

معنی‌شناسی واژگانی فرااصطلاحنامه و بازیابی اطلاعات

• ملوک‌السادات حسینی بهشتی^۱

گروه پژوهشی اصطلاح‌شناسی پژوهشگاه اطلاعات و مدارک علمی ایران

چکیده

این مقاله به نقش معناشناسی واژگانی در بازیابی اطلاعات می‌پردازد. ابتدا به بحثی روش‌شناختی در زمینه معناشناسی واژگانی و پیشرفت‌های حاصل در دهه‌های اخیر پرداخته و سپس کاربرد شیوه‌های اصطلاح‌شناختی و ساختار مفهومی در گنجینه واژگان (علوم پایه: ریاضی، فیزیک، شیمی، زیست‌شناسی و زمین‌شناسی، فنی - مهندسی و کشاورزی) را به‌عنوان زیرساختار ذخیره و بازیابی اطلاعات علمی - فنی تجزیه و تحلیل می‌کند. معناشناسی واژگانی مطالعه معانی واژگان و روابط میان آنهاست. درحالی‌که بسیاری از آثار تحقیقاتی در این زمینه، در مکاتب زبان‌شناسی تدوین شده و سبب ایجاد پیکره زبانی و مجموعه واژگانی قابل دسترس شده است، توسعه برخی از روش‌ها و شیوه‌های اجرایی آن با مسائل کاربردی زبان‌شناسی رایانه‌ای مرتبط است.

کلیدواژه: فرااصطلاحنامه‌ها، بازیابی اطلاعات، معناشناسی واژگان

مقدمه

به‌طور خلاصه این روابط عبارت‌اند از: تکرار یک واژه واحد (همان رابطه‌ای که در شیوه‌های سنتی بازیابی اطلاعات به کار می‌رود)، کاربرد مترادف یک واژه (مانند دیابت یا بیماری قند) و کاربرد روابط سلسله‌مراتبی واژه‌ها (مانند وسیله نقلیه و ماشین یا اتومبیل) و سرانجام عبارات ترکیبی (مانند گیاه‌پزشکی، شکم پایان، دماسیاب، پنجه‌گریبان).

شیوه دیگر، ارائه «زنجیره‌ای از واژگان»^۲ فرآیندی از استخراج کلیدواژه‌های مرتبط از یک متن است. به‌طور کلی یک متن ممکن است زنجیره‌های واژگانی متعددی را در برگیرد که هر یک بخشی از ساختار معنایی متن را توصیف می‌کند. این زنجیره‌های واژگانی به‌واسطه کاربرد هر منبع واژگانی که کلیدواژه‌ها را از نظر مفهومی به هم مرتبط می‌سازند، قابل شکل‌گیری است. کاربرد اولیه آن در اصطلاحنامه راجت موریس^۳ (۱۹۹۱) و نیز اکثر زنجیره‌های واژگانی متداول در آثار چاپمن (۱۹۹۲)، استرماند (۱۹۹۷)، هیرست (۱۹۹۸)، برازلی (۱۹۹۷)، گرین (۱۹۹۹)، الحیمی (۱۹۹۸) و نیز کاربرد آن در شبکه واژگانی بر اساس یافته‌های بکویث (۱۹۹۱) و فل باوم (۱۹۹۸) دیده می‌شود.

«شبکه واژگانی»^۴ ترکیبی است از مجموعه مترادف‌ها که هر یک شامل یک یا چند کلمه دارای معانی مشترک می‌باشند. بدین ترتیب ممکن است یک واژه در چند مجموعه مترادف به‌تناسب تعداد مفاهیمی که گویای آن هستند، ظاهر شود. مجموعه مترادفات پیوندهای معنایی گوناگون را در برمی‌گیرد که نشان‌دهنده روابط معنایی گوناگون هستند. برای مثال دو مجموعه مترادف می‌توانند با یک اتصال فرامتنی به یکدیگر متصل شوند که نشان‌دهنده آن است

قبل از هر چیز باید به این نکته توجه داشته باشیم که بسیاری از تحقیقات در زمینه بازیابی را می‌توان آثاری در زمینه معناشناسی واژگانی تلقی کرد. نظام‌های بازیابی رایانه‌ای اطلاعات با یافتن مشابهت زمینه موضوعی و با تعیین واژگان مشترک بین اسناد علمی و شناسایی کلیدواژه‌های موجود در این اسناد، رابطه بین آنها را تعیین می‌کنند. به‌کارگیری چنین نظام بازیابی مبتنی بر معناشناسی واژگانی در شکل‌گیری ساختار فرامتنی موفق خواهد بود. اگرچه چنین رویکردی تاکنون موفق بوده است، دو مشکل عمده در ارتباطات معنایی واژگان وجود دارد:

۱. «چند معنایی»^۵: وقتی یک کلمه واحد دارای معانی متعدد می‌باشد.
۲. «ترادف»^۶: وقتی کلمات متفاوت دارای یک معنای واحد می‌باشند.

هر دو پدیده مذکور، روابط شناسایی متداول در نظام بازیابی اطلاعات را مختل می‌سازد، زیرا گاهی استنباط می‌شود که نوعی رابطه میان دو واژه وجود دارد، در صورتی‌که واقعیت چنین نیست (چندمعنایی) و گاهی هیچ رابطه‌ای میان واژگان نیست، درحالی‌که واقعاً ارتباط وجود دارد (ترادف).

مرور بر نظریات متعدد در این زمینه نشان می‌دهد که پیوند و اتصال واژگان در یک متن که از نظر هالیدی و راکوایا (۱۹۷۶): «ارجاع به روابط معنایی موجود در یک متن تمایزدهنده (و بازنمودی) از متن است» که این مسئله نشان‌دهنده آن است که پیوند معنایی واژگان به‌واسطه کاربرد روابط معنایی واژگان ایجاد شده است.



روابط منطقی میان کلمات با در نظر گرفتن تعداد معینی از روابط معنایی میان مجموعه مترادفها در نقشه معنایی واژگانی شبکه واژگان محاسبه می‌شود

می‌شود. روابط منطقی میان کلمات با در نظر گرفتن تعداد معینی از روابط معنایی میان مجموعه مترادفها در نقشه معنایی واژگانی شبکه واژگان محاسبه می‌شود. در این زنجیره‌ها، به محض آنکه کلماتی از یک متن خوانش شوند، به مجموعه واژگانی که روابط مفهومی قوی‌تر دارند، اضافه می‌شوند. نقطه ضعفی که در این نوع زنجیره‌های واژگانی وجود دارد، آن است که به محض آنکه تصمیم بر اضافه کردن یک کلمه به زنجیره واژگانی گرفته شود، هرگونه رابطه معنایی میان واژگان از بین می‌رود.

الحلیمی و کارمان (۱۹۹۸) رویکرد متفاوتی را پیشنهاد می‌کنند که تشکیل نوعی «درخت واژگان» است. یک درخت واژگانی روابط موجود میان واژگان در جریان فرآیند ساخت زنجیره واژگانی را حفظ می‌کند. در بسیاری از موارد در این رویکرد تصمیمات درباره روابط معنایی واژه، مشابه آن چیزی است که در روش «هیرست» به کار رفته است. در فرآیند نمایه‌سازی مقالات پژوهشی و دیگر اسناد علمی این روابط مفهومی میان واژه‌ها به کار گرفته می‌شوند. تمام این رویکرد به دنبال سرعت بخشیدن به تصمیم‌گیری درباره تعلق هر واژه به زنجیره معنایی واژگان مربوط است.

برازیلی و الحداد (۱۹۹۷) رویکرد کاملاً متفاوتی را انتخاب کردند. زنجیره واژگانی آنها تمام تعبیرهای معنایی ممکن را تا شکل‌گیری تمام واژگان حفظ کردند. البته نگهداری تمام تعبیر گوناگون روابط معنایی، نیازمند ذخیره میزان زیادی اطلاعات می‌باشد. برازیلی و همکارانش این مشکل را بدین صورت حل کردند که ابتدا متن را تقسیم‌بندی (تقطیع) کرده، الگوریتم تشکیل زنجیره‌ها را در هر بخش ایجاد کرده و سپس میان زنجیره‌های مفهومی واژگانی بخش‌های

که واژگان مجموعه اصلی شاهدهی است بر واژگان مجموعه مترادف مقصد. این روابط همبستگی معنایی واژگانی مورد نظر هالیدی و راکیویا (۱۹۷۶) را به خوبی ترسیم می‌کند.

موريس نشان می‌دهد که زنجیره واژگانی بازیابی شده از متن، به انعکاس ساختار گفتمان متن گرایش دارد. بدین ترتیب که نقشه معنایی موضوعات و موضوعات فرعی مطرح در یک سند علمی، مشابه نقشه معنایی میزان وقوع عناصر مختلف در زنجیره‌های واژگانی یک سند رسمی است. همین ویژگی است که زنجیره‌های واژگانی را قابل استفاده در ساختار اتصالات فرامتنی در یک متن و بین متون می‌کند (موريس، ۱۹۹۱).

رویکردهای مختلف در کاربرد زنجیره‌های واژگانی

تاکنون الگوریتم‌های قابل اجرا برای استخراج زنجیره واژگان یک متن طراحی شده‌اند. استرماند الگوریتم ساده‌ای برای ساختار مجموعه زنجیره‌های واژگانی طراحی کرد. در این طراحی ابتدا تمام واژه‌های یک متن گردآوری می‌شوند. سپس برای هر واژه در یک متن، مجموعه‌ای از اصطلاحات نزدیک به واژه مذکور در «شبکه واژگانی» ایجاد می‌شوند و در نهایت زنجیره‌های واژگانی به‌واسطه یافتن اتصالات میان این مجموعه‌های اصطلاحات تعمیم یافته، ایجاد می‌شود (استرماند، ۱۹۹۴).

زنجیره‌های واژگانی دیگر مانند آنچه هیرست (۱۹۹۸) و گرین (۱۹۹۹) توصیف کرده‌اند، وجود دارد که روابط مفهومی واژه‌ها از نظر میزان ارزش اطلاعاتی آنها را محاسبه کرده‌اند. برای مثال تکرار و بسامد یک واژه به‌عنوان یک رابطه قوی‌تر از کاربرد مترادفها قلمداد

«شبکه واژگانی» ترکیبی است از مجموعه مترادف‌ها که هر یک شامل یک یا چند کلمه دارای معانی مشترک می‌باشند

واژگانی می‌شوند، برای مثال روابط محدودی میان اسم و فعل به صورت سلسله مراتبی وجود دارد.

۲. مهم‌تر آنکه واژگان موجود در شبکه واژگانی تثبیت شده و برخی از طبقات مهم اسم مانند اسامی خاص به کلی نادیده گرفته می‌شوند.

راه حل مشکل اول همان‌گونه که هیرست (۱۹۹۸) پیشنهاد کرده، کاربرد اصطلاحنامه سنتی است، زیرا ساختار اصطلاحنامه چنین روابطی را امکان‌پذیر می‌سازد. چنین اصطلاحنامه‌هایی در حال حاضر به شکل الکترونیکی موجود و در دسترس است. برای مشکل دوم نیازمند «منابع اطلاعاتی» علاوه بر «اطلاعات واژگانی» موجود هستیم. در آینده نیز باید درباره جمع‌آوری مجموعه‌ای از دانش واژگانی (شامل شبکه واژگانی، اصطلاحنامه‌ها و واژه‌نامه‌ها) و کاربرد مشترک میان آنها در تشکیل زنجیره‌های واژگانی و ایجاد ارتباط رایانه‌ای و خودکار، تحقیقات دقیق‌تری صورت پذیرد.

هستی‌شناسی^۱ و شبکه واژگانی

به نظر می‌رسد که ایجاد پایگاه دانش واژگانی مانند شبکه واژگانی می‌تواند ابزار کلیدی برای شکل‌گیری «هستی‌شناسی» محتوای یک متن علمی به‌ویژه در خصوص پایگاه اطلاعات علمی - فنی باشد. کاربرد نقشه معنایی بر اساس شکل‌گیری سلسله مراتبی واژگان تدوین شده است. هستی‌شناسی اغلب مفاهیمی را دربرمی‌گیرد که برای شناخت انسان از دنیای پیرامون، بنیادی و اساسی می‌باشند و دانش پیش نمونه‌ها را بازنمایی می‌کنند. اینکه ساختار هستی‌شناسی بر اساس اعم و اخص یا سطوح بالا و پایین طراحی می‌شوند، بیشتر به اهداف و حوزه مورد نظر بستگی دارد تا حل مشکلات مدیریت واژگان، زیرا تفاوت معنادار و واقعی میان این دو نوع هستی‌شناسی وجود دارد. هستی‌شناسی بر اساس سلسله مراتب و ویژگی‌های معنایی بر آن است تا حوزه معنایی خاصی را که دقیقاً مبتنی بر مفاهیم تعریف شده است، به دست آورد. برای مثال حوزه علوم پایه (ریاضی، فیزیک، شیمی، زیست‌شناسی و زمین‌شناسی) به راحتی فرمول‌بندی می‌شوند. زیرا در همه این حوزه‌ها نظام مفاهیم حاکم است که می‌توان آنها را با اصطلاحات علمی و تعریف دقیق بیان کرد. در چنین مواردی سعی می‌شود اشیاء در عالم مقال و نیز به‌طور محض و انتزاعی و با الگوهای آرمانی و ساده شده‌ای از پدیده‌های واقعی در جهان ارائه شود. برای مثال: «مثلاً چیزی نیست جز یک چند ضلعی دارای سه زاویه».

«پایگاه‌های اطلاعات واژگانی»^۱ همانند شبکه واژگان، شامل واژگان، اصطلاحنامه‌ها و فرهنگ‌های لغت است که به قاعده‌مند

مختلف اتصال برقرار نمودند. بدین ترتیب روشن است که کاربرد این رویکرد زمان بیشتری را به خود اختصاص می‌دهد.

یکی از مهم‌ترین ویژگی‌های کاربردی ایجاد زنجیره واژگانی این است که رفع ابهام مفهومی (معنایی) به‌عنوان تأثیر جانبی این فرآیند فراهم شود، زیرا هنگام افزودن واژه‌ها به زنجیره واژگانی، مجموعه مترادف‌های مشابه با اطلاعات واژگانی موجود که در زنجیره، تفاوت فاحش دارند، قابل پیوند نیستند. عناصر یک زنجیره واژگانی معرف بافت معنایی است که یک واژه در آن به کار رفته و بدین ترتیب معانی واژه‌ای که در این بافت غیرممکن است، مورد توجه قرار نمی‌گیرد.

ساختار فرامتنی خودکار

نگرش‌های متفاوتی در کاربرد زنجیره‌های واژگانی برای «تقسیم‌بندی» (تقطیع) متن وجود دارد. بسیاری از رویکردها تقسیم‌بندی سند علمی مبتنی بر شکل‌گیری زنجیره واژگانی را مستلزم کشف مرزبندی موضوعی و تحدید هر زنجیره واژگانی می‌دانند. نقطه‌ای که زنجیره‌ای پایان می‌یابد و زنجیره دیگری آغاز می‌شود، حد مناسبی برای مرزبندی موضوعی است. تصمیم‌گیری در تعیین مرزهای موضوعی بسیار مشابه تصمیم‌گیری درباره جایگزین کردن اتصال فرامتنی میان دو بخش از یک متن است. اساساً ما نیازمند تعیین میزان مشارکت هر زنجیره واژگانی در هر بخش از متن هستیم. بخش‌هایی که به زنجیره‌های گوناگون متصل می‌شوند و بیشترین وجه اشتراک را دارند، می‌توانند به‌عنوان نمونه‌های مناسب اتصال فرامتنی در نظر گرفته شوند. گرین (۱۹۹۹) آن را با محاسبه تراکم هر زنجیره واژگانی در هر بند متن فرمول‌بندی کرد. بندهایی که در آنها تعداد معنایی تراکم بالای زنجیره واژگانی مشارکت دارند، با یکدیگر متصل می‌شوند.

تشکیل اتصال (پیوند) میان متون

تلاش در ایجاد اتصال میان متون را می‌توان تلاش در کشف روابط معنایی میان متون تلقی کرد. البته نمی‌توان در درون یک متن روابطی به میزان روابط معنایی متعدد میان متون مختلف یافت. اما متونی که درباره موضوعات مشابه می‌باشند، زنجیره واژگانی مشابه را نیز دربرمی‌گیرند. یک روش ساده برای سنجش میزان پیوستگی معنایی، کاربرد ساده زنجیره‌های واژگانی برای ایجاد روابط معنایی بین دو متن است. متأسفانه این رویکرد بسیار کند و زمان‌بر می‌باشد و در رسیدن به نتایج موفق نبوده است، زیرا بین دو متن در مواردی روابط معنایی تصنعی ایجاد می‌کند. امکان دیگر، ارائه یک متن به‌واسطه مجموعه مترادف‌هاست که زنجیره‌های آن حاوی واژگان مشترک است و مشابهت آنها را در یک مجموعه قابل محاسبه می‌کند؛ بدین شکل که میزان مجموعه مترادف مشترک در اسناد علمی موجود را مشخص می‌نماید. از این‌رو نه تنها محتوای یک سند علمی ارائه می‌شود، بلکه ارتباط مفهومی آن نیز تعیین می‌شود. اساساً هر نوع زنجیره واژگانی که به‌منظور کاربرد در بازیابی اطلاعات تشکیل می‌شود، می‌تواند در شکل‌گیری روابط میان متون نیز به کار گرفته شود. محققان برخی از مشکلات کاربرد شبکه واژگانی را مورد مطالعه قرار داده‌اند:

۱. بسیاری از روش‌ها محدود به سلسله مراتب اسامی در شبکه

کردن معناشناسی واژگانی یعنی معنای واژه‌های یک یا چند زبان طبیعی می‌پردازد. این معانی بر اساس شبکه مفاهیم و مبتنی بر فهم مشترک اکثریت جامعه علمی و دارای تعاریف خاص و شامل معانی مختلف متداول و کاربردهای خاص و ویژه و ترکیبات گوناگون واژه‌ها مبتنی بر پیش نمونه‌ها و نیز مفاهیم پایه علوم است.

وب معناشناختی^{۱۱}

دیویس و همکاران (۲۰۰۷) نگاه جدیدی به آینده «شبکه جهانی اطلاعات» دارند که در آن تمام دانش موجود به صورت قاعده‌مند رمزگذاری می‌شوند تا عوامل هوشمند بتوانند به طریقی خودگردان (خودسامان) به همه انسان‌ها خدمات ارائه دهند. این کار مستلزم تعریف رسمی و قاعده‌مند هستان‌شناسی‌های مبتنی بر «وب» است تا بیان‌کننده دانش قابل فهم انسان و نیز عوامل هوشمند باشد و ارزش افزوده دانش و توسعه آن به صورت «فرا داده‌ها^{۱۲}» خواهد بود که در حال حاضر تعریف شده و قابل کاربرد در مجموعه‌های محدودی از تقاضاهاست. پیش‌بینی می‌شود که در دهه‌های آینده ساختار دانش مورد درخواست‌های بیشتر به صورت رسمی و به شکل نظام‌های هستان‌شناسی مبتنی بر «وب» رمزگذاری خواهد شد. اکنون تعریف و تدوین دانش مبتنی بر وب و ارائه آن به زبان‌های گوناگون، زمینه فعالی برای تحقیقات و مطالعات بیشتر است که حاصل آن تعداد قابل توجهی طرح‌های تحقیقاتی و تصویب استانداردهاست که در قالب «آر.دی.اف.^{۱۳}» و «اکس.ام.ال / آر.دی.اف.» و نقشه‌های معنایی به‌عنوان زبان‌های ارائه دانش به کار می‌روند.

جوامع مختلف علمی علاقه‌مندند که مفاهیم اصلی فعالیت‌های خود را تعریف و تثبیت کنند. علاوه بر این روابط واژگانی و ارتباطات معنایی که به صورت اصطلاحات استانداردسازی شده‌اند، نشان‌دهنده تقید این جوامع در کاربرد مفاهیم و زبان مشترک در سازمان‌ها و تبادل اطلاعات با عوامل هوشمند در وب معناشناختی است.

محیط تدوین اصطلاحنامه

تدوین اصطلاحنامه جامع، مصوب کمیسیون اطلاع‌رسانی شورای پژوهش‌های علمی کشور، بر مبنای ساختار واژگانی و روابط مفهومی صورت گرفته و تاکنون مراحل مختلفی را طی کرده است. در ۱۳۷۶ تدوین و طراحی ساختار سلسله مراتبی واژگان بر اساس روابط منطقی در حوزه‌های ریاضی، فیزیک، شیمی، زمین‌شناسی، زیست‌شناسی، مهندسی و کشاورزی به عهده گروه اصطلاح‌شناسی پژوهشگاه اطلاعات و مدارک علمی ایران گذاشته شد. ابتدا بر اساس تقسیم‌بندی

رشته‌ای و زیرمجموعه‌های آن، گروه‌های کاری در هر یک از رشته‌ها تشکیل شد و از فارغ‌التحصیلان رشته‌های مزبور دعوت به همکاری شد. از این‌رو برای هر رشته مدیر متخصص موضوعی تعیین شد و تدوین هر یک از اصطلاحنامه‌های تخصصی به صورت زیرمجموعه‌ای از طرح اصطلاحنامه جامع تعریف گردید و شروع به کار کرد. بودجه طرح نیز به تناسب فاز مطالعاتی و تدوین هر یک از اصطلاحنامه‌ها تقسیم‌بندی شد. شورای نظارتی اصطلاحنامه اعم از سرگروه‌های متخصصان موضوعی، زبان‌شناسان و اطلاع‌رسانان و مهندسان کامپیوتر تشکیل شد. ابتدا دوره آموزشی مبنای اصطلاح‌شناسی و واژه‌گزینی برای همکاران طرح برگزار شد و طی جلسات متعدد چگونگی تدوین اصطلاحنامه بحث و بررسی شد. پس از آن با همکاری مهندسان رایانه و متخصصان موضوعی، نقشه‌های معنایی و ترسیم نمودارهای خورشیدی زیرمجموعه‌های هر یک از رشته‌ها آغاز شد.

اولین مرحله، تعیین حوزه‌های مطالعاتی هر رشته با روندی از کل به جزء (قیاسی) انجام شد. در بررسی‌های اولیه مشخص گردید که تقسیم‌بندی‌های موجود در هر زیرمجموعه، در منابع مختلف مجازی و قراردادی است. زیرا حوزه‌های مطالعاتی مختلف هر رشته با یکدیگر همپوشانی زیادی دارند و بر اساس هدف خاصی که دنبال می‌کنند، تقسیم‌بندی‌های متفاوتی را ارائه کرده‌اند. بدین ترتیب با توجه به استانداردهای جهانی و بر اساس کتاب‌های معتبر درسی دانشگاهی، آرایه‌بندی‌ها (تاکسونومی) و نیز مجموعه واژگان مرکز نشر دانشگاهی، فرهنگستان زبان و ادب فارسی و نیز مجموعه واژگان بانک اطلاعات جامع مرکز و بر اساس قواعد حاکم بر روابط معنایی واژگان (روش استقرایی) ساختار هر یک از اصطلاحنامه‌های تخصصی شکل گرفت. در گزینش برابر نهاده‌های فارسی توصیفگرها از اولویت‌بندی قراردادی پیروی کرده و معادل‌های مرکز نشر دانشگاهی و فرهنگستان زبان و ادب فارسی در اولویت قرار گرفته، مگر آنکه با نظر متخصص موضوعی از نظر تداول عام بسیار محجور تشخیص داده می‌شد که در این صورت در جایگاه اصطلاح غیرمرجع قرار می‌گرفت. بنابراین اولویت‌بندی بر اصل اساسی گزینش توصیفگرهای موضوعی یعنی «تداول عام در بیان متخصصان موضوعی» صورت گرفت. زیرا هدف اصلی اصطلاحنامه تسهیل بازیابی اطلاعات علمی - فنی از طریق گزینش اصطلاحات مصطلح بود.

چگونگی کاربرد صورت جمع یا مفرد واژه نیز از آنجاکه نظر اغلب متخصصان امر اطلاع‌رسانی بر استعمال شکل مفرد اسامی در نمایه‌سازی و تدوین اصطلاحنامه‌های فارسی است، ما نیز سعی کردیم از این اصل پیروی کنیم، مگر در مواردی که با نظر متخصصان موضوعی ناگزیر از این امر بودیم. از جمله اسامی در آرایه‌بندی (تاکسونومی) گیاهی و جانوری و نیز شاخه و تیره که بر اساس معیارهای جهانی به شکل جمع انتخاب شده‌اند. برای مثال در اصطلاحنامه زیست‌شناسی، تیره «گل سرخیان^{۱۴}»، اگر به صورت مفرد «گل سرخ^{۱۵}» نوشته شود، دیگر نشان‌دهنده تیره نیست، بلکه اسم جنسی از آن تیره خواهد بود.

در مواردی از جمله اسامی ترکیبات شیمیایی و بعضی از اسامی

اساساً هر نوع زنجیره واژگانی که به منظور کاربرد در بازیابی اطلاعات تشکیل می‌شود، می‌تواند در شکل‌گیری روابط میان متون نیز به کار گرفته شود



ایجاد پایگاه دانش
واژگانی مانند
شبکه واژگانی
می تواند ابزار کلیدی
برای شکل گیری
«هستی شناسی»
محتوای یک متن
علمی باشد

روابط میان اصطلاحات عبارت‌اند از:
۱. رابطه ترجیحی. این رابطه ترجیح داشتن اصطلاحی را بر اصطلاح دیگر بیان می‌کند.
۲. رابطه سلسله مراتبی. این رابطه بیان‌کننده رابطه کلی و جزئی بین مفاهیم است که در واقع اصطلاحنامه‌ها را از واژه‌نامه‌های تخصصی و فرهنگ‌های لغت متمایز می‌کند. روابط سلسله مراتبی در این اصطلاحنامه‌ها تا سه سطح بالاتر/ پایین‌تر نشان داده شده است.
۳. رابطه همبستگی یا وابستگی. واژه‌نامه‌هایی که از نظر مفهوم به هم نزدیک‌ترند، ولی رابطه کلی و جزئی با هم ندارند، واژه‌های وابسته شمرده می‌شوند. این رابطه زمانی استفاده می‌شود که احتمال می‌رود به این نوع روابط در بازیابی اطلاعات نیازمند باشیم. در حقیقت ممکن است اصطلاحی که رابطه وابستگی با اصطلاح انتخاب شده را دارد، برای کاربر قابل توجه باشد و احتمال توفیق کاربر اطلاعاتی در دسترسی جامع‌تر به اطلاعات را فراهم سازد.
یادداشت دامنه^{۱۶}. چنانچه معنا یا کاربرد یک اصطلاح به‌طور کامل از خود اصطلاح یا از روابط آن استنباط نشود، «یادداشت دامنه» معنا و کاربرد دقیق آن را معلوم می‌کند. البته در اکثر موارد نیز به‌منظور بالا بردن سطح دانش و اهداف آموزشی از یادداشت دامنه استفاده می‌شود.

بخش دوم هر اصطلاحنامه «ساختار درختی» می‌باشد که در آن اصطلاحات در مجموعه‌ها و گروه‌های هم‌موضوع دسته‌بندی شده‌اند. این بخش تمام روابط سلسله مراتبی را از عام‌ترین به خاص‌ترین مفهوم در یک گروه اصطلاحات نمایش می‌دهد.
بخش سوم هر اصطلاحنامه شامل «نمایه‌گردان» است. این

گیاهی و جانوری یا در اصطلاحنامه زمین‌شناسی، اسامی جغرافیایی، حرف‌نویسی شده است. در این بخش شیوه‌نامه ضبط اعلام انگلیسی در فارسی نوشته‌ماندانا صدیق بهزادی و قواعد حرف‌نویسی آن به کار بسته شده و در مواردی بر اساس نظر متخصصان موضوعی یا کتاب‌های تخصصی موجود شکل مصطلح ترجیح داده شده است.
اسامی خاص در آرایه‌های سطوح بالاتر (سلسله‌ها، شاخه‌ها، رده‌ها و تیره‌های گیاهی و جانوری) با حروف بزرگ انگلیسی و در سایر موارد با حروف کوچک مشخص شده‌اند.

یکی از بزرگ‌ترین نقاط قوت در این طرح، وجود نرم‌افزار توانمند و در عین حال از نظر کاربرد ساده «قاموس» بوده است که از سوی متخصص نرم‌افزار طراحی شده است. در این نرم‌افزار هر سه نوع رابطه سلسله مراتبی^{۱۶}، وابستگی^{۱۷} و ترجیحی دو طرفه می‌باشند. اگرچه ابتدا در نرم‌افزار مذکور، ورود اطلاعات بر اساس روابط یک طرفه صورت می‌گیرد، در نهایت این برنامه در خروجی به‌صورت منطقی و صحیح همه ارتباطات معنایی را به‌صورت دو طرفه در ساختار الفبایی اصطلاحنامه به نمایش می‌گذارد. از دیگر خروجی‌های آن می‌توان به ساختار درختی، نمایه‌گردان^{۱۸} و واژه‌نامه اشاره کرد. در بخش نمایه‌الفبایی اصطلاحات، زیر هر اصطلاح مجموعه اطلاعات مربوط به آن قرار گرفته است. این اصطلاحات دو گونه‌اند:

۱. توصیفگرها: اصطلاحاتی که در ساختار اصطلاحنامه به کار رفته و برای نمایه‌سازی مورد استفاده قرار می‌گیرند.
۲. غیر توصیفگرها: اصطلاحاتی که در نمایه‌سازی از آنها استفاده نمی‌شود، بلکه با ارجاع به توصیفگر مناسب کاربران را راهنمایی می‌کنند.

نمای درختی اضافه کند، باید محل صحیح قرار گرفتن آن را در درخت واژگان تعیین نماید. با انتخاب هر واژه دیگر مشخصات آن در صفحه مشخصات واژه ظاهر خواهد شد.

۸. هر واژه اصطلاحنامه می‌تواند حداکثر سه ترجمه داشته باشد. برای هر یک از این فیله‌ها در هر اصطلاحنامه می‌توان هر زبانی را انتخاب کرد، مشروط بر اینکه از طریق سیستم عامل نصب شده بر کامپیوتر کاربر پشتیبانی شود. با استفاده از علائم اختیاری، این فیله شامل دسته‌بندی‌های گوناگونی از نظر معنایی یا موضوعی می‌باشد.

۹. واژه‌های غیر مرجح. این زیرصفحه شامل فیله «به جای» است. این فیله مترادف‌ها یا واژه‌های مرتبط با واژه مرجح فعلی را دربرمی‌گیرد که در نمایه‌سازی استفاده نمی‌شوند. از پنجره واژه‌های غیرمرجح برای اضافه کردن یا تغییر اطلاعات فرعی واژه‌های غیرمرجح که از روابط معنایی اصطلاحنامه دریافت شده‌اند، استفاده می‌گردد. واژه‌های غیرمرجح به ترتیب الفبایی و بر اساس استاندارد «یونی‌کد^{۲۰}» مرتب می‌شوند. هر واژه غیرمرجح نیز می‌تواند حداکثر سه ترجمه داشته باشد. برای هر یک از این فیله‌ها در هر اصطلاحنامه می‌توان هر زبانی را انتخاب کرد، مشروط بر اینکه از طریق سیستم عامل کامپیوتر کاربر پشتیبانی باشد. برای واژه‌های غیرمرجح عنوان وابسته وجود ندارد. این فیله نیز شامل دسته‌بندی‌های منظم عددی و غیر عددی است که اطلاعات واژگانی اصطلاحنامه را از نظر روابط معنایی یا موضوعی به دسته‌های گوناگونی تقسیم می‌کند.

۱۰. «فهرست بازدارنده^{۲۱}». در پنجره فهرست بازدارنده، فهرستی از کلمات که می‌توان آنها را از نمایه‌گردان حذف کرد، آمده است.

۱۱. نمایش چاپی. در پنجره مربوط به این فیله، گزینه‌های بسیاری برای تعیین شکل خروجی گزارش‌های چاپی «آر.تی.اف» قرار داده شده است که از طریق فهرست اختیارات می‌تواند کاربر به پیش‌فرض‌های چاپی دسترسی پیدا کند.

فرااصطلاحنامه^{۲۲}

اکنون که به لطف الهی و تلاش همکاران سخت‌کوش اصطلاحنامه جامع، به اهداف نهایی این طرح که همانا تدوین اصطلاحنامه‌های علوم پایه، مهندسی و کشاورزی بوده است، دست یافته‌ایم، برنامه‌ریزی اصطلاح‌شناسی و نرم‌افزاری تلفیق این اصطلاحنامه‌ها و شکل‌گیری «فرااصطلاحنامه» را پیش روی داریم که می‌تواند در بازیابی اطلاعات هماهنگی و یکدستی روابط معنایی و واژگان میان علوم مرتبط را فراهم سازد و از ریزش کاذب اطلاعات، ممانعت کند. فرااصطلاحنامه نوعی اصطلاحنامه است که باعث ادغام و تلفیق اصطلاحنامه‌های متعدد می‌شود و میان مجموعه واژگان‌های ملی و بین‌المللی ارجاعات متقابل ایجاد می‌کند. به منظور بازیابی در پایگاه‌های متعدد و چندگانه که به‌واسطه واژگان‌های کنترل‌شده متعدد نمایه شده‌اند، نوعی فرااصطلاحنامه ایجاد می‌شود. هر اصطلاح در فرااصطلاحنامه با برابر نهاده‌ها و اصطلاحات مرتبط از تمام اصطلاحنامه‌های تلفیق شده در آن به نمایش درمی‌آیند. فرآیند خودکار و رایانه‌ای در جست‌وجو و بازیابی توسعه‌یافته کاربر را قادر می‌سازد تا بازیابی را در نمایه‌های منفرد یا چندگانه و تلفیق‌شده با

پیش‌بینی می‌شود که در دهه‌های آینده ساختار دانش مورد درخواست‌های بیشتر به صورت رسمی و به شکل نظام‌های هستان‌شناسی مبتنی بر «وب» رمزگذاری خواهد شد

بخش به‌منظور دستیابی به واژه‌های دوم، سوم، و نظایر آن یا هر یک از واژه‌هایی که در ترکیب آن منظور شده است، به صورت افقی گردش می‌کند. فایده دیگر این فهرست‌ها آن است که ترتیب الفبایی آنها امکان دارد که تصادفاً واژه‌هایی را در یکجا گرد آورد که به جنبه‌های گوناگون موضوع واحدی مربوط می‌باشند، ولی در بخش الفبایی از هم جدا افتاده‌اند.

بخش چهارم اصطلاحنامه واژه‌نامه انگلیسی به فارسی است. در بخش اول اصطلاحنامه معادل انگلیسی هر توصیفگر آمده است. در این بخش این معادل‌ها به ترتیب الفبای انگلیسی ذکر شده‌اند.

نرم‌افزار تدوین اصطلاحنامه قاموس

محیط تدوین اصطلاحنامه راه‌های مختلفی را برای دستیابی به اهداف یکسان در اختیار می‌گذارد. مهم‌ترین هدف در طراحی این محیط تدوین اصطلاحنامه به شیوه‌ای آسان است:

۱. از طریق پنجره اصلی تمام ابزار تدوین اصطلاحنامه و دیگر بخش‌های محیط تدوین اصطلاحنامه به مجرد اجرای نرم‌افزار قاموس قابل دسترسی است که از فهرست انتخاب‌های مختلف تشکیل شده است و بر اساس برنامه میکروسافت اکسس و فرمت «آر.تی.اف.» تدوین شده است.

۲. فهرست انتخاب و ویرایش اصطلاحنامه شامل دستوراتی برای اضافه، حذف یا ویرایش اطلاعات واژگانی اصطلاحنامه است.

۳. فهرست بررسی که از دستورات آن برای شناسایی اشکالات ساختاری و روابط معنایی واژگان استفاده می‌شود. این فهرست در حالتی که اطلاعات واژگانی اصطلاحنامه در دست ویرایش باشند، غیر فعال خواهد بود.

۴. فهرست جست‌وجو برای جست‌وجوی واژگانی که در مجموعه اطلاعات اصطلاحنامه در دست ویرایش باشند، غیرفعال خواهد بود.

۵. فهرست نمایش که گزینه‌های مختلف واژه‌های غیرمرجح، فهرست بازدارنده و نظایر آن را به نمایش در می‌آورد.

۶. فهرست اختیارات که شامل دستوراتی برای دستیابی به پنجره‌های مربوط به اعمال تغییرات در گزینه‌های چاپی و نمایش الکترونیکی است.

۷. صفحه ساختار درختی که در آن کلیدواژه‌های مرجح و دیگر مشخصات آنها در این صفحه یکجا جمع شده است و شامل نمای ساختار درختی و تمام عنوان‌های اخص هر واژه مرجح در شاخه‌های پایین‌دستی آن واژه و نیز واژه‌های بالادستی هر واژه که بیانگر عناوین اعم آن واژه است، می‌باشد. هنگامی که کاربر می‌خواهد واژه‌ای به

**فرااصطلاح‌نامه نوعی اصطلاح‌نامه
است که باعث ادغام و تلفیق
اصطلاح‌نامه‌های متعدد می‌شود و میان
مجموعه واژگان‌های ملی و بین‌المللی
ارجاعات متقابل ایجاد می‌کند**

مهم‌ترین چالش در ایجاد وحدت رویه و شیوه جست‌وجو در مجموعه اصطلاحات، به واژگان‌های مختلفی مربوط است که تهیه‌کنندگان به کار می‌گیرند و آن فقدان واژگان کنترل شده و استاندارد یکسان است. معمولاً هریک از تهیه‌کنندگان واژگان خاص خود را برای توصیف محتوای منابع اطلاعاتی و طبقه‌بندی آنها به کار می‌گیرند. ساختار و ساماندهی متفاوت و اهداف گوناگون نیز به واژگان‌های متفاوت در یک رشته و گرایش علمی می‌انجامد و همین سبب می‌شود که جست‌وجو در منابع از سوی تهیه‌کنندگان متعدد با مشکل مواجه شده و کیفیت قابل قبول نداشته باشد، به طوری که کاربر قادر نخواهد بود با همه مجموعه‌های واژه‌های کنترل شده آشنا شده و تفاوت آنها را بداند و در جست‌وجو و بازیابی موفق عمل کند. تدوین و توسعه یک فرااصطلاح‌نامه به ایجاد پل ارتباطی میان اصطلاح‌نامه‌های مختلف کمک می‌کند. همچنین کاربران را قادر می‌سازد از امتیازات و مزایای فرآیند خودکارسازی جست‌وجو و بازیابی اطلاعات بهره‌مند شوند. این نگرش در تولید فرااصطلاح‌نامه سبب شناسایی فهرستی از واژگان و عبارات کلیدی در تحلیل رایانه‌ای متون علمی می‌شود و به دنبال آن تحلیل معناشناختی برای شناسایی مترادفها که برای گسترش دامنه اصطلاحات در جست‌وجو و بازیابی کاربران و با هدف گسترش بسامد واژگانی در گرایش‌های علمی متعدد به کار گرفته می‌شوند. کاربرد کلیدواژه‌ها نظام بازیابی اطلاعات، تبادل اطلاعات علمی و ساماندهی ذخیره و بازیابی اطلاعات را تسهیل می‌کند.

خودکارسازی تولید کلیدواژه‌ها و ایجاد گنجواژه‌ها در گرایش‌های علمی جدید که قبلاً برای آنها اصطلاح‌نامه‌ای وجود نداشته و ایجاد وحدت میان مجموعه فراداده‌ها، به کارگیری ساخت واژه رایانه‌ای را می‌طلبد و بر اساس آن الگوریتم خاص برای گزینش خودکار کلیدواژه‌ها ایجاد می‌شود. بدین منظور کوچک‌ترین مجموعه کلیدواژه‌ها را برای ایجاد دقت بیشتر و سنجش میانگین تراکم در ساخت خوشه‌ها و یافتن کلیدواژه‌ها یا خوشه‌های واژگانی بهینه با کمترین پراکندگی و نیز کاهش تعداد خوشه‌های واژگانی می‌یابیم و در چارچوب این طرح اولویت‌های الگوهای بهینه را تعیین می‌کنیم. شایان ذکر است که مبادله اطلاعات میان رویکردهای متفاوت فواید زیادی در ایجاد نسخه نهایی خواهد داشت. یکسان‌سازی نتایج حاصل از استخراج کلیدواژه‌ها و تحلیل معنایی نهفته در آنها سبب می‌شود که ابزاری برای اصطلاحات وابسته پیشنهادی در جست‌وجو و بازیابی به دست آید تا کاربران را در بازشناسی نیازهای اطلاعاتی واقعی‌شان و شناسایی مناسب‌ترین کلیدواژه‌ها برای جست‌وجو و بازیابی یاری

استفاده از اصطلاحات هریک از منابع واژگانی انجام دهد. فرااصطلاح‌نامه مفهوم - بنیاد است و هدف آن ایجاد وحدت میان نامگذاری معانی مختلف یک مفهوم است. این نامگذاری‌ها از واژگان‌ها و منابع واژگانی متعدد و رده‌بندی‌ها و اطلاعات پیرامون آن به دست آمده است. گسترش سریع میزان لغات یک فرااصطلاح‌نامه باعث شده است که به کارگیری رایانه برای سهولت در ایجاد آن مؤثر واقع شود. روزآمدی واژگان‌ها، بهره‌مندی از مجموعه واژه‌های متعدد، الگوریتم گزینش و بازنگری متخصص موضوعی را می‌طلبد. اصطلاح‌نامه‌های مختلف را می‌توان تحت شرایط ذیل در یک اصطلاح‌نامه تلفیق کرد:

۱. از ابتدا اصطلاح‌نامه‌های مختلف الزاماً بر اساس پیش‌زمینه نظری واحدی ساختار بندی شده باشند و این امر در گام نخست برای تلفیق کاملاً ضروری است.
 ۲. در گام دوم، شیوه تخصیص شناسه‌های یکسان به کار گرفته شوند و زمینه‌های بازنگری و رفع مشکلات شناسایی شده فراهم شود.
 ۳. در گام سوم، توسعه کنترل کیفیت دیدگاه‌ها و توافقات به عمل آمده، صورت پذیرد.
 ۴. گام چهارم، شناسایی زمینه‌های جابه‌جایی و دگرگونی ظریف و نامحسوس معانی و مفاهیم در هریک از گرایش‌های علمی است. این فرآیند نشان‌دهنده آگاهی ما نسبت به تفاوت‌های نامحسوس در حفظ و نگهداری شرایط محیطی و کاربرد دیدگاه‌های فلسفی بر تهیه و تولید واژگان است.
- تفاوت دیدگاه‌ها می‌تواند نشان‌دهنده چگونگی تأثیر بر توسعه و شکل فرااصطلاح‌نامه بر روند بازیابی اطلاعات باشد و آن را برای هرچه تخصصی‌تر شدن پیش‌برد، فرااصطلاح‌نامه گردآوری و تدوین نام‌ها، روابط مفهومی و اطلاعات وابسته از یک نظام بازیابی اطلاعات و بازتابی از دیدگاه‌های گوناگون محققان و متخصصان است.
- نامگذاری‌های گوناگون و نیز معانی مختلف در یک مفهوم گردآوری شده و از طریق اطلاعات افزوده و توصیف مختصات از یکدیگر متمایز می‌شوند. فرااصطلاح‌نامه فقط یک واژگان استاندارد شده نیست، بلکه ابزاری برای افزایش نهایی واژگان موجود و مفید است و به مثابه منبعی از دانش تخصصی در خدمت توسعه‌دهندگان اطلاعات علمی و نیز منبعی غنی برای متخصصان اطلاعات علمی می‌باشد. نظارت بر تناسب اجزای فرااصطلاح‌نامه در هر ویرایش جدید سبب توسعه و افزایش میزان مفاهیم یک فرااصطلاح‌نامه می‌شود. فناوری فرااصطلاح‌نامه برای فرمول‌بندی روابط معنایی میان ساختارهای دانش فنی، که به صورت مفاهیم و روابط میان آنها بیان می‌شود، از دیدگاه‌های گوناگون به کار گرفته می‌شود و کاربران را به توری، جست‌وجو و بازیابی میان رشته‌ای و گرایش‌های گوناگون علمی قادر می‌سازد. فرااصطلاح‌نامه شامل مجموعه‌ای از مفاهیم کلیدی در رأس و روابط معنایی میان مفاهیم همانند مترادفها، همنام‌ها و غیره و نیز ترسیم نقشه معنایی از منابع گوناگون اصطلاح‌نامه‌ها و مفاهیم کلیدی است که به نسبت تعداد کمی از اینها تألیف یا ابداع شده‌اند و آنهایی که ساخته شده‌اند، معمولاً به صورت دستی کدگذاری می‌شوند

19. Scope Note
20. Unicode
21. Stop list
22. Metathesaurus
23. Thesaurus Builder

منابع و ماخذ

1. Al-Halimi, R., Kasman, R. (1998). Temporal Indexing through Lexical Chaining. In WordNet: An Electronic Lexical Database, C. Fellbaum (editor), MIT Press, Cambridge, MA, 333-351.
2. Barzilay, R.; M. Elhadad (1997). Using Lexical Chains for Text Summarization. In ACL Workshop on Intelligent Scalable Text Summarization, July 1997.
3. Beckwith, R., et al (1991). WordNet: A Lexical Database Organized on Psycholinguistic Principle. In Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon, U. Zernik (editor), Lawrence Erlbaum Associates, 211-231.
4. Chapman, R.L., (1992) (editor). Roget's International Thesaurus (5. edition), Harper Collins.
5. Davies, J., Struder, R., Warren, P. (2007). Semantic Web Technologies, John Wiley & Sons
6. Fellbaum, C. (1998) (editor). WordNet: An Electronic Database. MIT Press.
7. Green, S. J. (1999). Lexical semantics and automatic hypertext construction. [on-line]. Available: <http://www.cs.brown.edu>
8. Halliday, M.A.K.; Ruqaiya, Hasan. (1976). Cohesion in English. Longman.
9. Hirst, G.; D., St-Onge (1998). Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms. In WordNet: An electronic lexical database, C. Fellbaum (editor), Cambridge, MA: The MIT Press.
10. Morris, J.; G. Hirst. (1991). Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text. In Computational Linguistics, 17(1) 21-48.
11. Stairmand, M.A. (1994). Lexical Chains, WordNet and Information Retrieval. Condensed version of Master's Thesis.

رساند. مجموعه عبارات به کار گرفته شده در این نظام از سه منبع اصلی نشأت می‌گیرد:

کلیدواژه‌های استخراج شده از طریق الگوریتم، نتایج حاصل از جست‌وجو در متون علمی مسبوق به سابقه، سرعنوان‌های موضوعی خاص منابع اطلاعاتی و کلیدواژه‌های مورد استفاده بازیابی گذشته کاربران اطلاعاتی.

این کلیدواژه‌ها و سرعنوان‌های موضوعی را با منابع سرمنشأ آنها مرتبط می‌سازیم. پرسش‌های کاربران مرتبط با منابع اطلاعاتی بازیابی شده، مورد مطالعه قرار می‌گیرد. تحلیل معناشناختی روی این منابع به منظور یافتن مترادف‌ها و همنام‌ها میان متون انجام می‌گیرد و به کارگیری خدماتی با استفاده از این ابزار در صفحات جست‌وجو گسترش می‌یابد. چنانچه اصطلاحات وابسته بر اساس پایگاه‌های اطلاعاتی مترادف‌ها پایهریزی شود، در صفحات متون بازیابی شده به‌طور متمایز نشان داده می‌شوند.

برای شکل‌گیری فرااصطلاحنامه در حال حاضر هریک از اصطلاحنامه‌های ترجمه و تألیف شده در پژوهشگاه برای ویرایش به متخصصان و استادان هریک از رشته‌های علمی مرکز نشر دانشگاهی ارائه شده و در مرحله بعد یکسان‌سازی آنها اجرا می‌شود. از نظر نرم‌افزاری نیز پیشرفت‌های قابل توجهی به‌دست آمده و نرم‌افزار «تزاروس بیلدر^{۳۳}» که توسعه یافته «قاموس» است، برای تدوین اصطلاحنامه چندزبانه از سوی کارشناسان رایانه‌ای این پژوهشگاه طراحی و ارائه شده است. امید می‌رود در آینده‌ای نه چندان دور، نتایج حاصل از تلاش دست‌اندرکاران طرح در این زمینه، به جامعه علمی کشور گزارش شود.

پی‌نوشت‌ها

1. Beheshti@irandoc.ac.ir
2. Polysemy
3. Synonymy
4. Lexical chaining
5. Rogrt's Thesaurus,
6. Wordnet
7. Lexical tree
8. Ontology
9. Upper-lower level
10. Lexical Knowledge Bases
11. Semantic Web
12. Metadata
13. RDF, XML/RDF
14. Rosaceae
15. Rose
16. Hierarchical Relation
17. Associative relation
18. Rotated Descriptor Display