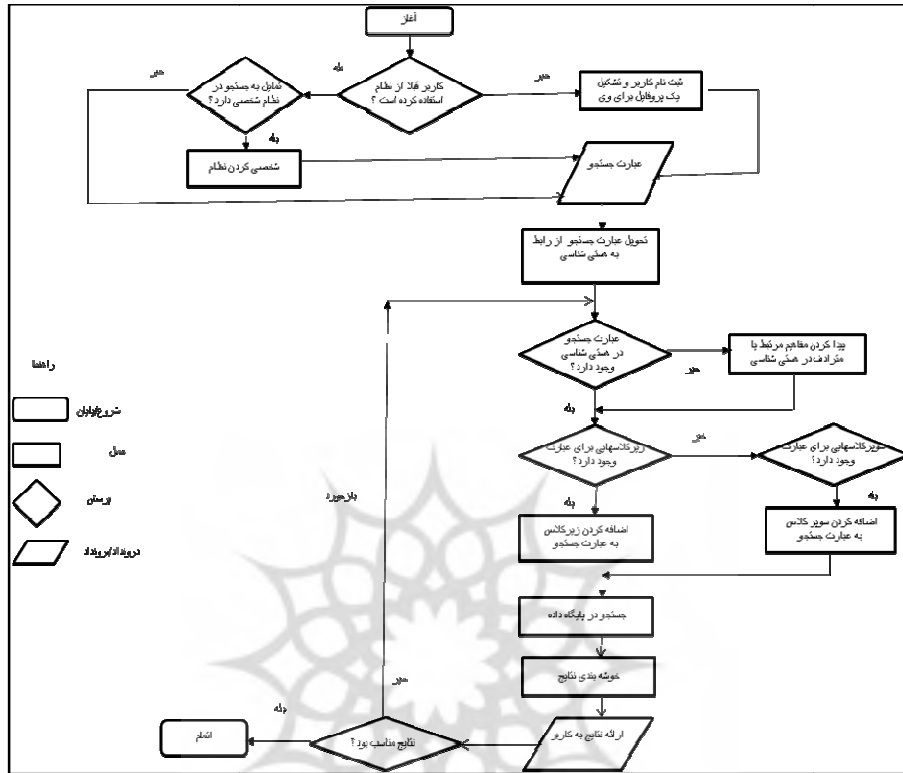
در این الگو، با اولین مراجعه کاربر به نظام، پروفایلی برای وی ساخته و همه رفتارهای جستجوی وی توسط هستی‌شناسی ذخیره می‌شود. ساخت پروفایل با کمک ماشین یادگیری هستی‌شناسی صورت می‌گیرد. در واقع، پروفایل هر کاربر از طریق تحلیل رفتارهای وی (مانند مرور، جستجو، زمان صرف‌شده برای هر صفحه، تحلیل محتوای هر صفحه جهت استخراج کلمات کلیدی آن، و وزن‌دهی و ثبت کلیدواژه‌ها در هستی‌شناسی کاربر) انجام می‌گیرد. از

این پس و در مراجعات بعدی، نظام بازیابی اختصاصی شده کاربر به وی معرفی می‌شود تا در صورت تمایل جستجو را بر مبنای نظام اختصاصی شده خود انجام دهد (شکل ۸).



شکل ۸. فرایند بازیابی اطلاعات مبتنی بر هستی‌شناسی

زمانی که پرسش از طریق رابط به هستی‌شناسی داده شد، بخش منطق هستی‌شناسی عبارت پرسش را تحلیل و با کمک فنون پردازش زبان طبیعی، کلیدواژه‌های پرسش را تفکیک می‌نماید و آن را به عبارت‌های معنایی تبدیل می‌کند. سپس، این عبارت‌ها از طریق واحد بازخورد به کاربر برگردانده می‌شود و با قضاوت کاربر در مورد آنها، براساس نیاز وی عبارت یا عبارت‌هایی به نظام برگردانده می‌شود. عمل تعامل و بازخوردگیری تا جایی ادامه پیدا می‌کند که کاربر به نیاز دقیق خود پی ببرد. در این مرحله، هستی‌شناسی از طریق انطباق عبارت‌های برگردانده شده با مدارک نمایه‌سازی شده توسط هستی‌شناسی، مدارک مرتبط را بازیابی می‌کند. در مرحله بعد، مدارک بازیابی شده مرتبط، از طریق هستی‌شناسی، سازماندهی و به صورت گروه‌های موضوعی به کاربر برگردانده می‌شود (شکل ۹).



شکل ۹. الگوریتم فرایند بازیابی اطلاعات مبتنی بر هستی‌شناسی

از ویژگی‌های این مدل نسبت به سایر مدل‌ها این است که علاوه بر اینکه نقاط قوت مدل‌های قبلی را حفظ کرده است، به نظام بازیابی تعاملی که در آن از طریق رابط کاربر مناسب، تعامل مطلوبی بین نظام و کاربر برقرار می‌شود، توجه ویژه‌ای شده است. همچنین، این مدل شخصی کردن نظام بازیابی اطلاعات مبتنی بر هستی‌شناسی و خوشه‌بندی مدارک بازیابی شده را نیز مورد توجه قرار داده است.

۴. نتیجه‌گیری

در این مقاله تلاش شد تا از طریق مرور متون، مهمترین کاربردهای هستی‌شناسی‌ها در بازیابی اطلاعات مورد توجه قرار گیرد و بر مبنای آن مدلی مفهومی ارائه شود. اگرچه این کاربردها می‌توانند نشان‌دهنده نقش مهم هستی‌شناسی‌ها در نظام‌های بازیابی اطلاعات آتی باشند و امیدواری زیادی به تولیدکنندگان و مصرف‌کنندگان نظام‌های بازیابی برای بهره‌گیری

بیشتر از این نظام‌ها را بدهند، هنوز هم چالش‌ها و مسائلی در این زمینه مطرح هستند: هزینه تولید هستی‌شناسی‌ها بسیار بالاست و همکاری گروهی از متخصصان کامپیوتر، زبان‌شناسان، متخصصان موضوعی، و کتابداران را می‌طلبد. همچنین، استفاده از هستی‌شناسی در نظام‌های بازیابی اطلاعات تاکنون به صورت آزمایشگاهی و تعداد بسیار محدودی در قالب تجاری ساخته شده‌اند. حجم فزاینده اطلاعات در محیط‌های الکترونیک، شکل‌های ساختاریافته متنی و اسناد چندرسانه‌ای از دیگر چالش‌های پیش روی استفاده از هستی‌شناسی‌ها در محیط‌های واقعی است. نکته دیگر اینکه با توجه به نقش آتی هستی‌شناسی‌ها در نظام‌های بازیابی اطلاعات از یک طرف و مطالعات بسیاری که در این زمینه به زبان‌های مختلف انجام می‌گیرد از سوی دیگر، به نظر می‌رسد که مطالعات مربوط به هستی‌شناسی در نظام‌های بازیابی فارسی می‌تواند از جنبه‌های مختلف به عنوان موضوعات پژوهشی در حوزه کتابداری و اطلاع‌رسانی مطرح باشد. برخی از زمینه‌های پیشنهادی عبارتند از:

- مهندسی مجدد هستی‌شناسی‌ها با استفاده از اصطلاحنامه‌های فارسی؛
- امکان استخراج روابط معنایی متون فارسی زبان با استفاده از هستی‌شناسی؛
- کشف میزان پیدایی متون از طریق نظام‌های بازیابی اطلاعات مبتنی بر هستی‌شناسی؛
- نمایه‌سازی مفهومی متون فارسی با استفاده از هستی‌شناسی‌ها؛
- امکان‌سنجی به کارگیری هستی‌شناسی‌ها در فهرست‌های پیوسته کتابخانه‌ای؛
- امکان به کارگیری هستی‌شناسی‌ها در موتورهای جستجوی فارسی؛ و
- امکان به کارگیری هستی‌شناسی‌ها در راهنماها.

سخن آخر اینکه در این مقاله تلاش شد تا به‌طور مختصر، کاربردهای بالقوه هستی‌شناسی‌ها در بازیابی اطلاعات مورد توجه قرار گیرد و بر مبنای آن، یک الگوی مفهومی برای نظام بازیابی اطلاعات مبتنی بر هستی‌شناسی ترسیم شود. امید است این الگو بتواند به فهم و درک بهتر نقش هستی‌شناسی در نظام‌های بازیابی اطلاعات کمک کند.

۵. منابع

- شریف، عاطفه. ۱۳۸۷. کاربرد هستی‌شناسی‌ها در نظام مدیریت دانش. فصلنامه کتابداری و اطلاع‌رسانی ۴۳ (۱۱): ۹۷-۱۱۶.
- صفری، مهدی. ۱۳۸۳. مدل‌سازی مفهومی در بازنمون رسم دانش: شناختی از هستی‌شناسی در هوش مصنوعی و نظام‌های اطلاعاتی. اطلاع‌شناسی. ۱ (۴): ۷۳-۱۰۴.
- صنعت‌جو، اعظم. ۱۳۸۴. ضرورت بازنگری در ساختار اصطلاحنامه‌ها: بررسی عدم کارایی اصطلاحنامه‌ها در محیط اطلاعاتی جدید و قابلیت‌های هستی‌شناسی در مقایسه با آن. فصلنامه کتاب ۴ (۶۴): ۷۹-۹۲.

- Albertoni, R., A. Riccardo, B. Alessio. 2005.. Semantic Web and Information Visualization. [http://semedia.deit.univpm.it/ swap2004/cameraready/albertoni.pdf](http://semedia.deit.univpm.it/swap2004/cameraready/albertoni.pdf) (accessed 8 Jan. 2010).
- Andreou, A. 2005. Ontologies and query expansion. *Master of Science School of Informatics. University of Edinburgh*. [http:// www.inf.ed.ac.uk/publications/thesis/online/IM050335.pdf](http://www.inf.ed.ac.uk/publications/thesis/online/IM050335.pdf) (accessed 8 Jan. 2010).
- Biocaster Ontology. <http://biocaster.nii.ac.jp/index.php?page=ontology> (accessed 8 Jan. 2010).
- Chen, M. 2009. Exploring the use of ontological relations in information retrieval. [http://nora.lis.uiuc.edu/ images/iConferences/iConference2009_poster_Final.pdf](http://nora.lis.uiuc.edu/images/iConferences/iConference2009_poster_Final.pdf) (accessed 8 Jan. 2010).
- Dan, W. U. et al. 2006. Role of Ontology in information retrieval . *Journal of Electronic Science and Technology of China* 4 (2). <http://www.xb.uestc.edu.cn/uploads/Jestc060212.pdf> (accessed 8 Jan. 2010).
- Ding, Y., Y. Ding, and S. Foo. 2001. Ontology research and development. Part 1: a review of ontology generation. *Journal of Information Science* 28 (2): 123-136.
- Gauch, S. Susan, G., C. Jason, and A. Pretschner. 2003. Ontology-based personalized search and browser. *Web Intelligence and Agent Systems* 1 (3-4). <http://portal.acm.org/citation.cfm?id=1016421> (accessed 8 Jan. 2010).
- Hernandez, N. 2006. Use and validation of the IAU astronomy thesaurus in Ontologies. [http://www.eso.org/sci/libraries/ IAUWGLib/iau06/thesaurus_IAU2006.pdf](http://www.eso.org/sci/libraries/IAUWGLib/iau06/thesaurus_IAU2006.pdf) (accessed 8 Jan. 2010).
- Lingpeng, Y. et al. 2004. Chinese information retrieval based on terms and Ontology. *Working notes of NTCIR-4, Tokyo 2-4 June*. <http://research.nii.ac.jp/ntcir-ws4/NTCIR4-WN/CLIR/NTCIR4WN-CLIR-YangLP.pdf> (accessed 8 Jan. 2010).
- Milstead, J. L. 1998. Use of Thesauri in the Full-Text Environment. <http://www.baysideindexing.com/Milstead/useof.htm> (accessed 8 Jan. 2010).
- Wikipedia. 2009. Ontology (information Science). [http://en.wikipedia.org/wiki/Ontology_\(information_science\)](http://en.wikipedia.org/wiki/Ontology_(information_science)) (accessed 8 Jan. 2010).
- Parlic, J. Ontology based information retrieval. www.people.tuke.sk/jan.paralic/papers/IIS03.pdf (accessed 8 Jan. 2010).
- Pretschner, A., and S. Gauch. 1999. Ontology based personalized search. In Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence, *November 08-10*. <http://portal.acm.org/citation.cfm?id=853649&dl=GUIDE&coll=GUIDE&CFID=60065494&CFTOKEN=31073854> (accessed 8 Jan. 2010).
- Soergel, D. 1999. The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science*. October 1999; 50 (12): 1119-1120. <http://www.dsoergel.com/cv/B70.pdf> (accessed 8 Jan. 2010).
- The Gene Ontology. <http://www.geneontology.org/> (accessed 8 Jan. 2010).
- Vallet, D., M. Fernández, and P. Castells. 2005. An Ontology-based information retrieval model. In *Second European Semantic Web Conference, (ESWC '05)*. www.acemedia.org/aceMedia/files/document/.../2005/eswc05-uam.pdf (accessed 8 Jan. 2010).
- Word Net Search-3. <http://wordnetweb.princeton.edu/perl/webwn> (accessed 8 Jan. 2010).

The Role of Ontology in Information Retrieval: Reviewing Current Research and Representing a Conceptual Model

Iranian Journal of
**Information
Processing &
Management**

Mahdieh Mirzabeigi*

PhD Student in LIS, Ferdowsi University of Mashhad

Iranian Research Institute Iranian
For Science and Technology
ISSN 2251-8223
eISSN 2251-8231
Indexed in LISA, SCOPUS & ISC
special issue: on Information Storage,
retrieval and Management (winter 2012)

Abstract: Inefficiency of thesauri and other information representation tools in electronic environment have forced librarians to revise the structure of these tools. So they have tried to develop other information organization tools such as ontology. In this paper, the performance of ontology in information retrieval was investigated. In addition, by reviewing two basic ontology-based information retrieval models- Lingpeng model and Dan Model- a new conceptual model was introduced.

Keywords: Ontology, information retrieval, conceptual model, information retrieval model

پژوهشگاه علوم انسانی و مطالعات فرهنگی
رتال جامع علوم انسانی

*Corresponding author: mmirzabeigi@gmail.com