

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



وزارت علوم، تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران

پژوهشگاه علوم و فناوری اطلاعات ایران

پژوهشکده علم سنجی و تحلیل اطلاعات

رساله دکترای رشته علم اطلاعات و دانش‌شناسی

طراحی مدل مفهومی بازیابی اطلاعات کاربرمدار مبتنی بر هستی‌شناسی

(مورد مطالعه پایگاه پایان‌نامه‌های پژوهشگاه علوم و فناوری اطلاعات ایران)

نگارش

المیرا کریمی

استادان راهنما

محمود بابایی

ملوک السادات حسینی بهشتی

مهر ۱۳۹۸

تقدیم به مهربان فرشتگانی که:

سحطات ناب باور بودن، لذت و غرور دانستن، جسارت خواستن، عظمت رسیدن و تمام تجربه های یکتا و زیبای زندگیم، مدیون حضور
سبز آنهاست...

پاسکزاری

بر خود لازم می‌دانم از استادان فرزانه جناب آقای دکتر بابایی و سرکار خانم دکتر بهشتی که به عنوان استاد راهنما در مراحل مختلف این پایان‌نامه همواره با سعه صدر و گشاده رویی در کنار من بودند و در طول مدت تحصیل از راهنمایی‌های اخلاقی و علمی ایشان بهره‌جسته‌ام تشکر و قدردانی نمایم

چکیده

هدف: با توجه به گسترش وب ۳ و آغاز مباحث مربوط به معنا، جست‌وجو و بازیابی اطلاعات کلیدواژه‌ای پاسخگوی نیاز کاربران نیست. یکی از راه‌حل‌ها برای حل مشکل بازیابی اطلاعات کلیدواژه‌ای استفاده از هستی‌شناسی‌ها و روابط معنایی موجود در آنها است. این پژوهش در پی بهبود بازیابی اطلاعات کاربرمدار با استفاده از روابط معنایی موجود در هستی‌شناسی‌ها است.

روش‌شناسی پژوهش: این پژوهش با مطالعه ترکیبی سعی در شناسایی ویژگی‌ها و مولفه‌هایی دارد که چگونگی استفاده از روابط معنایی موجود در هستی‌شناسی‌ها را به نمایش می‌گذارد. این پژوهش توجه ویژه‌ای به اصلاح جست‌وجوی کاربر و پیشنهاد برای جست‌وجوی کاربر دارد. برای این کار ابتدا نظام‌های بازیابی اطلاعاتی متفاوت مورد بررسی قرار گرفته، ویژگی‌های معنایی مورد استفاده در آنها به تصویر کشیده شد، این نظام‌های بازیابی اطلاعات مواردی بودند که از اصطلاحنامه و هستی‌شناسی در بازیابی اطلاعات استفاده می‌کردند و ۵ مورد بودند. سپس با غنی‌سازی روابط معنایی موجود در اصطلاحنامه Inspec، با استفاده از روابط موجود در هستی‌شناسی OIS در حوزه اطلاع‌رسانی و چکیده‌پایان‌نامه‌های گنج، روابط معنایی حوزه علم اطلاع‌رسانی شناسایی شد و براساس آن، لاگ جست‌وجوی کاربران در این حوزه تحلیل شد. در انتها نیز براساس ویژگی‌های استخراج شده از مراحل پیشین از طریق مصاحبه با متخصصان حوزه بازیابی اطلاعات و اصطلاح‌شناسی، تا رسیدن به اشباع نظری، مدل مفهومی بازیابی اطلاعات تکمیل شد. اعتبار سنجی مدل مفهومی، با مطالعه دلفی در دو دور انجام شد. در نهایت با انجام آزمایش کنترل شده مدل مفهومی طراحی شده بررسی شد.

یافته‌ها: در مرحله اول، ۵ مؤلفه اصلی و ۳۰ مؤلفه فرعی معنایی در نظام‌های بازیابی اطلاعات شناسایی شد. در مرحله بعد، مفاهیم و روابط معنایی حوزه اطلاع‌رسانی که شامل ۲۳۷ مفهوم و ۵۰ رابطه معنایی است، شناسایی شد، با بررسی ویژگی‌های معنایی برگرفته از متون و روابط معنایی شناسایی شده در مرحله قبل، روابط معنایی موجود در پرس‌وجوهای کاربر (شامل پرس‌وجوهای اولیه و پرس‌وجوهای اصلاح شده) شناسایی شد. مشخص شد که کاربران انواع خاصی از روابط معنایی که در حوزه اطلاع‌رسانی در این پژوهش شناسایی شده بوده، به کار می‌برند. در مرحله بعد، با مصاحبه با متخصصان، ویژگی‌های مدل مفهومی، شامل ۲ بخش و ۱۶ مؤلفه فرعی شناسایی شد. در مرحله بعد با ترکیب نتایج مراحل پیشین، مدل مفهومی اولیه شناسایی شد و نهایتاً با دو دور دلفی، مدل مفهومی شامل ۳ بخش، ۲۳ مؤلفه اصلی و ۳۷ مؤلفه فرعی تأیید. هر یک از این بخش‌ها به بعدی از ابعاد نظام بازیابی اطلاعات با تمرکز بر پیشنهاد برای اصلاح جست‌وجو توجه دارد. آزمایش کنترل شده‌ای که انجام شد، کارایی مدل مفهومی طراحی شده را تأیید کرد.

نتیجه‌گیری: روابط معنایی استفاده شده برای پیشنهاد برای اصلاح جست‌وجو براساس جست‌وجوهای کاربر شکل گرفته است، دسته اولیه این روابط، روابط معنایی هستند که کاربران به طور کلی در اصلاح جست‌وجوهای خود از آنها استفاده کرده‌اند. برای محدودسازی این روابط به گونه‌ای که ارائه آنها باعث سردرگمی بیشتر کاربران نشود، براساس متون و براساس نتایج مرحله تحلیل لاگ روابط خاصی به آنها ارائه می‌شود. علاوه براین، هر یک از کاربران در پرس‌وجوهای خود به نوعی از تاکتیک‌های مختلف استفاده می‌کنند بر این اساس نیز،

دسته دیگری از روابط معنایی برای پیشنهاد برای اصلاح پرس وجو به کاربران ارائه می شود. با توجه به این که یافته های پژوهش کارآمد بودن این مدل و مؤلفه های آن را در افزایش میزان ربط کاربران نشان داد، می توان برای استفاده از این مدل زیرساخت های لازم را فراهم کرد و در نهایت نتیجه آن را در محیط واقعی به نظاره نشست.

کلیدواژه ها: بازیابی اطلاعات، هستی شناسی، روابط معنایی، مطالعه کاربر، اصلاح پرس وجو، پیشنهاد اصلاح پرس وجو

فهرست نوشتار

۱۷	۱. معرفی پژوهش
۱۷	۱-۱. مقدمه و معرفی موضوع
۲۰	۲-۱. بیان مسئله
۲۴	۳-۱. ضرورت و اهمیت پژوهش
۲۵	۴-۱. اهداف پژوهش
۲۷	۵-۱. پرسش‌های پژوهش
۲۷	۶-۱. دانش‌افزایی و جنبه نوآوری پژوهش
۲۸	۷-۱. روش پژوهش
۳۱	۲. مبانی نظری و پیشینه پژوهش
۳۱	۲-۱. مقدمه
۳۲	۲-۲. بازیابی اطلاعات
۳۶	۲-۱-۲. فرآیند بازیابی اطلاعات
۳۷	۲-۲-۲. پرس و جو در بازیابی اطلاعات
۴۰	۲-۲-۳. مشکلات بازیابی اطلاعات
۴۴	۲-۳. هستی‌شناسی
۴۵	۲-۳-۱. تعریف هستی‌شناسی
۵۲	۲-۴. مقایسه اصطلاحنامه و هستی‌شناسی
۵۴	۲-۵. هستی‌شناسی در بازیابی اطلاعات
۶۲	۲-۶. پیشینه پژوهش
۶۳	۲-۶-۱. جستجو و گسترش آن
۶۷	۲-۶-۲. سیستم پیشنهاد

۶۹	۳-۶-۲ کاربردمداری در بازیابی اطلاعات
۷۵	۷-۲ استنتاج از پیشینه‌ها
۷۵	۸-۲ نظریه پردازي
۷۸	۹-۲ مدل پیش ساخته پژوهش
۸۴	۳. مواد و روش پژوهش
۸۴	۱-۳. مقدمه
۸۴	۲-۳. روش شناسی پژوهش
۸۶	۳-۳ جامعه و نمونه پژوهش
۸۶	۱-۳-۳ هستی شناسی Inspec
۸۷	۲-۳-۳ نظام‌های بازیابی اطلاعات
۸۸	۳-۳-۳ ذی‌نفعان پایگاه اطلاعاتی پایان‌نامه‌های پژوهشگاه علوم و فناوری اطلاعات ایران
۸۹	۴-۳ شیوه‌ها و ابزارهای گردآوری داده
۸۹	۱-۴-۳ مشاهده
۹۰	۲-۴-۳ مصاحبه
۹۰	۵-۳ تجزیه و تحلیل داده‌های پژوهش
۹۱	۶-۳ مدل مفهومی روش شناسی پژوهش
۹۲	۷-۳ آروایی و پایایی ابزارهای گردآوری اطلاعات
۹۴	۸-۳ مراحل و فرایند اجرای پژوهش
۹۴	۱-۸-۳ مرحله اول- مشاهده
۹۸	۲-۸-۳ مرحله دوم- مصاحبه با ذی‌نفعان
۹۹	۳-۸-۳ مرحله سوم- تعیین اعتبار مدل مفهومی
۹۹	۴-۸-۳ مرحله چهارم- آزمایش کنترل‌شده سنجش مدل مفهومی در پایگاه پایان‌نامه‌های ایرانداک
۱۰۱	۴. ارائه یافته‌ها
۱۰۱	۱-۴-۴ مقدمه

۱۰۱	۴-۲ مشاهده نظام‌های بازیابی اطلاعات موجود
۱۰۱	۴-۲-۱ معیارهای ویژگی‌های معنایی
۱۰۴	۴-۲-۲ معرفی نظام‌های بازیابی اطلاعات
۱۰۷	۴-۲-۳ نتایج بررسی نظام‌های بازیابی اطلاعات
۱۱۶	۴-۲-۴ بررسی موتور جستجوی گوگل
۱۲۱	۴-۳ تکمیل روابط معنایی حوزه اطلاع‌رسانی
۱۲۲	۴-۳-۱ بررسی اولیه هستی‌شناسی پایه
۱۲۴	۴-۳-۲ بررسی اولیه هستی‌شناسی
۱۴۰	۴-۴ مشاهده رفتار اصلاح پرس و جوی کاربر
۱۴۲	۴-۴-۱ تحلیل ویژگی‌های معنایی لاگ‌ها
۱۴۷	۴-۵ مصاحبه با متخصصان
۱۵۵	۴-۶ طراحی مدل مفهومی بازیابی اطلاعات کاربرمدار مبتنی بر هستی‌شناسی
۱۶۲	۴-۷ اعتبارسنجی مدل مفهومی
۱۶۷	۴-۸ آزمایش کنترل شده
۱۶۸	۴-۸-۱ معیار مفاهیم
۱۶۹	۴-۸-۲ معیار روابط
۱۷۱	۴-۸-۳ معیار کیفیت پیشنهادها
۱۷۵	۵. بحث و جمع‌بندی
۱۷۵	۵-۱ مقدمه
۱۷۵	۵-۲ خلاصه پژوهش
۱۷۷	۵-۳ پاسخ به پرسش‌های اساسی پژوهش
۱۸۵	۵-۴ مقایسه نتایج پژوهش با پژوهش‌های پیشین
۱۸۸	۵-۵ پیشنهادها
۱۸۸	۵-۵-۱ پیشنهادهایی برای پژوهش

۵-۲-۵ پیشنهادهایی برای اجرا

۱۸۹

۵-۶ نتیجه گیری

۱۹۰

فهرست جدول‌ها

۷۰	جدول ۱-۲. تحلیل مطالعات حوزه بازیابی اطلاعات و هستی‌شناسی
۹۷	جدول ۳-۱. لیست نظام‌های مورد نظر برای بررسی ویژگی‌های معنایی
۱۰۲	جدول ۴-۱. سیاهه واری و ویژگی‌های معنایی
۱۰۸	جدول ۴-۲. معیارهای جستجوی مفهومی در نظام‌های بازیابی اطلاعات
۱۱۰	جدول ۴-۳. معیارهای پیشنهاد برای جستجو در نظام‌های بازیابی اطلاعات
۱۱۱	جدول ۴-۴. معیارهای جستجوی مترادف‌ها در نظام‌های بازیابی اطلاعات
۱۱۲	جدول ۴-۵ تا کتیک‌های جستجوی حمایت‌شونده از سوی هستی‌شناسی (García and Sicilia 2003)
۱۱۳	جدول ۴-۶. معیارهای استفاده از هستی‌شناسی در نظام‌های بازیابی اطلاعات
۱۱۴	جدول ۴-۷. معیارهای نمایش در نظام‌های بازیابی اطلاعات
۱۲۹	جدول ۴-۸. روابط معنایی انتخاب‌شده برای تقویت
۱۳۱	جدول ۴-۹. نمونه‌ای از مفاهیم و روابط بین آنها
۱۳۳	جدول ۴-۱۰. نمونه‌ای از روابط معنایی استخراج‌شده در چهریزه «قلمرو»
۱۳۴	جدول ۴-۱۲. نمونه‌ای از روابط معنایی استخراج‌شده برای چهریزه «روش»
۱۳۴	جدول ۴-۱۳. روابط معنایی استخراج‌شده برای چهریزه «قواعد»
۱۳۴	جدول ۴-۱۴. روابط معنایی استخراج‌شده برای چهریزه «فلسفه و نظریه»
۱۳۵	جدول ۴-۱۵. روابط معنایی استخراج‌شده برای چهریزه «فعالیت»
۱۳۵	جدول ۴-۱۶. روابط معنایی استخراج‌شده برای چهریزه «منابع»
۱۳۵	جدول ۴-۱۷. روابط معنایی استخراج‌شده برای چهریزه «مطالعات»
۱۳۶	جدول ۴-۱۸. روابط معنایی استخراج‌شده برای چهریزه «نقش‌ها»
۱۳۶	جدول ۴-۱۹. روابط معنایی استخراج‌شده برای چهریزه «اجتماعی»
۱۳۶	جدول ۴-۲۰. روابط معنایی تعریف‌شده برای چهریزه «ابزار»
۱۳۷	جدول ۴-۲۱. نمونه‌ای از روابط استخراج‌شده خارج از چهریزه‌ها
۱۴۳	جدول ۴-۲۲. وضعیت هم‌رخدادی واژگان دارای ارتباط معنایی اصطلاحنامه Inspec در میان کاربران پایگاه اطلاعات گنج
۱۴۴	جدول ۴-۲۳. وضعیت استفاده کاربران پایگاه اطلاعات گنج از تاکتیک‌های حامی هستی‌شناسی
۱۴۵	جدول ۴-۲۴. نمونه‌ای از استفاده از روابط معنایی در اصلاح جستجوهای کاربران پایگاه اطلاعات گنج

جدول ۴-۲۵	درصد مطابقت روابط معنایی مورد استفاده در اصلاح جستجوها با روابط معنایی در اصطلاحنامه تقویت شده ۱۴۷
جدول ۴-۲۶	نتایج دور اول دلفی درباره اجزای مدل مفهومی بازیابی اطلاعات کاربرمدار مبتنی بر هستی‌شناسی
جدول ۴-۲۷	نمونه‌هایی از پرسش‌های جستجو و کلیدواژه‌های آنها
جدول ۴-۲۸	توزیع فراوانی نظرات کاربران در معیار مفاهیم
جدول ۴-۲۹	توزیع فراوانی نظرات کاربران در مورد معیار «روابط»
جدول ۴-۳۰	توزیع فراوانی نظرات کاربران در مورد معیار کیفیت پیشنهادها
جدول ۴-۳۱	درصد فراوانی معیار کیفیت نتایج در گنج بدون پیشنهاد
جدول ۴-۳۲	درصد فراوانی معیار کیفیت نتایج در گنج با ارائه پیشنهاد
جدول ۵-۱	مؤلفه‌های مدل مفهومی بازیابی اطلاعات کاربرمدار مبتنی بر هستی‌شناسی

فهرست نمودارها

۶۲	نمودار ۱-۲. حوزه های پژوهشی بازیابی اطلاعات و هستی شناسی (برگرفته از هنان، ۲۰۰۲)
۶۳	نمودار ۲-۲. دسته بندی مطالعات هستی شناسی در بازیابی اطلاعات
۷۹	نمودار ۲-۳. مدل پیش ساخته پژوهش
۹۲	نمودار ۳-۱ مدل مفهومی روش شناسی پژوهش
۱۲۲	نمودار ۴-۱ نمای کلی تکمیل روابط معنایی در حوزه علم اطلاعات
۱۷۶	نمودار ۵-۱ مشکلات بازیابی اطلاعات

فهرست پیوست‌ها

۲۰۱	پیوست الف- پرسش‌های مصاحبه
۲۰۴	پیوست ب: نمونه لاگ جستجوی کاربران
۲۲۱	پیوست ج- پرسشنامه دور اول دلفی
۲۲۶	پیوست د- پرسشنامه دور دوم دلفی
۲۳۲	پیوست ه- پرسشنامه کاربر- بررسی کارایی هستی‌شناسی
۲۳۵	پیوست و- پرسشنامه کاربر- آزمایش کنترل‌شده

۱. معرفی پژوهش

۱-۱. مقدمه و معرفی موضوع

در طی سال‌ها پیشرفت در وب و شکل‌گیری انواع آن، و نهایتاً ابداع وب معنایی باعث شده که اطلاعات به گونه‌ای سازماندهی و ذخیره شوند که جستجو و بازیابی اطلاعات با رایانه قابل پردازش باشد. (کشاورز ۱۳۸۶). از نظر برنرزی و همکارانش (۲۰۰۱) هستی‌شناسی^۱ یکی از لایه‌های وب معنایی و ستون فقرات آن به شمار می‌آید. این ابزار وب معنایی، مدلی انتزاعی از جهان واقع است که مفاهیم و روابط بین آنها را نمایش می‌دهد. پایگاه دانش مفهومی^۲ کاربردهای بسیاری، مانند شبکه‌های جهان گستر معنایی^۳، موتورهای جستجو، تجارت الکترونیکی، پردازش زبان طبیعی، استخراج و بازیابی اطلاعات و مواردی شبیه اینها دارند. (شمس فرد، عبدالله زاده بارفروش ۱۳۸۱). تفاوت لغات و تفاوت نوع استفاده از آنها باعث ایجاد چندگانگی‌ها و در نتیجه آشفتگی در فضای وب شده است، اما از آنجا که هستی‌شناسی، در وب، درک مشترکی از یک حوزه را ارائه می‌دهد، بر تفاوت لغات غلبه می‌کند.

در واقع آن چه موجب اهمیت یافتن هستی‌شناسی شده است، نبود استانداردهایی برای ارتباط نحوی و معنایی بین مفاهیم هم از دید انسان و هم از دید رایانه است. با توجه به این که حجم اطلاعات در سال‌های اخیر در حال افزایش است و موتورهای کاوش موجود در وب چندان کارایی ندارند، نیاز به هستی‌شناسی‌ها روز به روز بیشتر احساس می‌شود. (صفری ۱۳۸۳). هستی‌شناسی، مفاهیم حوزه مرتبط، ویژگی‌های آنها و روابط ممکن میان مفاهیم و خصیصه‌ها را مشخص می‌کند (ژاکوب^۴ ۱۳۸۴). ساختار موجود در هستی‌شناسی‌ها، ساختار سلسله‌مراتبی است. نمونه این ساختار را می‌توان در سرعنوان‌های موضوعی کتابخانه کنگره مشاهده کرد. به عنوان مثال، هستی‌شناسی مشخص می‌کند که مؤلف نوعی پدیدآورنده است. همان‌طور که اشاره شد افراد اصطلاحات را به معانی مختلف به کار می‌برند و ممکن است برای مفهوم از اصطلاحات متفاوتی استفاده کنند. هستی‌شناسی امکان استفاده دقیق و آسان را از واژه‌نامه‌ها فراهم می‌کند به عبارت دیگر، هستی‌شناسی وب، از یک نگاه کلی،

¹ Ontology

² Conceptual Knowledge

³ Semantic Web

⁴ Jacob

مجموعه‌ای از تعریف مفاهیم و ویژگی‌ها و محدودیت‌های موجود در نحوه به کارگیری این مفاهیم و ویژگی‌هاست (جمالی مهموئی ۱۳۸۲). هستی‌شناسی‌ها بنیاد مفهومی معناداری برای دستیابی به هدف وب معنایی، تهیه می‌کنند. به گفته ژاکوب (۱۳۸۴) طرح‌های ابرداشته‌ای که مجموعه ویژگی‌های مفهومی و فیزیکی منابع مناسب برای کاربران را فراهم می‌کنند، نوعی از هستی‌شناسی هستند. به عنوان مثال دابلین کور که مجموعه اصطلاحات را برای توصیف منابع ارائه می‌دهد، به این دلیل که مدل رسمی مفهومی را اعمال می‌کند، نوعی هستی‌شناسی است. در واقع مدل رسمی مفهومی پردازش خودکار اطلاعات را در میان نظام‌های رایانه‌ای تسهیل می‌کند و موجبات ظهور وب معنایی را فراهم می‌کند.

در حوزه کتابداری و اطلاع‌رسانی، هستی‌شناسی عمدتاً به عنوان ابزاری معناشناسی به کار گرفته شده است، یکی از انواع ابزارهای معنایی از دیرباز، اصطلاحنامه‌ها بوده‌اند که از مفاهیم و روابط بین آنها استفاده می‌کردند، اما، اصطلاحنامه‌ها محدودیت‌هایی در ترسیم روابط معنایی دارند و این مسئله سبب استفاده از ساختار هستی‌شناسی‌ها شده است. (صنعت جو و فتحیان ۱۳۹۱).

با این که هستی‌شناسی‌ها شباهت بسیار زیادی به اصطلاحنامه‌ها و دیگر طرح‌های رده‌بندی دارند اما تفاوت‌های آنها با اصطلاحنامه‌ها و طرح‌های رده‌بندی، باعث شده از آنها استفاده بیشتری شود. (Ding and Foo 2002a). با وجود شباهت‌ها، یکی دانستن هستی‌شناسی‌ها با اصطلاحنامه‌ها و طرح‌های رده‌بندی باعث نادیده گرفتن کاربرد عملی آنها می‌شود و این باعث بهره نگرفتن از ویژگی‌های متمایز کننده آنها می‌شود.

تفاوت اساسی هستی‌شناسی با سایر واژگان در سطح انتزاع، مفهومی بودن، قابلیت بازنمایی ماشینی و سطح بازنمون آنها است (Ding and Foo 2002a). در اصطلاحنامه روابط معنایی بین اصطلاحات به شکل اعم، اخص و وابسته مشخص می‌شود و در طرح‌های رده‌بندی، این روابط در ساختار سلسله‌مراتبی نمود پیدا می‌کند. در هستی‌شناسی، روابط غنی‌تر و دقیق‌تر از آن چیزی است که در اصطلاحنامه‌ها وجود دارد و تعداد و نوع این روابط براساس نیاز و حوزه خاص متفاوت است. هستی‌شناسی می‌تواند لایه‌های مختلفی داشته باشد، به این ترتیب که هستی‌شناسی سطح بالا که مفاهیم عام را ارائه می‌کند و هستی‌شناسی سطح پایین که خاص‌تر است. هدف هستی‌شناسی فراهم آوردن امکان ارتباط میان انسان و رایانه است و سطح استانداردسازی و قابلیت استفاده مجدد از دانش را بالا می‌برد.

⁵ Metadata scheme

با توجه به کاربرد هستی‌شناسی در زمینه‌های مختلف، در زمینه بازیابی اطلاعات مطالعه کابل و همکاران (۲۰۰۴) نشان داد که زمانی که واژگان با استفاده از هستی‌شناسی سازماندهی می‌شوند بازیابی اطلاعات از وجوه مختلف بهبود پیدا می‌کند. (صفری، ۱۳۸۳)

از طرفی، کاربران نظام‌های بازیابی اطلاعات در بیشتر مواقع، از پیدا کردن منابع مرتبط مورد نیاز از میان هزاران منبع موجود در وب گلایه دارند. یکی از راه‌حل‌های این مسئله و کسب رضایت کاربران، گسترش نظامی است که بتواند جستجویی را انجام دهد که در آن بیشترین مطالب مرتبط و کمترین مطالب نامرتب 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. راه‌حل 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. هر یک از این ابزارها توانستند تا حدودی به سازمان‌دهی منابع و بازیابی بهتر کمک کنند، اما همان‌گونه که میلستد (۱۹۹۸) 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.

پژوهش‌ها نشان می‌دهند روابط خاص موجود در اصطلاحنامه‌ها 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 این که 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 برای تطابق با محیط‌های الکترونیکی وادار کرد که از نتایج آن ظهور ابزارهای دیگر سازمان‌دهی اطلاعات همچون هستی‌شناسی‌ها است. با توجه به آنچه گفته شد، پژوهش حاضر در پی آن است که از ویژگی‌ها و مشخصه‌های این نوع ابزار معنایی بهره گیرد و از آن در بازیابی اطلاعات به شکل کاربرمدار استفاده نماید.

۲-۱. بیان مسئله

کتابخانه‌ها در میان اولین مؤسسه‌ای هستند که نظام‌های بازیابی اطلاعات را برای بازیابی اطلاعات پذیرفتند. (Baeza-Yates and Ribeiro-Neto 1999). هدف نظام‌های بازیابی اطلاعات، شناسایی مدارک یا داده‌هایی است که متناسب با نیاز اطلاعات خاص هستند. بازیابی اطلاعات یک عمل ساده و مکانیکی نیست،

⁶ Kabel et al

در واقع این عمل، کاری پیچیده و عمیق است که میان کسانی که اطلاعات را تولید می‌کنند و افرادی که اطلاعات را مصرف می‌کنند ارتباط برقرار می‌کند. بنابراین، نظام‌های بازیابی اطلاعات برای ایجاد این ارتباط شکل گرفته‌اند. (حریری، ۱۳۶۱). در بازیابی اطلاعات سنتی، یکی از مهم‌ترین مهارت‌هایی که کاربران برای رسیدن به نتیجه مطلوب نیاز دارند، فرمول‌بندی سؤال است که طبق نظر بیزا-یتس و ریبیرو-نتو^۷ (۱۹۹۹) بیشتر کاربران آن را سخت می‌دانند و تبدیل نیاز اطلاعاتی خود به سؤالی که متناسب با نظام بازیابی اطلاعات مربوطه باشد یکی از مشکلات افراد در بازیابی اطلاعات است.

افزون بر مهارت کاربران، بازیابی اطلاعات موجود بر روی شبکه اینترنت، وابسته به دقت و توانایی‌های سیستم‌های بازیابی اطلاعات است. ابزارهای بازیابی اطلاعات موجود همچون اصطلاحنامه‌ها، نظام‌های نمایه‌سازی، موتورهای جستجو و مانند اینها، برای بازیابی اطلاعات مرتبط محدودیت‌هایی دارند (Ding et al 2007). این محدودیت‌ها از این روست که نظام‌های بازیابی اطلاعات نمی‌توانند معنای مدنظر مخاطب و هدف او و علاوه بر این، محتوای سند را درک کنند. برای کاهش این نقص، نظام‌های بازیابی سنتی گسترش یافته‌اند. برخی از این گسترش‌ها، بر پایه استفاده از ویژگی‌های معنایی اصطلاحات مثل ارتباط بین آنها است. (خسروی و وظیفه دوست، ۲۰۰۷).

اصطلاحنامه‌ها، یکی از معروف‌ترین و رایج‌ترین ابزارهایی هستند که در بازیابی مفاهیم حوزه‌های موضوعی کمک زیادی هم به کاربران و هم به طراحان سیستم‌ها می‌کنند. هدف اصطلاحنامه برگرداندن زبان طبیعی مدارک به زبان کنترل‌شده است که باعث ایجاد هماهنگی میان نظرات نویسندگان و فراهم‌کنندگان و کاربران اطلاعات می‌شود. کمبودهایی نیز در اصطلاحنامه‌ها وجود دارد. اصلی‌ترین مشکل اصطلاحنامه‌ها محدودیت مجموعه ارتباطات است. این موضوع گاهی رابطه‌های واقعی را پنهان می‌سازد و ارتباطات مبهمی بین اصطلاحات به وجود می‌آورد. این ارتباطات مبهم منجر به مسائلی می‌شود که وقتی که کاربردهای هوشمند جدید به کار می‌رود، نیاز به استنتاج دارد. اگرچه تلاش‌های زیادی برای تدوین و ترجمه اصطلاحنامه‌های تخصصی فارسی شده است و خروجی‌های نسبتاً خوبی هم داشته است، اما متأسفانه به کارگیری بهینه و خودکار اصطلاحنامه‌های رایانه‌ای در بازیابی اطلاعات و کاربرد اصطلاحنامه به‌عنوان ابزاری در بازیابی اطلاعات رایانه‌ای مشکلاتی را دارا است. رویکردهای سنتی بازیابی اطلاعات در اصطلاحنامه‌ها به دلیل تمرکز بر ساختار نحوی مفاهیم و عدم توجه کافی به معانی محتوا نیز به دلیل ارائه بازنمونی ساده و مبهم از محتوای اطلاعات و نیاز اطلاعاتی کاربر، قادر نیستند که کاربر را در بازیابی نیاز اطلاعاتی خود و یافتن اطلاعات مرتبط یاری دهند (Stojnavic 2003).

⁷ Baeza-Yates and Ribeiro-Neto

هستی‌شناسی‌ها ابزار جدید برای بازنمون‌دانش هستند که به نظر می‌رسد می‌توان با استفاده از آن، بخشی از مشکلات اصطلاح‌نامه‌ها را برطرف نمود. (فتحیان، ۱۳۹۰). در ده سال گذشته، هستی‌شناسی‌ها از یک پارادایم مفهوم‌سازی به یک فرآیند مدل‌سازی معنایی، به‌ویژه در بافت وب معنایی، ارتقا یافته‌اند. از نقطه‌نظر بازیابی اطلاعات، هستی‌شناسی‌ها یک قالب قابل‌فهم توسط ماشین از توصیف و محتوا را فراهم ساخته‌اند؛ نظیر آنچه فرآیند بازیابی اطلاعات با معنی کردن محتوا می‌تواند فراهم کند. (صنعت جو، ۱۳۹۱).

چان و ولین (۲۰۰۴) به شباهت‌ها و تفاوت‌های اصطلاح‌نامه و هستی‌شناسی اشاره می‌کنند. طبق نظر آنها، اصطلاح‌نامه تعداد محدودی روابط دارد که می‌تواند در ارائه دانش خاص در حوزه خاص رسایی نداشته باشد. قابل توجه‌ترین تفاوت‌های این دو بر پایه ظرفیت زبان توصیف دانش دامنه‌ای است. از آنجا که هستی‌شناسی‌ها ساختار مشابهی با اصطلاح‌نامه‌ها دارند، به عنوان ابزار کمکی در بازنمون‌دانش و بازیابی مفاهیم موضوعی استفاده می‌شوند، اما ساختار در هستی‌شناسی‌ها غنی‌تر از اصطلاح‌نامه‌ها است. بنابراین آنها فرایند بازیابی را مفهوم‌سازی می‌کنند و نقش مهم‌تری در کاربردهای مشابه با اصطلاح‌نامه دارند. هستی‌شناسی‌ها، زمینه معنایی غنی را برای توصیف واحدهای معنا و تنوع تفسیرها ایجاد می‌کنند و با کاهش ابهام مفاهیم، طرح‌های استنتاجی قدرتمندی برای استدلال و طرح پرسش‌های جستجو ارائه می‌نمایند (Castells, Fernandez, Vallet, Mylonas and Avritthis 2005). دینگ (۲۰۰۱) معتقد است هستی‌شناسی بر مبنای کارکردهای استنتاجی اطلاعات مرتبط را بازیابی می‌کند و قادر است مفهوم بازیابی هوشمند اطلاعات را تحقق بخشد.

از طرفی، هدف بازیابی اطلاعات پیدا کردن اطلاعات مربوط به نیاز اطلاعاتی کاربران است و چون فرایند بازیابی اطلاعات بسیار مبهم است و کاربر با مخزن اطلاعات و یا نحوه جستجو آشنایی ندارد، نیاز خود را در قالب کلیدواژه مطرح می‌کند و لازم است برای از بین بردن شکاف بین معنای محتوا و معنای جستجوی کاربر، تعامل کاربر با نظام بازیابی اطلاعات بیشتر شود. هستی‌شناسی‌های موضوعی می‌توانند در نظام بازیابی اطلاعات به کار روند و با تفسیر مفهومی معنای جستجوی کاربر، امکان تعامل با کاربر را فراهم می‌کنند. به نظر می‌رسد هستی‌شناسی می‌تواند ویژگی‌های کاربرمداری نظام‌های اطلاعاتی و فرآیند بازیابی اطلاعات را افزایش دهد.

در پژوهشگاه علوم و فناوری اطلاعات ایران، پایگاه‌های اطلاعاتی متنوعی وجود دارد که بازیابی اطلاعات موجود در این پایگاه‌ها از راه‌های مختلفی انجام می‌گیرد. اصطلاح‌نامه‌ها در این پژوهشگاه تنها برای نمایه‌سازی اطلاعات استفاده می‌شوند. تعدادی از اصطلاح‌نامه‌های موجود تألیفی و تعدادی ترجمه شده‌اند. علاوه بر این، فاصله زمانی روزآمدسازی این اصطلاح‌نامه‌ها بسیار زیاد است. گرچه در حال حاضر اصطلاح‌نامه‌های علوم پایه

روزآمدسازی شده و تحت وب در دسترس همگان قرار گرفته است. از طرفی بازیابی اطلاعات در این پایگاه اطلاعاتی پایان‌نامه‌های پژوهشگاه علوم و فناوری اطلاعات ایران با مشکلاتی مخصوصاً در حوزه‌های مربوط به معنا مانند مترادف‌ها، املاهای مختلف، کلمات دارای معانی خاص، کلمات دارای معانی متعدد و مشکلات مربوط به مسیریابی مفهومی مربوط به روابط هم‌ارز و سلسله‌مراتبی مواجه است. از طرفی، نظام بازیابی اطلاعات مبتنی بر این پایگاه‌ها، تعامل کمی با کاربر دارند به طوری که حتی امکان پیشنهاد به کاربر برای اصلاح پرس‌وجو را ندارند. با توجه به آن چه گفته شده و براساس مطالعاتی که پیش از این انجام شده است به نظر می‌رسد یکی از راه‌حل‌های موجود برای حل این مسائل استفاده از هستی‌شناسی در بازیابی اطلاعات است.

طبق آنچه گفته شد، اگرچه تاکنون طرح‌هایی در ایران به تبدیل اصطلاح‌نامه‌ها به هستی‌شناسی‌ها پرداخته‌اند (حسینی بهشتی، ۱۳۹۱؛ خسروی و وظیفه دوست، ۲۰۰۷ و فتحیان، ۱۳۹۰)، اما تاکنون این هستی‌شناسی‌ها به‌ندرت در سیستم بازیابی اطلاعات و برای رفع مشکلات مربوط به بازیابی اطلاعات، به کار گرفته شده‌اند و افزون بر آن مدلی که بتواند متناسب با نیازهای کاربر و براساس الگوها و رفتارهای واقعی کاربران پیش برود و به نوعی کاربرمداری را برای بازیابی اطلاعات به کار گیرد و مشکلات مربوط به بازیابی آنها را حل کند، تاکنون ارائه نشده است. این پژوهش در پی آن است که با روش پژوهش ترکیبی دریابد چگونه هستی‌شناسی‌ها و روابط معنایی آنها در سیستم‌های بازیابی اطلاعات و با بررسی رفتار اصلاح پرس و جوی کاربر و استفاده آنها از روابط معنایی در جستجوی خود، می‌توانند به بازیابی بهتر و کاربرمدارتر اطلاعات کمک کنند.

۳- ضرورت و اهمیت پژوهش

یکی از مهم‌ترین اهداف هر کتابخانه و مرکز اطلاع‌رسانی، برطرف کردن نیازهای جامعه استفاده‌کننده است. یکی از ابزارهایی که کتابخانه را به این مهم نزدیک می‌کند استفاده از یک نظام بازیابی کارآمد است. نظام‌های بازیابی اطلاعات به بازنمون، ذخیره‌سازی و بازیابی اطلاعات و دانش مرتبط با نیازها و سؤال‌های خاص کاربران می‌پردازند. کاربر در قالب یک پرسش، با واژه‌هایی از زبان طبیعی نیاز خود را مطرح می‌کند. بازیابی اطلاعات، فرآیندی دارای ابهام است. این ابهام به مبهم بودن عناصر دخیل در آن یعنی زبان طبیعی، نیاز اطلاعاتی، ربط و مانند این‌ها مرتبط است. (کیانی، ۱۳۸۴)