
بررسی شیوه تدوین اصطلاحنامه‌های پیوسته

(۱) مهدی شقاقی

چکیده: در این مقاله شیوه‌های تدوین و ساخت اصطلاحنامه‌های پیوسته (آنلاین) از دو رویکرد کلی-ساخت اصطلاحنامه به کمک رایانه، و ساخت رایانه‌ای اصطلاحنامه-مورد بررسی قرار گرفته است. در بحث مربوط به ساخت اصطلاحنامه پیوسته به کمک رایانه، استفاده از موتور استدلال گر برای تعیین روابط اصطلاحنامه‌ای و نیز رویکرد ذهنی برای ساخت اصطلاحنامه‌های پیوسته مورد مطالعه قرار گرفته و مراحل عمل گرایانه برای ساخت اصطلاحنامه پیوسته بررسی شده است. در رویکرد رایانه‌ای برای ساخت اصطلاحنامه‌های پیوسته، روش‌های متنی، روش‌های برگرفته از سلسله‌مراتب وردنت، خوشه‌بندی، و روش‌های نقشه خودسامان بررسی شده است.

کلیدواژه‌ها: اصطلاحنامه، اصطلاحنامه پیوسته، تدوین، شیوه، رایانه

۱. مقدمه

کلمه «تزاروس»^۱ (اصطلاحنامه) را برای اولین بار «روژه»^۲ برای واژه‌نامه خود با عنوان «اصطلاحنامه عبارات و لغات انگلیسی»^۳ به کار برد. این واژه‌نامه نوعی طبقه‌بندی واژگانی بود که در آن، واژه‌ها براساس مفهومشان طبقه‌بندی شده بودند و می‌شد از آن به عنوان یک فرهنگ لغات مترادف و متضاد استفاده کرد (Glichrist 2003). «هلن برونسون»^۴ اولین کسی بود که به کاربرد اصطلاحنامه به عنوان شبکه‌ای از معانی برای حل مشکل بازیابی اطلاعات اشاره کرد. روالی که پس از گسترش این طرز تفکر طی شد موجب شکوفایی اصطلاحنامه سنتی برای نمایه‌سازی و بازیابی اطلاعات گردید؛ نمایه‌سازی و بازیابی‌ای که ابتدا از کارت‌های منگنه بهره می‌برد، ولی بعدها از پشتیبانی رایانه بهره‌مند شد. اگر چه اصطلاحنامه در ابتدا به وسیله رایانه‌هایی که برنامه‌ها را به صورت گروهی اجرا می‌کردند، به کار گرفته

می‌شد، ولی بعدها به عصر جستجوی پیوسته پا گذاشت و بدین ترتیب به یاری اینترنت و صفحات وب مبتنی بر زبان نشانه‌گذاری «اچ‌تی‌ام‌ال»، اصطلاحنامه‌های پیوسته ظهور کردند. از آنجا که بازیابی اطلاعات به نمایه‌ساز و جستجوگر و مسیرهای طی‌شده‌ی وی در جستجو تکیه داشت و هر دوی آن‌ها توسط شبکه‌ای از روابط سلسله‌مراتبی، همابندی، و هم‌ارزی میان اصطلاحات هدایت می‌شدند، نوع جدیدی از اصطلاحنامه به نام اصطلاحنامه جستجو ظهور نمود. این اصطلاحنامه‌ها در واقع سیاهه‌ای از اصطلاحات بودند که ویژگی اصلی آن‌ها خوشه‌بندی مترادف‌ها بود. پس از ظهور نوشتجات و آثار میان‌رشته‌ای، مشکل بازیابی این آثار ظهور کرد و بدین ترتیب ایجاد سازگاری میان اصطلاحنامه‌ها و تدوین ابر-اصطلاحنامه‌ها ذهن متخصصان را به خود مشغول نمود. راه‌حلی‌هایی چون بازنمایی^۵، استفاده از زبان‌های میانجی، تغییر خودکار، یا تلفیق پیشنهاد شد (Gilchrist 2003). امروزه به عقیده متخصصان، اصطلاحنامه‌های پیوسته تغییر نقش پیدا کرده‌اند و به عنوان ابزاری برای بازیابی اطلاعات و توسعه سؤال جستجو کاربرد یافته‌اند (Shiri and Revie 2000). بنابراین کتابخانه‌های دیجیتالی و پایگاه‌های اطلاعاتی، هم‌اکنون آغاز به تلفیق این گونه ابزارها با هدف توسعه جامعیت و مانعیت در بازیابی‌های کاربران و کمک به آنان در توسعه سؤال جستجو کرده‌اند. بسیاری دیگر نیز در حال گسترش روابط معنایی و غنی‌سازی اصطلاحنامه زیربنایی پایگاه‌های خود هستند. بدین ترتیب اصطلاحنامه‌های پیوسته روز به روز از دید کاربران مخفی‌تر می‌شوند و نحوه تدوین آن‌ها از روش‌های دستی به سمت روش‌های خودکار، و روابط اصطلاحنامه‌ای از روابط ذهنی و پیش‌ساخته انسانی به سمت روابط منطقی و روابط برخاسته از آمارهای مبتنی بر متن که به‌وسیله رایانه جمع‌آوری می‌شود تغییر پیدا می‌کنند.

۲. تدوین اصطلاحنامه‌های پیوسته

برای تدوین اصطلاحنامه‌های پیوسته با هر شکل نمایش یا هر کاربردی، دو رویکرد کلی وجود دارد: تدوین اصطلاحنامه به کمک رایانه^۶، و تدوین اصطلاحنامه رایانه‌ای شده^۷. عده‌ای دیگر همین دسته‌بندی را با عنوان «تدوین اصطلاحنامه به صورت کنترل‌شده»^۸ و «کنترل نشده»^۹ مطرح می‌کنند (Hodge and Austin 2002). تدوین اصطلاحنامه‌های پیوسته به صورت خودکار و خوشه‌بندی لغات و اصطلاحات بر اساس الگوریتم‌های برخاسته از پیکره متن،

بیش تر در پایگاه‌های اطلاعاتی که از یک مرورگر برای نمایه‌سازی صفحه‌ها و اسناد خود استفاده می‌کنند، موضوعیت پیدا می‌کند. از آنجا که موتورهای جستجو بیشتر از انطباق واژه به واژه برای بازیابی اطلاعات استفاده می‌کنند و اسنادی را در اولویت بازیابی قرار می‌دهند که بیش‌ترین وقوع کلمات جستجو را داشته باشند (و این باعث کاهش ربط، و بازیابی‌های ناخواسته می‌شود)، از نرم‌افزارهای خاصی برای خوشه‌بندی اصطلاحات موجود در پیکره متون و صفحه‌های موجود در پایگاه به منظور طبقه‌بندی اصطلاحات و تولید اصطلاحنامه استفاده می‌شود. این اصطلاحنامه‌ها حالت پنهان دارند و با ورود اسناد و صفحه‌های جدید، اصطلاحات جدید آن‌ها نیز به وسیله الگوریتم‌های خاص، خوشه‌بندی می‌شوند و جایگاه خود را در اصطلاحنامه پیدا می‌کنند. بدین ترتیب سؤال جستجو در بافت یک شبکه معنایی پردازش می‌شود و امکان بازیابی پاسخ‌های با ربط، افزایش می‌یابد (Hodge and Austin 2002). در روش‌های غیر خودکار، بیش‌تر گروهی از متخصصان، مسئولیت جمع‌آوری و طبقه‌بندی اصطلاحات را بر عهده می‌گیرند و سپس از یک نرم‌افزار سازگار با وب برای نمایش و ویرایش اصطلاحنامه استفاده می‌کنند. اصطلاحنامه به دست آمده مبنای نمایه‌سازی یا جستجوی اطلاعات در پایگاه اطلاعاتی خواهد بود.

۲-۱. تدوین اصطلاحنامه به کمک رایانه

در تدوین اصطلاحنامه به کمک رایانه، با توجه به ایجاد و توسعه روابط میان اصطلاحات می‌توان دو رویکرد کلی قایل شد. در رویکرد نخست، برای توسعه روابط میان اصطلاحات از یک موتور استدلال‌گر استفاده می‌شود. در رویکرد دوم، خود متخصصان روابط میان اصطلاحات را تعیین می‌کنند (که این رویکرد بسیار ذهنی است).

۲-۱-۱. استفاده از موتور استدلال‌گر^{۱۰}

در این رویکرد برای پیاده‌سازی روابط اصطلاحنامه‌ای میان اصطلاحات، از منطق‌های خاصی استفاده می‌شود که الگوریتم آن در یک برنامه نرم‌افزاری، تعریف شده است. بدین ترتیب با استفاده از عملگرهای منطقی از پیش تعریف شده، روابط میان اصطلاحاتی که از سوی متخصصان وارد سامانه شده‌اند، تعیین می‌شود. از این روش برای بازنمایی اصطلاحنامه‌های دستی به منظور یکپارچه‌سازی، منطقی‌سازی، و ارائه تفسیری جامع و یکدست از طبقات

مفاهیم استفاده می‌شود. همچنین این رویکرد برای ایجاد اصطلاحنامه در یک حوزه میان-رشته‌ای که از اصطلاحات چند حوزه مختلف استفاده می‌کند، به کار می‌رود (Bechofer and Goble 2001). نمونه آن، استفاده از «منطق توصیف»^{۱۱} در بازنمایی اصطلاحنامه «آی‌سی‌اُم»^{۱۲} است که در ادامه به شرح نمونه‌ای از کاربرد منطق توصیف در بازنمایی اصطلاحنامه یادشده خواهیم پرداخت.

«منطق توصیف» از خانواده زبان‌های بازنمایی دانش مبتنی بر طبقه‌بندی است که از «کال-وان»^{۱۳} ریشه گرفته است و تدوین و بازنمایی مدل‌های مفهومی را ممکن می‌کند. مدل منطق توصیف بر پایه معانی مفاهیم استوار است به طوری که طبقات اشیای دارای ویژگی‌های مشابه، افرادی که نمونه‌های آن مفاهیم‌اند، و نقش‌هایی که روابط میان افراد هستند را نشان می‌دهد. دو اصل اساسی یک منطق توصیف، ایده شمول مفهومی^{۱۴} و ایده طبقه‌بندی است. یک مفهوم زمانی به وسیله مفهومی دیگر احاطه می‌شود که همه موارد و اجزای آن، ضرورتاً موارد و اجزای مفهوم احاطه‌کننده باشند. شمول مفهومی، ایجاد سلسله‌مراتب دارای طبقه‌بندی را که تعاریف مفهومی در آن از عام به خاص چیده شده‌اند، میسر می‌کند (Bechofer and Goble 2001).

مدل منطق توصیف، بر پایه مجموعه‌ای از تعاریف پیشین از مفاهیم استوار است و بر روابط شمول مفهومی میان آن تعاریف پیشین تکیه دارد. این شبیه به سلسله‌مراتب‌های موجود در اصطلاحنامه‌های سنتی است. با این تفاوت که این زبان، تعداد اندکی از مفاهیم را در اختیار می‌نهد که این مفاهیم، عملگرهایی را برای ترکیب مفاهیم و نقش‌های پیشین به منظور ایجاد تعاریف جدیدی از مفهوم تشکیل می‌دهند و نیز مجموعه‌ای از استدلال‌ها را در اختیار ما قرار می‌دهد تا بتوانیم تفاسیر مربوط به آن ترکیبات را درک کنیم. این استدلال‌ها، مکانیسم‌های استنباطی هستند که می‌توانند دانش ضمنی را از آنچه که از سوی کاربران، دانش آشکار فرض می‌شود، استنباط کنند. همچنین اگر ترکیب جدیدی ساخته شود، با استفاده از این استدلال‌ها، مکان یا جایگاه آن در سلسله‌مراتب شمول مفهومی به طور خودکار تعیین می‌شود. این استدلال‌ها بر قواعد منطقی روشنی استوارند و شامل تعدادی اصل بدیهی^{۱۵} هستند، که روابط پیشین بین مفاهیم را تعریف می‌کنند و اطلاعات اضافه‌ای در مورد ترکیب‌ها ارائه

می‌دهند (Bechofer and Goble 2001). در جدول‌های ۱ و ۲ عملگرهای پایه و اصول بدیهی آورده شده‌اند.

جدول ۱ عملگرهای پایه در منطق توصیف (Bechofer and Goble 2001)

عملگر	وجه عینی	وجه انتزاعی	معناشناسی
رأس	TOP	\top	Δ'
انتها	BOTTOM	\perp	ϕ
همبستگی	(and C D)	$C \cap D$	$\{c c \in C' \cap D'\}$
انفصال	(or C D)	$C \cup D$	$\{c c \in C' \cup D'\}$
سلب	(not C D)	$\neg C$	$\{c c \notin C'\}$
وجودی	(some R C)	$\exists R : C$	$\{c R'(c) \cap C' \neq 0\}$
جامع	(all R C)	$\forall R : C$	$\{c R'(c) \subseteq C'\}$
حداکثر	(at most n R C)	$\leq nR : C$	$\{c R'(c) \cap C' \leq N\}$
حداقل	(at least n R C)	$\geq nR : C$	$\{c R'(c) \cap C' \geq N\}$
همسان	(exact n R C)	$= nR : C$	$\{c R'(c) \cap C' = N\}$
نقش معکوس	(inv R)	R^{-1}	$(R^{-1})'(c) = \{d c \in R'(d)\}$

« کمیته بین‌المللی موزه‌ها و مجموعه‌های پوشاک » (آی‌سی‌اُم)^{۱۶} دارای واژه‌نامه‌ای از اصطلاحات پایه است که برای توصیف انواع لباس‌ها استفاده می‌شود. این واژه‌نامه به سه زیر مجموعه اصلی به نام پوشاک مردانه، پوشاک زنانه، و پوشاک بچگانه، و هر یک از این بخش‌ها به زیربخش‌هایی چون لباس‌های اصلی، بالاپوش‌ها، زیرپوش‌ها، و ... تقسیم می‌شوند. بدین ترتیب موارد تکراری زیادی در این واژه‌نامه وجود دارد و باید با این زیربخش‌های تکراری به صورت ترکیبی برخورد شود. سلسله‌مراتب پوشاک در اصطلاحنامه هنر و معماری نیز مشابه آن است و تقسیمات مشابهی با توجه به شکل و کاربرد لباس دارد. عده‌ای از متخصصان روی این اصطلاحنامه کار کرده و آن را دوباره طراحی نموده‌اند. به علاوه،

اصطلاحاتی از اصطلاحنامه هنر و معماری نیز در طرح وارد شده است. این اصطلاحنامه از پایین به بالا ساخته شده و در تدوین آن از مجموعه‌ای از روابط پیشین برای تعریف مفاهیم به صورت ترکیبی استفاده شده. برای بازنمایی این واژه‌نامه با مدل منطق توصیف، استفاده از موتور استدلال منطق توصیف ضروری است. استفاده از مدل منطق توصیف، روابط سلسله‌مراتبی، هم‌ارزی، و هم‌ایندی را به شکلی شفاف ایجاد می‌کند.

جدول ۲ اصول بدیهی منطق توصیف (Bechofer and Goble 2001)

معناشناسی	وجه انتزاعی	وجه عینی	بدیهیات
$CN^I \subseteq \Delta^I$	CN	(تعریف مفهوم) (CN)	معرفی
$RN^I : \Delta^I \rightarrow \rho(\Delta^I)$	RN	(تعریف نقش) (RN)	معرفی
$\forall c \in \Delta^I, R^I(c) \leq I$	Func R	(R عملکردی)	عملکردی
$\forall c, d, e \in \Delta^I, d \in R^I(c) \wedge e \in R^I(D) \Rightarrow e \in R^I(c)$	Trans R	(R انتقال‌پذیر)	انتقال‌پذیری
$RN^I \subseteq RS^I$	$RS \subseteq RN$	(دلالت نقش) (S و R)	شمول
$C^I \subseteq D^I$	$C \subseteq D$	(دلالت نقش) (D و C)	شمول
$C^I = D^I$	$C = D$	(D و C) (همسنگ)	هم‌سنگی

برای روشن شدن مباحث مطرح‌شده، نمونه‌ای از ایجاد روابط سلسله‌مراتبی را مثال می‌زنیم. واژه‌نامه «آی‌سی‌آم» توصیف‌های مختلفی از لباس‌ها را با توجه به اجزایشان ارائه می‌دهد. مثلاً یک لباس ممکن است از سه جزء کت، جلیقه، و شلوار تشکیل شود، یا این‌که یک کت، یک شلوار، یا یک جلیقه باشد. می‌توانیم از عملگرهایی مثل «یا»^{۱۷}، «و»^{۱۸}، و «دقیقاً»^{۱۹} همراه با محدودکننده‌های عددی استفاده کنیم تا توصیف‌های زیر را از مفاهیم ارائه دهیم:

لباس ۱ = (ترکیب «و» دقیقاً سه فقره تشکیل یابنده)

(برخی تشکیل یافته از کت)

(برخی تشکیل یافته از شلوار)

(برخی تشکیل یافته از کت بارانی))

لباس ۲ = (ترکیب «و» دقیقاً دو فقره تشکیل یابنده)

(برخی تشکیل یافته از کت)

(برخی تشکیل یافته از شلوار))

لباس ۳ = (ترکیب «و» دقیقاً دو فقره تشکیل یابنده)

(برخی تشکیل یافته از کت)

(برخی تشکیل یافته از کت بارانی))

لباس ۴ = (ترکیب «و» دقیقاً دو فقره تشکیل یابنده)

(برخی تشکیل یافته از ژاکت)

(برخی تشکیل یافته از شلوار))

لباس = («یا» لباس ۱، لباس ۲، لباس ۳، لباس ۴)

با توجه به توصیف‌های بالا، شمول معنایی روابط اعم و اخص میان مفاهیم به راحتی قابل فهم است، به طوری که به آسانی می‌توان استنباط کرد که لباس ۲ نوعی از لباس ۱ است؛ و نیز توصیف «برخی تشکیل یافته از شلوار» توصیفی برای لباس ۲ می‌باشد. بنابراین مفهومی که حاوی توصیف مربوط به کت و شلوار است، زیر طبقه لباس ۲ قرار خواهد گرفت. برای مثال اصطلاحی مانند «تاکسیدو»^{۲۰} از اصطلاحنامه هنر و معماری به این طرح بازنمایی شده، اضافه گردید که در یادداشت دامنه به عنوان لباسی که ترکیبی از ژاکت شام و شلوار است، توصیف شده و این در حالی است که ژاکت شام، نوعی کت است و ژاکت نیست. اگر ما تعاریف زیر را به صورت منطقی توصیف وارد پایگاه داده‌های خود کنیم، متوجه می‌شویم که «تاکسیدو» به عنوان نوعی از لباس ۲ طبقه‌بندی می‌شود. در اینجا از اصل بدیهی «شمول»^{۲۱} استفاده شده است.

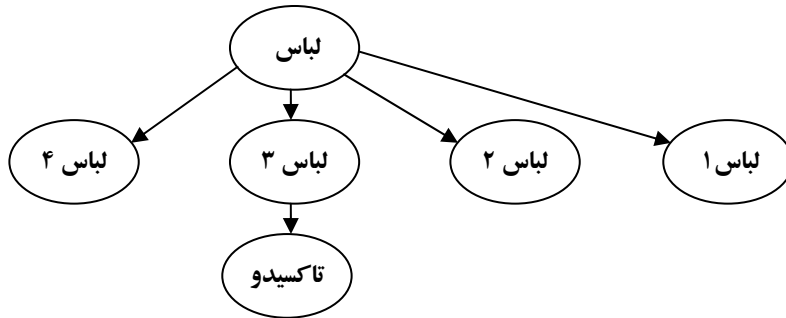
ژاکت شام \supseteq کت

تاکسیدو \supseteq (ترکیب «و» دقیقاً دو فقره تشکیل یابنده)

(برخی تشکیل شده از ژاکت شام)

(برخی تشکیل شده از شلوار))

سلسله مراتب به دست آمده در شکل زیر نشان داده شده است:



نمودار ۱ سلسله مراتب لباس ها با استفاده از منطق توصیف

این مثال نشان می دهد که چگونه رابطه میان اجزای تشکیل دهنده یک ترکیب می تواند رابطه میان آن ترکیب و توصیف های دیگر را تحت تأثیر قرار دهد. در این مثال رابطه میان ژاکت شام و کت، موجب رابطه میان «تاکسیدو» و «لباس ۲» شده است. اگر ما بخواهیم فرض کنیم که ژاکت شام نوعی ژاکت باشد و نه نوعی کت، کافی است تا اصل بدیهی خود را تغییر دهیم:

ژاکت شام \supseteq کت

به

ژاکت شام \supseteq ژاکت

بدین ترتیب، ژاکت شام نوعی از «لباس ۴» خواهد بود و نه نوعی از «لباس ۲». شایان ذکر است که طبقه بندی به صورت خود کار انجام می شود و طراح مجبور نیست که مفاهیم ترکیب شده را به طور مستقیم تعیین محل نماید. علاوه بر این، طبقه بندی به شکل پویا انجام می شود و نیازی به معرفی همه ترکیب های مورد نیاز قبل از استفاده در طبقه بندی نیست.

مشکل یک اصطلاحنامه تازه ایجادشده، در مورد طبقه‌بندی، چندگانه است. در بسیاری موارد مجاز شمردن دیدگاه‌های مختلف از یک مفهوم، ممکن است مفید باشد. مثلاً در اصطلاحنامه هنر و معماری اصطلاح «کت بارانی» زیر طبقه پوشاک از نظر شکل/بالاپوش‌ها/آورکت‌ها/بارانی‌ها ظاهر می‌شود. ممکن است از دیدگاهی دیگر، «کت بارانی» به عنوان یک پوشاک محافظ نیز در نظر گرفته شود. در یادداشت دامنه «پوشاک محافظ» دستوری وجود دارد با این مضمون که «برای لباس‌هایی که برای حفاظت از آب و هوا پوشیده می‌شود، اصطلاح «بالاپوش‌ها» یا اصطلاحات ذیل آن باید استفاده شود». چنین سازوکاری در جستجوهای پیوسته قابل استفاده نیست. بهتر است که طبقه‌بندی چندگانه‌ای وجود داشته باشد تا اصطلاح «کت بارانی» هم به عنوان نوعی از لباس‌های بیرونی و هم به عنوان نوعی از لباس محافظ در نظر گرفته شود و در مکان‌های مختلفی از سلسله‌مراتب ظاهر شود. استفاده از اصول بدیهی‌ای مانند شمول در منطق توصیف (چنانچه بیان شد)، چنین اجازه‌ای را به ما می‌دهد (Bechofer and Goble 2001).

۲-۱-۲. تنظیم ذهنی روابط اصطلاحنامه‌ای

در این رویکرد برای تنظیم روابط بین اصطلاحات از منطق یا استدلال استاندارد استفاده نمی‌شود، بلکه مبتنی بر ذهنیات گروهی متخصص، روابط میان اصطلاحات را تعریف می‌کنند. این رویکرد خود می‌تواند به دو روش استقرایی و قیاسی تقسیم شود. در شیوه استقرایی، فرایند تدوین اصطلاحنامه با ورود مدارک جدید به پایگاه اطلاعاتی مدارک انجام می‌شود. اصطلاحات در طول کار و با مستندسازی اصطلاحات نمایه‌ای تخصیص یافته به اسناد، در اصطلاحنامه جا پیدا می‌کنند و اصطلاحنامه به تدریج گسترده تر می‌شود. در شیوه قیاسی، کل اصطلاحات یک حوزه تخصصی به وسیله جمعی از متخصصان آن حوزه استخراج می‌شود و روابط میان آن‌ها تنظیم می‌شود بدون توجه به این که آیا آن اصطلاحات در بانک اطلاعات اسناد به کار رفته است یا خیر. در تدوین اصطلاحنامه‌های پیوسته‌ای که قرار است روابط بین اصطلاحات در آن‌ها به صورت ذهنی تنظیم شود، به طور معمول از ترکیبی از روش‌های بیان شده استفاده می‌شود. در این حالت، در ابتدا اصطلاحات پایه یک حوزه به وسیله متخصصان آن حوزه جمع‌آوری، و روابط میان آن‌ها تنظیم می‌شود. سپس از یک نرم‌افزار سازگار با وب استفاده می‌شود تا اصطلاحنامه در وب قابل جست‌وجو و مرور باشد و نیز

اصطلاحات جدید موجود در مدارک جدید پایگاه، با نظارت متخصصان وارد نرم‌افزار می‌شود و در محل مناسب خود قرار می‌گیرد. به عنوان مثال می‌توان از اصطلاحنامه موسسه ملی سرطان یاد نمود که در آن ۱۲ متخصص موضوعی با استفاده از محیط ایجاد اصطلاحات کمپانی آپولون و نرم‌افزار مدیریت جریان کار^{۲۲}، برای تضمین دقت و یکپارچگی اصطلاحنامه کار می‌کنند. طراح اصلی که در حال کار در پایگاه داده‌های محلی با نسخه‌های کنونی اصطلاحنامه است، مجموعه‌ای از فهرست‌های کاری را که مفاهیم نیازمند به ویرایش به‌وسیله هر طراح را شرح داده است، ایجاد می‌کند. فهرست‌های کاری به‌وسیله نرم‌افزار مدیریت جریان کار آپولون ارائه می‌شود و وظایف هر یک از طراحان متخصص در آن تعیین شده است. هر طراح، یک نسخه محلی از اصطلاحنامه را دارد که به محض ایجاد تغییرات در آن به‌وسیله وی، تغییرات به نرم‌افزار مدیریت جریان کار باز فرستاده می‌شود. این تغییرات برای تشخیص تناقض‌هایی که ممکن است در این فرایند مشاهده شود، تجزیه و تحلیل می‌شوند و اگر به تصحیحاتی نیاز باشد به عهده متخصصان گذاشته می‌شود. محدوده کاری جدیدی شامل تغییرات قطعی شده به طراحان ارائه می‌شود تا پایگاه داده‌های خود را روزآمد کنند (Golbech et al. 2003).

«مک کلوج»^{۲۳} (2005) ده مرحله برای تدوین اصطلاحنامه با رویکرد غیر خودکار و به صورت عمل‌گرایانه برمی‌شمرد که در اینجا با اندکی تصرف و تلخیص ذکر می‌شود:

- تعیین ضرورت: قبل از تصمیم به ایجاد اصطلاحنامه جدید باید منابع موجود را ارزیابی نمود تا مطمئن شد که آیا این کار واقعاً ضروری است یا خیر. بسیاری از سازمان‌ها بدون آمادگی و دانش قبلی و پیش از شناسایی اصطلاحات موضوعی حوزه خود که ممکن است نیازهای آن‌ها را برآورده کند، به ایجاد طرح‌های موضوعی جدیدی اقدام می‌کنند که در عمل به حجم هنگفت و متنوعی از اصطلاحات منتهی می‌شود که باعث تحمیل محدودیت به جستجوی کاربران می‌شود. با این حال، در یک حوزه تخصصی و فنی یا در سطوح بسیار بالایی از خاص بودن در یک مجموعه، ممکن است اصطلاحنامه موجود ناکافی باشد. در این شرایط، تنها راه قابل پیش‌بینی، ایجاد یک اصطلاحنامه جدید است. کلمه جدید به معنای نوآوری در ایجاد اصطلاحنامه نیست، زیرا بسیار محتمل است که بیش‌تر واژه‌های اصطلاحنامه شامل تلفیقاتی از اصطلاحات شناخته شده‌ای باشند که در جاهای دیگر نیز به کار رفته‌اند. در

جایی که اصطلاحنامه جدید مورد نیاز است، در ابتدا سلسله‌ای از اصطلاحنامه‌ها، رده‌بندی‌ها، و مجموعه اصطلاحات موجود در آن حوزه موضوعی را باید ملاحظه کرد. ممکن است ترکیبی از دو یا چند منبع موجود کافی باشد. در چنین مواردی، مفاهیم مجزا باید در قالب یک ابراصطلاحنامه واحد ترکیب شوند. بدون تلاش برای ایجاد یک ابزار تلفیقی، تعامل کاربران با سامانه‌های اطلاعاتی دچار مشکل خواهد شد، به طوری که استفاده از طرح‌های موضوعی جدا از هم برای جستجو و مرور کاربران، مناسب نخواهد بود. بنابراین باید با متخصصانی از آن حوزه موضوعی درباره سازگار بودن اصطلاحنامه در دست تدوین با مجموعه‌های پیوسته یا کتابخانه‌های مشاوره شود. شاید آن‌ها در حال استفاده از سیاهه‌های محلی از اصطلاحاتی باشند که می‌تواند به راحتی به یک اصطلاحنامه استاندارد وارد، و با آن هماهنگ شود.

- توجه به اثربخشی هزینه: پس از بررسی ضرورت ایجاد اصطلاحنامه جدید، باید ابزار انتخاب و طبقه‌بندی اصطلاحات را معین کرد، زیرا این ابزار باید نیاز به تحقیقات دامنه‌دار قبل از و در هنگام ایجاد اصطلاحنامه، و نیاز به بازبینی در هنگام انتشار و استفاده از اصطلاحنامه را کاهش دهد و به هر چه کم کردن هزینه‌ها کمک کند. همان‌طور که «جی هوج» و «آستین» پیشنهاد می‌کنند، استفاده از ابزارهای خودکار برای استخراج و خوشه‌بندی اصطلاحات، هزینه‌های کلی را کاهش خواهد داد (Hodge and Austin 2002). با این حال استفاده از این ابزارها برای طراحی چندین اصطلاحنامه شاید قابل توجه باشد، اما برای یک اصطلاحنامه هزینه‌بر است.

- انتخاب اصطلاح: پس از توجه ضرورت باید به انتخاب اصطلاحات پرداخت. در اینجا پیش فرض این است که از رویکرد غیر خودکار برای تدوین اصطلاحنامه استفاده خواهد شد (روش‌های خودکار در بخش بعد توضیح داده خواهند شد). برای شروع، یکی از راه‌های زیر انتخاب می‌شوند:

۱. فصل‌ها یا اصطلاحات رأس اصطلاحنامه انتخاب شوند و سپس اصطلاحات به تدریج گردآوری شوند و به آن‌ها تخصیص یابند.
۲. اصطلاحات بر اساس موضوع جمع‌آوری شوند و اصطلاحات رأس مناسب در مرحله بعد انتخاب شوند و به هر یک از دسته‌های موضوعی اختصاص یابند.

با توجه به حوزه‌های موضوعی مختلف، هر یک از این دو روش ممکن است مناسب باشند. اما باید گفت که در طول ایجاد اصطلاحنامه اگر اصطلاحات رأس از قبل تعیین شده باشند، به علت این که اصطلاحات به مرور اضافه می‌شوند، دامنه سرعنوان‌های رأس ممکن است تغییر کند. بنابراین ضروری است در سرتاسر کار از رویکردی انعطاف‌پذیر در سرتاسر کار استفاده شود. صرف نظر از این که کدامیک از دو رویکرد پیش‌گفته استفاده شود، آغاز کار با جمع‌آوری اصطلاحات منفرد از منابع (مانند مقالات دانشگاهی، صورتجلسات کنفرانس‌ها، اطلاعات اداری مانند اسناد دولتی، و وبسایت‌های سازمانی) است. در مرحله جمع‌آوری، شکل‌های مختلف اصطلاحات مشابه و مترادف‌ها را باید یادداشت نمود تا به‌عنوان اصطلاحات پذیرفته‌نشده وارد اصطلاحنامه شوند. این اصطلاحات برای ایجاد روابط هم‌ارزی میان اصطلاحات، جمع‌آوری می‌شوند تا از شکل‌های پذیرفته‌نشده به شکل‌های پذیرفته‌شده ارجاع داده شود. به منظور تشخیص مترادف‌ها باید از متخصصان موضوعی برای ایجاد تدوین تعاریف رسمی برای اصطلاحاتی که به نظر می‌رسد بیانگر مفهوم واحدی هستند، نظرخواهی کرد. با توجه به ماهیت هم‌نویسه‌ها^{۲۴} این اصطلاحات نیز در این مرحله یادداشت می‌شوند. یافتن جای دقیق یک اصطلاح، تعیین میزان اخص بودن آن، و قرار گرفتن آن در محل مناسب نیز اهمیت دارد. تعیین درجه اخص بودن و قرار دادن اصطلاح در محل مناسب منجر به ایجاد روابط سلسله‌مراتبی میان اصطلاحاتی خواهد شد که مربوط به یک مقوله هستند. به‌طور معمول، استحکام اصطلاحنامه با این نوع رابطه سنجیده می‌شود. سلسله‌مراتب‌ها به‌طور معمول بر اساس سه نوع رابطه بین اصطلاحات به وجود می‌آیند که عبارت‌اند از: روابط جنس و نوع^{۲۵}، کل و جزء^{۲۶}، و روابط مصداقی^{۲۷}. تشخیص این نوع روابط و نیز درجه اخص بودن اصطلاحات برای ایجاد سلسله‌مراتب‌ها به عهده متخصصان موضوعی است. اصطلاحات باید به تعدادی از منابع مجموعه قابل اعمال باشند و نباید بقدری اخص باشند که فقط به یک یا دو فقره مربوط شوند. نوع دیگر رابطه‌ای که در اصطلاحنامه‌ها وجود دارد رابطه هم‌ایندی^{۲۸} است که طی آن، اصطلاحاتی که به لحاظ مفهومی به هم مرتبط هستند به یکدیگر ارجاع داده می‌شوند. رابطه هم‌ایندی رابطه‌ای بین اصطلاحات است که نه از نوع هم‌ارز و نه از نوع اعم و اخص است؛ ولی در عین حال حضور یکی، دیگری را به ذهن متبادر می‌کند. تعیین این روابط

در اصطلاحنامه‌ها بسیار ذهنی، و از مسائل بغرنج در تدوین اصطلاحنامه‌ها است. تعداد این نوع روابط نیز بسیار زیاد است و بنابراین باید مرز استفاده از این روابط به دقت مشخص شود.

- ساختار و نشانه‌گذاری: ساختار سلسله‌مراتبی به همراه روابط بین اصطلاحات، باید به خوبی میزان اخص بودن را نشان دهد، به گونه‌ای که اگر سطح اخص بودن ناکافی است یا در آن زیاده‌روی شده است بتوان آن را اصلاح کرد. باید درباره وجود نشانه‌های عددی، الفبایی، یا عددی-الفبایی در اصطلاحنامه تصمیم‌گیری نمود. اگر به وجود نشانه‌ای نیاز بود، به دلیل جایجایی اصطلاحات در طول ایجاد اصطلاحنامه، این مرحله باید یکی از مرحله‌های نهایی کار باشد. ساختار و نشانه‌گذاری باید برای روزآمدسازی و حفاظت آسان اصطلاحنامه از انعطاف لازم برخوردار باشند؛ یعنی به آسانی با توجه به برخی حذفیات یا گسترش‌های اصطلاحنامه، قابل کاهش یا گسترش باشند.

- استانداردها: بهتر است تدوین اصطلاحنامه با تکیه بر استاندارد شناخته‌شده‌ای انجام شود. از استانداردهای معروف برای اصطلاحنامه‌های یک‌زبانه «انسی/نیزو، زد ۳۹/۱۹» (2005)، و برای اصطلاحنامه‌های چندزبانه «ایزو ۵۹۶۴» می‌باشند که خط‌مشی‌هایی را برای تدوین، قالب‌بندی، و مدیریت واژگان کنترل‌شده ارائه می‌دهند. شکل‌های دستوری، استعمال اصطلاحات مفرد و جمع، اختصارات، سرواژه‌ها، و نقطه‌گذاری‌ها از مواردی هستند که در یک استاندارد جامع به آن‌ها اشاره شده. حتی اگر اصطلاحنامه با هدف استفاده تنها در درون یک سازمان نیز طراحی شود باید با استانداردها منطبق باشد؛ زیرا در آینده ممکن است به عنوان مرجع گروه دیگری برای تدوین اصطلاحنامه از آن استفاده شود.

- مشاوره: گروهی از افراد متخصص باید برای نظارت بر انتخاب اصطلاحات و مسائل ساختاری، بویژه روابط معنایی در نظر گرفته شوند. این گروه همچنین وظیفه جمع‌آوری بازخوردهای مربوط به نسخه پیش‌نویس اصطلاحنامه را بر عهده دارند. به دلیل گوناگونی بازخوردها، بهتر است آن‌ها را به وسیله پست الکترونیکی یا گروه‌های مباحثه جمع‌آوری نمود. چون رسیدن به اتفاق نظر در مورد جایگاه برخی واژه‌ها بسیار مشکل است، بهتر است با توجه به حوزه‌های خاص اصطلاحنامه، گروه‌های تخصصی فرعی برای این کار ایجاد شوند و فردی به عنوان سروراستار بر کار آن‌ها نظارت نماید. در هنگام برخورد با اختلاف سلیقه‌ها و عقاید، سروراستار تصمیم نهایی را خواهد گرفت.

- **نرم افزار:** در خرید نرم افزار باید توجه داشت که اصطلاحنامه‌ای که قرار است با این نرم افزار حفاظت، روزآمد، و نمایش داده شود، باید برای کاربران دسترس پذیر، و کاربرپسند باشد. بهتر است که نرم افزار معروف و شناخته شده‌ای خریداری شود: اما در غیر این صورت باید مواد و منابع آموزشی لازم، نرم افزار را پشتیبانی کنند. نرم افزار باید به راحتی روزآمد شود و بتوان در مواقع لازم، مواردی را از آن حذف یا به آن اضافه نمود. «گانزمان»^{۲۹} در ضمیمه مقاله‌ای با عنوان «معیارهای ارزیابی نرم افزارهای اصطلاحنامه‌ای» سیاهه‌ای برای ارزیابی نرم افزارهای اصطلاحنامه‌ای آورده است که در آن، معیارهای ارزیابی یک نرم افزار اصطلاحنامه‌ای را از دو جنبه - معیارهای عمومی و معیارهای عملکردی - مطرح نموده است. در اینجا به اختصار به پاره‌ای از این معیارها اشاره می‌شود.

معیارهای عمومی مواردی چون سازگاری سخت افزاری، میزان حافظه لازم برای نرم افزار، زبان برنامه نویسی، سازنده نرم افزار، سال انتشار نسخه، قیمت، خدمات پشتیبانی، مقبولیت در میان عموم، دستنامه آموزشی نرم افزار و کاربرپسند بودن آن، همخوانی زبان پیام‌ها و شکل‌های گفتگوی نرم افزار با زبان کاربران، امنیت نرم افزار، و نحوه گرفتن فایل‌های پشتیبان را شامل می‌گردند. معیارهای عملکردی نرم افزار عبارت‌اند از فضا و تسهیلاتی که نرم افزار برای ورود اصطلاحات، نشانه‌ها، و یادداشت‌های دامنه در اختیار می‌نهد، روابط اصطلاحنامه‌ای و تعداد آن‌ها که میان اصطلاحات یک واژه‌نامه و چند واژه‌نامه مجاز می‌شمرد، مجاز شمردن روابط میان واژه‌نامه‌های چندزبانه، آسانی وارد کردن داده‌ها در نرم افزار، آسانی تغییر در اصطلاحات و روابط، آسانی حذف واژه‌ها و روابط بدون به هم خوردن روابط دیگر، کنترل جامعیت از راه پذیرفتن ورود واژه‌ها و روابط تکراری، کاربرپسند بودن محیط نمایش اصطلاحنامه در صفحه نمایش، امکان گرفتن خروجی چاپی با دامنه اختیارات بسیار منعطف، امکان گرفتن خروجی با شکل‌های الکترونیکی مانند ایزو^{۳۰}، امکان نمایش سلسله‌مراتبی، معنایی، درختی، و...، امکان جستجو و مرور با ساختارهای معنایی، آسانی ورود اصطلاحات جستجو، قابلیت انتخاب کلمات جستجو از صفحه نمایش اصطلاحنامه، کنترل عبارت‌های ورودی جستجو، جابه‌جا کردن کلمات جستجوی غیرمجاز با اصطلاحات مجاز، و گرفتن آمارهای استفاده از توصیفگرها و کلمات جستجو (Ganzmann 1990).

- **نمایش در وب:** نمایش اصطلاحنامه در محیط پیوسته برای جست‌وجو در پایگاه‌های اطلاعاتی مسئله مهمی است. برای چگونگی نمایش اصطلاحنامه در وب، پرسش‌های زیر کمک خواهند کرد تا تعیین کنیم که اصطلاحنامه چگونه باید در وب به نمایش درآید که بیش‌ترین کارایی را داشته باشد:

- آیا کاربران برای یافتن اصطلاحات، به جستجو و مرور نیاز دارند؟ آیا به هر گزینه جستجو و مرور نیاز دارند یا یکی از آن‌ها کافی است؟
- آیا کاربران مایل‌اند همه اصطلاحنامه و سلسله‌مراتب‌ها را یکجا در صفحه نمایش ببینند یا روی قسمت‌های خاصی تمرکز دارند؟ آیا سلسله‌مراتب‌های خاصی از اصطلاحنامه می‌توانند در پنجره دیگری باز شوند تا این قسمت‌های خاص اصطلاحنامه در داخل بافت کلی اصطلاحنامه دیده شوند؟
- کاربران خواستار کدامیک از نمایه‌های اصطلاحنامه هستند، نمایه الفبایی یا نمایه همه اصطلاحات موجود در اصطلاحنامه؟
- آیا نمایه‌سازها با نسخه مبتنی بر وب کار خواهند کرد یا آن‌ها پشت صحنه نمایش اصطلاحنامه خواهند بود؟

از مهم‌ترین ویژگی‌های نمایش اصطلاحنامه، توجه به دسترس‌پذیری است. برای نمایش اصطلاحنامه، استفاده از خط‌مشی‌های «دبلیوسی‌تری»^{۳۱} (امکانی که در سال ۲۰۰۵ برای دسترس‌پذیرسازی منابع و اصطلاحنامه‌ها در وب به وجود آمده) پیشنهاد می‌شود، زیرا «دبلیوسی‌تری» تلاشی است برای این که وب‌سایت‌ها و تسهیلاتی را برای افراد دارای معلولیت‌های مختلف فراهم کنند تا آنان بتوانند به راحتی اصطلاحنامه را بپذیرند، بفهمند، مرور کنند، و با آن تعامل داشته باشند.

- **ارزیابی:** ارزیابی اصطلاحنامه یکی از مراحل انتهایی فرایند تدوین اصطلاحنامه است. برای ارزیابی، نمایه‌سازان باید از اصطلاحنامه برای نمایه‌سازی حجم نمونه‌ای از مواد موجود استفاده کنند. با این کار، هرگونه شکاف موجود در مجموعه‌های خاص و مدخل‌های نامناسب در اصطلاحنامه شناسایی می‌شوند. ارزیابی اصطلاحنامه توسط کاربران نیز می‌تواند مفید باشد. برای این کار باید از کاربران خواسته شود جستجوهای را انجام دهند تا مشخص شود که آیا صفحه نمایش جستجو و تسهیلات جستجو را به آسانی پیدا می‌کنند و این که در عمل از چه

نوع اصطلاحاتی استفاده می‌نمایند؟ هر گونه اختلاف میان اصطلاحات استاندارد و اصطلاحاتی که بیش‌تر به وسیله کاربران استفاده شده است، آشکار خواهد شد و بدین ترتیب شاید نیاز باشد تا مترادف‌های دیگری به اصطلاحنامه اضافه شوند.

- حفاظت و روزآمدسازی: اصطلاحنامه باید به صورت منظم و نظام‌مند روزآمد شود، زیرا روزآمدسازی نامرتب اصطلاحنامه برای نمایه‌سازان و کاربران مشکلاتی ایجاد خواهد کرد. وجود یک گروه مشاور در رأس کار روزآمدسازی برای کنترل و تأیید ورود اصطلاحات جدید، ضروری است. بدین ترتیب وارد شدن اصطلاحات جدید به طور مرتب کنترل و مستند می‌شود تا از بروز تناقضات احتمالی پیشگیری شود.

۲-۲. تدوین رایانه‌ای اصطلاحنامه (تدوین خودکار)

در این روش‌ها فرایند تعریف روابط اصطلاحنامه‌ای و طبقه‌بندی اصطلاحات با استفاده از برنامه‌های رایانه‌ای و به‌طور خودکار انجام می‌گیرد. بهره‌وری در هزینه‌ها و حداقل بودن رسوخ ذهنیت در روابط اصطلاحنامه‌ای، از جمله مهم‌ترین مزایای این شیوه است. با توجه به این که عمل تطبیق دادن پرسش جستجو با اصطلاحات متن به دلیل روش ساده تطبیق واژه به واژه کلیدواژه‌های متن، در بیش‌تر موارد نارسا است و موتورهای جستجو، به‌طور معمول اسنادی را بازیابی می‌کنند که اصطلاح مورد جستجو، بیش‌ترین وقوع را در آن صفحه داشته باشد، بنابراین روشی برای پردازش خودکار متون، رسیدگی کردن به تعابیر موجود در نوشتجات، توجه به بازیابی با استفاده از کاستن فضای جستجو، و مقایسه خودکار تشابه مدارک به وسیله آمارهای برخاسته از پیکره متن، لازم خواهد بود (Hodge and Austin, 2002). تدوین اصطلاحنامه‌های خودکار بویژه با روش‌های خوشه‌بندی^{۳۲}، همه خواسته‌های فوق را پوشش می‌دهد. در زیر به چند روش موجود برای تولید خودکار اصطلاحنامه خواهیم پرداخت.

۲-۲-۱. روش‌های متنی

این روش‌ها از همسایه‌های موضعی کلمات^{۳۳} یک متن، برای پایه‌گذاری تشابه واژگانی و نسبت دادن گروه‌های کلمات مترادف، یا برای ابهام‌زدایی از کلمات چندمعنا^{۳۴} استفاده

می‌کنند. دو مرحله برای این کار وجود دارد: مرحله خود-راه‌انداز تکرار شونده^{۳۵}، و مرحله طبقه‌بندی خود کار^{۳۶}. همه نمونه‌های یک کلمه چند معنا در یک متن شناسایی می‌شوند. تعدادی از نمونه‌ها به صورت نماینده از هر مجموعه مفهوم‌ها انتخاب، و برای پروردن الگوریتم طبقه‌بندی غیر خود کار از آن‌ها استفاده می‌شود. باقیمانده این مجموعه مفاهیم در یک طبقه‌بندی غیر خود کار پرورده می‌شوند. مفاهیمی از اطلاعات متن (مانند موضوع مدرک) ممکن است به این طبقه‌بندی کننده اضافه شوند. سپس این طبقه‌بندی کننده می‌تواند در یک وضعیت خود کار قرار گیرد تا برای طبقه‌بندی کردن نمونه‌های جدید از آن استفاده شود. به اعتقاد برخی نویسندگان، معنای یک مدرک بیش تر با کلمات اسمی بیان می‌شود و نه با کلمات فعلی. بنابراین این روش فقط برای اسم‌ها مناسب است و قبل از این که در عمل پردازش، این اسم‌ها شناسایی شوند، نیاز است تا از پیکره متن، برآورد کوچکی صورت گیرد. این روش که مراحل اولیه آن، غیر خود کار و مراحل بعدی آن خود کار است، نیاز به یک روش یادگیری مقدماتی برای طبقه‌بندی به وسیله دخالت انسان دارد که ممکن است زمان‌بر باشد.

رویکرد دیگری که در این روش‌ها وجود دارد و به وسیله «اسکاتز» و «پدرسون»^{۳۷} پیشنهاد شده، سلسله‌مراتبی واژگانی است که از آمارهای متنی بهره می‌برد و برای پروردن الگوریتم‌ها نیازی به اطلاعات پیشین از معانی مختلف کلمات نیست. این الگوریتم یک ماتریس قرینه اصطلاح در اصطلاح ایجاد می‌کند که در آن، تعداد دفعاتی که کلمات «i» و «j» با هم در یک پنجره واقع در اطراف کلمه «i» در پیکره متن وقوع می‌یابند را ثبت می‌کند («i» و «j» نمادهای کلمه‌ای از همه کلمات پیکره متن هستند). از تجزیه تک‌ارزشی^{۳۸} برای کم کردن بزرگ‌شدگی ماتریس استفاده می‌شود. با استفاده از این روش، بردار فشرده‌ای برای هر فقره ایجاد می‌شود تا آن بردار، مشخصه همسایگی هم‌وقوعی^{۳۹} آن فقره باشد. این بردارهای هم‌وقوعی فشرده، با استفاده از الگوریتم خوشه‌بندی تراکمی خوشه‌بندی می‌شوند تا یک سلسله‌مراتب واژگانی تولید کنند. این رویکرد، کلمات را نه بر اساس طبقه‌بندی‌های از پیش پردازش شده اصطلاحنامه‌ای مانند قبل، بلکه طبق تشابه کنترل نشده آن‌ها گروه‌بندی می‌کند. با این حال، کم کردن بزرگ‌شدگی (بعدیت) بردارها با فشرده‌سازی آن‌ها، باعث ایجاد پیچیدگی در پردازش می‌شود و از آنجا که این بردارها نمایانگر فرامفاهیم هستند و صرفاً

نمایانگر کلمات نیستند، این کار ممکن است باعث از دست رفتن اطلاعات شود. این روش، مسئول مجاورت هم‌وقوعی‌های کلمه^{۴۰} نیست. هم‌وقوعی‌ها طبق همسایگی دو اصطلاح در پنجره متن وزن نمی‌شوند.

۲-۲-۲. روش‌های برگرفته از سلسله‌مراتب «وردنت»^{۴۱}

این روش‌ها از طبقه‌بندی سلسله‌مراتبی مترادف‌ها، اسامی عام^{۴۲}، و دگرنام‌های^{۴۳} «وردنت» که به وسیله انسان ایجاد شده‌اند بهره می‌برند تا تشابه کلمه‌ای و مناسب‌ترین معنای کلمه را تخمین بزنند. مثلاً عده‌ای از متخصصان از سلسله‌مراتب دگرنام‌ها، اسامی عام و مترادف‌های «وردنت» برای تعیین تشابه کلمه‌ای بر طبق فاصله اصطلاحات در سلسله‌مراتب، بهره برده‌اند. تشابه، با فاصله نسبت عکس دارد؛ یعنی هر چه فاصله اصطلاحات در سلسله‌مراتب بیش‌تر باشد، تشابه کم‌تر می‌شود و بالعکس. با این حال، فاصله پیوندهای رده‌ای^{۴۴} به دلیل متراکم‌تر بودن برخی رده‌ها در سلسله‌مراتب، نسبت به بقیه تغییر می‌کند. باید خاطر نشان ساخت که این تکنیک بر یک سلسله‌مراتب واژگانی پیش‌ساخته استوار است و فقط می‌تواند کلمات حاضر در این سلسله‌مراتب را در رابطه با نوشتجات دیگر پردازش کند و نمی‌تواند تشابه‌ها را در مورد کلمات جدید برآورد کند و حدس بزند.

۲-۲-۳. خوشه‌بندی

یک الگوریتم خوشه‌بندی خودکار، خوشه‌های لغات و انواع روابط میان آن‌ها را نه از روی طبقه‌بندی‌های از پیش تعیین شده، بلکه به‌طور مستقیم از داده‌های توزیعی استخراج می‌کند. با توجه به این امر، گروهی از متخصصان از یک الگوریتم خوشه‌بندی جدید برای توزیع‌های احتمال استفاده برده‌اند تا کلمات را بر اساس همجواری آن‌ها با کلمات دیگر، در روابط گرامری خاص گروه‌بندی کنند. در این رویکرد، کلمات بر حسب توزیع‌شان به عنوان مفعول‌های مستقیم فعل‌ها و نیز بر اساس احتمال عضویت یک کلمه در یک خوشه که با $P(c/w)$ برای هر کلمه نشان داده می‌شود (و نه بر اساس طبقه‌بندی دشوار بولی)، طبقه‌بندی می‌شوند. خوشه‌ها به وسیله یک عامل انقباض از پیش تعیین شده^{۴۵}، به تدریج منقبض می‌شوند و خوشه‌های کلی بزرگ به تدریج با افزایش عامل انقباض، به خوشه‌های کوچک‌تر می‌شکنند. هنگامی که عامل انقباض افزایش می‌یابد، از عمل شکستن خوشه‌ها یک خوشه‌بندی سلسله‌مراتبی ایجاد می‌شود. این رویکرد محدود به روابط گرامری خاصی است و به یک

پیش-پردازشگر نیاز دارد تا پیکره گرامری متن را تجزیه کند و به اجزای کلام، برچسب تخصیص دهد. برنامه‌نویسان و متخصصان این حوزه اظهار کرده‌اند که این روش به ارزیابی و بررسی بیش‌تری نیاز دارد.

۲-۲-۴. روش‌های نقشه خودسامان^{۴۶}

در این روش‌ها، بردارهای کلمه‌ای یا بردارهای سندی^{۴۷}، فضای برداری ورودی نقشه خودسامان را برای طراحی نقشه ریخت‌شناختی^{۴۸}، استنباط تشابه^{۴۹}، و طبقه‌بندی کلمات و اسناد شکل می‌دهند. هدف در این روش‌ها طراحی نقشه ریخت‌شناختی تشابه متنی است که از اطلاعات متنی برای استخراج روابط معنایی استفاده می‌شود. هر کلمه در یک مجموعه اصطلاحات ۲۹ کلمه‌ای با یک بردار هم‌وقوعی ۵۸ عنصری همراه است. مقدار n آمین عنصر در بردار هم‌وقوعی، تعداد دفعاتی که n آمین کلمه این مجموعه اصطلاحات، جلوتر واقع شده است را منعکس می‌کند و $(n+29)$ آمین عنصر، تعداد دفعاتی که n آمین کلمه این مجموعه اصطلاحات به دنبال این کلیدواژه آمده است را نشان می‌دهد و n ، عددی بین ۱ و ۲۹ است. بردارهای ۵۸ عنصری، بردارهای ورودی را برای یک شبکه نقشه خودسامان تشکیل می‌دهند. کل نقشه خودسامان با تعیین شبیه‌ترین واحدها یا بخش‌ها به هر بردار ورودی، برچسب می‌خورد. این برچسب‌های کلمه‌ای به صورت ریخت‌شناختی بر اساس تشابه لغوی و معنایی چیده می‌شوند (Hodge and Austin 2002).

۲-۳. تدوین اصطلاحنامه به کمک رایانه و تدوین خودکار اصطلاحنامه: بررسی و مقایسه

در این بخش، از اصطلاحنامه‌هایی که روابط اصطلاحنامه‌ای در آن‌ها به صورت ذهنی ایجاد می‌شوند یا عنوان «اصطلاحنامه‌های دستی»^{۵۰} یاد می‌کنیم. در این اصطلاحنامه‌ها، مراحل بعد از انتخاب و تدوین روابط اصطلاحنامه‌ای به کمک رایانه انجام می‌شوند. این اصطلاحنامه‌ها به طور معمول اصطلاحنامه‌هایی هستند که برای اولین بار برای یک حوزه موضوعی به وسیله گروهی از متخصصان به روش‌های قیاسی، استقرایی، و یا ترکیبی از آن دو تدوین می‌شوند. این اصطلاحنامه‌ها حالت پایه و کل‌نگر دارند و به اصطلاحنامه کلان (ماکروتراروس)^{۵۱} معروف هستند. از آنجا که روزآمد نگه‌داشتن این نوع اصطلاحنامه‌ها در محیط وب بدون

استفاده از سازوکار منطقی خاصی که روابط اصطلاحنامه‌ای را به صورت خودکار تولید کند، بسیار وقت گیر و پرهزینه است و همچنین، این اصطلاحنامه‌ها اصطلاحات بسیار تخصصی یک حوزه را پوشش نمی‌دهند، از این اصطلاحنامه‌ها در مراحل بعد برای انتخاب اصطلاحات و به کارگیری آن‌ها در منطق‌های توصیف برای ایجاد اصطلاحنامه‌های قابل روزآمدسازی سازگار با وب و هستی‌شناسی‌ها^{۵۲} استفاده می‌کنند. شکل چاپی این نوع اصطلاحنامه‌ها در واقع منبع بسیار مناسبی برای حفظ سرمایه‌های زبانی یک ملت به شمار می‌روند.

از مواردی که همواره به عنوان نقطه ضعف در این نوع تدوین اصطلاحنامه یاد می‌شود، این است که طبقه‌بندی کلمات در روش‌های دستی تدوین اصطلاحنامه، بیش‌تر اوقات غیردقیق است و کلمات به لحاظ معناشناختی از یکدیگر به‌طور مشخص قابل تمیز نیستند. همچنین، طبقه‌بندی کلمات بر اساس ذهنیات اصطلاحنامه‌نویسان و معیارهای آن‌ها برای طبقه‌بندی انجام می‌شود، که همواره شفاف و یکدست نیست (Takenobu, Makoto, and Hozumi 1995).

اصطلاحنامه‌هایی که به کمک رایانه و به وسیله منطق توصیف ایجاد می‌شوند، دارای ساختار و روابط بسیار منطقی هستند که پایه آن، قواعد صریح جبری و ریاضی می‌باشند. روابطی را که به وسیله یک زیربنای منطقی ایجاد شده‌اند می‌توان برای ایجاد هستی‌شناسی‌ها به کار برد، زیرا هستی‌شناسی‌ها (که به تعریف خیلی ساده، شکل توسعه یافته اصطلاحنامه‌ها هستند) دارای روابط منطقی غنی‌تری نسبت به روابط میان واژه‌های اصطلاحنامه می‌باشند. حوزه‌هایی مانند پزشکی که در آن‌ها، اصطلاحات تخصصی از یکدستی مناسبی میان متخصصان برخوردارند و کار انتخاب اصطلاحات کار دشواری نیست، از منطق‌های توصیف برای تدوین اصطلاحنامه و سپس هستی‌شناسی استفاده می‌کنند. اما در رشته‌ها و حوزه‌هایی که یکدستی مناسبی در اصطلاحات تخصصی آن‌ها میان متخصصان آن حوزه وجود ندارد و از واژگان متفاوتی برای بیان مفاهیم یکسان استفاده می‌شود، برای اجتناب از انتخاب و آرایش ذهنی کلمات در اصطلاحنامه، امروزه روش‌های خودکار برای تدوین اصطلاحنامه این حوزه‌ها مناسب‌تر به نظر می‌رسد.

یکی از چالش‌های عمده در استفاده از منطق‌های توصیف در تدوین اصطلاحنامه‌ها و حفاظت و توسعه اصطلاحنامه‌های حجیمی که به این صورت تدوین می‌شوند عبارت است از

ویرایش پراکنده و گاه و بیگاه اصطلاحنامه که ممکن است به وسیله افراد مختلفی صورت گیرد و این امر ممکن است مانعی برای یکدستی اصطلاحنامه باشد. چالش دیگر این است که متخصصان موضوعی بدون دانستن منطق توصیف زیربنایی اصطلاحنامه، دست به ویرایش اصطلاحنامه می‌زنند که این امر نیز می‌تواند یکدستی اصطلاحنامه را با مشکل مواجه کند (Hartel et al. 2005).

اصطلاحنامه‌هایی که به‌طور خودکار تهیه می‌شوند، معمولاً اصطلاحنامه‌هایی هستند که به‌صورت یک زیرسازه فرعی در موتورهای جستجو عمل می‌کنند. از آنجا که موتورهای جستجو معمولاً از انطباق واژه به واژه برای پاسخ‌گویی به پرسش جستجو و از فراوانی کلمه یا کلمات جستجو در صفحات وب برای رتبه‌بندی آن‌ها استفاده می‌کنند، اصطلاحنامه‌هایی لازم است تا پرسش‌های کاربران به وسیله آن‌ها یکدست شوند و اصطلاحات پذیرفته‌شده و موجود در پیکره متن که در اصطلاحنامه طبقه‌بندی شده‌اند، مورد جستجو قرار گیرند (Hodge and Austin 2002). این اصطلاحنامه‌ها که از راه خوشه‌بندی لغات و اصطلاحات بر اساس الگوریتم‌های برخاسته از پیکره متن حاصل می‌شوند، برای این موتورهای جستجو حالت زیربنایی دارند؛ بدین ترتیب که کلمات جستجو ابتدا وارد این اصطلاحنامه زیربنایی می‌شوند و در آنجا مورد تحلیل قرار می‌گیرند. اگر اصطلاح جستجو واژه پذیرفته‌نشده‌ای بود، شکل پذیرفته‌شده آن به وسیله اصطلاحنامه به‌طور خودکار انتخاب می‌شود و اصطلاحات اعم و اخص آن نیز انتخاب می‌گردند. در این حالت صفحات وب، دیگر بر اساس یک اصطلاح ناپخته جستجو که به وسیله کاربر وارد سامانه بازایی اطلاعات شده است، مورد تحلیل قرار نمی‌گیرند؛ بلکه بر اساس مجموعه اصطلاحاتی که از اصطلاحنامه زیربنایی موتور جستجو استخراج شده‌اند مورد تحلیل قرار می‌گیرند. در این حالت، صفحات وب با یک مدخل اصطلاحنامه‌ای مقایسه می‌شوند، و نه فقط با یک اصطلاح خاص. رتبه‌بندی پاسخ‌ها نیز معمولاً بر اساس میزان همپوشانی صفحات وب با سلسله‌مراتب‌های مدخل مورد جستجو انجام می‌گیرد. خود اصطلاحنامه نیز به تدریج و به‌طور خودکار با ورود اسناد جدید و بر اساس الگوریتم‌های خاص برخاسته از آمارهای پیکره متن، روزآمد می‌شود و اصطلاحات جدید به تدریج جای خود را در سلسله‌مراتب‌های اصطلاحنامه پیدا می‌کنند. روابط اصطلاحنامه‌ای در این نوع اصطلاحنامه‌ها از زیرساخت منطقی برخوردار نیستند و بدین ترتیب این اصطلاحنامه‌ها

را نمی توان به هستی شناسی های پر قدرتی برای تجزیه و تحلیل معناشناختی صفحات وب تبدیل کرد؛ بلکه ابزارهایی برای بازیابی اطلاعات در سامانه های حجیم اطلاعاتی باقی می مانند.

از نقاط قوت این روش این است که نوع اصطلاحاتی که محققان آن حوزه به کار می برند، پایه فرایند نمایه سازی و تخصیص اصطلاحات نمایه ای به اسناد و نیز تدوین اصطلاحنامه و تعیین روابط اصطلاحنامه ای می باشد (Lassi 2002).

۳. نتیجه گیری

اصطلاحنامه های پیوسته یا به کمک رایانه تدوین و بازنمایی می شوند یا به صورت رایانه ای تولید می گردند. در این دو رویکرد کلی تدوین، حرکت از ذهنیت به عینیت است، بدین ترتیب که روابط اصطلاحنامه ای ممکن است توسط متخصصان تنظیم شوند و از رایانه برای بازنمایی آن ها استفاده گردد که در این صورت، رسوخ ذهنیت در تدوین اصطلاحنامه، بیشینه است. ممکن است روابط به وسیله استدلال گره های خاصی که مبتنی بر قواعد ریاضی و جبر هستند تولید شوند که در این صورت، رسوخ ذهنیت در تنظیم روابط بسیار کاهش می یابد، ولی انتخاب اصطلاحات همچنان به وسیله انسان صورت می گیرد. در روش های رایانه ای تولید اصطلاحنامه، ممکن است برای تولید خود کار اصطلاحنامه و تشخیص مترادف ها و همایند های کلمات، از همسایگی موضعی اصطلاحات متن به عنوان الگو به وسیله انسان استفاده گردد که در این صورت نیز رسوخ ذهنیت در تدوین الگو قابل مشاهده است، ولی نسبت آن بسیار کم تر شده. در روش های جدید تولید خود کار اصطلاحنامه که از آمارهای متنی و الگوریتم های پیچیده برای انتخاب اصطلاحات و تشخیص روابط اصطلاحنامه ای و خوشه بندی لغات استفاده می شود، رسوخ ذهنیت محدود به تدوین الگوریتم های تولید اصطلاحنامه است و پس از این مرحله، اصطلاحنامه به صورت خود کار و بدون دخالت انسان تولید می شود. هر دوی این دو رویکرد کلی می توانند معایب و مزایایی داشته باشند که با توجه به حوزه ها و موقعیت های مختلف، باید رویکرد مناسب برای تدوین اصطلاحنامه انتخاب شود. اصطلاحنامه هایی که به صورت خود کار تولید می شوند ممکن است با نیازهای کاربران و اصطلاحات مورد نظر آن ها برای انجام جست و جو و بازیابی فاصله داشته باشند و نیازهای خاص آنان لحاظ نشده

باشد. همچنین برای حوزه‌های فعالیت کوچک، مقرون به صرفه نیستند. در عوض سرعت تولید، یکدستی روابط اصطلاحنامه‌ای، قابلیت روزآمدسازی، ویرایش سریع و کارآمد، و بازنمایی مناسب این نوع اصطلاحنامه‌ها در وب، از نقاط قوت آن‌ها محسوب می‌شوند.

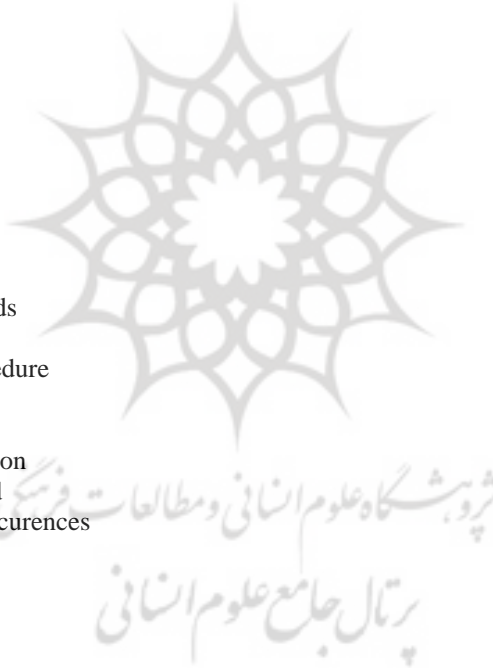
۴. منابع

1. Bechhofer, S. and C. Goble. 2001. Thesaurus construction through knowledge representation. *Data & knowledge engineering* 37(1): 25-45.
2. Ganzmann, Jochen. 1990. Criteria for the evaluation of thesaurus software. *International classification* 17(3/4): 148-157.
3. Gilchrist, Alen. 2003. Thesauri, taxonomies and ontologies-an ethymological note. *Journal of documentation* 59(1): 7-18.
4. Golbeck, J., G. Fragoso, F. Hartel, J. Hendler, B. Parsia, J. Oberthaler. 2003. The National Cancer Institute's thesaurus and ontology. *Journal of web semantics* 1(1): 75-80.
5. Hartel, F. W., S. de Coronado, R. Dionne, G. Fragoso, J. Golbeck. 2005. Modeling a description logic vocabulary for cancer research. *Journal of biomedical informatics* 38(2): 114-129.
6. Hodge, V. J. and J. Austin. 2002. Hierarchical word clustering-Automatic thesaurus generation. *Neurocomputing* 48(1): 819-846.
7. Lassi, Monica. 2002. *Automatic thesaurus construction*. A paper written within the GSLT course Linguistic Resources, 1-10. <http://www.adm.hb.se/~mol/gslt/thesauri.pdf> (accessed August 20, 2007).
8. MacCulloch, E. 2005. Thesauri: practical guidance for construction. *Library review* 54(7): 403-409.
9. Shiri, A. A. and C. Revie. 2000. Thesauri on the web: current developments and trends. *Online information review* 24(4): 273-280.
10. Takenobu T., I. Makoto, and T. Hozumi. 1995. Automatic Thesaurus Construction based on Grammatical Relations. In *proceedings of 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, ed. C. Mellish, vol 2: 1308-1313. Montreal, Canada. <http://dli.iiit.ac.in/ijcai/IJCAI-95-VOL2/PDF/037.pdf> (accessed August 20, 2007).

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پیشینه‌ها
پرتال جامع علوم انسانی

1. thesaurus
2. Roget
3. Thesaurus of English words and phrases
4. Helen Brownson

5. mapping
6. Computer Asisted Thesaurus Construction
7. Computerized Thesaurus Construction
8. supervized
9. unsupervized
10. Reasoning Engine
11. Description logic (DL)
12. ICOM
13. KL-one
14. subsumption
15. axiom
16. International Committee for the Museums and Collections of Custome (ICOM)
17. Or operator
18. And operator
19. Exact operator
20. Tuxedo
21. inclusion
22. Workflow Manager Software
23. MacCulloch
- 24 . homograph
25. generic
26. whole-part
27. instance
28. associative relations
29. Jochen Ganzmann
30. ISO
31. WC3
32. clustering
33. local neighbourhood of words
34. polysemic words
35. iterative bootstrapping precedure
36. unsupervized catagorization
37. Schutz and Pederson
38. singular-valued decomposition
39. co-occurrence neighborhood
40. proximity of the word co-occurences
41. WordNet Hierarchy
42. hyponyms
43. metonyms
44. taxonomic links
45. deterministic annealing
46. Self-organizing Map Methods (SOM)



47. word vectors or document vectors
48. topological map
49. similarity
50. handcrafted thesauri
51. macro - thesaurus
52. ontologies

۱) کارشناس ارشد کتابداری و اطلاع‌رسانی از دانشگاه تهران
پست الکترونیکی: arasto456@yahoo.com

