

کاربرد اصطلاح‌شناسی و واژه‌گزینی در نمایه‌سازی ماشینی و بازیابی اطلاعات

نوشته: ملوک‌السادات حسینی بهشتی*

چکیده

در عصر اطلاعات، اطلاعات علمی-فنی زیربنای هرگونه تصمیمات جدی اجرایی و تحقیقاتی محسوب می‌گردد. موج اطلاعات در سالهای اخیر به شکل چشمگیری گسترش یافته و در زمانی قابل پیش‌بینی میزان این اطلاعات رشد سریع‌تری خواهد یافت. از این رو ضرورت سازماندهی این گونه اطلاعات اهمیت بیشتری می‌یابد در این میان آنچه حائز اهمیت است روشهای دستیابی به آن می‌باشد. اطلاعات مرتبط بایستی به سرعت و در حداقل زمان قابل دسترسی گردد. سیستمهای رایانه‌ای شامل نظام مدیریت اطلاعات، نظام خبره، بانکهای اطلاعاتی و نظامهای پردازش و بازیابی اطلاعات است. هر چند نظامهای اطلاعاتی گوناگونی در این زمینه مورد استفاده قرار می‌گیرند لکن در طراحی آنها مشکلات و سئوالات مشترکی مطرح می‌باشند:

۱- چگونگی سازماندهی دانش موجود و نتایج

تحقیقات علمی

۲- چگونگی بازیابی اطلاعات و هدایت کاربر در

جهت استفاده صحیح از نظام اطلاعاتی

بنابراین هر نظام بازیابی اطلاعاتی متشکل از انبوهی از اسناد و مدارک و نیز حجم قابل توجهی از پرسش‌ها و سئوالات است و شیوه‌هایی را دربر می‌گیرد که به تناسب هر یک از پرسش‌های جستجوگر، اسناد و مدارک متناسب و مرتبط با آن را بازیابی و در اختیار قرار دهد. این شیوه عبارتند از:

۱- ذخیره اطلاعات (به کارگیری عناصری جهت

بازیابی و ارائه اطلاعات ذخیره شده)

این بخش محتوای مدرک علمی را توصیف نموده و به شکلی ارائه می‌دهد که براساس ویژگیهای آن اسناد و مدارک بازیابی می‌شوند. یکی از این روشها، کاربرد اصطلاحنامه و تعیین توصیفگرها براساس نظام منطقی مفاهیم و ارائه محتوای مدارک است.

* عضو هیئت علمی مرکز اطلاعات و مدارک علمی ایران

۲- بازیابی اطلاعات (به کارگیری عناصری جهت**تعیین نیازهای اطلاعاتی کاربران)**

این بخش می‌تواند سؤالات و نیازهای اطلاعاتی استفاده‌کنندگان را به گونه‌ای شکل دهد که بواسطه اطلاعات حاصل از نیازسنجی استفاده‌کنندگان، مدارک مناسب بازیابی شوند.

۳- اشاعه اطلاعات (به کارگیری عناصری جهت ارائه**و اشاعه اطلاعات بازیابی شده)**

در این بخش نتایج حاصل از جستجوی اطلاعات مورد نیاز کاربر در اختیار وی قرار می‌گیرد.

در این مقاله به طور ویژه، مبحث واژه‌گزینی در نمایه‌سازی مدارک و اسناد به منظور بازیابی اطلاعات و در نتیجه فرآیند ترجمه اسناد به زبان نمایه‌سازی و کاربرد اصطلاحنامه‌ها به عنوان پشتوانه نمایه‌سازی ماشینی با رویکرد به فعالیتهای اصطلاح‌شناختی در مرکز اطلاعات و مدارک علمی ایران مورد بحث و بررسی قرار می‌گیرد.

کلیدواژه‌ها: اصطلاح‌شناسی / واژه‌گزینی / نمایه‌سازی ماشینی / بازیابی اطلاعات

مقدمه

از دیرزمان اصطلاحات و اصطلاح‌شناسی^۱ نقش تعیین‌کننده‌ای در تبادل نظرات علمی میان متخصصان رشته‌های گوناگون ایفاء نموده و پیامدهای مفیدی را دربر داشته است. منجمله:

۱- شناخت اهمیت اصطلاحات جهت تبادل نظرات علمی-فنی؛

۲- آگاهی بر ضرورت نظام‌مند نمودن این اصطلاحات بعنوان ابزار شناخت علوم و زبانهای تخصصی، که بدون هماهنگی و دخل و تصرف بوجود آمده‌اند و حاصل آن رخدادی چون تشتت آراء در شکل‌گیری مفاهیم و نامگذاری آنها و نیز عدم شکل‌گیری نظام مفاهیم بوده که بعنوان عناصر زیربنایی زبانهای

تخصصی و در نتیجه ساختار یک گرایش علمی محسوب می‌گردند.

مهمتر از همه، انگیزه متخصصان رشته‌های علمی که نیاز روزافزون به داشتن ابزاری غنی و کارآمد جهت تبادل اطلاعات علمی، بویژه ساخت اصطلاحات و هماهنگ‌سازی و یکدست نمودن آنها سبب پیشرفت علم اصطلاح‌شناسی در قرن گذشته شده و دو گرایش عمده را پدید آورده است:

۱- توسعه و هماهنگی اصول و مبانی نظری اصطلاح‌شناسی

۲- تحقیقات پیرامون زبانهای تخصصی براساس مبانی زبانشناختی

نیاز روزافزون به ترجمه متون تخصصی و کاربرد دستگاههای الکترونیکی سبب گردیده است تا اولین بانکهای اطلاعات واژگانی و اصطلاح‌شناسی در مؤسسات ترجمه ملی و بین‌المللی دایر گردند.

بموازات این فعالیتهای تحقیقات میان رشته‌ای در زمینه‌های علمی مشترک میان اصطلاح‌شناسی و علم اطلاع‌رسانی با هدف ساخت ابزارها و شیوه‌های کارآمد جهت تبادل سریعتر و دقیقتر اطلاعات علمی، آغاز گشت.

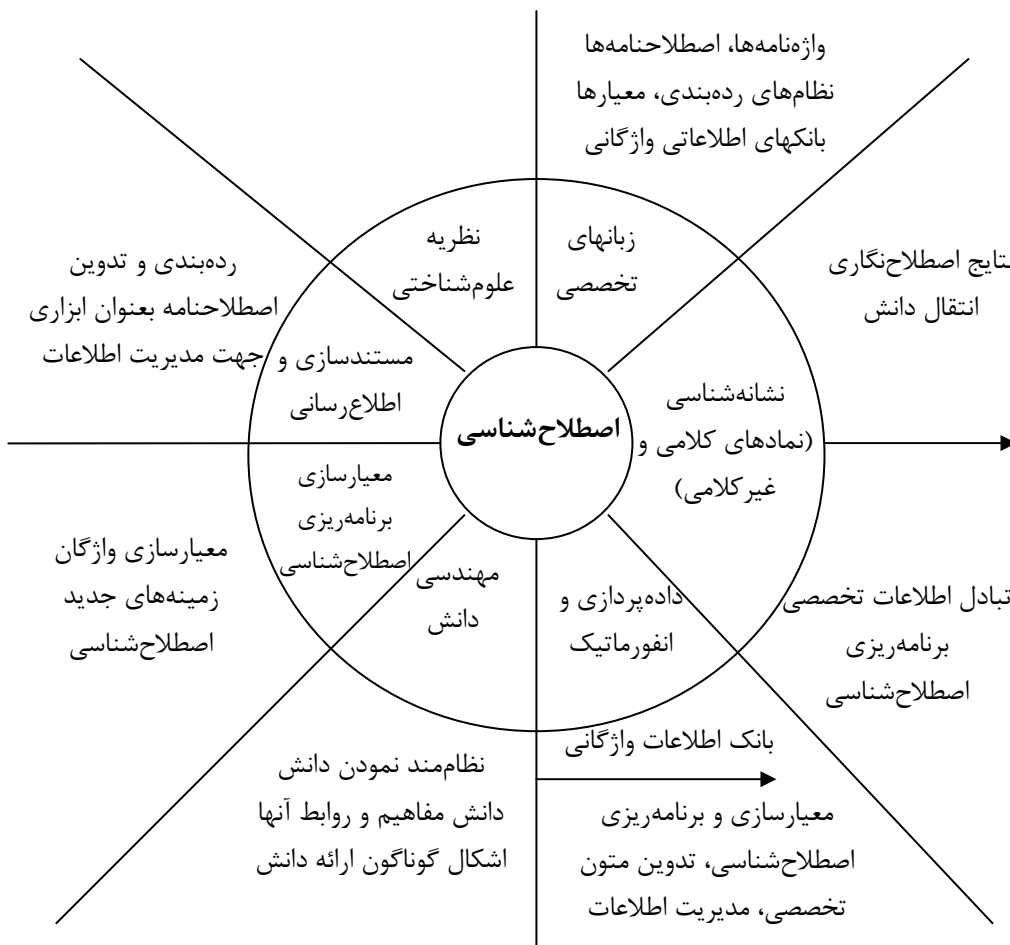
مستندسازی و اشاعه اطلاعات علمی-فنی بطور روزافزون نیاز به تحقیقات اصطلاح‌شناسی یافت و فن‌آوری اطلاعات نیز در زمینه پیشرفت علم اصطلاح‌شناسی به کار گرفته شد، در نتیجه ضرورت همکاری متقابل این دو زمینه علمی آشکار گردید.

از آخرین نتایج تحقیقات «ووستر»^۲ بنیانگذار علم اصطلاح‌شناسی تاکنون بیش از ۲۰ سال می‌گذرد و باید اذعان داشت که پیشرفت علم اصطلاح‌شناسی و ارائه نظرات جدید در رشته‌های مختلف علمی، در علم اصطلاح‌شناسی نیز بویژه در زمینه تدوین شیوه‌های جدید، گامهای اساسی برداشته است، بطوری که زمینه‌های دیگر علمی را نیز فراگرفته است مانند:

علوم شناختی^۳، منطق و هستی‌شناسی^۴، اصطلاح‌نگاری^۵، داده‌پردازی^۶، علوم اطلاع‌رسانی^۷، مدیریت اطلاعات^۸ و مهندسی دانش^۹.

ارتباط میان علم اصطلاح‌شناسی و دیگر علوم را

می‌توان به شکل ارتباطات متقابل ذیل نشان داد:



بانکهای اطلاعات واژگانی و برنامه‌ریزی زبان و اصطلاح‌شناسی بکار گرفته می‌شوند.

داده‌پردازی^{۱۴} و انفورماتیک^{۱۵}

ماشینی شدن ثبت واژگان و اصطلاح‌نگاری رایانه‌ای تسهیلاتی را در این امر بهمراه داشته است:

- روزآمد کردن سریع داده‌های واژگانی و دسترس‌پذیر نمودن آنها برای کاربران
 - ایجاد امکان بکارگیری سریع این اصطلاحات در بانکهای اطلاعات علمی- فنی
- هرچند داده‌پردازی و انفورماتیک بطور مستقیم در مبانی نظری اصطلاح‌شناسی بمعنی اخص آن دخیل نبوده، اما کاربرد آنها در تحقیقات واژه‌شناسی و اصطلاح‌شناسی در بسیاری از زمینه‌ها تأثیر فراوان داشته است. از قبیل:

نتایج حاصل از فعالیت اصطلاح‌شناسی در زبانهای تخصصی بصورت واژه‌نامه، بانکهای اطلاعات واژگانی، معیارها، رده‌بندی موضوعی و اصطلاحنامه‌ها ارائه می‌شود. (لارن، مایکینگ، پیش، ۱۹۹۸)

نشانه‌شناسی^{۱۰}

در اصطلاح‌شناسی برای ارائه و بیان اشیاء و مفاهیم از نمادهای کلامی و غیر کلامی استفاده می‌گردد. بدین منظور نیاز به شناخت مباحثی از قبیل واژه‌شناسی^{۱۱}، معنی‌شناسی^{۱۲}، واژگان‌نگاری^{۱۳}، روانشناسی و جامعه‌شناسی زبان (ساخت واژه و نامگذاری مفاهیم در واژه‌نامه‌های تخصصی، پذیرش اصطلاحات از سوی گویشوران زبان و غیره) ضروری است که در تدوین متون تخصصی و ترجمه شفاهی و کتبی و نیز شکل‌گیری

ارتباط میان اصطلاح‌شناسی و معیارسازی بصورت رسمی و سازمان یافته در اواخر قرن ۱۹ و اوایل قرن ۲۰ میلادی صورت گرفت و مستندات فراوانی در این زمینه وجود دارد.

در اینجا بایستی توجه داشت که معیارسازی نیاز به تغییر و تحولات بنیادی و پذیرش تقاضاها و روابط جدید را در پی دارد که لازمه آن بررسی و شناخت مداوم مبانی نظری و شیوه‌های کاربردی است که در دهه‌های هشتاد و نود قرن بیستم میلادی در سطح ملی و بین‌المللی ابعاد گسترده‌ای یافته است. بموازات این فعالیتها یک نوع هماهنگ‌سازی و معیار نمودن شیوه‌ها در گروههای کوچکتر کاری و تخصصی نیز با هدف ایجاد ابزار ارتباطی مشترک و واحد در حال انجام است. جوامع بزرگتر زبانی نیز به این نتیجه رسیده‌اند که چنانچه بخواهند زبان بومی آنان در انتقال دانش که اکنون از مرزها عبور کرده و با سرعت فراوان وارد جوامع گوناگون بشری می‌شود، مؤثر واقع گردد، بایستی در نامگذاری مفاهیم جدید علمی-فنی با همان سرعت هماهنگی و یکدستی ایجاد گردد تا متخصصان رشته‌های علمی بتوانند از زبان بومی خود جهت تبادل نظر و دانش فنی بهره کافی ببرند.

مستندسازی و اطلاع‌رسانی

ارتباط این رشته علمی با اصطلاح‌شناسی از قدمت فراوان برخوردار است. در ابعاد نظری این دو رشته بواسطه طبقه‌بندی و رده‌بندی دانش به یکدیگر نزدیک شده‌اند.

تحقیقات وسیع پیرامون روابط میان اشیاء و مفاهیم و نیز نامگذاری آنها را شامل می‌شود که در علوم اطلاع‌رسانی بعنوان ابزار کار طبقه‌بندی و تدوین اصطلاحنامه ظهور می‌یابند.

از سوی دیگر در جستجوی اطلاعات بمنظور یافتن مفاهیم علمی و نامگذاری آنها، کاربرد اصطلاح‌شناسی و هماهنگ‌سازی اصطلاحات علمی-فنی ضرورت می‌یابد.

- شیوه‌های کاربرد اصطلاح‌نگاری
- تعاریف عناصر و اجزاء اصطلاح‌شناسی
- سازماندهی این عناصر در قالبهای رایانه‌ای
- تسهیل در تبادل اطلاعات داده‌های واژگانی
- معیارسازی شیوه‌های اطلاع‌رسانی
- کاربرد معیارها در برنامه‌ریزی زبان و اصطلاح‌شناسی
- شکل‌گیری بانکهای اطلاعات واژگانی، تدوین متون تخصصی و جستجوی اطلاعات از راه دور و تسهیل در امر ترجمه و تبادل اطلاعات علمی-فنی

مهندسی دانش

از آغاز ارتباط میان اصطلاح‌شناسی و مهندسی

دانش حدود یک دهه می‌گذرد.

اولین ارتباط این دو حوزه علمی در سال ۱۹۸۷ در کنگره «**اصطلاح‌شناسی و مهندسی دانش**» در شهر «**تبریز**» اتریش برقرار شد. در این کنگره مقالات فراوانی با هدف بررسی ارتباط میان این دو رشته علمی ارائه گردید. از آنجا که «**مهندسی دانش**» به دنبال این هدف است که دانش را به کمک برنامه‌ریزی رایانه‌ای و شبکه‌**اطلاعاتی** انتقال دهد، لازم است که این دانش بصورت نظام‌مند و دارای ساختار ویژه درآید. و همین نقطه مشترک میان این دو رشته علمی محسوب می‌گردد، زیرا اصطلاح‌شناسی در سطح ارائه مفاهیم بصورت نظام‌مند عمل می‌کند که این خود مقدم بر تمامی شیوه‌های ارائه دانش محسوب می‌گردد. بنابراین نقطه اشتراک این دو علم، دریافت علم و ارائه دانش بصورت **نشانه‌گذاری** (نمادهای کلامی و غیر کلامی) می‌باشد.

معیارسازی و برنامه‌ریزی زبان

این توسعه پشتیبانی فن‌آوری اطلاعات را می‌طلبد و تلفیق نظریه اصطلاح‌شناسی و فلسفه علم را با مهندسی دانش ضروری می‌نماید.

انتقال از پردازش داده‌ها بسوی پردازش دانش به این نگرش منجر شد که اجزاء زیربنایی دانش از داده‌های منفرد تشکیل نیافته است، بلکه بیشتر از یک ساختار پیچیده و ترکیبی شکل یافته، بدین معنی که مفاهیم بعنوان مجموعه‌ای از ویژگیها با همدیگر ارتباط یافته، بصورت یک زمینه موضوعی با روابط نظام‌مند مفاهیم گوناگون آن تحقق می‌یابد.

این گونه پردازش دانش «مفهوم‌بنیاد»، مزیت‌های روشنی نسبت به پردازش معنی‌شناختی اطلاعات دارد و وظیفه‌اش خلاصه نمودن تمامی معانی و تعاریف زبان‌شناختی به شکل ارائه الگوهای نحوی می‌باشد.

براین اساس، اصطلاح‌شناسی، علمی است میان رشته‌ای، که موضوع تحقیق آن از سویی به شیء و مفهوم، شیوه تعریف و نامگذاری مفاهیم و از سوی دیگر به سازماندهی مفاهیم یک رشته به شکل نظام‌مند و کاربرد آنها در زمینه‌های گوناگون علمی مربوط است.

امروزه اصطلاح‌شناسی بعنوان یک گرایش مستقل و با دیدگاه شناختی ویژه و هدفگذاری زمینه‌ها و کاربردهای ویژه مطرح می‌گردد و هیچ یک از رشته‌های وابسته این علم نمی‌توانند مشکلات ارتباطات را بدون تأثیر و کمک رشته‌های دیگر حل نمایند و بنابراین میان آنها نوعی وجوه اشتراک دیده می‌شود.

همچنین باید تأکید گردد که مشارکت تک‌تک این علوم بصورت اجزاء و یک کل، رسیدن به هدف نهایی اصطلاح‌شناسی یعنی ایجاد ارتباطات تخصصی را امکان‌پذیر می‌نماید و سهولت می‌بخشد.

اصطلاحنامه در نظام‌های بازیابی اطلاعات

اصطلاحنامه، مجموعه‌ای از واژگان زبان نمایه‌ای کنترل شده است که بطور منظم سازمان یافته تا روابط استنباطی میان مفاهیم را مشخص نماید. اصطلاحنامه در نظام‌های بازیابی اطلاعات، از فهرست‌نویسی^{۱۹} در مراکز

در «مدیریت اطلاعات» نیز اصطلاح‌شناسی اهمیت یافته زیرا در این حوزه نیز بدون هماهنگ‌سازی اصطلاحات تخصصی، ساماندهی فعالیتها مشکل می‌نماید. از اینرو «اصطلاح‌شناسی» در مدیریت کارآمد انبوه اطلاعات روزافزون ویژگی کلیدی می‌یابد، بویژه آنکه امروزه با شیوه‌های ماشینی متداول به اجرا درمی‌آید.

علوم شناختی

مدت زمان کوتاهی است که ارتباط مستقیم میان علوم شناختی و اصطلاح‌شناسی ایجاد شده است. از آنجا که علوم شناختی در روشن نمودن روابط میان تک‌تک علوم می‌کوشد تا جایگاه علوم جدید را در ساختار کلی علوم بیابد، در علم اصطلاح‌شناسی نیز اینگونه مسائل مورد بحث و بررسی قرار گرفته و در این زمینه در گردهمایی‌های ۱۹۶۹ و ۱۹۷۱ که بدین منظور در مسکو برگزار گردید مقالات بر محور: جایگاه اصطلاح‌شناسی در نظام علوم و مشکلات نشانه‌شناسی در زبانهای تخصصی، اصطلاح‌شناسی و اطلاع‌رسانی ارائه گردید.

ارتباط میان علوم شناختی و اصطلاح‌شناسی در سالهای اخیر توسط دانشمندان رشته فلسفه علم یعنی «اوزر»^{۱۶} مورد بحث و بررسی قرار گرفت. در مقاله‌ای که در سال ۱۹۹۲ ارائه نمود، روابط تنگاتنگ اصطلاح‌شناسی و علوم شناختی را بیان نمود و اثبات کرد که منطق مفاهیم و بیان مفهوم پیشینه این علم محسوب می‌گردند و منجر به «فرا اصطلاح‌شناسی»^{۱۷} می‌گردد. همچنین ارتباط میان منطق بویژه منطق مفاهیم، فلسفه علم، فلسفه زبان و علوم شناختی و تأثیر آنها بر «تحول مفاهیم»^{۱۸} و فرا اصطلاح‌شناسی را اثبات می‌نماید. (لارن، مایکینگ، پیش، ۱۹۹۸)

این شیوه نظارت بر تحول مفاهیم نشانگر نوآوری قابل توجه در تحقیقات اصطلاح‌شناسی است، زیرا زمینه‌ای موردنیاز و ضروری برای پردازش انبوه سیل‌آسای اطلاعات، بعنوان فرآیند علمی شتاب دهنده در توسعه تکنولوژی جدید اطلاعات محسوب می‌گردد.

با گسترش رو به رشد نظامهای اطلاعاتی درونخطی شامل اینترنت که از متن آزاد به عنوان ابزار بازیابی استفاده می‌کنند (یعنی زبان طبیعی و واژه‌های عناوین، چکیده‌ها و متن کامل اسناد) مفید بودن زبان کنترل شده (یعنی نظامهای رده‌بندی، سرعنوان‌های موضوعی و اصطلاحنامه‌ها) باید مورد ارزیابی مجدد قرار گیرد. نقش اصطلاحنامه در حال تغییر است، اما احتمال دارد به عنوان ابزار بازیابی کارآمد باقی بماند و تغییر در برخی ویژگی‌های آن و ایجاد تحول در شیوه استفاده از این ویژگی‌ها، حاکی از گسترش استفاده مفید اصطلاحنامه خواهد بود تا کاهش آن. (ایچیسن، گیلکریست، بادن، ۲۰۰۰)

در نظام‌های نمایه‌سازی دستی، مفید بودن زبان کنترل شده چه در نظامهای پیش‌همارا و چه در نظامهای پس‌همارا به اثبات رسیده است. زیرا نظامهای دستی قابلیت پرداختن به زبان پردازش نشده را به خاطر تعداد گسترده واژگان ندارند. بنابراین اصطلاحنامه در نمایه‌های چاپی و رایانه‌ای برای کنترل مترادف‌ها و ایجاد مبنای ارجاعات کاربرد دارد. در نظامهای پس‌همارای ماشین‌خوان زبان طبیعی اغلب تنها زبان بازیابی است، اما پایگاههای اطلاعاتی وجود دارند که از زبان کنترل شده استفاده می‌کنند و یا از زبانهای طبیعی و کنترل شده توأم بهره می‌جویند. در برخی نظامهای ترکیبی، زبان کنترل شده آسان و غیراخص عمدتاً به عنوان ابزار بهبودی جامعیت عمل می‌کند، در صورتی که در نظامهای دیگری زبان کنترل شده نسبتاً اخص و پیچیده است.

با گسترش پایگاههای اطلاعاتی تمام‌متن، اصطلاحنامه نقش جدیدی در زمینه نمایه‌سازی و بازیابی اطلاعات ایفا می‌کند. نمایه‌سازی واژگان کنترل شده ممکن است برای تقویت نظامهای بازیابی هوشمند تمام متن با فنون زبانی و اطلاعات معناساختی و آماری مورد استفاده قرار گیرد.

چنانچه پایگاه اطلاعاتی کوچک بوده و به کندی رشد می‌کند، یک اصطلاحنامه گسترده و مفصل، فقط تجملی و گرانبه محسوب می‌شود. مگر آنکه حوزه موضوعی تحت پوشش، بسیار پیچیده باشد. در نظامهای تمام متن میزان داده‌های پردازش شده و قابل دسترس

اطلاع‌رسانی و کتابخانه‌ها تا اینترنت، مورد استفاده قرار می‌گیرد. هدف اصلی اصطلاحنامه، بازیابی اطلاعات (نمایه‌سازی^{۲۰} و کاوش^{۲۱}) است و در درجات بعد کمک به درک کلی یک حوزه موضوعی، فراهم کردن «نقشه‌های معنایی»^{۲۲} از طریق ترسیم نظام مفاهیم و کمک در ارائه تعاریف اصطلاحات است. همچنین تولید فهرست‌های کلیدواژه‌ای^{۲۳} که مبنای برنامه‌ریزی، اولویت‌بندی و سایر وظایف مدیریت پژوهش را تشکیل می‌دهند.

هم اکنون آنچه رایج‌تر است استفاده گردآورندگان پایگاههای اطلاعاتی از اصطلاحنامه در نمایه‌سازی رایانه‌ای است و کاربرانی که از اصطلاحنامه بعنوان ابزار بازیابی استفاده می‌کنند از آن برای جستجوی اطلاعات بهره خواهند برد. این شیوه در استفاده از نظامهای درونخطی و لوح فشرده توسط رابط‌های متخصص مرسوم است. در هر صورت اصطلاحنامه باید در شکل چاپی و یا به شکل رایانه‌ای در اختیار تمام کاربران قرار گیرد.

شکل چاپی آن مستلزم انتخاب فکری اصطلاحات است و شکل رایانه‌ای امکان‌پذیرش خودکار را فراهم می‌کند. کاربرد اصطلاحنامه سنتی بر روی اینترنت و برخی فراشبکه‌ها از تحولات جدید است و پیشرفتی طبیعی برای تولیدکنندگان پایگاههای اطلاعاتی محسوب می‌شود. نسل جدیدی از نرم‌افزارها با استفاده از فنونی نظیر استنباط زبانشناختی (عمدتاً با بکارگیری تحلیل آماری و قواعد معناساختی) هم اکنون قادر به گزینش موضوعی خودکار یا تحلیل متن اسناد در یک پایگاه اطلاعاتی هستند هرچند هنوز مداخله انسان کماکان در آموزش نرم‌افزارها، ویراستاری و نظم بخشی به نتایج ضرورت دارد. بنابراین اصول اساسی رده‌بندی موضوعی و گردآوری واژگان اصطلاحنامه همچنان معتبر است.

ارزش کاربرد اصطلاحنامه در نمایه‌سازی، گردآوری مجموعه‌ای غنی از اصطلاحات، مترادف‌ها و اصطلاحات اعم و اخص برای افزایش احتمال بازیابی دقیق و موفقیت‌آمیز است و این امر مستلزم بازنمایی خودکار نظام بازیابی از اصطلاح نامرجح به اصطلاح مرجح است که لازمه آن وجود اصطلاحنامه به شکل الکترونیکی است.

«راهنمای اصطلاحنامه»^{۲۷} که بصورت کتابشناسی در سال ۱۹۹۲ برای کمیسیون جوامع اروپایی تدوین شد، بسیاری از اصطلاحنامه‌ها را بر روی اینترنت دسترس‌پذیر نموده است. دستیابی رایگان از مزایای قابل توجه سایت‌های اینترنتی است.

با بررسی این اصطلاحنامه‌ها می‌توان پی برد که آیا زبان نمایه‌سازی موجود قابل انطباق با نیازهای مرکز اطلاع‌رسانی داخلی هست یا خیر. در غیر این صورت ضرورت دارد که اصطلاحنامه‌ای هدفمند در چارچوب نظام محور و براساس رشته‌های مورد نظر تدوین گردد. حتی در این صورت نیز می‌توان از اصطلاحنامه‌های موجود و یا بخشهائی مناسب برای استخراج، ویرایش و کاربرد در نظام جدید استفاده نمود.

اصطلاحنامه و عناصر طراحی نظام بازیابی اطلاعات

زبان نمایه‌سازی فقط شامل اصطلاحات یا واژگان نمایه‌سازی نمی‌شود بلکه ابزار زبان نمایه‌سازی خاصی را دربر می‌گیرد. ابزاری که بازیابی اسناد مرتبط را بهبود می‌بخشد (تعداد، شکل دستوری، ترتیب واژگانی و غیره) و شیوه‌هایی که از آوردن اطلاعات غیرمرتبط ممانعت بعمل آورند:

- کنترل مترادف‌ها و شبه‌مترادف‌ها بصورت ارجاع مترادف‌های غیرمرجح به اصطلاح مرجح از پراکندگی مترادف‌ها و شبه‌مترادف‌ها در پایگاه‌های اطلاعاتی ممانعت بعمل می‌آورد.
- با هدایت کاوشگر مفاهیم و اصطلاحات اخص به اعم یعنی نزدیکترین اصطلاح نمایه‌ای موجود برای ارائه این مفاهیم جامعیت می‌بخشد.
- با ارائه روابط ساختاری مفاهیم اعم از سلسله مراتبی یا غیر سلسله مراتبی جامعیت حاصل می‌گردد و کاوش را با ارائه طبقه‌های دارای ارتباط بسیار نزدیک توسعه می‌دهد. از این طریق مانعیت نیز با پیشنهاد اصطلاحات اخص تر بهبود می‌یابد.

برای بازیابی بیشتر است و در نتیجه جامعیت احتمالاً بالاتر بوده و مانعیت از طریق بررسی مکرر رایانه بهبود می‌یابد. چنانچه اطلاعات به بیش از یک زبان وارد نظام شده و بازیابی شود به اصطلاحنامه دوزبانه و یا چندزبانه نیاز است و مسایل ویژه تدوین اصطلاحنامه که در این وضعیت مطرح می‌شود باید در نظر گرفته شود.

در پایگاه‌های اطلاعاتی گسترده که بسیار مورد استفاده قرار می‌گیرند، اصطلاحنامه‌ای پیشرفته برای نمایه‌سازی و بازیابی توجیه‌پذیر است، تا با پرداختن به سؤالات جزئی‌تر، سرعت و کیفیت بازیابی را بهبود بخشد. منابع مالی عاملی کلیدی است، اگرچه غالباً خارج از نظارت نظام اطلاعاتی است. نوع و گستره فعالیت سازمان ذیربط اعم از آنکه دانشگاهی، تجاری یا صنعتی باشد، بر منابع مالی قابل دسترس، تأثیر خواهد گذاشت. در جایی که منابع محدود و طرح‌های اطلاعاتی در معرض کاهش منابع مالی هستند، صرفه‌جویی ممکن است به شکل انطباق با اصطلاحنامه‌های موجود صورت پذیرد. چنانچه الزاماً بایستی اصطلاحنامه‌ای جدید تألیف گردد، هزینه‌ها را می‌توان بالاچار با کاهش گستره اصطلاحنامه، تعدیل ویژگی اخص بودن و ساده کردن ساختار آن، تقلیل داد. خرید اصطلاحنامه بجای تدوین آن و نیز انطباق آن با نیازهای داخلی و استخدام مشاور علمی برای تدوین یا انطباق یک اصطلاحنامه در خصوص طراحی و اجرای پروژه اصطلاحنامه نیز در دنیا رایج می‌باشد. (همان منبع)

تدوین اصطلاحنامه جدید یا انطباق

در حال حاضر کتابشناسی جامعی به صورت چاپی یا درونخطی در خصوص تمامی اصطلاحنامه‌های موجود، وجود ندارد. با این حال می‌توان به فصلنامه «سازماندهی دانش»^{۲۴} منتشر شده توسط «مجمع بین‌المللی سازماندهی دانش»^{۲۵} شامل اصطلاحنامه‌های موجود و طرح‌های رده‌بندی مراجعه نمود. همچنین «مرکز منابع اطلاعاتی» «سلیب»^{۲۶} مجموعه‌ای از اصطلاحنامه‌ها را نگهداری می‌کند، هرچند این مجموعه نیز کامل نیست، اما می‌تواند بعنوان مرجع مورد بهره‌برداری قرار گیرد.

□ تدوین اصطلاحنامه‌ای پیچیده با ساختاری بسیار منظم و نظام‌مند شامل واژگان اخص و اصطلاحات مرکب، بکارگیری و حفظ و نگهداری آن مستلزم تلاش فراوان است و هزینه‌های بالایی را در پی خواهد داشت.

کنترل واژگان

□ در اصطلاحنامه نظارت بر اصطلاحات به شیوه‌های مختلف محقق می‌شود. ابتدا شکل اصطلاح نظارت می‌شود چه این کار شکل دستوری، املائی، جمع و مفرد و اختصارات را دربرگیرد و چه شکل مرکب اصطلاح را.

□ شیوه^{۲۷} دیگر گزینش بین دو یا چند مترادف موجود برای بیان یک مفهوم است.

□ شیوه^{۲۸} سوم تصمیم‌گیری در خصوص پذیرش و نحوه برخورد با انواع خاصی از اصطلاحات نظیر **واژه‌های قرضی^{۲۸}**، **واژه‌های عامیانه^{۲۹}**، اسامی تجاری و اسامی خاص است.

□ در شیوه چهارم، معنی اصطلاح که در یک فرهنگ ممکن است با توضیحات گوناگون همراه باشد، عملاً به یک معنا محدود می‌گردد که مؤثرترین معنا در اصطلاحنامه خاص است. این محدودیت با اضافه نمودن تعاریف و یادداشتهای دامنه و عبارات توضیحگر در خصوص همنام‌ها در اصطلاحنامه مورد اشاره قرار می‌گیرد.

در خصوص کنترل واژگان در استاندارد بین‌المللی **ایزو^{۳۰}**، استاندارد **بریتانیا^{۳۱}** و استاندارد **آمریکا^{۳۲}** قواعدی تدوین یافته ارائه شده است که اکثر این قواعد در هر سه استاندارد یکسان می‌باشد. (همان منبع)

گزینش اصطلاحات

اصطلاحات شامل واژه‌های قرضی، نوواژه‌ها^{۳۳}، واژه‌های عامیانه، اصطلاحات علمی، اسامی مکانها و اسامی خاص می‌شود:

□ ابزار مانعیت شامل اخص بودن زبان نمایه‌سازی است. هر قدر جزئیات و تعداد اصطلاحات بیشتر باشد، موضوع اسناد و مدارک علمی با دقت بیشتری توصیف می‌یابد. اخص بودن قابلیت‌های مانعیت نظام را کنترل می‌کند و مهارت بیشتری در نمایه‌سازی و بازیابی می‌طلبد.

□ همارایی جزء قوی‌ترین ابزار مانعیت است. با افزایش تعداد اصطلاحات در ترکیب با یکدیگر در نمایه‌سازی و کاوش، مفاهیم موردنیاز با دقت بیشتری تعریف می‌شود و اطلاعات حشو حذف می‌گردد.

□ اصطلاحات مرکب، که نوعی پیش‌همارایی اصطلاح است، سبب می‌گردد که موضوع مدرک به طور جزئی‌تر شناسایی و بدینسان از بازیابی اسناد نامربوط جلوگیری شود.

□ روابط ساختاری در اصطلاحنامه سبب می‌گردد که نمایه‌ساز و کاوشگر از مفاهیم گسترده به اصطلاحات مشخص‌تر هدایت شود که این امر بازیابی را محدود کرده و اطلاعات حشو را حذف می‌کند.

□ ویژگی‌های اخص بودن واژگان، سلسله مراتب و کنترل مترادف‌ها را می‌توان بخش ضروری ساختار اصطلاحنامه دانست.

□ در هنگام طراحی یک اصطلاحنامه، باید به خاطر داشت که بکارگیری یک ابزار مانعیت، نظیر ویژگی بالای اخص بودن و اصطلاحات مرکب، عملکرد جامعیت را تقلیل خواهد داد. در صورتی که کاهش اخص بودن واژگان و کاربرد اصطلاحات تکواژه‌ای، جامعیت را بهبود خواهد بخشید اما به ضرر مانعیت تمام خواهد شد.

□ ترکیب و تعامل بین زبانهای کنترل شده و طبیعی می‌تواند فقدان اخص بودن در زبان کنترل شده را جبران نماید و مانعیت را بهبود بخشد. درحالیکه زبان کنترل شده می‌تواند جامعیت را از طریق کنترل مترادف‌ها و شبه‌مترادف‌ها و نمایش رابطه‌ها بهبود بخشد.

۱- واژه‌های قرضی

که از زبانهای دیگر قرض گرفته شده و در زبان قرض گیرنده بخوبی تثبیت شده‌اند باید در اصطلاحنامه پذیرفته شوند.

چنانچه ترجمه این اصطلاحات وجود داشته باشد ولی به طور رایج مورد استفاده قرار نگیرد، با اصطلاح ترجمه شده باید بصورت اصطلاح نامرچج برخورد گردد و بین اصطلاح و واژه قرضی ارجاع متقابلی با استفاده از علائم خاص ایجاد گردد.

هر زمان که استفاده از ترجمه یا نزدیکترین معادل از واژه قرضی متداول تر شد، جایگاه واژه قرضی تغییر کرده و به اصطلاح نامرچج تقلیل می‌یابد.

۲- نوواژه‌ها، اصطلاحات عامیانه و زبان حرفه‌ای

نوواژه‌ها، اصطلاحات حرفه‌ای و عامیانه اغلب مفاهیم جدیدی را دربر می‌گیرند که در یک رشته تخصصی خاص و یا جامعه زبانی کوچکتر شکل می‌گیرند. چنانچه جایگزینی که بطور گسترده توسط کاربران مورد استفاده قرار گیرد وجود نداشته باشد، نوواژه، اصطلاح عامیانه یا حرفه‌ای بعنوان یک توصیفگر مورد قبول قرار می‌گیرد. در صورتی که اصطلاحی عامیانه بعنوان گزینه‌ای برای اصطلاحی موجود و تثبیت شده، آنقدر شناخته شده باشد که ممکن است به عنوان نقطه دستیابی کاربر مورد استفاده قرار گیرد، باید آنرا بعنوان اصطلاح مدخل پذیرفت.

۳- اسامی عامیانه و اسامی تجاری

استانداردها توصیه می‌کنند که در جایی که اسم عامیانه معادلی وجود دارد، باید از آن بجای اسم تجاری استفاده کرد، زیرا اسم عامیانه مفهوم را بطور دقیق‌تر بازگو می‌کند. اسم تجاری باید فقط زمانی به عنوان اصطلاح نامرچج پذیرفته شود که احتمال داده شود کاربر به آن نیاز داشته باشد.

بنابراین این امکان وجود دارد که هم اسم عامیانه و هم اسم تجاری بعنوان اصطلاحات نمایه‌سازی در

اصطلاحنامه آورده شود که محصولات گوناگون و نیز نشانهای خاص را دربرگیرند. در اینصورت با نامهای تجاری مانند اصطلاحات اخص برخورد می‌شود.

۴- اسامی مشهور و اسامی علمی

توصیه استانداردها در جایی که امکان انتخاب بین اسامی مشهور و علمی وجود دارد، گزینش شکلی است که بیشتر احتمال می‌رود کاربران اصطلاحنامه به دنبال آن باشند.

۵- اسامی مکانها

در جایی که برای یک کشور یا منطقه جغرافیایی درون یک جامعه تک زبانی به دلایل مختلف بیش از یک اسم انتخاب می‌گردد، باید آن اسمی را به عنوان اصطلاح مرجح تعیین کرد که نزد کاربران اصطلاحنامه آشناتر است. در صورتی که شرایط یکسانی حاکم باشد، باید اسم رسمی را به اسم مشهور ترجیح داد. چنانچه مراجع استاندارد وجود دارد باید درباره اشکال رسمی به آنها رجوع شود.

۶- اسامی خاص مؤسسات، افراد و غیره

هرچند اسامی خاص افراد، مؤسسات، سازمانها، فرآیندها، انواع تجهیزات، قوانین، عناوین و مکانها نقاط دستیابی مهمی برای بازیابی اطلاعات هستند، لکن اغلب یا از اصطلاحنامه حذف و یا تعداد آنها محدود می‌گردد و در فایل‌های مجزا نگهداری می‌شوند که شکل اسامی خاص را برای فهرستنویسی استاندارد تعیین می‌نماید، اما روابط ساختاری را مشخص نمی‌کند. میزان نیاز دستیابی به اسامی خاص، گنجاندن اسامی را در اصطلاحنامه اصلی تعیین می‌کند. مثلاً اصطلاحنامه‌ای در حوزه روابط بین‌الملل، بایستی دربرگیرنده اسامی کشورها باشد زیرا بخش لاینفک این حوزه تخصصی محسوب می‌گردد. چنانچه اسامی خاص در اصطلاحنامه‌ای گنجانده شود باید شکل اسامی با قوانین و دستورالعمل‌های شناخته شده

سلول (زیست‌شناسی) Cell (biology)
 سلول (تجهیزات الکتریکی) Biological cells
 سلول‌های الکتریکی Electric cells Cells (electric)

در این زمینه نیز باید قواعد استاندارد را رعایت شود. گاهی برای تکمیل معنایی که با ساختار اصطلاحنامه انتقال می‌یابد به تعاریف محدود یا مبسوط نیاز می‌باشد. تعاریف غالباً در اصطلاحنامه‌های علوم اجتماعی و علوم انسانی ضرورت می‌یابد تا واژه‌های غیردقیق را که بیشتر در این حوزه یافت می‌شوند توضیح دهند. نماد SN (ی.د.) برای نشان دادن یادداشت دامنه شامل تعاریف نیز می‌گردد که بمنظور نشان دادن محدودیت معنایی یا طیف موضوعات زیر پوشش به کار برده می‌شود.

۸- مترادف‌ها^{۳۴}

استانداردها مترادف‌ها را به عنوان اصطلاحاتی که معانی‌شان را می‌توان در طیف گسترده‌ای از متن‌ها یکسان تلقی کرد می‌دانند، بطوری که این اصطلاحات عملاً قابل جایگزینی با یکدیگر هستند. مترادف‌ها در زبان‌شناسی عمومی رایج و پذیرفته شده نیستند، اما در واژگان علمی وجود دارند، زیرا اسامی تجاری و اسامی مشهور و گونه‌های مختلف یک واژه بسته به افکار عمومی و کاربرد محلی یا ریشه واژه دارد.

در تمامی حوزه‌ها، نه فقط حوزه‌های علمی، مترادف‌های بیشتری در زبان کنترل شده وجود دارد تا زبان طبیعی، چرا که معانی اصطلاحات عملاً در زبان کنترل شده محدود است. گزینش اصطلاحات مرجح از مترادف‌های ممکن همیشه باید تحت تأثیر نیازهای کاربران اصطلاحنامه باشد. این اصطلاحات یا بطور مفصل مورد بحث قرار می‌گیرند و یا از نقطه نظر رایج یا تخصصی و یا علمی متفاوت خواهند بود. هر یک از گزینه‌ها که انتخاب شود، باید در سراسر اصطلاحنامه به طور یکسان اعمال گردد.

۹- شبه مترادف‌ها^{۳۵}

فهرست‌نویسی تطابق داشته باشد. استانداردها بویژه استاندارد ایزو توصیه می‌کند که از قواعد زیر پیروی شود:

□ اسامی مؤسسات ملی و محلی که فعالیت تجاری را به یک زبان انجام می‌دهند باید به اشکال ترجمه نشده خود وارد اصطلاحنامه شوند. در صورت وجود شکل ترجمه شده باید به آن ارجاع شود.

□ اسامی سازمانهای بین‌المللی یا سازمانهای محلی که اسناد را به بیش از یک زبان منتشر می‌کنند باید به صورتی که برای کاربران اصطلاحنامه آشنا تر است ارائه شوند. در صورتی که احتمال می‌رود اصطلاحنامه توسط اعضاء جوامع زبانی دیگر نیز مورد استفاده قرار گیرد، ارجاع‌هایی از اشکال دیگر داده شود.

□ اسامی فردی باید به اشکال اصلی خود ارائه شوند، مگر آنکه شکل محلی اسم خصوصاً شخصیت‌های تاریخی که در سطح بین‌الملل شناخته شده‌اند، نزد کاربران اصطلاحنامه آشنا تر باشد. در چنین وضعیتی اسم محلی، ترجیح داده می‌شود و ارجاعی از شکل اصلی داده می‌شود.

۷- همنام‌ها و هم‌آواها

همنام‌ها کلماتی هستند که با کلمه‌ای دیگر دارای یک املاء هستند، اما منشاء و معنای آنها متفاوت است. مانند: سلول Cells
 که به معنای میکروسیستم‌های بیولوژیکی و نیز تجهیزات الکترونیکی است.

- همنام‌ها ممکن است تلفظ متفاوتی نیز داشته باشند.

همنام‌ها گاهی اوقات به عنوان کلمات چندمعنایی یا به عنوان هم‌آوا شناخته می‌شوند. به طور کلی هم‌آواها را می‌توان به عنوان کلمه‌ای تعریف کرد که برای اشاره به مفاهیم متفاوتی مورد استفاده قرار می‌گیرند یا بطور اخص بعنوان کلماتی تعریف می‌گردند که یک صدا اما معانی متفاوت دارند.

در چنین مواردی روش معمول ابهام‌زدایی، اضافه کردن توضیحگر است که در داخل پرانتز قرار می‌گیرد:

می‌کنند. در اینصورت کاوشگر می‌تواند هر اصطلاحی از حلقه را وارد کند و به طور خودکار تمامی اعضاء دیگر حلقه را بازیابی نماید.

نمایه‌سازی ماشینی

تدوین خودکار اصطلاحنامه یکی از اجزاء حوزه گسترده‌تر بازیابی خودکار اطلاعات است که نمایه‌سازی خودکار، چکیده‌نویسی خودکار و رده‌بندی خودکار اسناد را شامل می‌شود.

در تمام این عملیات فنون آماری یا زبان‌شناسی محاسبه‌ای جایگزین فرآیندهای فکری انسان می‌شود. نمایه‌سازی خودکار استخراجی، یعنی گزینش خودکار اصطلاحات برای نشان دادن محتوای اسناد است. شیوه دیگر نمایه‌سازی که بسیار نویدبخش است نمایه‌سازی خودکار تخصیصی یا نمایه‌سازی با ماشین^{۳۶} است. با استفاده از این فن، کدهای رده‌بندی و اصطلاحات نمایه‌سازی از یک اصطلاحنامه تدوین یافته توسط انسان در صورتی به سندی اختصاص می‌یابد که این کدها و اصطلاحات، با عبارات و کلماتی که طور خودکار استخراج شده سازگار باشند.

پیشرفت در این فن‌آوری امکان استفاده از نرم‌افزار قوی تحلیل متن را به منظور پیشنهاد اصطلاحات برای نمایه‌سازی فراهم می‌کند، بدون اینکه الزاماً ایجاد هزینه‌های بیشتر را ضروری نماید. تعداد قابل ملاحظه‌ای از اسنادی که قبلاً فهرست شده برای آموزش نرم‌افزار تحلیل متن به کار گرفته می‌شود که این نرم‌افزار بعداً اصطلاحات نمایه‌ای موردنظر را به منظور تولید نمایه‌سازی خودکار به اسنادی برای مرور نمایه‌ساز تخصیص می‌دهد.

امروزه هنوز کاربرد فنون خودکار تدوین اصطلاحنامه‌ها به طور گسترده به اجرا درنیامده است، اگرچه تحقیق با نظامهای آزمایشی ادامه دارد. در تولید خودکار اصطلاحنامه، رابطه‌ها بین اصطلاحات نمایه‌سازی عمدتاً به صورت آماری شناسایی می‌شوند تا به صورت معناشناختی. در تولید خودکار اصطلاحنامه برای ایجاد رده‌های اصطلاحنامه تحلیل میزان حضور اصطلاحات در

شبه مترادف‌ها، که به آنها نیمه مترادف نیز گفته می‌شود، به عنوان اصطلاحاتی تعریف می‌شوند که معانی آنها به طور کلی در کاربرد رایج متفاوت در نظر گرفته می‌شوند، اما به شیوه‌ای با آنها برخورد می‌شود که گویا آنها برای اهداف نمایه‌سازی متفاوت هستند. شبه مترادف‌ها شامل اصطلاحاتی می‌شوند که دارای همپوشانی قابل توجهی باشند. مانند: افراد با استعداد- تیزهوشان پذیرش مجموعه‌های اصطلاحات به عنوان شبه مترادف از حوزه موضوعی زیر پوشش اصطلاحنامه متأثر است:

در اصطلاحنامه‌ای تخصصی «بهداشت شغلی»، اصطلاحات «آسیب صنعتی» و «آسیب شغلی» به عنوان مترادف در نظر گرفته نمی‌شوند، حال آنکه در اصطلاحنامه‌ای عمومی تر درباره «علوم بهداشت»، این اصطلاحات به عنوان شبه مترادف پذیرفته می‌شوند. باید از ابزار شبه مترادف به عنوان وسیله کاهش حجم واژگان و از طریق کنار هم قرار دادن اصطلاحاتی که باید به آنها به عنوان اصطلاحات نمایه‌سازی، مگر در حوزه‌های موضوعی حاشیه‌ای، اجتناب نمود.

شبه مترادف‌ها ممکن است متضادها، یعنی اصطلاحاتی که نشانگر نقطه نظرات متفاوتی از یک پیوستار ویژگی است را نیز دربرگیرند. مثال: سوادآموزی/ بیسوادی

اگر احتمالاً کاوشی در خصوص مثلاً «سوادآموزی» صورت پذیرد همیشه بررسی اسنادی را دربر خواهد داشت که هم درباره «سوادآموزی» و هم «بیسوادی» بحث خواهند کرد، با این اصطلاحات باید بصورت شبه مترادف رفتار کرد. هدف برخی اصطلاحنامه‌های کاوش، افزایش جامعیت از طریق فهرست کردن حداکثر اصطلاحات جایگزین است. در این موارد انتخاب اصطلاحاتی مرجح از گروهی از اصطلاحاتی که نشانگر یک مفهوم است لزومی ندارد. اکنون برخی بسته‌های نرم‌افزاری مجهز به واژه‌های الکترونیکی برای این هدف وارد بازار می‌شوند، حال آنکه بسته‌های دیگر آنچه موسوم به حلقه‌های مترادف خودکار است ارائه

عملیات ریاضی مفروضات ساده و مطمئنی را ارائه می‌دهد. فرض بر آن است که مدارک براساس میزان ربط در هنگام بازیابی مورد ارزیابی قرار گیرند. این روش مدل مزبور را قابل محاسبه و داده‌های آنرا قابل تجزیه و تحلیل می‌نماید. همچنین مزیت آن سادگی آن است. از طرف دیگر فرض آن است که قاعده فوق‌الذکر از واقعیات پیروی می‌کند.

این روشها خصوصاً تعیین می‌کنند که چه داده‌هایی موردنیاز است و چطور این داده‌های احتمالی محاسبه می‌شود (یعنی یک مدرک اصطلاحات احتمالی را داراست یا با یک سؤال که اصطلاحات احتمالی را شامل می‌شود، مرتبط است).

این روشها مفهوم ربط را در تعداد پیشامدها جستجو می‌کند. هرچند این الگوی محاسباتی از پایه نظری مناسبی برخوردار است، اما به نتایج دقیقی در مقایسه با روشهای آماری نمی‌انجامد. اساساً باید توجه داشت که نظامهای نمایه‌سازی رایانه‌ای در اقتصاد اطلاعات با رویکرد تجاری هنوز به طور نسبی و ندرتاً کاربرد دارند. نظامهایی که گاهی ما را به استثناءهای بسیار ارجاع می‌دهد. این روشها در بخشهای وسیعی از نمایه‌سازی ماشینی، از تعداد لغات بهره می‌گیرند. جهت‌یابی شدید به سمت استفاده از آمار باعث می‌شود توصیفگرهای مرتبط با متن همیشه با معیارهای دلخواه سنجیده نشوند و به همین دلیل بدنبال آن کنترل توسط انسان ضرورت می‌یابد.

خصوصاً با توجه به موج عظیم اطلاعات، ابزارهای نمایه‌سازی خودکار نیاز به توسعه و ارتقاء داشته تا کیفیت بازیابی از سوی کاربران قابل پذیرش باشد.

در روشهای زبانشناختی به کمک تجزیه و تحلیل ریخت‌شناسی لغات^{۳۹}، نحو^{۴۰} و معناشناسی مدرک، توصیفگرها برای نمایه‌سازی انتخاب می‌شوند و مقدمات جستجو در متن را فراهم می‌آورند. آنچه جستجو می‌شود اصطلاحات زبان طبیعی بر مبنای قواعدی خاص به شکل نظامهای رایانه‌ای و آماده‌سازی همه اطلاعات مرتبط جهت تجزیه و تحلیل آنهاست.

تجزیه و تحلیل معنانشناختی در سطح متون مشترک در مدارک صورت می‌پذیرد و می‌کوشد ارتباطات

اسناد نظام صورت می‌گیرد و در این کار از بررسی وزن اصطلاحات بر مبنای الگوی فاصله برداری که توسط «سالتن»^{۳۷} طراحی شده، و یک الگوریتم دسته‌بندی استفاده می‌شود.

الگوی فاصله برداری بسامد یک اصطلاح در سندی (یعنی تعداد دفعاتی که اصطلاح در یک سند ظاهر می‌شود) را تعیین می‌کند. اصطلاحاتی که با بسامد بالایی در اسناد ظاهر می‌شوند حاکی از آن است که اصطلاح به میزان بالایی به سند مربوط می‌شود. بسامد سندی معکوس تعداد کل اسنادی است که اصطلاح در آنها ظاهر می‌شود. اصطلاحاتی که در اسناد کمتری در کل پایگاه اطلاعاتی ظاهر شوند (اصطلاح اخص‌تر) بسامد بالاتری دارند.

بنابراین محاسبه کثرت لغات براساس سه مرحله زیر بنا شده است:

- ۱- تشخیص کثرت حقیقی هر مفهوم برای هر مدرک
- ۲- تشخیص کثرت هر مفهوم در مجموعه مدارک مشترک در یک پایگاه اطلاعاتی
- ۳- تفکیک همه مفاهیمی که در حد نهایت ارزش قرار دارند و یا پائین‌ترین حد ارزش را دارا می‌باشند.

برای تشخیص توصیفگرها دو شیوه عملکرد امکان‌پذیر است:

- ۱- پس از مقایسه مدارک با اصطلاحات لیست **بازدارنده**^{۳۸}، همه لغاتی که در لیست بازدارنده قرار دارند انتخاب و حذف می‌شوند.
- ۲- مفاهیم باقیمانده براساس ریشه لغت تفکیک می‌شوند. (کایزر، ۱۹۹۳)

حضور واژگان و در نتیجه میزان ارتباط موضوعی اسناد و مدارک در روشهای مبتنی بر احتمالات تعیین احتمال، اساس همه تحقیقات می‌باشد. عمل نمایه‌سازی که با بکارگیری احتمالات انجام می‌شود یک تخمین احتمالی است برای آنکه توزیع توصیفگرها صورت پذیرد. این ارتباط می‌تواند مطابق خواست کاربر باشد. این شیوه با

در این نظام رایانه به واسطه فهرست کلیدواژه‌های موجود در انجام نمایه‌سازی مدارک، کلیدواژه‌ها و اصطلاحات معیار را به نمایه‌ساز یادآور می‌شود. اصطلاحاتی که گزینش می‌شوند توسط نمایه‌ساز بازمینی و توصیفگرهایی که برنامه رایانه‌ای قادر به تعیین آن نبوده را اضافه می‌کند و توصیفگرهای غلط با توصیفگرهای صحیح جایگزین می‌شوند.

اگر متن ماشین خوان باشد، رایانه می‌تواند با استفاده از بسامد واژه‌ها و واژه‌نامه موجود به طور خودکار نمایه‌سازی را انجام دهد. یکی از معایب آشکار استفاده از بسامد واژگان در گزینش اصطلاحات این است که حتی پس از استفاده از واژه‌نامه، برخی واژه‌ها که به دفعات در مدرک یافت می‌شوند، ممکن است وجه تمایز خوبی برای تشخیص مدرک از دیگر مدارک اطلاعاتی نباشند و در نتیجه بصورت تخصصی بازیابی انجام نگردد.

در **نمایه‌سازی رایانه‌ای تخصصی**، برای گزینش اصطلاحات از ابزارهای کنترل واژگان مانند اصطلاحنامه استفاده می‌شود، هرچند نمایه‌سازی تخصصی و اجرای آن بصورت خودکار بسیار پیچیده خواهد بود. بدین منظور اصطلاحنامه بصورت نظام‌مند ارتباطات معنایی مفاهیم را ایجاد می‌نماید و ساختار اصطلاحنامه بعنوان پشتوانه‌ای جهت ذخیره و بازیابی بانک اطلاعاتی قرار می‌گیرد.

امروزه چون از ابتدا منابع و مدارک به شکل الکترونیکی ارائه می‌شوند و بحث بر سر ایجاد کتابخانه‌های الکترونیکی است، کاربرد نمایه‌سازی رایانه‌ای تأثیر به‌سزایی در کاهش نیاز به کارکنان و متخصصان رشته‌های گوناگون دارد و دستیابی اطلاعات از راه دور، کاهش زمان آموزش و دستیابی همزمان افراد به اطلاعات جدید از مزایای کاربرد نمایه‌سازی رایانه‌ای است. در آینده‌ای نه چندان دور با توجه به گسترش روزافزون تولید اطلاعات در رسانه‌های الکترونیکی به جای تولید منابع کاغذی، نمایه‌سازی رایانه‌ای رواج بیشتر خواهد یافت و مستلزم آن است که کتابخانه‌ها و مراکز اطلاع‌رسانی کشور نیز زمینه ایجاد چنین شرایطی را فراهم نمایند.

معنایی در مدرک را بازشناسد و از دانش قراین بهره‌گیر تا یک متن را در واحدهای بسته به معنای آنها تجزیه نماید. استفاده از روشهای زبانشناختی این امکان را بوجود می‌آورد که محتوای اصلی یک مدرک به خوبی شناسایی شود. مزیت جامعیت بالاتر و بسیار کاربرپسند بودن نتیجه استفاده از آنهاست.

نتیجه

بازیابی سریع و دقیق مدارک علمی متضمن نمایه‌سازی اصولی و صحیح مدارک موجود در کتابخانه‌ها و مراکز اسناد می‌باشد. لکن تربیت و آموزش نمایه‌ساز متخصص و آشنا به شیوه‌های درست نمایه‌سازی و ذخیره و بازیابی اطلاعات به سالها کار و تجربه نیاز دارد.

نمایه‌سازانی که در زمینه‌های گوناگون علمی-تخصصی از تبحر کافی برخوردار باشند به ندرت یافت می‌شوند و هزینه تربیت نمایه‌سازان مختلف در رشته‌های گوناگون بسیار بالا خواهد بود. بنابراین استفاده از رایانه در نمایه‌سازی می‌تواند شیوه‌ای نوین در حل این معضل مبتلا به مراکز اطلاع‌رسانی باشد. تلاش در یافتن شیوه‌های نمایه‌سازی ماشینی و تطبیق آنها با زبان فارسی و بکارگیری آن در نمایه‌سازی ماشینی مستلزم همکاری اساتید فن در رشته‌های زبانشناسی، کامپیوتر و اطلاع‌رسانی است.

مطالعات در مورد بازیابی اطلاعات با تحلیل بر روی متن آزاد شکل گرفته و به سوی تحقیق در مورد کاوش اطلاعات به شیوه تحلیل هوشمندانه متن به دو شیوه **آماری-تحلیلی** و **تحلیل نحوی-معنایی** متن پیش می‌رود. نمایه‌سازی رایانه‌ای مستلزم آن است که متن موردنظر ماشین خوان باشد و رایانه بتواند کلمات را بصورت انفرادی و جداگانه شناسایی نماید و براساس قوانینی میان این کلمات و عبارات گزینش نماید.

در اغلب نظامهای نمایه‌سازی خودکار، رایانه بطور کامل جانشین انسان نمی‌شود، بلکه نمایه‌سازی توسط انسان و به کمک رایانه انجام می‌پذیرد که اصطلاحاً آنرا «**نمایه‌سازی به کمک رایانه**» می‌نامند.

- 1- Terminology
- 2- Wuester, E.
- 3- Cognitive Sciences
- 4- Logics & Ontology
- 5- Terminography
- 6- Informatics
- 7- Informations science
- 8- Information management
- 9- Knowledge engineering
- 10- Semiotics
- 11- Lexicology
- 12- Semantics
- 13- Lexicography
- 14- Data Processing
- 15- Informatics
- 16- Oeser, E.
- 17- Metaterminology
- 18- Concept dynamics
- 19- Catalourging
- 20- Indexing
- 21- Search
- 22- Semantic Maps
- 23- Keywords Indexes
- 24- Knowledge Organization
- 25- International Society for Knowledge Organization (ISKO)
- 26- Aslib
- 27- Thesaurus Guide
- 28- Loan words
- 29- Slang words
- 30- ISO 2288, 1986
- 31- BS 5723, 1987
- 32- ANSI/NISO 239.19, 1994
- 33- Neologisms
- 34- Synonyms
- 35- Quasi-Synonyms
- 36- Machine Aided Indexing (MAI)
- 37- Salton
- 38- Stop list
- 39- Morphology
- 40- Syntax

منابع

- Aitchison, J. Gilchrist, A & Bawden, D. (2000). *Thesaurus construction and use: A practical manual, Aslib, 4.ed.*
- Batori, I.S. Hahn, U. Pinkal, M. Wahlster, W. (1988). *Computerlinguistik u.ihre theoretischen Grundlagen, proceeding, Springer Verlag, Saarbruecken, Germany.*
- Buder, M; Rehfeld, W. Seeger, T. (1990). *Grundlagen der praktischen Information u. Dokumentation, K.G.Saur, Germany.*
- Felber, H. (1982): *Terminology and Knowledge Engineering*, TKE 87, Indeks Verlag, Germany.
- Felber, H. (1984). *Terminology Manual*, Infoterm, Unesco, Paris, France.
- Kaiser, A. (1993). *Computer-unterstuetztes Indexieren in Intelligenten Information Retrieval System*. Doktorat Diss. Wien, Austria.
- Seelbach, D. (1979). *Computerlinguistik u. Dokumentation*, Verlag Dokumentation, Muenchen, Germany.
- Lauren, Ch. Myking, J. Picht, H. (1998). *Terminologie unter der Lupe*. Termnet, Vienna, Austrulia.