

ساختواره، اصطلاح‌شناسی، و مهندسی دانش

ملوک‌السادات حسینی بهشتی (دانشکده ادبیات و علوم انسانی، دانشگاه تهران)

مقدمه

در مهندسی دانش^۱ که به تدوین قواعد دانش فنی^۲ لازم برای حل مسائل ذهنی می‌انجامد، با برنامه‌ریزی رایانه‌ای و ایجاد شبکه اطلاعات، اطلاعات علمی - فنی مبادله می‌شود. بدین منظور لازم است که این دانش به صورت نظام‌مند و دارای ساختار ویژه درآید و این وجه اشتراک اصطلاح‌شناسی و مهندسی دانش است؛ زیرا، در اصطلاح‌شناسی، در سطح مفهوم‌سازی نظام‌مند عمل می‌شود که مقدم بر همه مراحل بیان علمی است. سازماندهی دانش فنی و ساماندهی مفاهیم علمی از مهم‌ترین مباحث اصطلاح‌شناسی است. مفهوم^۳، به حیث واحد دانش فنی، از مختصات معنایی^۳ یعنی اجزای شناخت تشکیل شده است. مفهوم بُعدی توصیفی و بُعدی تعریفی دارد. بُعد محتوایی هر مفهوم دانش اصطلاح‌شناختی آن تلقی می‌شود.

دانش اصطلاح‌شناختی، در سطحی دیگر، به صورت نظام مفاهیم درمی‌آید و، در این سطح، روابط عناصر سازنده دستگاه مفاهیم را نشان می‌دهد. در سطح نظام مفاهیم، سازماندهی دانش فنی با نظام‌مند ساختن گروهی از مفاهیم در رشته‌ای

1) knowledge engineering

2) concept

3) semantic features

تخصّصی صورت می‌پذیرد و، بر این اساس، اصطلاح‌نامه‌های تخصّصی مدوّن می‌شوند. در هر مدخل اصطلاح‌نامه، ساماندهی دانش فنی به صورت مجموعه‌ای از اطلاعات ساختاری و معنایی درمی‌آید که نظام مفاهیم و روابط آنها - مفهوم کلّ، مفاهیم جزء، و مفاهیم وابسته - را شامل است. در رشته‌های تخصّصی، سازماندهی دانش فنی به ایجاد روابط مفاهیم اکتفا نمی‌کند. مهم این است که نظام مفاهیم ساماندهی شود و نمودار گردد. به منظور درک ساختار دانش فنی حوزه‌ای تخصّصی، به ناچار مفاهیم را در نظامی کلان باید فراهم آورد و هر دسته از آنها را در طبقه‌ای جای داد. رده‌بندی^۵ مفاهیم با اهداف گوناگونی طرّاحی می‌شود. برای مستندسازی مدارک علمی - فنی، رده‌بندی مباحث علمی به صورت نظام‌های مفاهیم در قالب اصطلاح‌نامه‌های تخصّصی سازماندهی می‌شود. در اصطلاح‌نامه‌ها، کلیدواژه‌ها با توجه به شکل و محتوای معنایی بافت زبانی متون علمی تعیین می‌شوند و روابط سلسله‌مراتبی آنها به صورت کلّ، جزء، و وابسته ارائه می‌گردد. در نمایه‌سازی و بازیابی اطلاعات علمی - فنی، کلیدواژه‌ها توصیفگر^۶ مدارک علمی اند. سازماندهی اصطلاح‌شناختی دانش فنی، در انتقال دستگاه مفهومی تخصّصی، کاربردهای متفاوتی می‌یابد. در اینجا، ساختار مفهومی و دانش مفهومی در رشته‌ای تخصّصی، مقدّمه تبادل اطلاعات تخصّصی است برای فعالیتهای تخصّصی یا حلّ مسئله‌ای علمی و تبادل زبان تخصّصی برای برقراری ارتباط و انتقال اطلاعات علمی - فنی که شامل ترجمه فنی نیز می‌شود. در این کاربرد، واژه‌نامه‌ها، فرهنگ‌ها، سرعنوان‌های موضوعی، و اصطلاح‌نامه‌ها، به عنوان ابزار مستندسازی و اطلاع‌رسانی، غیرمستقیم، کلید دانش فنی در سطح مفهومی و دیگر سطوح پیچیده آن می‌شوند و استفاده از آنها برای برقراری ارتباط در زمینه دانش تخصّصی اجتناب‌ناپذیر است. همچنین سازماندهی اصطلاح‌شناختی دانش در همه سطوح، برای خلق مدوّن بی‌واسطه یا اقتباسی نظام‌های دانش‌بنیاد در سطح مفاهیم، نقش زیربنایی دارد.

4) thesaurus

5) taxonomy

6) keyword

7) descriptor

بانک اطلاعات واژگان علمی فارسی

در جامعه اطلاعاتی ایران، به رغم تلاش‌های صورت گرفته، بازیابی اطلاعات علمی - فنی و استفاده از آنها با مشکلات و کمبودهای جدی روبه‌روست. بخشی از این مشکلات به آشفتگی‌هایی مربوط می‌شود که در مرحله تولید اطلاعات وجود دارد و بخشی دیگر به نابسامانی‌هایی که در مرحله ذخیره‌سازی اطلاعات پدید می‌آید. یکی از بهترین راه‌حل‌های رفع این معضلات تهیه و کاربرد اصطلاح‌نامه یا گنج‌واژه است که هدفش برگرداندن زبان مدارک به زبان ساخته و پرداخته شده در تبادل اطلاعات و دانش فنی است - زبانی که تبادل اطلاعات تولیدکنندگان و کاربران دانش فنی را با کمترین هزینه، نیرو، و زمان و با بیشترین دقت و کیفیت میسر سازد. هرچند تاکنون در جهت تدوین و ترجمه اصطلاح‌نامه‌های تخصصی به زبان فارسی قدم‌های مؤثری برداشته شده و نتایج نسبتاً خوبی هم داشته، استفاده بهینه از اصطلاح‌نامه به عنوان ابزاری توانمند در نظام بازیابی رایانه‌ای به حد مطلوب نرسیده است.

مهندسی دانش فنی است که در آن از رشته‌های علمی زبان‌شناسی و هستی‌شناسی^۸ همچنین از مهندسی زبان^۹ استفاده می‌شود و یکی از اهداف آن ساماندهی اطلاعات علمی پراکنده است (Farghaly 2003; Georgiev 2007). در این فن، نمایه‌سازی برای سازماندهی اطلاعات علمی صورت می‌گیرد. در نمایه‌سازی، متخصصان رشته‌های علمی و فنی، زبان‌شناسان، و متخصصان علوم رایانه‌ای سهیم‌اند. اصطلاح‌نامه تخصصی ابزار کار نمایه‌ساز و چکیده‌نویس است و، با حذف آن، ورود اطلاعات کاذب و بی‌نقش به نمایه در حدی خواهد بود که ارزش اطلاعاتی آن را بسیار تنزل خواهد داد. محققان و رساله‌نویسان رشته‌های تخصصی نیز، برای ارتقای سطح علمی نوشته‌های خود، به اصطلاح‌نامه نیاز جدی دارند. آن دسته از اطلاع‌رسانان و برنامه‌نویسان رایانه‌ای که با نرم‌افزارهای نگارش متون، فرهنگ لغات، یا مترجمان هوشمند سروکار دارند از اصطلاح‌نامه‌های تخصصی، به عنوان ابزاری مؤثر 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

8) ontology

9) language engineering

نقشی شایسته ایفا کنند. اصطلاح‌نامه کارآمد حاصل گردآوری و تنظیم واژه‌های علمی و تخصصی است. سازمان‌های چکیده‌نویسی و نمایه‌سازی اطلاعات انبوه تولیدات علمی را به ناچار پردازش می‌کنند؛ در نتیجه، فرآورده این پردازش بهترین منبع واژگان تخصصی رشته‌های علمی و گنجینه واژگان برای تولید اصطلاح‌نامه‌ها شمرده می‌شود. برای تحلیل زبان‌شناختی، اعم از ساختاری و معنایی، ۳۹,۰۰۰ واژه از ۱۱۵,۰۰۰ واژه علوم پایه، از بانک اطلاعات واژگان علمی پژوهشگاه اطلاعات و مدارک علمی ایران، شامل اصطلاح‌نامه‌های علوم پایه، کشاورزی، و فنی - مهندسی، پیکره زبانی^{۱۰} اختیار شد. این مجموعه واژگان شامل اصطلاحات رشته‌های شیمی، فیزیک، زمین‌شناسی، زیست‌شناسی، و ریاضی است که بسیاری از آنها در اصطلاح‌نامه‌های این رشته‌ها مشترک‌اند. واژه‌ها از متون علمی - تخصصی استخراج شده‌اند و سابقه انتشار دارند. هر یک از این اصطلاحات به صورت مدخل درآمده و زیرمجموعه‌های آن، بر اساس نظام روابط منطقی مفاهیم، با استفاده از کلیدواژه‌های نمایه‌های متون علمی تعیین شده‌اند - متونی از قبیل 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 واژه‌نامه‌های کتب درسی دانشگاهی هر یک از رشته‌های تخصصی.

برای بالا بردن دقت پردازش آن مجموعه واژگان، مراحل زیر طی شد:

- واژه‌های مشترک تکراری حذف شد.

- واژه‌ها و اصطلاحاتی که صرفاً حرف‌نگاری واژه‌های بیگانه بوده (۹,۳۹۶ اصطلاح استخراج شده از مجموعه واژگان اصطلاح‌نامه‌ها)، شامل اسامی اشخاص، اسامی جغرافیایی، مواد شیمیایی، و اصطلاحات خاص هر یک از رشته‌های علوم پایه، تفکیک و در پرونده جداگانه‌ای ذخیره شده‌اند. مجموعه این واژه‌ها، که معادل فارسی ندارند و یا هنوز فرهنگستان زبان و ادب فارسی برای آنها معادل‌گزینی نکرده است، حاوی اطلاعات ارزشمندی است که نیاز به تحلیل جداگانه دارد. اسامی خاص، جغرافیایی، و شیمیایی که اصولاً نمی‌توانند معادل فارسی داشته باشند باید از اصطلاحات فنی

رشته‌ها جدا شوند. مجموعه اصطلاحات فنی را می‌توان به منظور معادل‌گزینی برای طرح درگروه‌های واژه‌گزینی فرهنگستان پیشنهاد کرد.

– بسامد^{۱۱} و نسبت درصد اقسام واژگان علمی فارسی شامل ۱۰,۹۵۵ اصطلاح به بسط^{۱۲} و غیربسط^{۱۳} و غیربسط به مرکب^{۱۴} (C) و مشتق^{۱۵} (D) و مشتق مرکب (DC) تقسیم، و بسامد و نسبت درصد هر یک از تقسیمات معین شد. (← جدول ۱)

جدول ۱. بسامد واژگان علمی بسط و غیربسط

فارسی	۱۰,۹۵۵	٪۵۳/۸۳
بیگانه	۹,۳۹۶	٪۴۶/۱۷
غیربسط	۷,۷۶۱	٪۷۰/۸۴
بسط	۳,۱۹۴	٪۲۹/۱۶
مشتق	۶,۸۳۵	٪۶۲/۳۹
مرکب	۹۲۶	٪۸/۴۵

در فرایند ذخیره‌سازی و بازیابی اطلاعات، واژه‌های مرجح^{۱۶} در کاربرد محققان و استادان و دانشجویان مقاطع تحصیلی تکمیلی، که کاربران اصلی اطلاعات علمی - فنی‌اند، ملاک‌گزینش در نمایه‌سازی بوده است. از این رو، بیشترین بسامد کاربردی ملاک مرجح بودن آن شمرده شده است.

– گروه واژه‌های مشتق، با استفاده از سلسله مقالات «شیوه‌ها و امکانات واژه‌سازی در زبان فارسی» (صادقی، ۱۳۷۰-۱۳۷۲) همچنین تحقیق درباره عملکرد معنایی پسوندهای فارسی بر اساس نظریه لیبر^{۱۷} (← رفیعی، ۱۳۸۶؛ Lieber 2004) تحلیل رایانه‌ای شد و بسامد کاربرد آنها در اصطلاحات علمی فارسی بر اساس فهرست پسوندها (S) و پیشوندها (P) گردآمده معین گردید. (← جدول‌های ۲ و ۳)

11) frequency

12) simplex

13) complex

14) compound

15) derived

16) preferred terms

17) R. Lieber

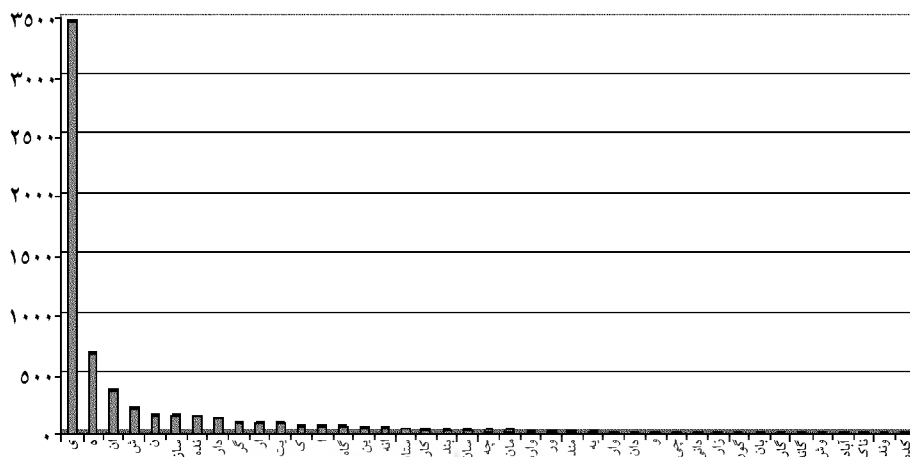
جدول ۲. بسامد پسوندهای فارسی در واژگان علمی

مان	۲۰	٪۰/۳۳	ی	۳۴۵۸	٪۰۵۶/۸۲
واره	۱۷	٪۰/۲۸	ه	۶۷۸	٪۰۱۱/۱۴
ور	۱۵	٪۰/۲۵	ان	۳۶۲	٪۰۵/۹۵
مند	۱۲	٪۰/۲۰	ش	۲۱۲	٪۰۳/۴۸
یه	۱۰	٪۰/۱۶	ن	۱۵۱	٪۰۲/۴۸
وار	۹	٪۰/۱۵	ساز	۱۴۹	٪۰۲/۴۵
دان	۸	٪۰/۱۳	نده	۱۴۲	٪۰۲/۳۳
و	۷	٪۰/۱۲	دار	۱۲۵	٪۰۲/۰۵
چی	۶	٪۰/۱۰	گر	۹۳	٪۰۱/۵۳
دانی	۶	٪۰/۱۰	ار	۸۶	٪۰۱/۴۱
زار	۶	٪۰/۱۰	یت	۸۳	٪۰۱/۳۶
گون	۵	٪۰/۰۸	ک	۶۱	٪۰۱
بان	۴	٪۰/۰۷	ا	۵۷	٪۰۰/۹۴
گار	۳	٪۰/۰۵	گاه	۵۷	٪۰۰/۹۴
گانه	۳	٪۰/۰۵	ین	۴۸	٪۰۰/۷۹
وش	۳	٪۰/۰۵	انه	۴۵	٪۰۰/۷۴
آباد	۱	٪۰/۰۲	ستان	۳۵	٪۰۰/۵۸
ناک	۱	٪۰/۰۲	کار	۳۳	٪۰۰/۵۴
وند	۱	٪۰/۰۲	بند	۲۶	٪۰۰/۴۳
کده	۱	٪۰/۰۲	سان	۲۵	٪۰۰/۴۱
			چه	۲۲	٪۰۰/۳۶

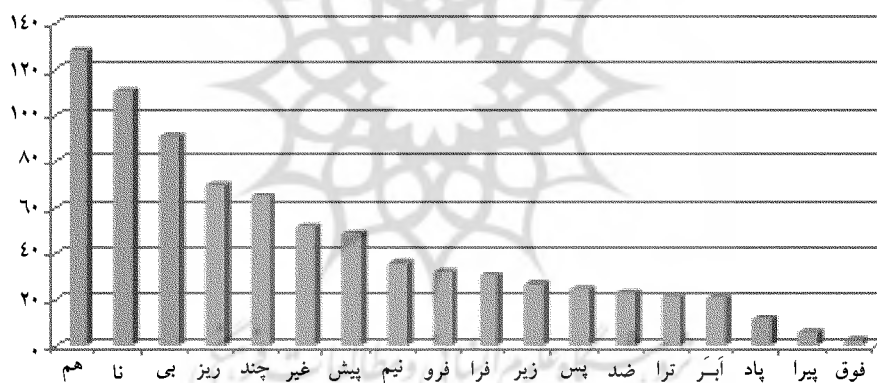
جدول ۳. بسامد کاربرد پیشنوندها در اصطلاحات علمی

فرا	۳۱	٪۰۴/۴۰	هم	۱۲۸	٪۰۱۸/۱۹
زیر	۲۷	٪۰۳/۸۳	نا	۱۱۱	٪۰۱۵/۷۷
پس	۲۵	٪۰۳/۵۵	بی	۹۱	٪۰۱۲/۹۳
ضد	۲۳	٪۰۲/۲۷	ریز	۷۰	٪۰۹/۹۴
ترا	۲۲	٪۰۳/۱۲	چند	۶۵	٪۰۹/۲۳
آبَر	۲۱	٪۰۲/۹۸	غیر	۵۲	٪۰۷/۳۹
پاد	۱۲	٪۰۱/۷۰	پیش	۴۹	٪۰۶/۹۶
پیرا	۶	٪۰/۸۵	نیم	۳۶	٪۰۵/۱۱
فوق	۳	٪۰/۴۳	فرو	۳۲	٪۰۴/۵۴

نمودار ۱. بسامد کاربرد پیشنهادهای فارسی در واژگان علمی



نمودار ۲. بسامد کاربرد پیشنهادهای فارسی در واژگان علمی

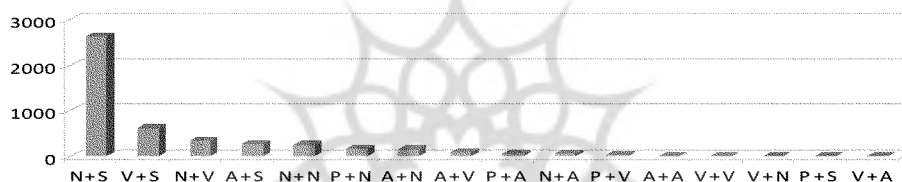


– اصطلاحات مرکب بر اساس ساختار درونی و معنایی آنها به سه گروه درون‌مرکز، برون‌مرکز، و متوازن (← طباطبائی ۱۳۸۲؛ Lieber 2004) تقسیم و اجزا و عناصر سازنده هر یک به تفکیک مقوله‌های دستوری فعل (V)، اسم (N)، صفت (A)، و قید (D) و نوع هم‌نشینی آنها معین شد. در نتیجه این تحلیل، واژه‌های مرکب، مشتق، مشتق مرکب، و ساختار درونی اصطلاحات فارسی فرمول‌بندی گردید. (← جدول‌های ۴ و ۵)

جدول ۴. بسامد واژه‌های دو‌عنصری

٪۱/۲۱	۵۷	P+A	٪۵۵/۵۸	۲۶۱۹	N+S
٪۱/۱۰	۵۲	N+A	٪۱۳/۱۸	۶۲۱	V+S
٪۰/۸۷	۴۱	P+V	٪۷/۴۵	۳۵۱	N+V
٪۰/۲۱	۱۰	A+A	٪۵/۷۵	۲۷۱	A+S
٪۰/۱۹	۹	V+V	٪۵/۵۰	۲۵۹	N+N
٪۰/۱۱	۵	V+N	٪۳/۴۴	۱۶۲	P+N
٪۰/۰۸	۴	P+S	٪۳/۲۹	۱۵۵	A+N
٪۰/۰۴	۲	V+A	٪۱/۹۹	۹۴	D+V

نمودار ۳. بسامد واژه‌های دو‌عنصری



در پژوهش حاضر، از تحلیل ساختاری اصطلاحات علمی، با برچسب‌دهی^{۱۸} عناصر سازنده واژگان علمی، حدود ۸۰ الگوی قاموسی^{۱۹} به دست آمد که بسامد برخی از اصطلاحات مرکب در کاربرد اصطلاحات علمی به شرح زیر است:

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

در ساختار اصطلاحات علمی زبان فارسی نمونه‌های سه‌عنصری 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 بسامد و کاربرد کمتری دارند.

18) tagging

19) lexical pattern

جدول ۵. بسامد واژه‌های سه‌عنصری

٪۰/۱۰	۳	V+N+S	٪۴۷/۶۵	۱۳۷۰	N+V+S
٪۰/۱۰	۳	V+V+S	٪۱۱/۳۰	۳۲۵	N+N+S
٪۰/۰۷	۲	A+S+A	٪۹/۶۳	۲۷۷	D+V+S
٪۰/۰۷	۲	A+S+S	٪۸/۹۷	۲۵۸	P+N+S
٪۰/۰۷	۲	D+V+V	٪۸/۱۷	۲۳۵	A+N+S
٪۰/۰۷	۲	N+A+V	٪۴/۵۶	۱۳۱	P+V+S
٪۰/۰۷	۲	N+P+V	٪۴/۲۱	۱۲۱	N+S+S
٪۰/۰۷	۲	P+A+N	٪۱/۱۸	۳۴	V+S+S
٪۰/۰۷	۲	P+A+V	٪۱/۱۱	۳۲	P+A+S
٪۰/۰۷	۲	P+N+V	٪۰/۶۶	۱۹	A+A+S
٪۰/۰۷	۲	V+S+N	٪۰/۴۹	۱۴	N+A+S
٪۰/۰۳	۱	A+A+N	٪۰/۳۱	۹	N+S+V
٪۰/۰۳	۱	N+N+N	٪۰/۲۸	۸	N+N+V
٪۰/۰۳	۱	V+A+S	٪۰/۲۴	۷	A+N+V
٪۰/۰۳	۱	V+P+S	٪۰/۱۰	۳	N+A+N
٪۰/۰۳	۱	V+S+V	٪۰/۱۰	۳	N+S+N

نمودار ۴. بسامد واژه‌های سه‌عنصری



توصیف‌گرهای اصطلاح‌نامه، که در نمایه‌سازی و بازیابی اطلاعات کاربرد دارند، معمولاً شامل اسم و گروه اسمی‌اند. طبق استانداردهای جهانی (ISO 2788 و ISO 5964)، صفات عموماً به عنوان اصطلاحات نمایه‌سازی پذیرفته نیستند مگر آنکه در اصطلاحات مرکب به کار روند. قیدها نیز در اصطلاح‌نامه وارد نمی‌شوند مگر آنکه جزئی ترکیبی از اصطلاحات باشند. مشتقات فعلی نیز به صورت اسم مصدر و صفت مفعولی در اصطلاحات مرکب به کار می‌روند. اصطلاحات مرکب نسبت به اصطلاحات بسیط دقیق‌تر و مشخص‌تر و برای بازنمایی مفاهیم پیچیده کارآئی بیشتری دارند. این اصطلاحات مانع ریزش کاذب اطلاعات می‌شوند.

مدیریت اصطلاح‌نامه و نمایه‌سازی رایانه‌ای

اصطلاح‌نامه یا گنج‌واژه ابزار نمایه‌سازی و بازیابی اطلاعات در شبکه اطلاعات علمی - فنی است. واژگان اصطلاح‌نامه براساس اصطلاحات برگرفته از زبان طبیعی شکل می‌گیرد، همانند کدهای هوشمند که در رده‌بندی موضوعی به کار می‌روند. در نمایه‌سازی، هم‌نشینی واژه‌ها (پس‌همارایی) معمولاً یگانه راه ارائه کلاً شامل و اجزای آن است. کاربران، برای بازیابی مدارک، اصطلاحات نمایه‌سازی را، منفرد یا گروهی و به شیوه بولین^{۲۰}، به کار می‌برند. عملیات رایانه‌ای نیز می‌تواند به گونه‌ای انجام گیرد که، به منظور حداکثر بهره‌برداری، روابط پیش‌همارایی در واژگان اصطلاح‌نامه ایجاد گردد. اصطلاح‌نامه چیزی بیش از فهرست صرف اصطلاحات و، در واقع، نظامی ساختارمند و معنی‌دار از مفاهیم است. مواد آن براساس روابط معنایی نظام‌مند شده‌اند. اصطلاح‌نامه، به منظور تضمین کارآیی، باید شمار محدودی از حوزه‌های دانش را همراه با جزئیات کاملاً تخصصی پوشش دهد. این تخصصی بودن مشخصاً در گزینش اصطلاحات ذیل اصطلاح شامل جلوه‌گر می‌شود که با پیش‌همارایی، یعنی کاربرد واژه‌های مرکب و برقراری روابط منطقی میان اصطلاحات، انجام می‌پذیرد.

اصطلاح‌نامه چندزبانه امکان دستیابی به داده‌های مستند مشترک را برای گویشوران زبان‌های گوناگون فراهم می‌سازد. اصطلاحات از چند زبان طبیعی برای یک مفهوم

گزینش می‌شوند و از نظر زبان‌شناختی مترادف و هم‌معنی‌اند. در نتیجه، روابط معنایی، بر اساس قاعده دقیقاً همانند، از زبانی به زبان دیگر انتقال می‌یابد؛ زیرا، در واقع، واژه‌ها هستند که مفاهیم را به یکدیگر ربط می‌دهند. به رغم هزینه گزاف تدوین اصطلاح‌نامه، کاربرد آن در مراکز تحقیقاتی روز به روز گسترش می‌یابد. در حوزه اصطلاحات علمی و فنی از فهرست موضوعی نیز استفاده می‌شود. اصولاً سرعنوان‌های موضوعی مانند سرعنوان‌های پزشکی^{۲۱} به لحاظ ساختاری به اصطلاح‌نامه نزدیک شده‌اند. از آنجا که اصطلاح‌نامه‌ها، به لحاظ ساختاری، با رایانه سازگارترند و می‌توانند عمل ذخیره و بازیابی اطلاعات را آسان سازند، لذا به زودی شاهد ساختاری هماهنگ به صورت فرااصطلاح‌نامه^{۲۲} خواهیم بود. بهره‌گیری از رایانه در فهرست‌نویسی ضرورت این یکپارچگی و تلفیق را دوچندان کرده است. (Wright & Budin 2001) (→)

کاربرد اصطلاح‌نامه جامع در بازیابی اطلاعات

پس از تدوین اصطلاح‌نامه‌های علوم پایه (شیمی، فیزیک، زیست‌شناسی، زمین‌شناسی، ریاضی)، فنی - مهندسی، و کشاورزی بر اساس نرم‌افزار قاموس و تکامل یافته آن (گنج‌واژه‌ساز^{۲۳}) در گروه پژوهشی اصطلاح‌شناسی پژوهشگاه اطلاعات و مدارک علمی ایران، کاربرد آنها در نظام‌های اشاعه اطلاعات مطرح شد. درباره مزایای کاربرد اصطلاح‌نامه در نظام اطلاعاتی، بحث‌های زیادی شده است. در اینجا، به مهم‌ترین اثر آن، یعنی نزدیک شدن مفاهیم مورد استفاده پژوهشگران با مدخل‌های برگزیده نمایه‌سازان، اشاره می‌شود. می‌توان گفت سه بخش از نظام‌های اشاعه اطلاعات با اصطلاح‌نامه‌ها سروکار دارند:

- نظام‌هایی که در آنها به نوعی واژه‌گزینی صورت می‌گیرد. این بخش با فرایندهایی از قبیل اشاعه‌گزینشی اطلاعات^{۲۴} (گزینش موضوعی) و درج و ویرایش اطلاعات در پایگاه‌های گوناگون سروکار دارد. در این فرایند، کاربر، برخوردار از رابطه کاربری مناسب، باید بتواند اصطلاح مورد نظر خود را در اصطلاح‌نامه و پایگاه واژگان بیابد.

21) medical subject headings (MESH)

22) metathesaurus

23) Thesaurus builder

24) selective dissemination of information (SDI)

– نظام‌هایی از واژه‌های برگزیده اصطلاح‌نامه همچون بخش ارسال مقالات SDI که، در آن، کاربر می‌تواند، با توجه به ساختار درختی موضوعی اصطلاح‌نامه، شاخه‌ای را به همراه زیرشاخه‌های آن انتخاب کند. همچنین، در تولید فصلنامه‌های موضوعی، از واژه‌هایی به عنوان استاندارد باید استفاده شود که از اصطلاح‌نامه استخراج شده‌اند.

– نظام‌هایی که در آنها اطلاعات بازیابی می‌شوند. مهم‌ترین فرایند در این عملیات جست و جو در پایگاه‌هاست که، از طریق آن، کاربران می‌توانند هر موضوعی را به همراه زیرشاخه‌های آن بیابند.

در اصطلاحاتی که وارد نمایه‌سازی می‌شوند می‌توان دخل و تصرف کرد و به حذف اصطلاحات مهجور و بدون کاربرد یا تنزل دادن درجه آنها پرداخت. سابقه هر اصطلاح ذخیره‌شدنی است و تحوّل مفهوم آن را می‌توان با ارجاع نشان داد. ذخیره واژه‌هایی که جستجوگر آنها را به دستگاه وارد می‌کند، برای روزآمد ساختن اصطلاح‌نامه و تجدیدنظر در آن، پشتوانه مناسبی است. (Aitchison et al 2000) (→)

فرا اصطلاح‌نامه

فرا اصطلاح‌نامه از ادغام و تلفیق اصطلاح‌نامه‌های متعدّد پدید می‌آید و در مجموعه واژگان ملی و بین‌المللی ارجاعات متقابل ایجاد می‌کند. برای بازیابی در پایگاه‌های چندگانه که با واژه‌های برگزیده متعدّد نمایه شده‌اند، نوعی فرا اصطلاح‌نامه تهیه می‌شود. هر اصطلاح در فرا اصطلاح‌نامه با اصطلاحات متناظر آن در همه اصطلاح‌نامه‌ها نشان داده می‌شود که، کاربر را قادر می‌سازد، با استفاده از اصطلاحات هر یک از منابع واژگانی، بازیابی را در نمایه‌های منفرد یا چندگانه و تلفیق‌شده انجام دهد. فرا اصطلاح‌نامه مفهوم بنیاد است و هدف آن ایجاد وحدت میان نام‌گذاری کاربردهای متعدّد مفهوم واحد است. این نام‌گذاری‌ها از منابع واژگانی متعدّد و رده‌بندی‌ها و اطلاعات پیرامون آن به دست می‌آید. رایانه تدوین فرا اصطلاح‌نامه را، با توجه به گسترش سریع شمار واژه‌های آن آسان می‌سازد. روزآمدی واژگان، بهره‌گیری از مجموعه واژه‌های متعدّد، الگوریتم‌گزینش و بازنگری ایجاب می‌کند که در تدوین

فرا اصطلاح‌نامه از متخصص رشته‌ای استفاده شود. اصطلاح‌نامه‌های متعدد را می‌توان، تحت شرایط ذیل، در یک اصطلاح‌نامه تلفیق کرد:

– اصطلاح‌نامه‌ها الزاماً بر اساس پیش‌زمینه نظری واحدی ساختار بندی شده باشند؛
– در آنها، شناسه‌های یکسان به کار رفته باشد و زمینه رفع مشکلات شناسایی شده فراهم آمده باشد؛

– در کیفیت دیدگاه‌ها نظارت و بر سر آنها توافق شده باشد؛
– زمینه‌های تغییر و دگرگونی ظریف و نامحسوس معانی و مفاهیم در هر یک از گرایش‌های علمی شناسایی شده باشد.

فرا اصطلاح‌نامه تنها مجموعه‌ای از اصطلاحات استاندارد شده نیست بلکه ابزاری برای افزایش اصطلاحات مفید نیز هست و می‌تواند، به عنوان منبعی غنی، مورد استفاده متخصصان قرار گیرد. با واریسی تناسب اجزای فرا اصطلاح‌نامه در هر ویرایش، گنجینه آن را می‌توان غنی‌تر ساخت. از فناوری فرا اصطلاح‌نامه برای فرمول بندی روابط معنایی ساختارهای دانش فنی از دیدگاه‌های گوناگون استفاده می‌شود. فرا اصطلاح‌نامه جست و جو و بازیابی میان رشته‌ای در گرایش‌های گوناگون علمی را میسر می‌سازد. فرا اصطلاح‌نامه شامل مجموعه‌ای از مفاهیم کلیدی، روابط معنایی مفاهیم (مترادف‌ها، همنام‌ها و نظایر آنها)، نمودارهایی معنایی از اصطلاح‌نامه‌ها و مفاهیم کلیدی است که معدودی از آنها تألیف یا ابداع شده‌اند. (Wood 2004) →

هستی‌شناسی و شبکه‌واژگانی^{۲۵}

در جمع‌آوری مجموعه‌ای از دانش واژگانی شامل شبکه‌واژگانی، اصطلاح‌نامه‌ها، و واژه‌نامه‌ها و استفاده از آنها در تشکیل زنجیره‌های واژگانی و ایجاد ارتباط رایانه‌ای و خودکار، تحقیقات دقیق‌تری باید انجام پذیرد. ایجاد پایگاه دانش واژگانی مانند شبکه‌واژگانی، برای نشان دادن فرایند تکوین (هستی‌شناسی) متنی علمی به‌ویژه در ارتباط با پایگاه اطلاعات علمی - فنی، می‌تواند ابزار کارآمدی باشد.

نقشه مفهومی^{۲۶} بر اساس سلسله‌مراتب واژه‌ها مدون می‌شود. هستی‌شناسی غالباً مفاهیمی را در برمی‌گیرد که برای شناخت انسان از جهان پیرامون اساسی و بازنمون دانش‌پیش‌نمونه‌ها^{۲۷} باشند. این که ساختار هستی‌شناسی بر اساس شامل و مشمول و یا سطوح بالا و پایین طراحی شود، بیشتر به اهداف و حوزه مورد نظر بستگی دارد تا حل مشکلات مدیریت اصطلاحات؛ زیرا میان این دو نوع هستی‌شناسی تفاوت ذی‌نقشی وجود دارد. در هستی‌شناسی بر اساس سلسله‌مراتب و ویژگی‌های معنایی، سعی بر آن است که حوزه معنایی خاصی مبتنی بر مفاهیم دقیق تعریف شده به دست آید. به عنوان مثال، در حوزه علوم پایه (ریاضی، فیزیک، شیمی، زیست‌شناسی، و زمین‌شناسی) فرمول‌بندی دانش به راحتی صورت می‌گیرد؛ زیرا، در این حوزه، نظام مفاهیمی حاکم است که می‌توان آنها را با اصطلاحات علمی دارای تعریف دقیق بیان کرد. در چنین مواردی، سعی می‌شود اشیاء در عالم مقال و انتزاعی با الگوهای مفهومی پیراسته و ساده‌شده‌ای بازنموده شوند. (→ GÓMEZ-PÉREZ et al 2004)

پایگاه‌های اطلاعاتی واژگانی از جمله شبکه واژگانی شامل واژگان، اصطلاح‌نامه‌ها و لغت‌نامه‌ها به قاعده‌مند ساختن معنی‌شناسی واژگانی می‌پردازند. معانی مبتنی‌اند بر شبکه مفاهیم و دریافت مشترک اکثریت جامعه علمی. مفاهیم تعاریفی دارند دربرگیرنده معانی متداول و کاربردهای ویژه و ترکیبات گوناگون مبتنی بر پیش‌نمونه‌ها و مفاهیم پایه‌ای علوم.

شبکه معنایی^{۲۸}

دیویس و همکاران (→ DAVIES et al 2007) نگاه تازه‌ای به آینده شبکه جهانی اطلاعات^{۲۹} دارند که، در آن، تمامی دانش موجود به صورت قاعده‌مند رمزگذاری می‌شود تا عوامل هوشمند بتوانند، خودسامان، خدمات اطلاعاتی ارائه دهند. لازمه سازمان دادن چنین شبکه‌ای آن است که از هستی‌شناسی‌های مبتنی بر «وب» تعریفی دقیق به دست داده شود که بیانگر دانش قابل فهم انسان و عوامل هوشمند باشد - دانشی که

26) concept map

27) prototype

28) semantic web

29) world wide web

ارزش افزوده و پرورش آن به صورت فراداده‌ها^{۳۰} نمودار خواهد شد و آن، در حال حاضر، در مجموعه‌های محدودی از تقاضاها تعریف شده و به کار بردنی است. پیش‌بینی می‌شود که، در دهه‌های آتی، ساختار دانش مطلوب بیشتر به صورت رسمی و نظام‌های هستی‌شناسی مبتنی بر وب درآید و ارائه آن به زبان‌های گوناگون زمینه‌ای فعال باشد برای تحقیقات و مطالعات بیشتر و شمار قابل ملاحظه‌ای از طرح‌های تحقیقاتی و تصویب استانداردها در قالب چارچوب توصیف منبع (RDF)^{۳۱} و زبان نشانه‌گذاری توسعه‌پذیر (XML)^{۳۲} و نقشه‌های معنایی به عنوان زبان‌های بیان دانش. XML قادر به بازنمایی پدیده‌هاست اما نمی‌تواند ساختار معنایی مفاهیم دربرگیرنده آنها را بیان کند؛ در حالی که RDF قادر است مؤلفه‌های معنایی را متمایز سازد. برای تمیز فاعل و مفعول و اقسام فعل می‌توان از نشانگر متحدالشکل منبع (URI)^{۳۳} استفاده کرد.

شبکه معنایی مجزا از وب جهانی نیست بلکه پرورده و تکامل یافته آن است و اطلاعات، در آن، با تحلیل معنایی و برای هماهنگی بیشتر انسان و رایانه ذخیره شده است. ویژگی وب جهانی بودن آن است که ایجاد شبکه معنایی و بیان روابط مفاهیم گوناگون را مقدور می‌سازد. شبکه معنایی، همچون اینترنت، به مکان خاصی محدود نخواهد بود. این شبکه، برای عملکرد دقیق‌تر رایانه، به ساختارمند کردن اطلاعات و اعمال قواعد معنایی نیاز دارد. بدین منظور، از الگوهای یکسان معنایی باید استفاده شود. توانایی تحلیل معنایی این شبکه امکان دسترسی به اطلاعات بسیار را فراهم می‌سازد. از این رو، این شبکه به زبان خاصی نباید تعلق داشته باشد بلکه باید بر اساس نقشه‌های مفهومی عمل کند. در این شبکه، از نظام‌های گوناگون بازنمایی دانش استفاده می‌شود و، با بهره‌گیری از شیوه‌های بیان ریاضی و مهندسی هستی‌شناسی، منطق آن، برای توصیف عناصر سازنده یک پدیده یا مؤلفه‌های معنایی، از توانایی بالایی برخوردار است و تحت حاکمیت زبانی خاص و، به تبع آن، فرهنگی خاص قرار نمی‌گیرد.

30) metadata

31) resource description framework

32) extensible markup language

33) uniform resource identifier

باری، جوامع گوناگون علمی علاقه‌مندند مفاهیم اصلی حوزه فعالیت خود را تعریف کنند و ثبات بخشند. در عین حال، روابط واژگان و ارتباطات معنایی که به صورت اصطلاحات استانداردسازی شده‌اند نمودار مقید بودن این جوامع به کاربرد مفاهیم و زبان مشترک در سازمان‌ها و در تبادل اطلاعات از طریق عوامل هوشمند در شبکه معنایی‌اند.

نتیجه

با توجه به نظریات جدید در حوزه مهندسی اصطلاح‌شناسی، به نظر می‌رسد لازم باشد اقداماتی به شرح زیر در زمینه فعالیت‌های واژه‌گزینی فرهنگستان زبان و ادب فارسی در دستور قرار گیرد:

- تلفیق اصطلاح‌نامه‌های علوم پایه و تشکیل بانک اطلاعات واژگان علمی؛
- تحلیل ساختاری و معنایی واژگان علوم پایه و استخراج واژه‌های مرکب، مشتق، و مشتق مرکب و محاسبه بسامد کاربرد آنها برای بازشناسی واژه‌های مرجح؛
- استخراج الگوهای ساختوازی واژگان علمی زبان فارسی و تعیین مقوله‌های دستوری و زیرمقوله‌های آنها؛
- ایجاد ارتباطات مفهومی در نظام مفاهیم و صورت‌بندی هستی‌شناختی آنها؛
- فراهم ساختن زمینه تدوین شبکه واژگانی و شبکه معنایی.

منابع

- رفیعی، عادل، نگاهی به عملکرد معنایی پسوند‌های اشتقاقی زبان فارسی، رساله دکتری دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبائی، تهران ۱۳۸۶.
- شقایق، ویدا، مبانی صرف، انتشارات سمت، تهران ۱۳۸۶.
- صادقی، علی‌اشرف، «شیوه‌ها و امکانات واژه‌سازی در زبان فارسی معاصر» (۱-۱۱)، نشر دانش، سال‌های یازدهم تا چهاردهم (۱۳۷۰-۱۳۷۲).
- طباطبائی، علاء‌الدین، اسم و صفت مرکب در زبان فارسی، مرکز نشر دانشگاهی، تهران ۱۳۸۲.

Airchison, J., A. Gilchrist and A. Bowden (2000), *Thesaurus Construction. A Practical Manual*, Aslib, London.

- DAVIES, J., R. STUDER and P. WARREN (2007), *Semantic Web Technologies*, John Wiley & Sons, New York.
- FARGHALY, A. (ed.) (2003), *Handbook for Language Engineers*, CSLI Publication, Stanford.
- GEORGIEV, H. (2007), *Language Engineering*, Continuum, London.
- GÓMEZ-PÉREZ et al. (2004), *Ontological Engineering*, Springer, New York.
- LIEBER, R. (2004), *Morphology and Lexical Semantics*. Cambridge University Press, Cambridge.
- WOOD, W. H. (2004), *Developing a Learner-Centered Metathesaurus for Science, Mathematics Engineering and Technology Education*, Final Report, (http://best.berkeley.edu/pubs/04_1202_P.pdf).
- WRIGHT S. E. and BUDIN, G. (2001), *Handbook of Terminology Management*, Vol. 2, John Benjamin, Amsterdam/ Philadelphia.

