

بسمه تعالی

ساختواره، اصطلاح‌شناسی و مهندسی دانش

Morphology, Terminology and Knowledge Engineering

نوشته: ملوک‌السادات حسینی بهشتی

Beheshti@irandoc.ac.ir

گروه پژوهشی اصطلاح‌شناسی
پژوهشگاه علوم و فناوری اطلاعات

چکیده

مهندسی دانش، به منظور سازماندهی اطلاعات علمی، به برنامه‌ریزی رایانه‌ای و ایجاد شبکه‌ی اطلاعاتی می‌پردازد. ساماندهی نام‌گذاری مفاهیم علمی، یکی از مهم‌ترین مباحث علم اصطلاح‌شناسی محسوب می‌گردد که در آن "مفهوم" به عنوان واحد دانش فنی تعریف شده است. به منظور ترسیم روابط مفهومی یک رشته تخصصی بایستی نظام مفاهیم زیررشته‌های مرتبط با آن را، در یک ساختار کلی تلفیق نمود. بر این اساس، اصطلاح‌نامه‌های تخصصی، تدوین یافته و در نهایت با یکدیگر ادغام می‌گردند. فرااصطلاح‌نامه نوعی اصطلاح‌نامه کلان است که باعث تلفیق اصطلاح‌نامه‌های متعدد موجود در سطح ملی و بین‌المللی می‌گردد و در بازیابی، ارجاعات متقابل، میان پایگاه‌های اطلاعات علمی گوناگون برقرار می‌نماید. "پایگاه‌های اطلاعات واژگانی" همانند "شبکه‌ی واژگان" شامل مجموعه‌ی واژگان، اصطلاح‌نامه‌ها و فرهنگ‌های لغت می‌شود که به قاعده‌مند نمودن معنی‌شناسی واژگانی یک یا چند زبان طبیعی می‌پردازد. در این میان، "هستی‌شناسی"ها می‌توانند در ساماندهی، نقش زیربنایی را ایفا نمایند و در آینده‌ای نه چندان دور "شبکه‌ی معنایی"ای را که توسعه یافته "شبکه جهانی اطلاعات" است، ایجاد نموده تا بازنمایی مفاهیم متعلق به زبان‌ها و فرهنگ‌های گوناگون به دور از حاکمیت مطلق یک زبان و یک فرهنگ فراهم گردد. در تحقیق حاضر با استفاده از بانک اطلاعات واژگان (تلفیق واژگان علمی اصطلاح‌نامه‌های علوم پایه) پژوهشگاه اطلاعات و مدارک علمی ایران که به عنوان پیکره‌ی زبانی انتخاب گردید و با توجه به نظریه‌ی ساختوازی و معنی‌شناسی واژگانی (لیبر ۲۰۰۴)، این واژگان مورد تجزیه و تحلیل آماری قرار گرفت و علاوه بر تعیین فراوانی هر یک از عناصر واژگان مرکب، مشتق و مرکب‌مشتق، الگوهای ساختوازی اصطلاحات علمی استخراج شد. سپس با تعیین روابط معنایی آنها، زمینه‌ی طراحی هستی‌شناسی و شبکه‌ی واژگانی این حوزه علمی برای برنامه‌ریزی آموزشی و پژوهشی و در نهایت طراحی نظام مفاهیم شبکه‌ی علمی کشور فراهم گردید.

کلیدواژه‌ها: ساختواژه، معنی‌شناسی واژگانی، اصطلاح‌شناسی، مهندسی زبان، شبکه‌ی واژگانی، هستی‌شناسی، شبکه‌ی معنایی، مهندسی دانش.

مهندسی دانش^۱، یعنی فرایندی که به استخراج، قاعده‌مندی و تبیین دانش فنی لازم برای حل مسائل ذهنی انسان خبره می‌انجامد، با برنامه‌ریزی رایانه‌ای و ایجاد شبکه‌ی اطلاعات به تبادل اطلاعات علمی- فنی می‌پردازد. بدین منظور لازم است که این دانش به صورت نظام‌مند و دارای ساختار ویژه درآید و این نقطه‌ی مشترک میان اصطلاح‌شناسی و مهندسی دانش محسوب می‌شود، زیرا اصطلاح‌شناسی در سطح ارائه‌ی مفاهیم به صورت نظام‌مند عمل می‌کند و این خود مقدم بر تمامی شیوه‌های ارائه‌ی دانش است. سازماندهی دانش فنی و ساماندهی مفاهیم علمی، یکی از مهم‌ترین مباحث علم اصطلاح‌شناسی محسوب می‌گردد. "مفهوم"^۲ که به عنوان واحد دانش فنی تعریف شده، از مولفه‌های معنایی^۳، یعنی اجزاء دانش تشکیل شده است. همچنین مفهوم، یک بعد توصیفی و یک بعد تعریفی دارد که خود شکلی از دانش را ترسیم می‌نماید. بر این اساس، بعد محتوایی و بیانی هر یک از مفاهیم به عنوان دانش اصطلاح‌شناختی یک مفهوم در نظر گرفته می‌شود.

نوعی دیگر از دانش اصطلاح‌شناختی، در نظام مفاهیم منعکس است و دانشی را در بر می‌گیرد که چگونگی ارتباط تک‌تک عناصر سازنده‌ی یک دستگاه مفاهیم با همدیگر را نشان می‌دهد. به منظور ترسیم نظام مفاهیم دانش ساختارمند یک رشته‌ی تخصصی بایستی مفاهیم زیررشته‌های مرتبط با آن را در یک ساختار کلی تلفیق و روابط میان آن‌ها را ایجاد نمود. در سطح نظام مفاهیم، سازماندهی دانش فنی به شکل نظام‌مند نمودن گروهی از مفاهیم، در یک رشته‌ی تخصصی صورت می‌پذیرد. بر این اساس اصطلاح‌نامه‌های تخصصی تدوین می‌یابند و به صورت فهرست‌های واژگانی ارائه می‌گردند. در هر مدخل واژگانی، این ساماندهی دانش فنی به صورت مجموعه‌ای از اطلاعات ساختاری و معنایی درآمده است که نظام مفاهیم و روابط میان آن‌ها مانند مفاهیم اعم، مفاهیم اخص و مفاهیم وابسته یا مرتبط را شامل می‌گردد. در سطح رشته‌های تخصصی، سازماندهی دانش فنی تنها به ایجاد ارتباطات میان مفاهیم اکتفا نمی‌کند. مهم این است که نظام‌های مفاهیم، ساماندهی و قابل رویت شده و به تناسب، با یکدیگر مرتبط شوند. به منظور درک و دریافت ساختار دانش فنی یک حوزه‌ی تخصصی، به ناچار بایستی موضوعات را در یک ساختار نظام کلان، جمع‌بندی نمود که چنین ساختار مفهومی را "طبقه‌بندی"^۴ می‌نامند. طبقه‌بندی مفاهیم با اهداف گوناگونی طراحی می‌شود. به عنوان مثال برای مستندسازی مدارک علمی- فنی، رده‌بندی موضوعات علمی را به شکل ارائه‌ی نظام‌های مفاهیم در قالب

¹ Knowledge Engineering

² concept

³ Semantic features

⁴ thesaurus

⁵ taxonomy

اصطلاح‌نامه‌های تخصصی سازماندهی می‌کنند. در اصطلاح‌نامه‌ها، کلیدواژه‌ها^۶ با توجه به شکل و محتوای معنایی بافت زبانی متون علمی تعیین می‌شوند و روابط سلسله‌مراتبی میان آن‌ها به شکل اصطلاح‌اعم، اخص و اصطلاح وابسته ارائه می‌گردد. کلیدواژه‌ها به عنوان "توصیفگر"^۷ مدارک علمی در نمایه‌سازی و بازیابی اطلاعات علمی- فنی عمل می‌کنند و در نتیجه هرکدام دارای پشتوانه انتشاراتی هستند. سازماندهی اصطلاح‌شناختی دانش فنی، در قالب انتقال یک دستگاه مفهومی در زمینه‌های تخصصی، کاربردهای متفاوتی می‌یابد. در این جا ساختار مفهومی و دانش مفهومی یک زمینه تخصصی، مقدمه‌ای بر تبادل اطلاعات تخصصی (فعالیت تخصصی و یا حل مسئله علمی) و تبادل زبان تخصصی (ایجاد ارتباط و انتقال اطلاعات علمی- فنی) است که شامل ترجمه فنی نیز می‌شود. در این زمینه کاربردی، واژه‌نامه‌ها، فرهنگ‌ها، سرعنوان‌های موضوعی و اصطلاح‌نامه‌ها به عنوان ابزار مستندسازی و اطلاع‌رسانی رابطه تنگاتنگی یافته و به طور غیر مستقیم کلید دانش فنی مستندسازی شده در سطح مفهومی و دیگر سطوح پیچیده آن می‌باشد که استفاده آن در زمینه ارتباط دانش تخصصی و اصطلاح‌شناختی به عنوان ارائه‌کننده تبادل اطلاعات فنی اجتناب‌ناپذیر است. همچنین سازماندهی اصطلاح‌شناختی دانش در تمامی این سطوح، برای خلق و ابداع نظام‌های دانش- بنیاد، در سطح مفاهیم، بدون واسطه و یا به شکل اقتباس شده، نقشی زیربنایی دارد.

۲- بانک اطلاعات واژگان علمی فارسی

در جامعه اطلاعاتی ایران با وجود تلاش‌های صورت گرفته، بازیابی و استفاده از اطلاعات علمی- فنی با مشکلات و کمبودهای جدی روبروست. بخشی از آنها به آشفتگی‌هایی مربوط می‌شود، که در مرحله تولید اطلاعات وجود دارد. بخشی دیگر به نابسامانی‌هایی که در مرحله ذخیره‌سازی اطلاعات به وجود می‌آید. از بهترین راه‌حل‌های رفع این معضلات، تهیه و کاربرد اصطلاح‌نامه یا گنجواژه است، که هدفش برگرداندن زبان طبیعی مدارک به زبان کنترل شده در تبادل اطلاعات و دانش فنی است و باعث انطباق نظرات تولیدکنندگان و کاربران دانش فنی می‌شود تا حداکثر اطلاعات با کمترین هزینه، نیرو و زمان (جامعیت) و با بیشترین دقت و کیفیت (مانعیت) ذخیره و سازماندهی و بازیابی شوند. اگرچه تاکنون در جهت تدوین و ترجمه اصطلاح‌نامه‌های تخصصی به زبان فارسی قدم‌های مؤثری برداشته شده و نتایج نسبتاً خوبی هم داشته، اما متأسفانه این اقدامات در جهت به‌کارگیری بهینه و خودکار اصطلاح‌نامه به عنوان ابزاری توانمند در نظام بازیابی رایانه‌ای به حد مطلوب نرسیده است. مهندسی دانش، علمی میان‌رشته‌ای شامل زبان‌شناسی،

⁶ keywords

⁷ descriptor

هستی‌شناسی^۸ و مهندسی زبان^۹ و یکی از اهداف آن ساماندهی اطلاعات علمی پراکنده است (فرقالی^{۱۰}، ۲۰۰۳ و جورجیف^{۱۱}، ۲۰۰۷). نمایه‌سازی نیز جهت سازماندهی اطلاعات علمی صورت می‌پذیرد و نمایه‌سازی، متخصصین موضوعی، زبان‌شناسان و متخصصین علوم رایانه در این امر دخیل هستند. در نمایه‌سازی و چکیده‌نویسی مدارک علمی نقش اصطلاح‌نامه تخصصی به عنوان ابزار کار نمایه‌ساز و متخصص موضوعی حائز اهمیت بسیار است، به شکلی که اگر آن را حذف کنیم، اطلاعات نمایه و ثبت و ضبط شده قابلیت بازیابی و بازنمایی چندانی نخواهد داشت و ریزش کاذب اطلاعات در حدی خواهد بود که ارزش اطلاعاتی نمایه‌های ایجاد شده را بسیار پایین خواهد آورد. همچنین محققان و رساله‌نویسان رشته‌های تخصصی به منظور معیارسازی ادبیات علمی خود به اصطلاح‌نامه نیاز وافر دارند. آن دسته از اطلاع‌رسانان و برنامه‌نویسان رایانه‌ای که با نرم‌افزارهای نگارش متون، فرهنگ لغات و یا مترجم‌های هوشمند سر و کار دارند، از اصطلاح‌نامه‌های تخصصی به عنوان ابزاری بسیار مفید در تحلیل اطلاعات حاصل از نتایج تحقیقاتی بهره‌ فراوان می‌برند و از این رو می‌توانند در سیاست‌گذاری‌های کلان تحقیقاتی کشور و ارتباط نظام‌مند بین دانشگاهها و مراکز تحقیقاتی و صنایع، نقش به‌سزایی ایفاء نمایند. یک اصطلاح‌نامه کارآمد حاصل جمع‌آوری و ساختاربندی واژه‌های علمی و تخصصی است. سازمانهای چکیده‌نویسی و نمایه‌سازی اطلاعات به دلیل حجم بالای تولیدات علمی، به پردازش آنها پرداخته و در نتیجه بهترین منبع واژگان تخصصی رشته‌های علمی و گنجینه واژگان برای تولید اصطلاح‌نامه‌ها محسوب می‌گردند.

برای تحلیل زبان‌شناختی اعم از ساختاری و معنایی، تعداد ۳۹ هزار واژه علوم پایه از بانک اطلاعات واژگان علمی پژوهشگاه اطلاعات و مدارک علمی ایران شامل اصطلاح‌نامه‌های علوم پایه، کشاورزی و فنی-مهندسی (در کل ۱۱۵ هزار واژه)، به عنوان "پیکره زبانی"^{۱۲} در نظر گرفته شد. این مجموعه واژگان شامل اصطلاحات رشته‌های شیمی، فیزیک، زمین‌شناسی، زیست‌شناسی و ریاضی است که تلفیق آنها موجب شد که بسیاری از این اصطلاحات در پنج اصطلاح‌نامه مذکور مشترک باشد. بنابراین برای بالا بردن دقت در تحلیل این واژگان مراحل زیر طی شد:

۱- در مرحله اول واژه‌هایی که در اصطلاح‌نامه‌های مذکور به صورت تکراری آمده است، از این مجموعه حذف گردید. قابل ذکر است که استخراج این کلیدواژه‌ها از متون علمی- تخصصی بوده و از پشتوانه انتشاراتی برخوردار است، زیرا ارائه هر یک از این اصطلاحات به صورت مدخل‌های واژگانی و

⁸ ontology

⁹ Language Engineering

¹⁰ Farghaly

¹¹ Georgiev

¹² corpus

زیرمجموعه آن بر اساس نظام روابط منطقی مفاهیم بر مبنای کلیدواژه‌هایی است که از نمایه‌سازی متون علمی شکل گرفته است و شامل پایان‌نامه‌های فارغ‌التحصیلان داخل و خارج از کشور، مقالات سمینارهای تخصصی، گزارش طرح‌های پژوهشی اجرا شده و در دست اجرای دانشگاهها و مراکز تحقیقاتی، مقالات مجلات علمی- پژوهشی و فهرست واژگان انتهای کتب درسی دانشگاهی هر یک از رشته‌های تخصصی بوده است.

۲- در مرحله دوم، واژه‌ها و اصطلاحاتی که صرفاً حرف‌نگاری واژه‌های لاتین (تعداد ۹۳۹۶ اصطلاح استخراج شده از مجموعه واژگان اصطلاح‌نامه‌ها) و شامل اسامی اشخاص، اسامی جغرافیایی، مواد شیمیایی و اصطلاحات خاص هر یک از رشته‌های علوم پایه بوده است، تفکیک و در پرونده جداگانه‌ای ذخیره شده است. این مجموعه واژگان که در واقع لاتین هستند و معادل فارسی ندارند و یا هنوز فرهنگستان زبان و ادب فارسی برای آنها معادل‌گزینی نکرده است، حاوی اطلاعات ارزشمندی است که نیاز به تحلیل جداگانه‌ای دارد. بایستی این مجموعه واژگان شامل اسامی خاص، اسامی جغرافیایی و اسامی شیمیایی که اصولاً نمی‌توانند معادل فارسی داشته باشند، از اصطلاحات فنی هر یک از رشته‌ها تفکیک شوند. مجموعه اصطلاحات فنی را می‌توان به عنوان واژه‌های پیشنهادی جهت معادل‌گزینی فارسی و مشورت با گروه‌های واژه‌گزینی فرهنگستان ارائه نمود.

۳- در مرحله سوم، واژگان علمی فارسی شامل ۱۰۹۵۵ اصطلاح از مجموعه واژگان علمی به نسبت بسیط^{۱۳} (۳۱۹۴ واژه) یا غیربسیط^{۱۴} (۷۷۶۱ واژه) دسته‌بندی و میزان بسامد (فراوانی)^{۱۵} آنها نسبت به کل استخراج شد (جدول ۱).

جدول ۱- بسامد واژگان علمی بسیط و غیر بسیط

فارسی	۱۰۹۵۵	%۵۳/۸۳
لاتین	۹۳۹۶	%۴۶/۱۷
غیر بسیط	۷۷۶۱	%۷۰/۸۴
بسیط	۳۱۹۴	%۲۹/۱۶
مشتق مرکب	۶۸۳۵	%۶۲/۳۹
مرکب	۹۲۶	%۸/۴۵

۴- در مرحله چهارم واژه‌های غیربسیط به مرکب^{۱۶} (C) و مشتق^{۱۷} (D) و مشتق مرکب (DC) تقسیم و بسامد هر کدام تعیین شد. تعداد واژه‌های مرکب شامل ۹۲۶ واژه و مشتق مرکب شامل ۶۸۳۵ واژه

¹³ simplex

¹⁴ complex

¹⁵ frequency

¹⁶ compound

می‌باشد. لازم به ذکر است که در فرآیند ذخیره و بازیابی اطلاعات واژه‌های "مرجع"^{۱۸} میان محققین، اساتید و دانشجویان تحصیلات تکمیلی که کاربران اصلی اطلاعات علمی- فنی می‌باشند، ملاک‌گزینش واژه‌ها در نمایه‌سازی مدارک علمی-فنی بوده‌است. از این رو بیشترین بسامد در کاربرد یک اصطلاح، نشانهٔ مرجع بودن آن است.

۵- گروه واژه‌های مشتق با الهام از پسوندهای مورد مطالعه در مقالات "شیوه‌ها و امکانات واژه‌سازی در زبان فارسی" (صادقی، ۷۲-۷۰) و نیز تحقیق "عملکرد معنایی پسوندهای فارسی بر اساس نظریهٔ لیبر"^{۱۹} (رفیعی، ۱۳۸۶) مورد تجزیه و تحلیل رایانه‌ای قرار گرفت. براساس فهرست پسوندها^{۲۰} (S) و پیشوندهای^{۲۱} (P) به دست آمده میزان بسامد و کاربرد آنها در اصطلاحات علمی فارسی تعیین شد (جدول ۲ و ۳).

جدول ۲- بسامد پسوندهای فارسی در واژگان علمی

مان	۲۰	%۰/۳۳	ی	۳۴۵۸	%۵۶/۸۲
واره	۱۷	%۰/۲۸	ه	۶۷۸	%۱۱/۱۴
ور	۱۵	%۰/۲۵	ان	۳۶۲	%۵/۹۵
مند	۱۲	%۰/۲۰	ش	۲۱۲	%۳/۴۸
یه	۱۰	%۰/۱۶	ن	۱۵۱	%۲/۴۸
وار	۹	%۰/۱۵	ساز	۱۴۹	%۲/۴۵
دان	۸	%۰/۱۳	نده	۱۴۲	%۲/۳۳
و	۷	%۰/۱۲	دار	۱۲۵	%۲/۰۵
چی	۶	%۰/۱۰	گر	۹۳	%۱/۵۳
دانی	۶	%۰/۱۰	ار	۸۶	%۱/۴۱
زار	۶	%۰/۱۰	یت	۸۳	%۱/۳۶
گون	۵	%۰/۰۸	ک	۶۱	%۱
بان	۴	%۰/۰۷	ا	۵۷	%۰/۹۴
گار	۳	%۰/۰۵	گاه	۵۷	%۰/۹۴
گانه	۳	%۰/۰۵	ین	۴۸	%۰/۷۹
وش	۳	%۰/۰۵	انه	۴۵	%۰/۷۴
آباد	۱	%۰/۰۲	ستان	۳۵	%۰/۵۸
ناک	۱	%۰/۰۲	کار	۳۳	%۰/۵۴
وند	۱	%۰/۰۲	بند	۲۶	%۰/۴۳
کده	۱	%۰/۰۲	سان	۲۵	%۰/۴۱

¹⁷ derived

¹⁸ Preferred Terms

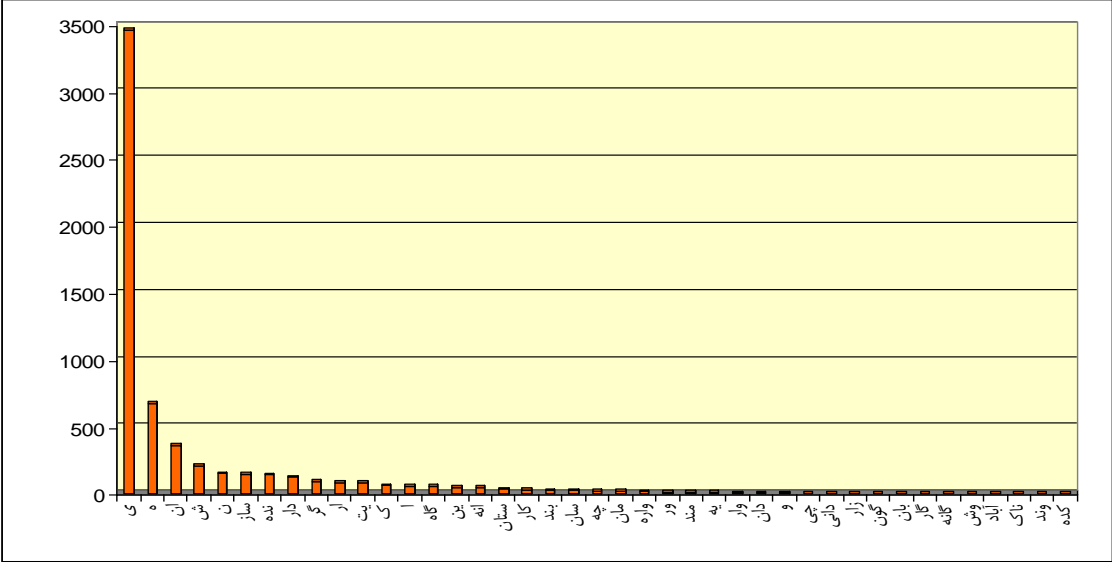
¹⁹ Lieber, R

²⁰ suffixes

²¹ prefixes

چه	۲۲	٪۳۶
----	----	-----

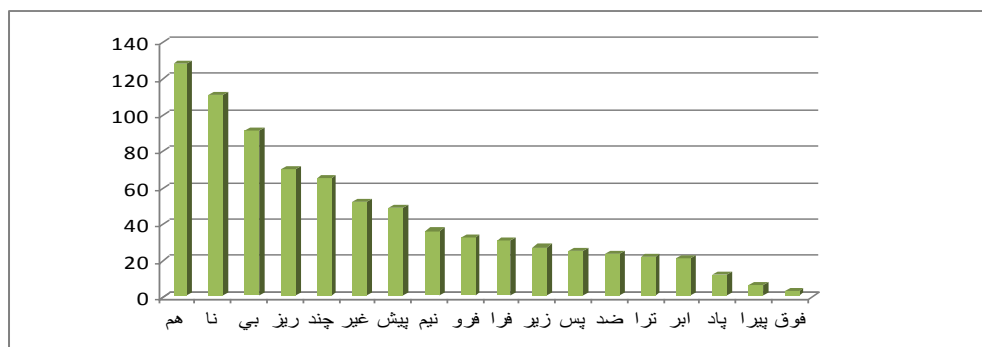
نمودار (۱) بسامد پسوندهای فارسی در واژگان علمی



جدول ۳- بسامد کاربرد "پیشوندها" در اصطلاحات علمی

هم	۱۲۸	٪۱۸/۱۹
نا	۱۱۱	٪۱۵/۷۷
بی	۹۱	٪۱۲/۹۳
رئز	۷۰	٪۹/۹۴
چند	۶۵	٪۹/۲۳
غیر	۵۲	٪۷/۳۹
پیش	۴۹	٪۶/۹۶
نیم	۳۶	٪۵/۱۱
فرو	۳۲	٪۴/۵۴
فرا	۳۱	٪۴/۴۰
زیر	۲۷	٪۳/۸۳
پس	۲۵	٪۳/۵۵
ضد	۲۳	٪۲/۲۷
ترا	۲۲	٪۳/۱۲
ابر	۲۱	٪۲/۹۸
پاد	۱۲	٪۱/۷۰
پیرا	۶	٪۰/۸۵
فوق	۳	٪۰/۴۳

نمودار (۲) بسامد کاربرد "پیشوندها" در اصطلاحات علمی

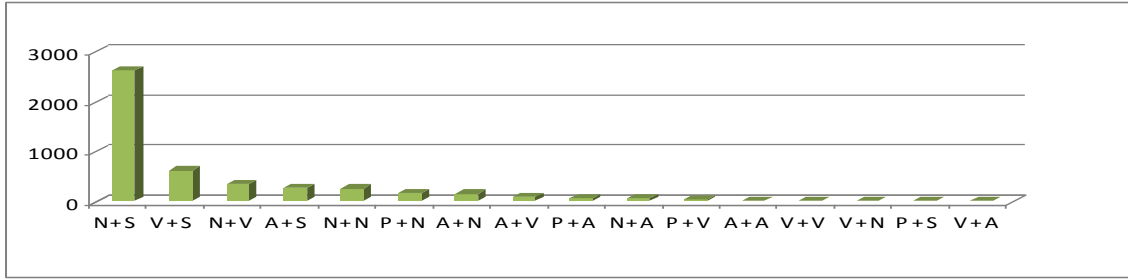


۶- اصطلاحات مرکب براساس ساختار درونی و معنایی آنها به سه گروه درون‌مرکز، برون‌مرکز و متوازن (طباطبایی، ۱۳۸۲ و لیبر، ۲۰۰۴) تقسیم و اجزاء و عناصر سازنده هر یک به نسبت مقولات دستوری فعل (V)، اسم (N)، صفت (A) و قید (D) و نوع همنشینی آنها تعیین گردید. در نتیجه این تجزیه و تحلیل فرمول‌بندی واژه‌های مرکب، مشتق و مشتق‌مرکب و ساختار درونی اصطلاحات فارسی تعیین شد (جدول ۴ و ۵).

جدول ۴- بسامد واژه‌های دو‌عنصری

N+S	2619	55.58%
V+S	621	13.18%
N+V	351	7.45%
A+S	271	5.75%
N+N	259	5.50%
P+N	162	3.44%
A+N	155	3.29%
D+V	94	1.99%
P+A	57	1.21%
N+A	52	1.10%
P+V	41	0.87%
A+A	10	0.21%
V+V	9	0.19%
V+N	5	0.11%
P+S	4	0.08%
V+A	2	0.04%

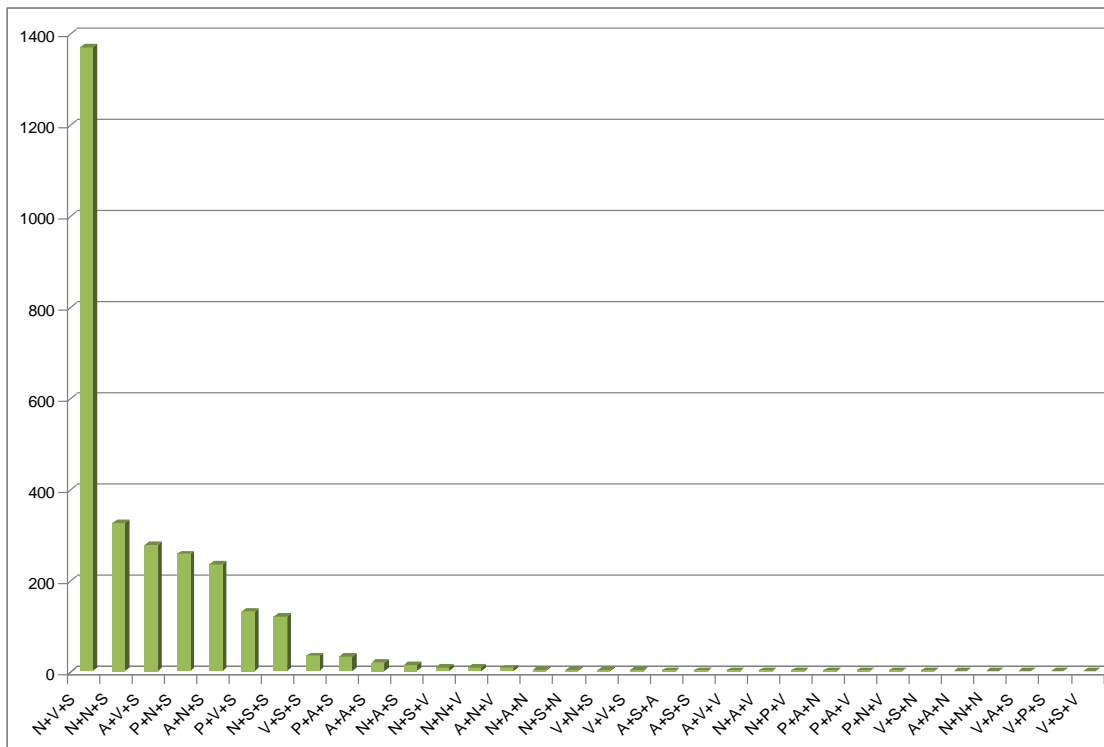
نمودار (۳) بسامد واژه‌های دو‌عنصری



جدول ۵- بسامد واژه‌های سه‌عنصری

N+V+S	1370	47.65%		N+S+V	9	0.31%		N+P+V	2	0.07%
N+N+S	325	11.30%		N+N+V	8	0.28%		P+A+N	2	0.07%
D+V+S	277	9.63%		A+N+V	7	0.24%		P+A+V	2	0.07%
P+N+S	258	8.97%		N+A+N	3	0.10%		P+N+V	2	0.07%
A+N+S	235	8.17%		N+S+N	3	0.10%		V+S+N	2	0.07%
P+V+S	131	4.56%		V+N+S	3	0.10%		A+A+N	1	0.03%
N+S+S	121	4.21%		V+V+S	3	0.10%		N+N+N	1	0.03%
V+S+S	34	1.18%		A+S+A	2	0.07%		V+A+S	1	0.03%
P+A+S	32	1.11%		A+S+S	2	0.07%		V+P+S	1	0.03%
A+A+S	19	0.66%		D+V+V	2	0.07%		V+S+V	1	0.03%
N+A+S	14	0.49%		N+A+V	2	0.07%				

نمودار (۴) بسامد واژه‌های سه‌عنصری



در پژوهش حاضر در تحلیل ساختاری اصطلاحات علمی با برچسب دهی^{۲۲} عناصر سازنده واژگان علمی حدود ۸۰ الگوی واژگانی^{۲۳} به دست آمده، که برخی از اصطلاحات مرکب به نسبت میزان بسامد آنها در کاربرد اصطلاحات علمی از این قرار است:

N+V, N+N, A+V, D+V, N+A, A+A, V+D, V+N, V+V

قابل توجه است، که در ساختار اصطلاحات علمی زبان فارسی نمونه‌های سه عنصری و یا بیشتر نیز وجود دارد مانند:

A+A+N, A+N+N, A+N+V, D+V+V, N+A+N, N+N+V

که البته از بسامد کمتری برخوردار هستند و در واقع کاربرد کمتری دارند. توصیفگرهای اصطلاح نامه که در نمایه‌سازی و بازیابی اطلاعات کاربرد دارند، معمولاً شامل اسم و عبارت اسمی می‌شوند. طبق استانداردهای جهانی (۲۷۸۸ و ۵۹۶۴) صفات عموماً به عنوان اصطلاحات نمایه‌سازی قابل قبول نیستند مگر آن‌که در اصطلاحات مرکب به کار روند. قیدها نیز از اصطلاحنامه حذف می‌شوند، مگر آن‌که این قیدها بخشی از اصطلاحات مرکب را تشکیل دهند. افعال نیز به صورت اسم فعل و صفت مفعولی در اصطلاحات مرکب استفاده می‌شوند. اصطلاحات مرکب نسبت به اصطلاحات بسیط دقیق‌تر و مشخص‌تر و برای بازنمایی مفاهیم پیچیده کارآیی بیشتری دارند. این اصطلاحات از ریزش کاذب اطلاعات جلوگیری کرده و متضمن مانعیت مطلوب هستند.

۳- مدیریت اصطلاحنامه و نمایه‌سازی رایانه‌ای

اصطلاحنامه یا گنج‌واژه، زبان مشترک میان نمایه‌سازی و بازیابی اطلاعات در شبکه اطلاعات علمی- فنی است. مدیریت اصطلاحنامه، این زبان مشترک را برای ارائه محتوای اسناد و مدارک (نمایه‌سازی) و کاربران اطلاعات، به منظور جستجوی موضوعات مورد علاقه محقق (بازیابی) پیرامون اسناد علمی به کار می‌گیرند. به عنوان یک زبان مستندسازی، واژگان اصطلاحنامه بر اساس اصطلاحات منتخب از زبان طبیعی شکل می‌گیرند، همانند کدهای هوشمند که در طبقه‌بندی و رده‌بندی موضوعی به کار می‌روند. در نمایه‌سازی هم‌نشینی واژه‌ها (معروف به پس‌همارایی) معمولاً تنها راه ارائه موضوعات اخص و اعم در یک سند می‌باشد. به منظور بازیابی مدارک، کاربران اطلاعاتی اصطلاحات نمایه‌سازی را به صورت منفرد یا گروهی و به شیوه

²² tagging

²³ Lexical pattern

"بولین" به کار می‌برند. عملیات رایانه‌ای نیز می‌تواند به گونه‌ای فراهم آید که به منظور حداکثر بهره‌برداری، روابط پیش‌همارایی در روابط واژگان اصطلاح‌نامه ایجاد گردد. اصطلاح‌نامه چیزی بیش از یک فهرست صرف اصطلاحات است و در واقع نظامی ساختارمند و معنی‌دار از مفاهیم را تشکیل می‌دهد. واژگان آن طبق روابط معنایی، نظام‌مند شده‌اند و یک اصطلاح‌نامه به منظور تضمین کارآیی، بایستی تعداد محدودی از حوزه‌های دانش را همراه با جزئیات کاملاً تخصصی پوشش دهد. این تخصصی بودن، مشخصاً در گزینش اصطلاحات اخص معرفی شده است که به واسطه پیش‌همارایی، یعنی به کارگیری واژه‌های مرکب و برقراری روابط منطقی میان اصطلاحات انجام می‌پذیرد. اصطلاح‌نامه چندزبانه، امکان دستیابی به داده‌های مستند مشترک را برای کاربران اطلاعاتی که گویشور زبانهای گوناگون هستند، فراهم می‌سازد. اصطلاحات، از میان چند زبان طبیعی در مقابل یک مفهوم گزینش شده و از نظر زبان‌شناختی مترادف و هم‌معنی می‌باشند. در نتیجه روابط معنایی بر اساس یک قاعده دقیقاً همانند از زبانی به زبان دیگر انتقال می‌یابد، زیرا در واقع واژه‌ها مفاهیم را با یکدیگر مرتبط می‌سازند. "علیرغم هزینه بالای تدوین اصطلاح‌نامه، کاربرد آن‌ها در مراکز تحقیقاتی، روز به روز گسترش می‌یابد. برای بهره‌گیری از اصطلاح‌نامه، در فهرست نویسی موضوعی نیز اقداماتی به عمل آمده است و اصولاً سرعنوان‌های موضوعی مانند سرعنوان‌های پزشکی (مش^{۲۴}) به لحاظ ساختاری به اصطلاح‌نامه نزدیک شده‌اند. از آنجا که ساختار اصطلاح‌نامه‌ها با رایانه سازگارتر بوده و می‌توانند عمل ذخیره و بازیابی اطلاعات را تسهیل نمایند، لذا به زودی شاهد ساختاری یکدست و هماهنگ به صورت فرااصطلاح‌نامه خواهیم بود. بهره‌گیری از رایانه در امر فهرست‌نویسی، لزوم این اصلاحات، یکپارچگی و تلفیق آن را چند برابر کرده است." (رایت و بودین، ۲۰۰۱^{۲۵}).

۴- کاربرد اصطلاح‌نامه جامع در بازیابی اطلاعات

پس از تدوین اصطلاح‌نامه‌های علوم پایه (شیمی، فیزیک، زیست‌شناسی، زمین‌شناسی و ریاضی)، فنی-مهندسی و کشاورزی براساس نرم‌افزار "قاموس" و تکامل یافته آن "تزاروس بیلدر"^{۲۶} در گروه پژوهشی اصطلاح‌شناسی پژوهشگاه اطلاعات و مدارک علمی ایران، به کارگیری آن در نظام‌های اشاعه اطلاعات مطرح شد. در مورد مزایای کاربرد اصطلاح‌نامه در نظام اطلاعاتی، بحث‌های زیادی مطرح شده است، اما در این جا به اختصار به مهم‌ترین اثر آن، یعنی نزدیک شدن مفاهیم به کار گرفته شده توسط کاربران، به مفاهیم به کار گرفته شده

²⁴ MESH

²⁵ Wright, S.E. & Budin, G.

²⁶ Thesaurus builder

توسط نمایه سازان اشاره می‌گردد. بنابراین می‌توان گفت سه بخش کلی از این نظام‌ها از اصطلاح‌نامه تاثیر می‌پذیرند:

۱- نظام‌هایی که به نوعی در آن‌ها واژه‌گزینی می‌گردد. این بخش شامل فرآیندهایی از قبیل اس.دی.آی.^{۲۷}، نظام اشاعه‌گزینی اطلاعات (بخش انتخاب موضوعی) و ورود و ویرایش اطلاعات در پایگاه‌های مختلف است. در این فرآیند لازم است تا کاربر با در اختیار داشتن رابطه کاربری مناسب بتواند در میان موضوعات اصطلاح‌نامه و پایگاه واژگان جستجو کرده و از میان آن‌ها واژه‌های مورد نظر خود را انتخاب و ثبت نماید.

۲- نظام‌هایی که باید از واژگان گزیده شده اصطلاح‌نامه استفاده کنند، که می‌توان از آن جمله به بخش ارسال مقالات فرآیند اس.دی.آی اشاره نمود. در این بخش، کاربر می‌تواند با توجه به ساختار درختی موضوعات اصطلاح‌نامه، یک شاخه را به همراه زیرشاخه‌های آن انتخاب نماید. هم‌چنین در فرآیند تولید فصلنامه‌های موضوعی، بایستی از واژه‌هایی که از اصطلاح‌نامه استخراج شده‌اند به عنوان واژه‌های استاندارد استفاده گردد.

۳- نظام‌هایی که در آن کاربران از اصطلاح‌نامه برای بازیابی اطلاعات استفاده می‌کنند. مهم‌ترین فرآیند این بخش جستجو در پایگاه‌هاست که به وسیله آن، کاربران می‌توانند موضوعات را به همراه زیرشاخه‌های آن جستجو نمایند.

۴- بسامد اصطلاحات به کارگرفته شده در نمایه‌سازی و بازیابی، قابل نظارت است و می‌توان به حذف و یا نامرئج کردن اصطلاحات مهجور و بدون کاربرد پرداخت.

۵- سابقه هر اصطلاح قابل ذخیره است و تغییراتی که در طی زمان ایجاد شده یعنی تاریخ تحول مفاهیم قابل ارجاع به آخرین انتخاب است.

۶- ذخیره واژه‌هایی که توسط جستجوگر به نظام وارد می‌شود، پشتوانه مناسبی برای روزآمد ساختن اصطلاح‌نامه و تجدید نظر در آن است (ایچسن^{۲۸} و دیگران، ۲۰۰۰).

۵- فرآیند اصطلاح‌نامه^{۲۹}

فرآیند اصطلاح‌نامه نوعی اصطلاح‌نامه است که باعث ادغام و تلفیق اصطلاح‌نامه‌های متعدد می‌گردد و میان مجموعه واژگان ملی و بین‌المللی، ارجاعات متقابل ایجاد می‌کند. به منظور بازیابی در پایگاه‌های متعدد و

²⁷ Selective dissemination of Information (SDI)

²⁸ Aitchison

²⁹ Metathesaurus

چندگانه که به واسطه واژگان کنترل شده متعدد نمایه شده‌اند، نوعی فرااصطلاح‌نامه ایجاد می‌شود. هر اصطلاح در فرااصطلاح‌نامه با برابرنامه‌ها و اصطلاحات مرتبط از تمامی اصطلاح‌نامه‌های تلفیق شده در آن به نمایش در می‌آیند. فرایند خودکار در جستجو و بازیابی، توسعه یافته، کاربر را قادر می‌سازد تا بازیابی را در نمایه‌های منفرد یا چندگانه و تلفیق شده و با استفاده از اصطلاحات هر یک از منابع واژگانی انجام دهد. فرااصطلاح‌نامه مفهوم-بنیاد است و هدف آن ایجاد وحدت میان نام‌گذاری معانی مختلف یک مفهوم است. این نام‌گذاری‌ها از 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 جهت سهولت در ایجاد آن مؤثر واقع شود. روزآمدی واژگان، بهره‌مندی از مجموعه‌ی واژه‌های متعدد، الگوریتم‌گزینش و بازنگری متخصص موضوعی را می‌طلبد. اصطلاح‌نامه‌های مختلف را می‌توان تحت شرایط ذیل در یک اصطلاح‌نامه تلفیق نمود:

- ۱- از ابتدا اصطلاح‌نامه‌های مختلف الزاما بر اساس پیش‌زمینه نظری واحدی ساختاربندی شده باشند. این امر در گام نخست برای تلفیق کاملا ضروری است.
- ۲- در گام دوم، شیوه تخصیص شناسه‌های یکسان به کار گرفته شود و زمینه‌های بازیابی و رفع مشکلات شناسایی شده فراهم گردد.
- ۳- در گام سوم، توسعه نظارت کیفیت دیدگاه‌ها و توافقات به عمل آمده صورت پذیرد.
- ۴- در گام چهارم، شناسایی زمینه‌های جایابی و دگرگونی ظریف و نامحسوس معانی و مفاهیم در هر یک از گرایش‌های علمی ضروری است.

فرااصطلاح‌نامه تنها یک واژگان استاندارد شده نیست، بلکه ابزاری برای افزایش نهایی واژگان موجود و مفید است و به عنوان منبعی غنی برای متخصصین اطلاعات علمی مورد استفاده قرار می‌گیرد. نظارت بر تناسب اجزاء فرااصطلاح‌نامه در هر ویرایش جدید، سبب توسعه و افزایش میزان مفاهیم یک فرااصطلاح‌نامه می‌گردد. فناوری فرااصطلاح‌نامه برای فرمول‌بندی روابط معنایی میان ساختارهای دانش فنی (به صورت مفاهیم و روابط سلسله‌مراتبی آن‌ها) از دیدگاه‌های گوناگون به کار گرفته می‌شود و کاربران را قادر به تورق، جستجو و بازیابی میان‌رشته‌ای در گرایش‌های گوناگون علمی می‌سازد. فرااصطلاح‌نامه دربرگیرنده مجموعه‌ای از مفاهیم کلیدی در راس 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 تالیف یا ابداع شده‌اند (وود^{۳۰}، ۲۰۰۴).

³⁰ Wood

۶- هستی‌شناسی و شبکه‌واژگانی^{۳۱}

در آینده نیز بایستی در مورد جمع‌آوری مجموعه‌ای از دانش واژگانی شامل "شبکه‌واژگانی"، اصطلاح‌نامه‌ها و واژه‌نامه‌ها و کاربرد مشترک میان آن‌ها در تشکیل زنجیره‌های واژگانی و ایجاد ارتباط رایانه‌ای و خودکار، تحقیقات دقیق‌تری انجام پذیرد. ایجاد "پایگاه دانش واژگانی" مانند شبکه‌واژگانی می‌تواند به عنوان ابزار کلیدی، جهت شکل‌گیری "هستی‌شناسی" محتوای یک متن علمی به‌ویژه مرتبط با پایگاه اطلاعات علمی- فنی باشد.

"نقشه مفهومی"^{۳۲} براساس شکل‌گیری سلسله‌مراتبی واژگان تدوین شده‌است. هستی‌شناسی، اغلب مفاهیمی را در برمی‌گیرد که برای شناخت انسان از دنیای پیرامون، بنیادی و اساسی می‌باشند و دانش "پیش‌نمونه‌ها"^{۳۳} را بازنمایی می‌کنند. این‌که ساختار هستی‌شناسی، براساس اعم و اخص و یا سطوح بالا و پایین طراحی شوند، بیشتر به اهداف و حوزه مورد نظر بستگی دارد تا حل مشکلات مدیریت واژگان، زیرا تفاوتی معنادار و واقعی میان این دو نوع هستی‌شناسی وجود دارد. هستی‌شناسی براساس سلسله‌مراتب و ویژگی‌های معنایی سعی دارد حوزه معنایی خاصی را که مبتنی بر مفاهیم دقیق تعریف شده است به دست آورد. به عنوان مثال در حوزه علوم پایه (ریاضی، فیزیک، شیمی، زیست‌شناسی و زمین‌شناسی) دانش مربوطه به راحتی فرمول‌بندی می‌شوند، زیرا در همه این حوزه‌ها نظام مفاهیم حاکم است، که می‌توان آن‌ها را به واسطه اصطلاحات علمی با تعریف دقیق بیان نمود. در چنین مواردی سعی می‌شود اشیاء در عالم مقال و نیز به‌طور محض و انتزاعی و با الگوهای آرمانی و ساده‌شده‌ای از پدیده‌های واقعی در جهان ارائه شوند (گومز-پرز^{۳۴}، ۲۰۰۴).

"پایگاه‌های اطلاعات واژگانی" همانند شبکه‌واژگانی، شامل واژگان، اصطلاح‌نامه‌ها و فرهنگ‌های لغت است که به قاعده‌مند نمودن معنی‌شناسی واژگانی یعنی معنی واژه‌های یک یا چند زبان طبیعی می‌پردازد. این معانی براساس شبکه مفاهیم و مبتنی بر فهم مشترک اکثریت جامعه علمی و دارای تعاریف خاص و

³¹ WordNet

³² Concept Map

³³ Prototypes

³⁴ Gómez-Pérez

درب‌گیرنده معانی مختلف متداول و کاربردهای خاص و ویژه و ترکیبات گوناگون واژه‌های مبتنی بر "پیش‌نمونه‌ها" و نیز مفاهیم پایه علوم است.

۷- شبکه معنایی^{۳۵}

دیویس^{۳۶} و همکاران (۲۰۰۷) نگاه جدیدی به آینده "شبکه جهانی اطلاعات"^{۳۷} دارند که در آن تمامی دانش موجود به صورت قاعده‌مند رمزگذاری می‌شود تا عوامل هوشمند بتوانند به طریقی خودسامان به همه انسان‌ها خدمات اطلاعاتی ارائه دهند. این کار مستلزم تعریف رسمی و قاعده‌مند هستی‌شناسی‌های مبتنی بر "وب" است تا بیان‌کننده دانش قابل فهم انسان و نیز عوامل هوشمند باشد که ارزش افزوده دانش و توسعه آن به صورت "فراداده‌ها"^{۳۸} خواهد بود و آن در حال حاضر تعریف شده و قابل کاربرد در مجموعه‌های محدودی از تقاضاهاست. پیش‌بینی می‌شود که در دهه‌های آتی، ساختار دانش مورد درخواست، بیشتر به صورت رسمی و به شکل نظام‌های هستی‌شناسی مبتنی بر "وب" و ارائه آن‌ها به زبان‌های گوناگون، زمینه فعالی برای تحقیقات و مطالعات بیشتر باشد که حاصل آن تعداد قابل توجهی طرح‌های تحقیقاتی و تصویب استانداردهاست که در قالب "آر.دی.اف"^{۳۹} و "اکس.ام.ال/آر.دی.اف"^{۴۰} و نقشه‌های معنایی به عنوان زبان‌های ارائه دانش به کار می‌روند. "اکس.ام.ال" قادر به بازنمایی پدیده‌هاست، اما نمی‌تواند ساختار معنایی آن را بازگو کند در حالی که "آر.دی.اف" قادر به تمایز مولفه‌های معنایی است. برای تشخیص فاعل، مفعول و نیز اقسام فعل می‌توان از "یو.آر.آی"^{۴۱} استفاده نمود. "شبکه معنایی" شبکه‌ای مجزا از "وب" جهانی نیست، بلکه توسعه یافته آن است که در آن اطلاعات با تجزیه و تحلیل معنایی و به منظور هماهنگی بیشتر انسان و رایانه ایجاد شده است. ویژگی خاص "وب"، جهانی بودن آن است که ایجاد "شبکه معنایی" ارتباطات مفاهیم گوناگون را مقدور می‌سازد. همانند اینترنت، شبکه معنایی نیز محدود به مکان خاصی نخواهد بود. این شبکه برای عملکرد دقیق‌تر رایانه نیاز به ساختارمند نمودن اطلاعات و نیز اعمال قواعد معنایی می‌باشد. بدین منظور بایستی از الگوهای یکسان معنایی استفاده شود. به این صورت، توانایی تجزیه و تحلیل معنایی این شبکه، امکان دسترسی میزان زیادی از اطلاعات را فراهم می‌سازد. از این رو این شبکه نبایستی به یک زبان خاص

³⁵ Semantic Web

³⁶ Davies

³⁷ World Wide Web

³⁸ Metadata

³⁹ Resource description framework(RDF)

⁴⁰ Extensible Markup Language(XML)

⁴¹ Uniform resource identifier(URI)

تعلق داشته باشد، بلکه براساس نقشه‌های مفهومی عمل کند. در نتیجه از نظام‌های گوناگون بازنمایی دانش استفاده می‌نماید و با استفاده از شیوه‌های بیان ریاضی و مهندسی هستی‌شناسی، منطق آن از توانایی بالایی برای توصیف عناصر سازنده یک پدیده و یا مولفه‌های معنایی برخوردار می‌باشد و به این شکل، تحت حاکمیت یک زبان خاص و در پی آن یک فرهنگ خاص نیز قرار نمی‌گیرد.

جوامع مختلف علمی، علاقه‌مند هستند که مفاهیم اصلی فعالیت‌های خود را تعریف و تثبیت نمایند. علاوه بر این، روابط واژگانی و ارتباطات معنایی که به صورت اصطلاحات، استانداردسازی شده‌اند، نشانگر تقید این جوامع در کاربرد مفاهیم و زبان مشترک در سازمان‌ها و تبادل اطلاعات با عوامل هوشمند در "شبکه معنایی" است.

۸- نتایج

الف- گردآوری واژگان علمی و معادل‌یابی فارسی طبق واژه‌نامه‌های نشر دانشگاهی و واژه‌های مصوب گروه‌های واژه‌گزینی فرهنگستان زبان و ادب فارسی.

ب- تلفیق اصطلاح‌نامه‌های علوم پایه و تشکیل بانک اطلاعات واژگان علمی.

ج- ویرایش صوری و محتوایی واژگان براساس "شیوه‌نامه نگارش فارسی" فرهنگستان.

د- تجزیه و تحلیل ساختاری و معنایی واژگان علوم پایه و استخراج واژه‌های مرکب، مشتق و مشتق مرکب و محاسبه میزان بسامد کاربرد آنها به عنوان واژه مرجح.

ه- استخراج الگوهای ساختارهای واژگان علمی زبان فارسی (حدود ۸۰ الگو) و تعیین مقوله‌ها و زیرمقوله‌ها.

و- ایجاد ارتباطات مفهومی در نظام مفاهیم و شکل‌گیری هستی‌شناسی.

ز- فراهم نمودن زمینه شکل‌گیری شبکه واژگانی و شبکه معنایی.

۹- منابع

رفیعی، عادل (۱۳۸۶)، نگاهی به عملکرد معنایی پسوندهای اشتقاقی زبان فارسی، رساله دکتری دانشکده ادبیات فارسی و زبانهای خارجی، دانشگاه علامه طباطبایی.

شقایق، ویدا (۱۳۸۶)، مبانی صرف، تهران، انتشارات سمت.

صادقی، علی اشرف (۱۳۷۲-۱۳۷۰)، "شیوه‌ها و امکانات واژه‌سازی در زبان فارسی معاصر" (۱-۱۱)، تهران، مرکز نشر دانشگاهی.

طباطبایی، علاءالدین (۱۳۸۲)، *اسم و صفت مرکب در زبان فارسی*، تهران، مرکز نشر دانشگاهی

- Aitchison, J., A.Gilchrist, and A. Bowden (2000). *Thesaurus Construction. A Practical Manual*. London. Aslib.
- Budin, G. and E. Oeser (1997). *Beitraege zur Terminologie und Wissenstechnik*. Vienna. TermNet.
- Cruse, D. A.(1995). *Lexical Semantics*. Cambridge Uni.Press
- Davies, J., R. Studer and P. Warren(2007). *Semantic Web Technologies*. John Wiley&Sons.
- Farghaly, A. (ed.),(2003). *Handbook for Language Engineers*. CSLT.Publ.
- Georgiev, H. (2007). *Language Engineering*. Continuum.
- Gómez-Pérez et al.(2004). *Ontological Engineering*. Springer
- Green, S.J.(1999). *Lexical Semantics and automatic hypertext Construction*.
<http://www.cs.brown.edu>.
- Lauren, Ch., J. Myking and H. Picht(1998). *Terminologie unter der Lupe*. Vienna. TermNet.
- Lieber, R.(2004). *Morphology and Lexical Semantics*. Cambridge Uni. Press.
- Morris, J. And G. Hirst(1991). Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text. *In Computational Linguistics, 17(1) 21-48*.
- Saint-Dizier, P. and E.Viegas (1995). *Computational Lexical Semantics*. Cambridge Uni. Press.
- Stairmand, M.A.(1994). *Lexical Chains, WordNet and Information Retrieval*.
Condensed Version of Master's Thesis
- Wood, W.H.(2004). *Developing a Learner- Centered Metathesaurus for Science, Mathematics Engineering and Technology Education*, Final Report.
- Wright, S.E. and Budin, G. (2001). *Handbook of Terminology Management*.Vol.2. Amsterdam/Philadelphia. John Benjamin