

بسم الله الرحمن الرحيم



وزارت علوم، تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

طرح پژوهشی

مفهوم‌شناسی و روش تدوین فرااصطلاح‌نامه

مجری

ملوک‌السادات حسینی بهشتی

حقوق: پژوهشگاه علوم و فناوری اطلاعات ایران

طرح پژوهشی مفهومی‌شناسی و روش تدوین فرااصطلاح‌نامه پیرو قرارداد شماره ۳۳/۱۵۰۴۱ مورخ ۱۳۹۷/۱۲/۲۱ میان پژوهشگاه علوم و فناوری اطلاعات ایران (کارفرما) و خانم ملوک‌السادات حسینی بهشتی اجرا شده است. گزارش حاضر یک جلد از مستندات این طرح است.

این گزارش و تمامی حقوق مادی آن بر اساس قانون حمایت حقوق مؤلفان و مصنفان و هنرمندان، مصوب ۱۳۴۸ و اصلاحیه‌های بعدی آن و همچنین آیین‌نامه‌های اجرایی این قانون متعلق به پژوهشگاه علوم و فناوری اطلاعات ایران و هرگونه استفاده از تمامی یا پاره‌ای از آن، شامل: نقل قول، تکثیر، انتشار، کاربرد نتایج، تکمیل و مانند آن‌ها به صورت چاپی، الکترونیکی یا وسایل دیگر فقط با اجازه کتبی پژوهشگاه امکان‌پذیر است. نقل قول در حد هزار واژه در انتشارات علمی مانند کتاب و مقاله با درج اطلاعات کامل کتابشناختی، نیازی به مجوز پژوهشگاه ندارد.

صفحه اصلی گواهی انجام طرح، در این صفحه قرار گیرد

مفهوم‌شناسی و روش تدوین فرااصطلاح‌نامه

مدیر طرح پژوهشی: ملوک‌السادات حسینی بهشتی، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

نشانی: تهران، خیابان انقلاب، چهارراه فلسطین، شماره ۱۰۹۰، طبقه سوم، اتاق ۳۱۸

صندوق پستی: ۱۳۱۵۷۷۳۳۱۴ تلفن: ۰۲۱-۶۶۴۹۴۹۸۰ (داخلی ۳۲۰)، دورنگار: ۰۲۱-۶۶۴۶۲۲۵۴

وبگاه: <http://www.irandoc.ac.ir>

رایانامه: beheshti@irandoc.ac.ir

چکیده

فرااصطلاح‌نامه ترکیبی است از زبان‌های ساختاریافته موجود که از طریق برقراری پیوند، ادغام و تلفیق آن‌ها به‌دست آمده است. اصطلاح‌نامه‌ها با یکدیگر مطابقت می‌یابند و از هر مفهوم در یک اصطلاح‌نامه، نشانگرهایی به نزدیک‌ترین مفهوم معادل در اصطلاح‌نامه‌های دیگر ایجاد می‌شود. این امر امکان ترجمه پرسش کاربر به کاوشی برای یک پایگاه اطلاعاتی مشخص را که با اصطلاح‌نامه تخصصی نمایه‌سازی شده است، فراهم می‌سازد. واژگان نیز ادغام می‌شوند و یک مبنای دانش غنی‌تر و جامع‌تر ایجاد می‌گردد. هر یک از واژگان ساختاریافته، واژه‌ای به عنوان «اصطلاح مرجع» برای مفهوم انتخاب می‌شود. فرااصطلاح‌نامه تنها یک مجموعه واژگان مقید نیست، بلکه ابزاری برای افزایش سودمندی و بهره‌وری از فرهنگ لغات و اصطلاح‌نامه‌های موجود است که به منظور هماهنگی میان زبان‌های ساختاریافته از طریق مطابقت‌دادن، ادغام و تلفیق اصطلاح‌نامه‌هاست که از ویژگی‌های فرااصطلاح‌نامه شمرده می‌شود. این هماهنگی، انتقال رکورد بین فیله‌های پایگاه‌های اطلاعاتی و نیز جست‌وجو در آن‌ها را سهولت می‌بخشد. بنابراین فرااصطلاح‌نامه پایگاه اطلاعات واژگانی جامع، بزرگ و گسترش‌یابنده چندمنظوره و چندزبانی حاصل از انطباق، اتصال، ادغام و تلفیق اصطلاح‌نامه‌ها و زبان‌های مقید موجود است که در اصل به قصد استفاده توسعه‌دهندگان آن تهیه و تدوین می‌شود. در عین حال می‌تواند برای سازندگان پایگاه‌های اطلاعاتی، کتابداران، متخصصان اطلاع‌رسانی و سایر علاقه‌مندان، مرجع مفیدی باشد. همچنین می‌تواند اطلاعات مورد نیاز برنامه‌های رایانه‌ای را برای ایجاد داده‌های استاندارد تأمین نماید، پرسش‌های کاربران را تفسیر و از طریق پالایش و جرح و تعدیل سؤالات با آن‌ها تعامل داشته باشد و اصطلاحات کاربران را به واژگان مورد استفاده در منابع مرتبط برگرداند. فرااصطلاح‌نامه معمولاً در حوزه معینی از دانش شکل می‌گیرد.

در ایران‌داک اصطلاح‌نامه‌های تخصصی در دهه ۱۳۸۰ تدوین شده است و در سال‌های ۹۷-۱۳۹۶ روزآمدسازی شده‌اند. علاوه بر اصطلاح‌نامه‌ها، مجموعه‌ای از واژه‌نامه‌ها و واژگان تخصصی را نیز تدوین و به چاپ رسانده است. برای جستجو و کاوش دقیق‌تر ضرورت دارد که این واژگان و اصطلاح‌نامه‌ها با یکدیگر تلفیق یابند تا از ریزش کاذب اطلاعات در جستجوی پایگاه گنج ممانعت به عمل آید و سرعت دستیابی هدفمند به اطلاعات تخصصی فراهم گردد. در این طرح ابتدا به بررسی منابع گوناگون با موضوع فرااصطلاح‌نامه و شناسایی مفاهیم وابسته به آن، بیان کاربرد و نیز مراحل تدوین آن پرداخته می‌شود.

کلیدواژه‌ها: اصطلاح‌نامه؛ فرااصطلاح‌نامه؛ هستی‌شناسی؛ وب معنایی؛ ایران‌داک.

فهرست مطالب

عنوان	شماره صفحه
چکیده	ج
فهرست جدول‌ها	ذ
فهرست شکل‌ها	ر
فصل اول: کلیات تحقیق	
۱-۱. مقدمه	۲
۲-۱. پیشینه پژوهش	۳
۳-۱. پردازش زبان طبیعی در بازیابی اطلاعات	۵
۴-۱. فرایند بازیابی اطلاعات	۸
فصل دوم: فراداده	
۱-۲. تعریف فراداده	۱۲
۱-۱-۲. انواع فراداده	۱۳
۲-۱-۲. دیجیتالی‌شدن فراداده	۱۹
۲-۲. پیدایش اینترنت	۲۰
۳-۲. استانداردهای فراداده	۲۵
۱-۳-۲. معناشناسی در فراداده	۲۶
۲-۳-۲. نحو	۲۷
۳-۳-۲. قوانین محتوا	۲۸
۴-۳-۲. واژگان کنترل‌شده	۲۹
۵-۳-۲. طبقه‌بندی	۲۹
۱-۵-۳-۲. طبقه‌بندی به‌عنوان فراداده: جهان در سلسله‌مراتب	۳۰
۲-۵-۳-۲. اصطلاح‌نامه: انعطاف‌پذیری در سلسله‌مراتب	۳۲

۳۴.....	۲-۳-۳. سلسله مراتب مسطح: رده بندی چهریزه ای
۳۶.....	۲-۴. فراداده ها در کتابخانه دیجیتالی
۳۶.....	۲-۴-۱. تاریخچه
۳۷.....	۲-۴-۲. تعریف
۳۷.....	۲-۴-۳. ضرورت فراداده
۳۸.....	۲-۴-۴. کارکردهای فراداده
۳۸.....	۲-۴-۵. مزایای فراداده
۳۹.....	۲-۴-۶. کاربردهای فراداده
۳۹.....	۲-۴-۷. عناصر و اجزای تشکیل دهنده فراداده
۴۰.....	۲-۴-۷-۱. نمونه یک رکورد فراداده ای
۴۰.....	۲-۴-۷-۲. دامنه بین المللی
۴۰.....	۲-۴-۷-۳. گسترش پذیری
۴۰.....	۲-۵. قالب داده های رده بندی
۴۱.....	۲-۵-۱. استانداردهای ساختار توصیفی منابع
۴۱.....	۲-۵-۲. استانداردهای کتابخانه دیجیتال
۴۲.....	۲-۶. پردازش هسته دوبلین چیست؟
۴۲.....	۲-۶-۱. اهداف هسته دوبلین
۴۲.....	۲-۶-۱-۱. سادگی ایجاد و نگهداری
۴۲.....	۲-۶-۱-۲. قابل فهم بودن برای همه (از لحاظ معناشناختی)
۴۳.....	۲-۶-۲. مجموعه عناصر فراداده ای هسته دوبلین
۴۳.....	۲-۶-۲-۱. عناصر سطح ساده
۴۴.....	۲-۶-۲-۲. عناصر سطح ویژه
۴۴.....	۲-۶-۲-۳. مارک در زبان نشانه گذاری گسترش پذیر
۴۵.....	۲-۶-۲-۴. طرح فراداده ای توصیف شیء
۴۵.....	۲-۶-۲-۵. مهم ترین مزایای این طرح عبارت است از:
۴۷.....	۲-۶-۳. تعریف مستندسازی
۴۷.....	۲-۶-۴. مزایای مدز
۴۷.....	۲-۶-۵. اهداف مدز
۴۸.....	۲-۷. استاندارد انتقال و کد گذاری فراداده ها

فصل سوم: فرااصطلاح‌نامه (متا تزاروس)

- ۱-۳. فرااصطلاح‌نامه (متا تزاروس) ۵۰
- ۱-۱-۳. اصطلاح‌نامه چیست؟ ۵۰
- ۲-۳. فرااصطلاح‌نامه چیست؟ ۵۲
- ۳-۳. تفاوت‌های اصطلاح‌نامه و فرااصطلاح‌نامه (متا تزاروس) ۵۳
- ۳-۴. فلسفه وجودی فرااصطلاح‌نامه ۵۴
- ۳-۵. ضرورت و فایده فرااصطلاح‌نامه‌ها ۵۵
- ۳-۶. زبان‌های فرااصطلاح‌نامه ۵۵
- ۳-۷. شبکه معناشناختی ۵۶
- ۳-۸. نقشه منابع اطلاعاتی ۵۷
- ۳-۹. فرهنگ اصطلاحات تخصصی ۵۷
- ۳-۱۰. ویژگی‌های فرااصطلاح‌نامه ۵۷
- ۳-۱۰-۱. دامنه فرااصطلاح‌نامه‌ها ۵۷
- ۳-۱۰-۲. حفظ مفاهیم و معانی از واژگان منبع ۵۷
- ۳-۱۰-۳. لزوم کاربرپسند کردن فرااصطلاح‌نامه ۵۷
- ۳-۱۰-۴. فرهنگ واژگان منبع ۵۸
- ۳-۱۰-۵. شمول مجموعه رموز استانداردهای بین‌الملل و ترمینولوژی ۵۸
- ۳-۱۰-۶. شمول زبان‌ها به غیر از زبان انگلیسی ۵۸
- ۳-۱۰-۷. مفاهیم، اسامی مفهومی و شناسه‌های آن‌ها ۵۸
- ۳-۱۰-۸. مفاهیم و شناسه‌های مفهومی ۵۹
- ۳-۱۰-۹. اسامی مفهومی و شناسه‌های زنجیره‌ای ۵۹
- ۳-۱۰-۱۰. اتم‌ها و شناسه‌های اتمی (شناسه‌های عنصری و عناصر) ۵۹
- ۳-۱۰-۱۱. واژه‌ها و شناسه‌های لغوی ۶۰
- ۳-۱۰-۱۲. شناسه‌های رابطه‌ای و روابط در فرااصطلاح‌نامه ۶۰
- ۳-۱۰-۱۳. اساس انواع روابط غیر مترادف‌ها ۶۰
- ۳-۱۰-۱۴. برچسب‌های روابط ۶۰
- ۳-۱۱. تدوین فرااصطلاح‌نامه ۶۰
- ۳-۱۱-۱. تلفیق اصطلاح‌نامه‌ها ۶۳
- ۳-۱۱-۲. توسعه و روزآمدسازی اصطلاح‌نامه‌های تخصصی ۶۴
- ۳-۱۱-۳. تدوین فرااصطلاح‌نامه ۶۵
- ۳-۱۱-۴. مراحل تلفیق اصطلاح‌نامه‌ها ۷۰
- ۳-۱۱-۵. نحوه پیاده‌سازی فرااصطلاح‌نامه ۷۲

فصل چهارم: تأثیر آنتولوژی‌ها بر اصطلاح‌نامه و فرااصطلاح‌نامه

۷۵.....	۱-۴. تأثیر آنتولوژی‌ها بر اصطلاح‌نامه و فرااصطلاح‌نامه
۷۸.....	۲-۴. وب معنایی
۷۸.....	۱-۲-۴. مقدمه
۷۹.....	۲-۲-۴. تعریف
۷۹.....	۱-۲-۲-۴. معرفی وب معنایی
۷۹.....	۳-۲-۴. وب معنایی فراتر از کلمات
۸۰.....	۴-۲-۴. اهداف وب معنایی
۸۱.....	۵-۲-۴. معماری وب معنایی
۸۹.....	۶-۲-۴. مروری مختصر بر عملکرد وب معنایی
۸۹.....	۷-۲-۴. نقش هستی‌شناسی در وب معنایی
۹۰.....	۸-۲-۴. اُ.دبلیو.ال زبان بازنمون هستی‌شناسی

فصل پنجم: جمع‌بندی

۹۲.....	۱-۵. جمع‌بندی
۹۵.....	فهرست منابع

فهرست جدول‌ها

عنوان	شماره صفحه
جدول ۱-۷. کارکردهای فراداده	۳۸
جدول ۱-۲. عناصر سطح ساده	۴۳
جدول ۲-۲. عناصر سطح ویژه	۴۴

فهرست شکل‌ها

عنوان	شماره صفحه
شکل ۱-۲. هرم اکوف	۱۶
شکل ۲-۲. پانزده عنصر ساده دوبلین کور	۲۲
شکل ۳-۲. طرح فراداده‌ای توصیف مستند MADS (Metadata Authority Description Schema)	۴۶
شکل ۱-۳. نمونه‌ای از نمایش اینترنتی سلسله‌مراتبی اصطلاح‌نامه یونسکو	۵۲
شکل ۲-۳. نمایش کامل مدرک مربوط به "employment" در GEMET	۵۴
شکل ۳-۳. UMLS	۵۶
شکل ۴-۳. جایگاه هستی‌شناسی در پلکان معنایی	۶۱
شکل ۵-۳. تلفیق اصطلاح‌نامه‌ها	۶۴
شکل ۶-۳. فرااصطلاح‌نامه «یوام ال اس» اطلاعات مربوط به هر یک از مفاهیم	۶۷
شکل ۱-۴. هستی‌شناسی علوم پایه	۷۶

فصل اول

کلیات تحقیق

۱-۱. مقدمه

فرااصطلاح‌نامه یک پایگاه اطلاعاتی واژگان چندزبانه دارای اهداف چندگانه و بسیار بزرگ است که شامل اطلاعاتی در مورد مفاهیم موضوعی خاص در یک زمینه علمی با اصطلاحات مختلف در آن رشته موضوعی و رابطه‌های معنایی میان آن‌ها می‌باشد. فرااصطلاح‌نامه از نسخه‌های الکترونیکی اصطلاح‌نامه‌های مختلف، رده‌بندی‌ها، مجموعه واژگان کنترل‌شده که در زمینه‌های موضوعی همان رشته وجود دارد و نیز نمایه‌ها و فهرست‌های متون آن رشته ساخته می‌شوند. فرااصطلاح‌نامه نوعی از اصطلاح‌نامه است که سعی در یکپارچه‌سازی اصطلاح‌نامه‌ها و واژگان موجود دارد. فرااصطلاح‌نامه دارای نرم‌افزاری هوشمند است. در واقع می‌توان گفت فرااصطلاح‌نامه، اصطلاح‌نامه اصطلاح‌نامه‌هاست. هر رکورد فرااصطلاح‌نامه شامل واژگان اصلی مستند شده است، که به‌عنوان مرجعی برای واژگان مبهم به کار می‌رود و همراه با واژگان مرتبط (وابسته) ذخیره شده است (یورلند، ۲۰۰۶).

فرااصطلاح‌نامه ترکیبی است از زبان‌های کنترل‌شده موجود که از طریق برقراری پیوند، ادغام و تلفیق آن‌ها به دست آمده است. اصطلاح‌نامه‌ها با یکدیگر مطابقت می‌یابند و از هر مفهوم در یک اصطلاح‌نامه، نشانگرهایی به نزدیک‌ترین مفهوم معادل در اصطلاح‌نامه‌های دیگر ایجاد می‌شود. این امر امکان ترجمه پرسش کاربر به کاوشی برای یک پایگاه اطلاعاتی مشخص را که با اصطلاح‌نامه به‌خصوصی نمایه‌سازی شده است، فراهم می‌سازد. واژگان نیز ادغام می‌شوند و یک مبنای دانش غنی‌تر و جامع‌تر ایجاد می‌گردد. در هر یک از واژگان ساختاریافته، اصطلاحی به‌عنوان مرجع برای مفهوم انتخاب می‌شود. فرااصطلاح‌نامه تنها مجموعه واژگان کنترل‌شده نیست، بلکه ابزاری برای افزایش سودمندی و بهره‌وری از فرهنگ لغات و اصطلاح‌نامه‌های موجود است که به‌منظور هماهنگی میان زبان‌های کنترل‌شده از طریق مطابقت‌دادن، ادغام و تلفیق اصطلاح‌نامه‌هاست که از ویژگی‌های فرااصطلاح‌نامه شمرده می‌شود. این هماهنگی، انتقال رکورد بین فیلدهای پایگاه‌های اطلاعاتی و نیز جستجو در آن‌ها را سهولت می‌بخشد. مانند: فرااصطلاح‌نامه پزشکی (مش) که جهت هماهنگی

ساختن زبان جستجوها در زمینه پزشکی از ادغام سرعنوان‌های موضوعی پزشکی و اصطلاح‌نامه‌های تخصصی پزشکی ایجاد شده است. در ایران‌داک اصطلاح‌نامه‌های تخصصی در دهه ۱۳۸۰ تدوین شده است و در سال‌های ۹۷-۱۳۹۶ روزآمدسازی شده‌اند. علاوه بر اصطلاح‌نامه‌ها، ایران‌داک مجموعه‌ای از واژه‌نامه‌ها و واژگان تخصصی را نیز تدوین و به چاپ رسانده است. برای جستجو و کاوش دقیق‌تر ضرورت دارد که این واژگان و اصطلاح‌نامه‌ها با یکدیگر تلفیق یابند تا از ریزش کاذب اطلاعات در جستجوی پایگاه گنج ممانعت به عمل آید و سرعت دستیابی هدفمند به اطلاعات تخصصی فراهم آید. بنابراین فرااصطلاح‌نامه پایگاه اطلاعات واژگانی جامع، بزرگ و گسترش‌یابنده چندمنظوره و چندزبانی حاصل از انطباق، اتصال، ادغام و تلفیق اصطلاح‌نامه‌ها و زبان‌های کنترل‌شده موجود است. فرااصطلاح‌نامه در اصل به قصد استفاده توسعه‌دهندگان آن تهیه و تدوین می‌شود. اما در عین حال می‌تواند برای سازندگان پایگاه‌های اطلاعاتی، کتابداران، متخصصان اطلاع‌رسانی و سایر علاقه‌مندان، مرجع مفیدی باشد. فرااصطلاح‌نامه همچنین می‌تواند اطلاعات مورد نیاز برنامه‌های کامپیوتری را برای ایجاد داده‌های استاندارد تأمین نماید، پرسش‌های استفاده‌کنندگان را تفسیر و از طریق پالایش و جرح و تعدیل سؤالات با آن‌ها تعامل برقرار کند و اصطلاحات استفاده‌کنندگان را به واژگان مورد استفاده در منابع ذی‌ربط برگرداند. فرااصطلاح‌نامه معمولاً در حوزه معینی از دانش شکل می‌گیرد. با توجه به اصطلاح‌نامه‌های علوم پایه، کشاورزی و فنی مهندسی در ایران‌داک می‌توان جهت توسعه بیشتر از واژه‌نامه‌های موجود نیز استفاده کرده و یک فرااصطلاح‌نامه استاندارد جهت پشتیبانی پایگاه‌های اطلاعات تخصصی ایران‌داک ایجاد نمود تا از ریزش کاذب اطلاعات در بازیابی دانش نهفته در پایگاه گنج جلوگیری به عمل آورد. در این طرح، ابتدا به بررسی منابع گوناگون با موضوع فرااصطلاح‌نامه و شناسایی مفهوم فرااصطلاح‌نامه و بیان کاربرد آن، و سپس به ارائه مراحل تدوین آن پرداخته می‌شود.

۱-۲. پیشینه پژوهش

سابقه تلفیق و انطباق اصطلاح‌نامه‌ها به دهه ۱۹۶۰ بازمی‌گردد. از جمله در سال ۱۹۶۹ در آمریکا در زمینه واژگان کشاورزی، در سال ۱۹۸۳ در ارتباط با اصطلاح‌نامه ناسا، در سال ۱۹۸۶ در حوزه پزشکی و در سال ۱۹۹۰ در حوزه هنر و معماری، تبدیل خودکار اصطلاحات و تلفیق و ادغام اصطلاح‌نامه‌های خرد و کلان در مجموعه‌های واژگانی وسیع صورت گرفته است و کم و بیش ادامه دارد.

مشهورترین نمونه‌های چنین طرح‌هایی، اصطلاح‌نامه ادغام‌شده نظام پزشکی واحد^۱ مربوط به کتابخانه ملی پزشکی ایالات متحده آمریکا است که ویرایش سال ۲۰۱۵ آن حدود یک میلیون مفهوم

^۱ UMLS

و بیش از دو میلیون نام مفهومی (در زمینه‌های پزشکی؛ زیست‌پزشکی، بهداشت و رشته‌های وابسته) دارد. این فرااصطلاح‌نامه، حاصل بیش از ۱۰۰ مجموعه اصطلاح‌شناسی، رده‌بندی، اصطلاح‌نامه و سایر منابع ذی‌ربط است که بعضی از آن‌ها ویرایش‌های متعددی نیز دارند. هدف از طرح نظام زبان پزشکی واحد کتابخانه ملی پزشکی ایالات متحده آمریکا فراهم‌نمودن امکان دسترسی یکسان به تعدادی از پایگاه‌های اطلاعاتی توزیع‌شده و ناهمخوانی است که از مجموعه اصطلاحات مختلفی استفاده می‌کنند. این طرح شامل چهار منبع فرااصطلاح‌نامه، شبکه معنایی، نقشه منابع اطلاعاتی و واژگان است. فرااصطلاح‌نامه ترکیبی است از زبان‌های کنترل‌شده موجود که در هر مجموعه واژگانی آن، یک نام مفهومی انتخاب می‌شود و تطابق با واژه‌های مترادف و مرتبط را نشان می‌دهد. شبکه معنایی، رده‌بندی دائمی همه مفاهیم را در فرااصطلاح‌نامه مطابق رده‌ها یا انواع معنایی و روابط بین این رده‌ها را فراهم می‌سازد. نقشه منابع اطلاعاتی نشان می‌دهد کدام پایگاه اطلاعاتی به سؤال خاصی مربوط می‌شود، اطلاعات درباره منابع خاص را برای استفاده‌کنندگان فراهم می‌سازد و به‌صورت خودکار به منابع وصل می‌شود و کار بازیابی را بر روی یک یا چند منبع انجام می‌دهد. واژگان آن نیز برای پردازش زبان طبیعی ایجاد شده و شامل اصطلاحات زیست‌پزشکی و بسیاری از واژه‌های گرایش‌های دیگر را دربرمی‌گیرد. کار انطباق اصطلاح‌نامه‌ها و سایر زبان‌های کنترل‌شده برای ساختن فرااصطلاح‌نامه، به‌ویژه از این نظر که اصطلاح‌نامه‌هایی که قرار است انطباق یابند، از جنبه ساختاری متفاوتند و بسیاری از آن‌ها به‌ویژه واژه‌نامه‌ها از استاندارد بین‌المللی پیروی نکرده‌اند، با دشواری صورت می‌گیرد. اگر بین اصطلاح‌نامه‌ها و واژه‌نامه‌ها زمینه‌های مشترکی وجود داشته باشد، این کار سریع‌تر و آسان‌تر انجام می‌شود.

امروزه با گسترش روزافزون نظام‌های اطلاعاتی پیوسته همچون اینترنت که از متن آزاد، یعنی زبان طبیعی و واژه‌های مندرج در عناوین، چکیده‌ها و متن کامل اسناد به‌عنوان ابزار بازیابی استفاده می‌کنند، نقش اصطلاح‌نامه‌ها تغییر یافته و تحول در برخی از ویژگی‌های آن و ایجاد تحول در شیوه استفاده از این ویژگی‌ها، حاکی از گسترش استفاده مفید اصطلاح‌نامه در ذخیره و بازیابی اطلاعات در محیط جدید دیجیتال می‌باشد. در این نظام‌ها اصطلاح‌نامه برای نمایه‌سازی واژگان کنترل‌شده و برای تقویت نظام‌های بازیابی هوشمند تمام‌متن با فنون زبانی و اطلاعات معنانشناختی و آماری مورد استفاده قرار می‌گیرد. بسیاری از پایگاه‌های اطلاعاتی از اصطلاح‌نامه‌ها و هم‌زمان از زبان طبیعی بهره می‌برند (ایچیسن و همکاران، ۱۳۸۵).

فرااصطلاح‌نامه فهرست منظم‌شده‌ای الفبایی از واژگان تخصصی است که در یک نظم آکادمیک یا گروهی از نظام‌های مرتبط است و روابط معنایی بیشتری نسبت به اصطلاح‌نامه را میان

واژگان نشان می‌دهد. این مجموعه واژگان برای یکپارچه‌سازی تعدادی از مجموعه واژگان کنترل‌شده گسترش پیدا کرده و بازایی را به‌طور مستقل امکان‌پذیر می‌سازد. نخستین نمونه در این زمینه فرااصطلاح‌نامه یوام‌ال‌اس است که توسط کتابخانه ملی پزشکی آمریکا تهیه شده است. این فرااصطلاح‌نامه بخشی از زبان واحد پزشکی است و در یک سیستم منفرد ترمنولوژی از علوم زیست‌شناسی ادغام شده است. فرااصطلاح‌نامه توسط مفاهیم یا معانی سازماندهی شده است که اسامی جانشین، مفاهیم مشابه و شناسایی روابط معنایی بین مفاهیم را به یکدیگر ربط می‌دهد. تمام مفاهیم در فرااصطلاح‌نامه حداقل به یک نوع شبکه مفهومی تعلق دارد و همین موجب می‌شود که رده‌بندی هماهنگی از تمام مفاهیم در فرااصطلاح‌نامه ایجاد شود که در سطح عمومی شبکه مفهومی ارائه می‌شود. بیشتر واژگان ترکیبی که اسامی مفهومی یا زنجیره‌ای در فرااصطلاح‌نامه ظاهر می‌شوند، در واژه‌نامه‌های تخصصی هم وجود دارند. فرااصطلاح‌نامه نوعی اصطلاح‌نامه اصطلاح‌نامه‌هاست که سعی در یکپارچه‌سازی اصطلاح‌نامه و مجموعه لغات واژگان تخصصی دارند (یورلند، ۲۰۰۶).

در ۳۰ سال گذشته بسیاری از زبان‌های کنترل‌شده انتشار یافته‌اند که مستقل از یکدیگرند و اگر چه در یک حوزه موضوعی واحد می‌باشند، اما از لحاظ ساختار دیدگاه و تخصصی‌شدن با هم تفاوت‌هایی دارند. این ناسازگاری‌ها، انتقال رکورد بین فیله‌های پایگاه‌های اطلاعاتی و نیز جستجو در آن‌ها را با مشکل روبه‌رو ساخته است. در نتیجه انطباق و یا ادغام اصطلاح‌نامه‌ها راهی برای غلبه کردن بر ناسازگاری می‌باشد. از جمله شیوه‌های انطباق زبان‌های کنترل‌شده ناسازگار، مطابقت‌دادن، ادغام و تلفیق است که از ویژگی‌های فرااصطلاح‌نامه نیز شمرده می‌شود. مطابقت‌دادن، اصطلاحات را به‌طور مستقیم از یک زبان کنترل‌شده به زبانی دیگر یا به واژگان‌های چندگانه پیوند می‌دهد. اصطلاحات یک زبان کنترل‌شده به اصطلاحات معادل در زبان دیگر اشاره می‌کنند و یک شبکه معنایی را ایجاد می‌نمایند. استفاده از فرااصطلاح‌نامه برای رسیدن به یک مفهوم صریح و آشکار در علوم مختلف امری تردیدناپذیر است. اما این به معنای حذف منابع واژگانی موجود در آن رشته موضوعی خاص نمی‌باشد، زیرا این منابع به مفاهیم کلی‌تر پرداخته‌اند و مفاهیم جزئی‌تر را دربرنمی‌گیرند. ولی فرااصطلاح‌نامه به تجزیه و ترکیب مفاهیم، اصطلاحات و مترادف‌های آن‌ها پرداخته و ما را قادر به رسیدن به معانی و مفاهیم خاص‌تر در آن حوزه موضوعی می‌سازد (ایچسن و همکاران، ۱۳۸۵).

۳-۱. پردازش زبان طبیعی در بازایی اطلاعات

تحقیق و پژوهش در زمینه پردازش زبان طبیعی در پی آن است تا برای این پرسش که انسان‌ها چگونه معنی یک جمله یا سند را درک می‌کنند، پاسخ مناسبی ارائه نماید. به راستی ما انسان‌ها برای تشخیص فاعل، مفعول، زمان وقوع فعل و نیز تعیین حقیقی یا فرضی بودن امور و پدیده‌ها، چه

راهکارهایی را مورد استفاده قرار می‌دهیم؟ درست است که واژه‌ها (اسامی، افعال، صفات، قیود) را باید عناصر سازنده معنی جمله به‌شمار آوریم، با این همه آنچه معنی نهایی متن را می‌سازد، رابطه‌ای است که بین واژه‌ها (در محدوده ساختار جمله، سند و نیز بافت جهان پیرامون) وجود دارد (مهراد، جعفر، محمدرضا فلاحتی فومنی، ۱۳۸۴).

برای استخراج معنی یک متن یا زبان گفتاری، دست کم باید از شش مرحله یا سطح گذشت، چرا که هیچ یک از سیستم‌های پردازش زبان طبیعی از تمامی این سطوح، آن هم به‌طور هم‌زمان، استفاده نمی‌کنند.

۱. سطح آوایی یا واج‌شناسی^۱

آواشناسی به مطالعه شیوه تلفظ واژه‌ها می‌پردازد. در سیستم‌های بازیابی اطلاعات و در متون مکتوب، این سطح مورد استفاده‌ای ندارد. با این همه، وجود آن در درک زبان گفتاری و نیز «سیستم‌های تشخیص صوت»^۲ ضروری است؟

۲. سطح تکواژه‌شناختی^۳

تکواژه یکی از واحدهای زبان‌شناسی است و به کوچک‌ترین جزء معنی‌دار زبان اطلاق می‌گردد. امروزه بسیاری از موتورهای کاوش جدید قادرند ریشه واژه‌ها را به گونه‌ای مطلوب تعیین کنند. این بدان معنی است که این نوع موتورها می‌توانند به‌طور خودکار، صورت مفرد و جمع یک واژه را در اختیارمان بگذارند. البته استفاده خودکار از تکواژه‌شناسی آن هم بدون درک، می‌تواند بازیابی اسناد و نتایجی کاملاً غلط را سبب گردد.

۳. سطح نحوی^۴

وقتی از ما خواسته می‌شود که جمله‌ای را تجزیه نماییم، در واقع به دنبال شناسایی، نقش هر واژه، در جمله و نیز روابط موجود بین هر واژه با سایر واژه‌ها هستیم. موقعیت واژه در جمله می‌تواند فاعل یا مفعول بودن آن را مشخص سازد. سیستم‌های پردازش زبان طبیعی در کامل‌ترین کاربردشان، از این نوع اطلاعات ساختاری نیز استفاده می‌کنند. این گونه سیستم‌ها علاوه بر تعیین فعل بودن یک کلمه قادرند نوع آن را نیز مشخص سازند.

۴. سطح معنایی^۵

سطح معنایی، معانی لغت‌نامه‌ای واژه‌ها و نیز معانی کلمات را در بافت جمله مورد بررسی قرار

^۱ Phonetic Level

^۲ Voice recognition Systems

^۳ Morphology Level

^۴ Syntactic Level

^۵ Semantic Level

می‌دهد، معناشناسی بیان می‌دارد که بیشتر واژه‌ها دارای بیش از یک معنی‌اند. با این همه می‌توان معنی مناسب را از طریق بازیابی سایر اجزاء جمله، تعیین نمود.

در سطح معنایی، بخشی از معنا کاملاً به بافت^۱ وابسته است. این بافت است که تعیین می‌کند هر یک از اجزاء در جمله به چه معناست. در زبان انگلیسی، معانی دقیق غالباً در گروه‌های اسمی نهفته است. در واقع برای جستجو در هر نوع سیستم بازیابی اطلاعات، استفاده از یک گروه اسمی مناسب، تکنیک بسیار خوبی است. سیستم پردازش زبان طبیعی می‌تواند با استفاده از هر دو سطح معنایی و نحوی و به صورت خودکار عبارت اسمی را تجزیه و تحلیل نمود.

۵. سطح گفت‌وگو^۲

این سطح، ساختار انواع متون را مورد بررسی قرار داده، از ساختار سند برای استخراج معانی بیشتر استفاده می‌کند. یک سند فنی با چکیده‌ای آغاز می‌گردد که در یک بند، تمام محتویات سند را به شکل فشرده در خود دارد. در واقع این محتویات چکیده است که در بدنه اثر گسترش می‌یابد. پردازش زبان طبیعی، با استفاده از این ساختار قابل پیش‌بینی شده و می‌تواند دریابد که بخش خاصی از اطلاعات موجود در یک سند دارای چه نقشی است.

۶. سطح کاربردشناسی^۳

سطح کاربردشناسی با مجموعه‌ای از دانش در خصوص جهان بیرون در ارتباط است که ریشه آن‌ها را باید خارج از محتویات مدرک / سند جستجو کرد. برخی از محققان بازیابی اطلاعات بر این عقیده‌اند که راه افزایش سطح دانش بیرونی (در خصوص جهان پیرامون) گردآوری تمامی دانسته‌های لازم در مورد جهان خارج و استفاده از آن به عنوان یک راهنمای مرجع در سیستم‌های اطلاعاتی است. بین تمامی سطوح ادراکی که تاکنون ذکر کردیم، رابطه وجود دارد. نقش واژه در ساختار جمله، انتخاب معانی معقول و مربوط را میسر می‌سازد. بافت و ساختار جمله، می‌توانند در این زمینه راهگشا باشند.

از آنجایی که تمامی این سطوح زبانی از الگوهای قابل تعریف خاصی پیروی می‌کنند، امکان واردنمودن نوعی نظام ادراک زبانی در سیستم‌های رایانه‌ای، آن هم با استفاده از آن تعاریف، دور از دسترس نخواهد بود. با این همه، هرچه سطح مورد بررسی بالاتر باشد، انجام این کار نیز به تبع آن سخت‌تر خواهد شد.

^۱ Context

^۲ Discourse Level

^۳ Pragmatic Level

۴-۱. فرایند بازیابی اطلاعات

به‌طور کلی، یک سیستم بازیابی متن از سه فایل مجزا تشکیل می‌شود. فایل رکوردهای کامل فرهنگ واژه^۱ فهرستی الفبایی از واژه‌های منحصربه‌فرد موجود در پایگاه و سرانجام فهرست معکوس^۲ حاوی تمامی مکان‌های ظهور یک واژه در فرهنگ واژه‌هاست. حتی در پایگاه‌های اطلاعاتی بسیار بزرگ نیز جستجو با موفقیت همراه خواهد بود، زیرا هر واژه‌ای در فرهنگ واژه می‌تواند برای انجام هر جستجویی نقطه شروع باشد.

پردازش زبان طبیعی را می‌توان با استفاده از یک یا تمامی سطوح شش‌گانه درک، به یک یا تمامی چهار مرحله: ۱. پردازش سند، ۲. پردازش پرسش، ۳. مطابقت‌دادن پرسش، ۴. رتبه‌بندی و ردیف کردن اضافه نمود. پردازش زبان طبیعی کامل، در تمامی مراحل و سطوح به تفسیر و ذخیره‌سازی معنی پرسش و سند می‌پردازد (مهراد، جعفر، محمدرضا فلاحی فومنی، ۱۳۸۴).

بسیاری از سیستم‌هایی که مدعی برخورداری از پردازش زبان طبیعی می‌باشند، از این نوع پردازش تنها در سطوح پایین درک استفاده می‌کنند که البته این امر بیشتر اوقات صرفاً به‌منظور تغییر پوشش صورت می‌گیرد. مثلاً، بیشتر موتورهای کاوش وب می‌توانند به‌صورت خودکار صورت‌های مفرد یا جمع واژه‌های پرسش را مشخص سازند. برخی قادرند الگوهای نحوی ساده را تغییر دهند که البته این امر از طریق تجزیه جملات یا عبارات انجام می‌شود. در این نوع موتورهای کاوش، از این تکنیک برای ذخیره‌سازی مدارک استفاده نمی‌شود. همین‌که به سطح معنایی برسیم، از تعداد سیستم‌هایی که بتوان آن‌ها را به‌حق پردازش زبان طبیعی نامید، به‌شدت کاسته خواهد شد، به‌خصوص اگر از دو منظر پردازش سند و پرسش به این قضیه نگاه کنیم.

پردازش سند می‌تواند در بهبود کیفیت بازیابی اطلاعات و به‌ویژه مدیریت دانش، نقش بسیار مهمی را ایفا نماید. در همین مرحله است که اطلاعات مربوط به هر سند استخراج و ذخیره می‌گردد. هرچه محتوای یک سند بیشتر تحلیل شود، به همان میزان نیز پایگاه دانش برای استفاده‌های آتی از قابلیت بیشتری برخوردار خواهد بود و در نتیجه سیستم هوشمندتری خواهیم داشت. امروزه موتورهای کاوش می‌توانند از پایگاه دانش در امر بازیابی اطلاعات استفاده نمایند. چنین پایگاهی می‌تواند به‌طور پیوسته اطلاعات جدیدی را در خود ذخیره نماید.

هر سیستم خودکاری با مشکلات و موانعی مواجه است و سیستم‌های بازیابی اطلاعات نیز از این امر مستثنی نیستند. زبان‌های مختلف دنیا، پیچیدگی‌ها و مسائل مختلفی را مطرح می‌کنند، مسائلی که بدون مراجعه به معنی واقعی یک کلمه، امکان حل آن‌ها وجود ندارد. مثلاً، چنانچه سیستم نتواند بین

¹ dictionary

² Inverted File

دانش به دست آمده از بافت متن و دانش جهان پیرامون، ارتباط خوبی برقرار نماید، موارد مبهم فراوانی به وجود خواهد آمد. پردازش زبان طبیعی می‌تواند برخی از مشکلات رایج موجود در سیستم‌های بازیابی اطلاعات را مرتفع سازد. امروزه بیشتر فناوری‌های جستجوی وبی و تجاری، از معنا و محتوای اطلاعاتی که خود جستجو می‌کنند بی‌خبرند. آن‌ها این کار را از طریق مطابقت دادن رشته‌هایی از حروف (کلمات) در پرسش با اسناد موجود در پایگاه به انجام می‌رسانند تا از این طریق بتوانند بهترین و دقیق‌ترین اطلاعات را بازیابی کنند. حتی سیستم‌های خبره^۱ نیز به انطباق الگوها و کلیدواژه‌ها می‌پردازد و نه معانی. سیستم‌های بولی به طرز خاصی با معضل ریزش کاذب اطلاعات دست به گریبانند. سیستم‌های آماری و سنتی نیز با چنین مشکلی مواجه‌اند، با این همه وضعیت سیستم‌های بولی بسیار بدتر است، چون در این سیستم‌ها به فراوانی و همجواری عبارت و واژگان توجهی نمی‌شود.

در تمام دنیا، به اسناد مربوط به موضوعی واحد، عبارات و اصطلاحات نمایه‌ای یکسانی اختصاص داده می‌شود. اما، واقعیت این است که ثبات و یکپارچگی در کار نمایه‌سازان در بهترین حالت به ۵۰ درصد می‌رسد. با توجه به این واقعیت باید اذعان کرد که به سیستم‌هایی نیازمندیم که بتوانند ضمن توجیه تنوع موجود، تمامی اسناد نمایه‌سازی شده را نیز بازیابی نمایند.

هدف یک سیستم بازیابی اطلاعات آن است که ضمن بازیابی مربوط‌ترین اطلاعات، موارد نامربوط و ناخواسته را از جریان جستجو حذف کند. ما این نوع توانایی را در سیستم‌های بازیابی اطلاعات از طریق دو عبارت دقت^۲ و میزان فراخوانی^۳ اندازه‌گیری می‌نماییم که البته بین این دو عبارت رابطه عکس برقرار است. بدین معنی که هرچه میزان دقت بیشتر باشد، منابع و اطلاعات کمتری فراخوانده می‌شود و هرچه میزان اطلاعات فراخوانده شده بیشتر باشد، به تبع آن نتایج جستجو از دقت کمتری برخوردار خواهد بود. سیستم‌های بولی صرفاً می‌توانند یکی از دو معیار دقت یا میزان فراخوانی را رعایت نمایند.

اما سیستم‌های آماری دقیقاً در سوی دیگر این طیف قرار دارند. این سیستم‌ها ضمن تأکید بر فراخوانی اطلاعات مجموعه‌های بزرگی را پدید می‌آورند. با این همه، سعی می‌کنند از طریق رتبه‌بندی مجموعه‌های بزرگ بازیابی شده، اطلاعات دقیقی را در اختیار کاربران قرار دهند که به موضوع نیز کاملاً مربوط باشد.

سیستم‌های پردازش زبان طبیعی کامل می‌توانند از طریق پالایش بازیابی آماری، بر دقت جستجو

^۱ Expert System

^۲ Precision

^۳ Recall

بیافزایند. این سیستم‌ها همچنین قادرند از طریق گسترش بازیابی بولی، فراخوانی اطلاعات را بهبود بخشند. آنان برای نیل به این هدف از شواهد و منابعی استفاده می‌کنند که به فرایند درک در زبان طبیعی مربوط است، زیرا در زبان طبیعی امکان تغییر سند و پرسش وجود دارد (مهراد، جعفر، محمد رضا فلاحتی، ۱۳۸۴).

فصل دوم

فراداده

۲-۱. تعریف فراداده^۱

ریشه‌شناسی لغات همیشه راه خوبی برای تعریف یک مفهوم است و فراداده هم از این قاعده مستثنی نیست. پیشینه این پیشوند^۲ از یونان باستان می‌آید و معمولاً در انگلیسی «درباره» ترجمه می‌شود و اغلب برای بیان ایده‌ای استفاده می‌شود که در برخی موارد خودپژواکی است. مثلاً در زبان‌شناسی، یک فرازبان^۳ زبانی است که زبان دیگری را توصیف می‌کند. در ریاضیات، یک فرانظریه^۴، نظریه‌ای در مورد نظریه دیگر است. در ادبیات فراداستان^۵ یک اثر داستانی، توجه خواننده را به آن، به عنوان داستان ترسیم می‌کند. بنابراین فراداده، معمولاً به عنوان داده‌هایی درباره داده تعریف شده است (گاردنر، ریچارد، ۱۳۹۶).

اگرچه اصطلاح فراداده تنها نیم قرن اصالت دارد، اما به وضوح بسیار قدیمی است و حیاتی قبل از اختراع رایانه دارد. کتابداران تنها گروهی هستند که در هزاران سال گذشته، فراداده را در مقادیر زیاد تولید کرده‌اند. آن‌ها مجموعه‌های تاریخی ثبت شده خود را فهرست می‌کردند. مفهوم فراداده به راحتی با روش کار یک کتابخانه سازگار است: «داده»^۶ از مجموعه‌های قرار داده شده در درون آن تشکیل می‌شود و «فراداده» فهرست‌های استفاده شده برای توصیف آن‌هاست. از آن‌جا که ایجاد فراداده همواره یک بخش اصلی از کار یک کتابدار بوده است، یافتن بسیاری از نوآوری‌ها در تاریخ، ریشه در کتابخانه دارد.

یکی از ویژگی‌های «فراداده» این است که ساخته و پرداخته انسان است و در طبیعت یافت نمی‌شود. شکل فراداده توسط انسان برای یک هدف خاص و یا حل یک مشکل خاص طراحی شده و شکلی که می‌گیرد متناسب با ریشه‌های آن است. فراداده، عینی نیست: همیشه گزاره‌ای در مورد

¹ Metadata

² Meta

³ Metalanguage

⁴ Metatheory

⁵ Metanarrative

⁶ data

جهان را می‌سازد که این گزاره ذهنی است و هر آنچه در آن هست یا از آن حذف شده، مرزهایش کجاست و از چه اصطلاحاتی برای توصیف استفاده می‌کند، را شامل می‌شود.

۲-۱-۱. انواع فراداده

فراداده می‌تواند بر اساس طبقه‌بندی پیچیده‌ای از اهداف خدمت‌رسانی کند. اما بسیاری از این طبقه‌بندی‌ها شامل سه دسته کلی هستند. این تقسیم‌بندی سه‌جانبه می‌تواند در درک هدف اصلی فراداده مفید باشد:

۱. اولین نوع که «فراداده توصیفی» خوانده می‌شود، شاید یکی از آشنا‌ترین شکل فراداده باشد

که با هدف کمک به ما در کشف و تعیین محل داده‌های مربوطه طراحی شده است.

در یک کتابخانه، فراداده توصیفی شکل اطلاعاتی یک رکورد را به ما نشان می‌دهد که ممکن است شامل جزئیاتی مانند نویسنده، عنوان کتاب، ناشر، زمان نشر و موضوع آن باشد. همچنین ممکن است اطلاعاتی که منجر به شناسایی دقیق می‌شود، مانند: شماره استاندارد بین‌المللی کتاب^۱ را نیز پوشش دهد.

به دلیل این که هدف اصلی فراداده توصیفی قادر ساختن ما در جابجایی یک موضوع است، اغلب فراداده بازیاب نامیده می‌شود. فراداده توصیفی باید به ما کمک کند تا آنچه دقیقاً به دنبالش هستیم بیابیم. مثلاً وقتی دنبال یک کتاب خاص و یا موضوع مورد علاقه‌مان هستیم، باید بتوانیم آن را پیدا کنیم.

در فهرست کتابخانه‌ای، کلمات کلیدی می‌تواند توصیف موضوع کتاب یا سند باشد. فراداده توصیفی ممکن است بسیار متنوع‌تر از آنچه که ما می‌شناسیم باشد. اما اغلب در اندازه و دامنه‌های دیگری که برای اجرای اکثریت سیستم‌های اطلاعاتی الکترونیکی مورد نیاز است، وجود دارد. هر کتابخانه باید پیشینه‌ها را حفظ و ذخیره کند تا بتواند در زمان تجویز مورد استفاده قرار گیرد. در محیط الکترونیکی، این فراداده در حجم بزرگ‌تر و اغلب از نظر محتوایی بسیار پیچیده‌تر است.

برخی از این فراداده‌ها اطلاعات فنی مورد نیازی هستند که به سیستم اجازه می‌دهند تا داده‌های دیجیتال را در یک فرم قابل استفاده به کار گیرد و خدمت‌رسانی کند. این نوع از فراداده که قاعدتاً فراداده فنی نامیده می‌شوند، هر چیزی را که سیستم باید برای ارائه صحیح یک شیء دیجیتال بداند را پوشش می‌دهد. مثلاً یک تصویر دیجیتال شامل جزئیاتی مانند: اندازه آن به پیکسل، فرمت فایل که با آن مطابقت دارد به اندازه پالت رنگ بالقوه‌ای که هر

^۱ ISBN

پیکسل ممکن است استفاده کند، هر گونه فشرده‌سازی مورد استفاده برای کاهش اندازه آن و بیت‌ها یا قطعات بیشتر.

برای انواع مختلف اشیاء دیجیتال، قرارداد فنی مختلف مورد نیاز است. آنچه برای یک تصویر ثابت مورد نیاز است، به طرز محسوسی با آنچه برای ویدئو، صدا یا متن مورد نیاز است، تفاوت دارد. این قرارداد معمولاً به شکل یک سیستم خودکار، اجرا و به‌طور کلی برای کاربر نامرئی است.

۲. یکی دیگر از انواع مهم «فراداده اجرایی»، اطلاعات مورد نیاز برای قادر ساختن سیستم برای به اجرا درآوردن حقوق مالکیت معنوی^۱ است. این فراداده حقوقی می‌تواند شامل عناصری مانند: کسانی که حقوق مالکیت معنوی دارند، کپی‌رایت که تحت آن حقوق مالکیت اعمال می‌شود و حقوقی که برای دسترسی به اطلاعات به آن‌ها داده می‌شود، باشد. البته حقوق مالکیت معنوی در خارج از دنیای دیجیتال اعمال می‌شود اما معمولاً در یک محیط الکترونیکی توصیف مالکیت، روشی که در آن انواع دسترسی به یک شیء دیجیتالی کنترل می‌شود، پیچیده‌تر از محیط متنی است.

۳. «فراداده حفاظت» نوع نهایی فراداده اجرایی، اطلاعات لازم برای اطمینان از قابل دسترسی و قابل استفاده بودن داده‌های دیجیتال در آینده است. این فراداده حفاظت، پیچیده و گسترده است، زیرا فرایند حفظ اطلاعات دیجیتال برای مدت زمان طولانی به خودی خود بسیار پیچیده است. سخت‌افزار و نرم‌افزار امروزی عمر بسیار کوتاهی نسبت به عمر مفید داده‌های ما دارند، میراث دیجیتال ما نیاز به تعمیر و نگهداری دقیق دارد تا برای مدت طولانی دوام بیاورد. پشتیبانی از فرایندها برای اطمینان، نیاز به حجم انبوهی از فراداده دارد و هر عملیاتی که بر روی آن انجام می‌شود، چگونگی انجام و فرد انجام‌دهنده و این که همه آن‌ها می‌توانند پس از سازندگان اصلی‌شان (ماشین، نرم‌افزار و انسان) مورد استفاده قرار گیرند را تضمین می‌کند. همانند بسیاری از فراداده‌های اجرایی، این فراداده نیز برای کاربر آشکار نیست، اما اغلب ضروری است. این فراداده، اطلاعاتی است که بین قطعات کوچک داده‌ها ارتباط برقرار می‌کند تا آن‌ها را به یک شیء پیچیده‌تر دیجیتالی تبدیل کند (گاردنر، ریچارد، ۱۳۹۶).

این سه نوع فراداده: توصیفی، اجرایی و ساختاری، نشان‌دهنده یک تعریف ساده است مبنی بر این که «اطلاعات تراکشی» تنها بخش کوچکی از آنچه ما می‌توانیم انجام دهیم را پوشش می‌دهد. فراداده

¹ Intellectual Property Right (IPR)

وجود دارد به دلیل این که مغز انسان محدودیت‌هایی دارد، اما مغز انسان حجم محدودی از اطلاعات را با دقت می‌تواند ذخیره و بازیابی کند و به روش‌های قیاس منطقی شبکه‌ای از مفاهیم در حافظه انتزاعی ایجاد می‌شود. روش توصیف هوش و حافظه از راه قیاس منطقی با استخراج ویژگی‌های کلیدی داده‌های آن‌ها برای حفظ در حافظه ما و ترسیم آن‌ها در یک فرم شناختی قابل کنترل طراحی شده است و الگوهایی از داده «در مورد» داده‌ها که در مغز ترسیم می‌شود و بازیابی آن‌ها را آسان‌تر می‌کند.

اما فراداده این امکان را فراهم می‌کند که داده‌ها برای تشکیل دانش^۱ با یکدیگر پیوند برقرار کنند و این دانش در آنچه ما به عنوان فرهنگ می‌شناسیم، تثبیت شود. تعریف دانش: دانش را به عنوان یک مفهوم در شاخه‌ای از فلسفه به عنوان باور صادق موجه تعریف می‌کنند. تعریفی سه‌جانبه که تا به امروز در بسیاری از مباحث معرفت‌شناسی مطرح است. در نیمه دوم قرن بیستم، نگرش‌های دیگری برای تعریف دانش ایجاد شد که آن را در ارتباط با اطلاعات^۲ در نظر می‌گیرند.

این تصور که دانش ناشی از اطلاعات، به طور طبیعی تأثیر زیادی در حوزه علم اطلاعات داشت. یکی از جالب‌ترین آن‌ها، نظریه سازمانی راسل اکوف^۳ است. در مقاله‌ای که وی با عنوان «از داده تا خرد»^۴ نوشت، «هرم اکوف»^۵، هرمی از این مفاهیم به هم پیوسته را که هر رده بالایی دیگری قرار داشت را تعریف کرد. هرم اکوف از پایین با کوچک‌ترین جزء آن یعنی یک واحد داده شروع می‌شود.

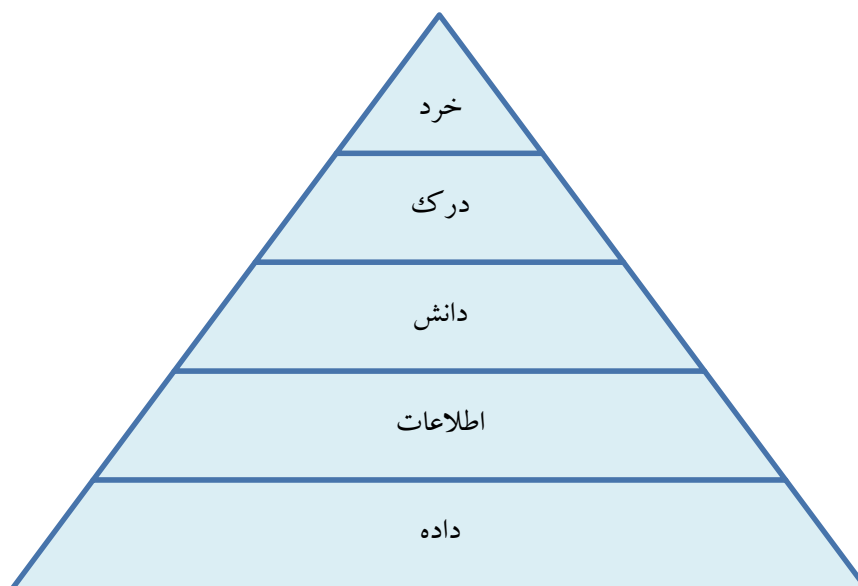
¹ Knowledge

² Information

³ Russel Akoff

⁴ From Data to Wisdom

⁵ Acroff Piromid



شکل ۱-۲. هرم اکوف

در این سطح، داده‌ها با هم ترکیب شده و روابط معنایی میان واحدها استنباط می‌شود، از داده‌ها به سمت اطلاعات حرکت می‌کنیم. اطلاعات به‌عنوان داده‌های ساختاریافته، پاسخگوی سؤالات اساسی در جهان می‌باشند. در سطح بعدی هرم، اطلاعات جمع‌آوری شد و پیوند معنادار میان اجزاء آن برقرار می‌شود و از ترکیب آن‌ها دانش به‌دست می‌آید. بدین منظور بایستی الگوهایی در نظر گرفته شوند که خود معنادار باشند. شکل این الگو و روابط میان آن‌ها پایه‌های دانش را تشکیل می‌دهند. در واقع بایستی دانش را به شکل اطلاعات در بافت دید. یک واحد از اطلاعات، یک گزاره جدا شده در مورد جهان نیست. اما تعامل اطلاعات در بافت، معناهای جدیدی را ایجاد می‌کند. تغییر و تجزیه و تحلیل این الگوها به‌گونه‌ای است که ما را از سطح دانش به سطح درک^۱ سوق می‌دهد. در حرکت از اطلاعات به دانش، نقش فراداده کاملاً آشکار است. این مرحله در مورد فراداده اطلاعات در بافت و پیوندهای میان اجزای آن برای ایجاد معانی جدید است. همه فراداده‌ها می‌توانند این کار را انجام دهند مثلاً برچسب‌زدن یک کتاب به نام نویسنده آن، پیوند به کتاب‌های دیگر همان نویسنده را ایجاد می‌کند. این پیوندهای معنایی هستند که معنای اطلاعات را به هم متصل می‌کنند و بدین طریق شبکه‌ای از مفاهیم و معانی برای ما ترسیم می‌شود. حرکت از دانش به درک، لایه انتزاعی دیگری از این شبکه فراداده است همان‌گونه که گفته شد «درک» فرایند تجزیه و تحلیل الگوها برای استخراج آنهایی است که در سطح بالاتر جهت پاسخگویی به سؤالات جدید ایجاد می‌شود.

به همین ترتیب، وقتی ما به سطح معقول‌تری از «خرد»^۲ حرکت می‌کنیم، از واحدهای جدیدی از

¹ Understanding

² Wisdom

فرا داده با هدف پاسخگویی به سؤالات درست و نادرست در سطح بالاتری از «سطح درک» استفاده می‌کنیم.

فرا داده به وضوح نقش حیاتی در جمع‌آوری داده‌ها در دانش و درک (و شاید حتی حکمت) دارد. اما سطح بالاتری از هرم اکوف نیز می‌تواند وجود داشته باشد که آن فرهنگ است. در اینجا نیز فرا داده نقشی اساسی به عهده خواهد داشت (گاردنر، ریچارد، ۱۳۹۶، ص ۴۴).

در اواسط قرن نوزدهم شاهد یک شکاف عمیق بین دنیای کتابداری و آرشیو بودیم. یکی از آن‌ها که امروزه همچنان ادامه دارد، رویکرد متفاوت فرا داده است. این شکاف مدیون ناتالیس دی. ویلی^۱، رئیس امور اداری در بخش بایگانی ملی فرانسه است. وی در سال ۱۸۴۱ پیشنهاد کرد که مرتب‌سازی آرشیو باید به روش قدیمی فهرست‌ها (مجموعه‌ها) که در کتابخانه‌ها مرسوم بوده است بازگردد. ممکن است این شکاف عمیق به نظر نرسد، اما نسبت به راهی که کتابداران با فرا داده‌ها و مجموعه‌ها در پیش گرفته بودند، فاصله را بیشتر می‌کرد. آن‌ها در موجودی‌هایشان بر روی ویژگی‌های فردی تمرکز دارند. فهرست‌های آن‌ها بیشتر یا کمتر جزئیات را توصیف می‌کند، اما به طور خاص علاقه‌مند به این که کجا به عنوان یک کل جا می‌گیرد، نیستند.

در مقابل آرشیویست‌ها از زمان دی. ویلی به قاعده کلی او، که برگرفته از یک اصطلاح فرانسوی است به نام «احترام به منشأ»^۲ را، یعنی تمرکز بر توصیف مجموعه و نظم درونی آن توجه می‌کردند. آن‌ها بیشتر تمایل داشتند هر فایل یا سند را درون محتوا قرار دهند تا آن را توصیف کنند و نشان دهند هر بخش از مجموعه چگونه بازگوکننده یک کل است. در دنیای آرشیو، ثبت فرا داده، فهرست راهنمایی است که مجموعه را به عنوان یک موجودیت مفرد توصیف می‌کند. ایده فهرست‌های راهنما، کمک به پژوهشگران برای استفاده از آرشیو در پیدا کردن منابعی است که به آن علاقه‌مندند، ولی تمایلی به توصیف آن‌ها ندارند. به طور سنتی این ابزار شامل اطلاعاتی در مورد منشأ مجموعه، یادداشت زندگینامه در مورد افراد کلیدی (جمع‌آوری‌کننده و یا فردی که اسناد را نگهداری کرده) و شرحی از آرایش آن است. طبق معمول در جهان فرا داده به این بخش آخر معمولاً به عنوان رده‌بندی سلسله‌مراتبی از لایه‌ها «منشأ» در بالا تا «مورد» در پایین است.

اگر قرن نوزدهم یکی از اصلاحات بزرگ در فرا داده را پیش روی قرار داد، از طرف دیگر یکی از روش‌های قابل اجرا برای تقسیم کردن دانش بر اساس موضوع را برای ما به ارث گذاشت. با وجود تلاش‌های شجاعانه چهره‌هایی چون توماس جیمز در ایجاد نظم در جهان فهرست موضوعات، تا اواسط قرن نوزدهم بیشتر رویکردها نسبت به این موضوع منحصر به فرد و متناقض بودند. در این دوره

¹ Natalis de Wailly

² Respect to funds

علاقه‌مندی به طبقه‌بندی علوم به‌ویژه در علم شیمی (در قالب جدول تناوبی) و زیست‌شناسی (در کاربرد و استفاده از یک طرح طبقه‌بندی برای تمام موجودات زنده طراحی شده توسط جانورشناس سوئدی (کارل لینه)^۱ افزایش یافت. از این رو، کتابداران احساس نیاز بیشتری به اصلاح طبقه‌بندی موضوع به شکل سیستماتیک کردند. شناخته‌شده‌ترین و بادوام‌ترین این تلاش‌ها برای صعود در این رقابت، اثر ملویل دیویی^۲ آمریکایی بود. دیویی، سرکتابدار کتابخانه دانشگاه کلمبیا و بعد از آن کتابخانه ایالتی نیویورک در طی دو دهه آخر قرن ۱۹، رده‌بندی معروف خود یعنی رده‌بندی دهدهی دیویی^۳ را در سال ۱۸۷۶ نشر کرد.

بسیاری از کتابخانه‌های عمومی که از آن استفاده می‌کنند، برچسب عطفی روی کتاب می‌زنند که نشان‌دهنده شماره رده دیویی برای موضوع کتاب است. در حال حاضر در ویرایش بیست‌وسوم، این رده‌بندی دنیای دانش را به ده رده گسترده تقسیم کرده است. سپس رده‌ها به جزئیات بیشتری در یک سلسله‌مراتب دقیق همه‌جانبه تقسیم‌بندی شد. هر موضوع جایگاه اصلی خود را در شاخه‌ای از درخت موضوعات پیدا کرد. این رده‌بندی اجازه می‌دهد که کتاب‌ها در قفسه طبق موضوع و شماره‌ای که فهرست‌نویس بر اساس رده‌بندی دهدهی دیویی به منبع می‌دهد، چیده شود. بخش بزرگی از تاریخ فراداده در اوایل قرن بیستم، این است که پیشرفت‌های بزرگ ۱۰۰ سال پیش باعث ایجاد و انتشار آن در یک مقیاس صنعتی شده است. اما در این دوره تغییر و نگرش عمیقی پدید آمد. روش جدید نگاه کردن به فراداده در یک سطح انتزاعی، حرکت از مدل سلسله‌مراتب که برای قرن‌ها تسلط داشت و رسیدن به مدل‌های جدیدی که برگرفته از محیط دیجیتال قرن ۲۱ بود.

با جاری شدن بیشتر اطلاعات در دنیا به‌ویژه کتابخانه‌ها، متمرکز کردن، ایجاد و ثبت فراداده برای مقابله با این هجوم بیشتر احساس شد. در ایالات متحده، کتابخانه کنگره در سال ۱۹۰۱ شروع به فروش فهرست برگه کرد، سرویسی که به سرعت توسط صدها کتابخانه برای صرفه‌جویی در وقت و هزینه به کار گرفته شد. در حال حاضر این کتابخانه‌ها به سادگی اطلاعات محلی خود را مانند شماره قفسه ضمیمه می‌کند، کارت‌ها و پرونده‌های آن‌ها مستقیماً در جعبه‌هایی با اندازه استاندارد اضافه می‌شوند. به جای صرف منابع خود در بازآفرینی فراداده از آنچه از قبل توسط کارشناسان کتابخانه‌های بزرگ کشوری ساخته شده است، استفاده می‌کنند و روند این مجموعه برای بیش از یک قرن و حتی در دوران الکترونیکی ادامه دارد.

^۱ Carl Linnaeus

^۲ Melvil Dewey

^۳ Dewey Decimal Classification (DDC)

۲-۱-۲. دیجیتالی‌شدن فراداده

البته قبل از ظهور آنچه ما امروزه به‌عنوان رایانه به رسمیت می‌شناسیم، برخی فراداده‌های نوشتاری و چاپی به فرمت دیجیتال که امروزه برای ذخیره‌سازی استفاده می‌شود، تغییر شکل دادند. در اینجا به دنبال اولین روش‌ها برای خودکارسازی فراداده هستیم.

فراداده برداشت و مفهومی از الگوهایی است که قبلاً تعیین شده است، الگویی که داده می‌تواند از آن ساخته شود و بعد از پردازش، دستگاه می‌تواند رموز را توسط سوراخ‌های روی کارت‌ها کشف و رمزگشایی کنند. بدین ترتیب فراداده از شکل انسان‌خوان به سمت درک الگویی شبیه پوششی از جوهر بر روی کاغذ یا پوست حرکت کرده است و در حال حاضر به الگوهای شبیه به کدهای دوتایی که قابل پردازش توسط ماشین و به اصطلاح ماشین‌خوان، تبدیل شده است. در نهایت در قرن بیستم الکترونیک جایگزین کارت‌های سوراخ‌شده مکانیکی شد و فراداده جایگاه خود را در ذخیره‌سازی مغناطیسی به شکل امروزی یافت. کتابداران با در پیش گرفتن چالش حرکت به سمت دیجیتالی کردن داده‌ها، یک‌بار دیگر خود را به‌عنوان پیشگامان این عرصه به اثبات رساندند.

در سال ۱۹۶۰ کتابخانه کنگره، به‌عنوان تهیه‌کننده کارت‌های فهرست‌نویسی چاپ‌شده برای هزاران کتابخانه و نگهداری از اولین فهرستگان کتاب، ایجاد فرمت جدیدی از به اشتراک‌گذاری اطلاعات به صورت الکترونیک را ضروری شناخت، نتیجه این تفکر نوآورانه، مقدمه‌ای برای فراگیرترین و بادوام‌ترین استاندارد فراداده تا به امروز به‌عنوان قالب مارک^۱ است. مارک مخفف فهرست‌نویسی ماشین‌خوان است، امکان دارد در ابتدا ساده به نظر برسد، اما این مجموعه دقیق برای انجام این کار طراحی شده است. این استاندارد برای تولید رکوردهای فهرست‌نویسی که قابل پردازش با رایانه بودند و جستجوی رایانه‌ای همچنین برای ویرایش، به‌روزرسانی، انتقال بین سیستم‌ها و به اشتراک‌گذاری در نظر گرفته شده است. این فرمت باعث شد فراداده با نرم‌افزار در تعامل و سازگار باشد تا فهرست‌ها بتوانند به شکل خودکار با یکدیگر تعامل و همکاری داشته باشند و این یک هدف جسورانه بود. این فرمت بیش از نیم قرن مورد استفاده قرار گرفته و توانایی‌های بیشتری را از خود نشان داده است (گاردنر، ریچارد، ۱۳۹۶، ص ۵۵).

قالب مارک نتیجه همکاری بین کتابداری و علم رایانه بود. نویسنده آن یکی از اولین برنامه‌نویسان رایانه هنریته آورام^۲ بود که کار خود را با آژانس امنیت ملی آمریکا، که اخیراً به فراداده‌ها علاقه‌مند شده بود، آغاز کرد. وی با کتابخانه کنگره برای حل این مسئله که چگونه فهرست‌نویسی می‌تواند به‌عنوان یک محصول ماشینی ارائه شود، همراه شد. او همراه دو تن از

^۱ Machine Readable Cataloging (MARC)

^۲ Henriette Avram

همکارانش یک فهرست کارتی بسیار گسترده بر روی دیوار نصب و هر بخش را با جزئیات تجزیه و تحلیل کرد. وی از این تحلیل، زمینه آنچه را امروزه به‌عنوان قالب مارک می‌شناسیم فراهم کرد. وی تصمیم گرفت به‌جای برچسب‌های متنی برای هر فیلد، از اعداد سه رقمی استفاده کند که معمولاً برچسب‌های مارک نامیده می‌شوند. استفاده از اعداد دلایل زیادی داشت، چون مخصوص به زبان خاصی نیست و برای پردازش رایانه‌ای آماده‌تر و در معرض خطا و اشتباه کمتری هستند. فهرست‌نویسان آموزش دیده این اعداد را به‌عنوان عناصر مرکزی می‌شناسند. ارقامی که پس از برخی از این اعداد سه رقمی ظاهر می‌شود، دستورالعملی برای چگونگی پردازش محتوای یک فیلد در زمان نمایش رکورد است. از ویژگی‌های نهایی رکورد مارک که «آورام» معرفی کرد، استفاده از زیربخش‌ها در داخل فیلدهای اصلی به‌عنوان فیلد فرعی شناخته شده است. یکی دیگر از ویژگی‌های فراموش شده مارک این است که فهرست‌های اولیه، فیلدهای جداگانه‌ای برای شناسه‌های اصلی و افزوده ایجاد می‌کنند. در یک فهرست سنتی، یک رکورد معمولاً نقطه ورود اولیه برای یک کتاب است. فرض بر این است که بیشتر کاربران تمایل دارند با نام نویسنده یک کتاب را جستجو کنند و این اغلب یک رکورد کامل برای کتاب است. امکان دارد به موارد کوتاه دیگر آن اشاره کنند، مثلاً اشکال مختلف نام نویسنده، اما قطعاً شناسه افزوده خواهند بود و اغلب با اطلاعات پراکنده‌تر نسبت به متن‌های اصلی خود هستند.

در یک فهرست رایانه‌ای، هر بخش از یک رکورد را می‌توان به‌عنوان یک نقطه دسترس استفاده کرد، بنابراین هیچ منطقی برای ارتقاء سطح یک فیلد به‌عنوان نقطه دسترسی و تعیین شناسه وجود ندارد. در هر صورت قوانین فهرست‌نویسی همچنان ادامه دارد و فهرست‌نویسان هنوز هم به‌دنبال قوانین پیچیده برای بخشی از یک رکورد هستند که با اهمیت‌تر است.

۲-۲. پیدایش اینترنت

در حالی که استاندارد مارک وارد دنیای کتابخانه شد، تغییرات فناورانه عمیقی در مسیر رشد انتقال اطلاعات در تمام جهان آغاز شده این روش به نام «انتقال بسته‌ای»^۱ توسعه داده شد که در آن، داده درون بلوک‌های کوچکی جمع‌آوری می‌شد و می‌توانست به‌طور مؤثر به شبکه‌های ارتباطی دیجیتال منتقل شود. طبیعتاً این روش جدید انتقال اطلاعات مورد توجه وزارت دفاع ایالات متحده آمریکا که در سال ۱۹۶۰ بودجه ایجاد شبکه‌ای بر اساس آن را تأمین کرد، قرار گرفت. آرپانت^۲، پایه و اساس آنچه ما به‌عنوان اینترنت می‌شناسیم، است. مفهوم شبکه‌ای از شبکه‌ها به‌طور گسترده در سال

^۱ Package transfer

^۲ Arpanet

۱۹۸۰ گسترش یافت. زمانی که دانشگاه‌ها و بعد از آن سرویس‌دهندگان تجاری دسترسی به آن را به دست آوردند. آن‌ها می‌توانستند این کارها را انجام دهند، زیرا منطبق با پروتکل‌های مشابه، مجموعه‌ای از قوانین مشابه برای بسته‌بندی، ارسال و دریافت داده بودند که در ابتدا توسط آشونت^۱ توسعه یافته و قاعدتاً به عنوان مجموعه پروتکل اینترنت شناخته شده بودند.

اینترنت رسانه‌ای شکوفا بود که به سرعت قبل از سال ۱۹۸۹ گسترش یافت، زمانی که تیم برنرزی با ایده ساده‌اش برای اشتراک داده‌ها از طریق اینترنت آمد که اکنون به عنوان وب جهان گستر یا به طور ساده، به عنوان وب می‌شناسیم.

مدیریت این حجم عظیم از داده‌ها نیازمند مقادیر عظیمی از فراداده‌های اجرایی، فنی و از طرف دیگر گزینش و نظارت است، چگونه می‌توانیم تولید و حفاظت داده‌ها را در چنین حجم بالایی داشته باشیم؟ شاید به همه آن‌ها نیاز نباشد. گوگل و دیگر موتورهای جستجو که به ما اجازه جستجوی دقیق در همه محتوای موجود در اینترنت را می‌دهند، چالشی از تصور مرسوم ایجاد می‌کنند که ما نیاز به فراداده‌ها برای ایجاد ادراک داده‌ها داریم.

در روزهای اولیه پیدایش اینترنت، به خصوص قبل از ظهور گوگل و دیگر موتورهای جستجو در نیمه دوم دهه ۱۹۹۰، این تفکر حاکم بود که برخی از کتابخانه‌های بزرگ نیاز به فهرست‌نویسی به شیوه دیگر کتابخانه‌ها دارند.

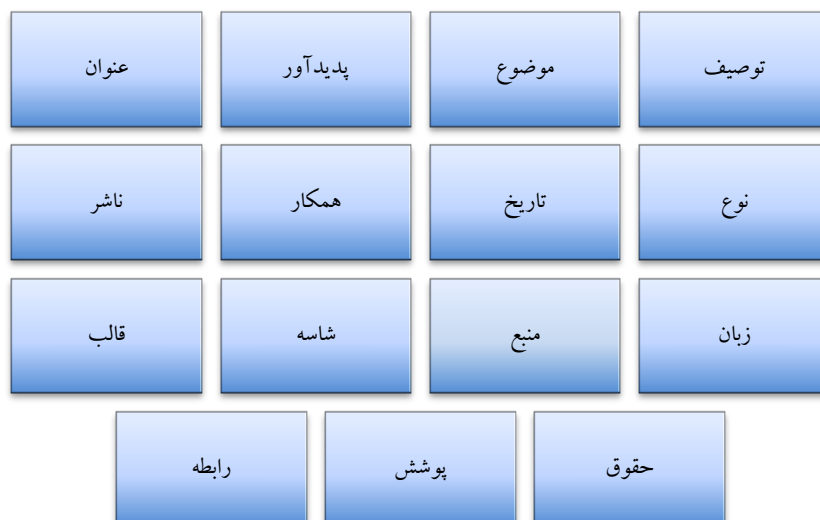
برای توصیف منابع اینترنتی، کنسرسیونی منطقی‌تر، استانداردتر و مخصوص اینترنت لازم بوده چیزی شبیه به رکورد مارک برای کتاب در قفسه کتابخانه. این استاندارد، مشابه با رکورد مارک، به منابع آنلاین اجازه می‌دهد تا یافت شوند، رکوردها به اشتراک گذاشته و پرتال‌های بزرگی مشابه به فهرستگان جهانی کتابخانه برای این منابع ساخته شوند.

تلاش برای شبیه‌سازی مارک کار را سخت کرد چون اشیاء در اینترنت بسیار متنوع‌تر از آنچه در کتابخانه یافت می‌شود هستند. بنابراین تصمیم‌گیری در مورد مجموعه‌ای از عناصری که بتوانند هر چیزی را در این آشفته بازار اداره کنند، سخت بود.

با توجه به این چالش‌ها مجموعه‌ای از کارشناسان از دنیای کتابخانه، علوم رایانه، موزه‌ها، خدمات اطلاعات آنلاین و آرشیو، اوایل سال ۱۹۹۵ در دوبلین و اوهایو یکدیگر را ملاقات کردند. به جای تلاش در تقلید از رکورد پیچیده مارک و ایجاد یک مجموعه عظیمی از عناصر برای مقابله با هر چیزی که از طریق وب پیدا شده است آن‌ها هدفی را برای ساده کردن و به حداقل رساندن اشیاء انتخاب کردند. قصد آن‌ها تشخیص مجموعه هسته از عناصری بود که می‌تواند برای هر شیء فیزیکی

¹ Ushondt

یا دیجیتالی به کار برده شود. استاندارد دی که آن‌ها به دست آوردند نام «هسته دوبلین»^۱ به خود گرفت.



شکل ۲-۲. پانزده عنصر ساده دوبلین کور

هسته دوبلین در ساده‌ترین شکل شامل تنها ۱۵ عنصر است که هر کدام به‌طور گسترده تعریف شده است و هدفش قابل فهم بودن برای همه است نه فقط برای یک متخصص فراداده.

سادگی این مجموعه عناصر، نقطه قوت هسته دوبلین است.

رکوردهای هسته دوبلین می‌تواند به‌صورت فیزیکی (یک کتاب، یک لباس، یک آباژور...) و دیجیتالی نسخه دیجیتالی از این‌ها باشد.

سادگی هسته دوبلین ثابت می‌کند که این هدف ارزشمند است و می‌تواند در تمام گوشه و کنار وب یافت شود.

ظهور اطلاعات دیجیتالی به اندازه همتای خود حائز اهمیت است. سیستم‌های بسیار کمی که در خدمت دیجیتالی کردن اشیاء هستند از کتاب‌های الکترونیک، فایل‌های صوتی و ویدئو می‌توانند بدون مجموعه فراوانی از عملیات فراداده‌ای پشت صفحه، عملیاتی شوند. با توسعه تکنولوژی رسانه‌های دیجیتال، نیاز به استانداردهای فراداده برای انجام این کار احساس شد. یکی از انواع اشیاء که در اینترنت محبوبیت پیدا کرد، تصاویر ثابت بودند. به خصوص زمانی که مرورگرهای گرافیکی اولیه موزایک^۲ و نت اسکپ^۳ در دهه ۱۹۹۰ پدیدار شدند. تکنیک‌های فشرده‌سازی جدید مانند فرمت تصویر ثابت جی‌پگ^۴ که هنوز به‌صورت گسترده استفاده می‌شوند، امکان کاهش اندازه فایل تصاویر

^۱ Dublin Core

^۲ Mosaic

^۳ Net Scape

^۴ Joint Holographic Expeto Group (JPEG)

را به طرز چشم‌گیری می‌دهد (گاردنر، ریچارد، ۱۳۹۶).

در جهان آرشیو نگرش‌های متفاوتی به فراداده تا به امروز باقی مانده است. آرشیویست‌ها به جای پذیرش مارک برای توصیف مجموعه استاندارد فراداده خود را داشتند که به‌عنوان استاندارد تبادل داده توصیف آرشیو کدگذاری^۱ شناخته شده است. این استاندارد ساختار یک چاپ سنتی را می‌گیرد و آن را به شکل ماشین‌خوان ترجمه می‌کند. همان‌طور که انتظار می‌رود یک فایل ای‌ای‌دی، بخشی از تاریخچه، اطلاعات زندگینامه‌ای افراد، شرح روایتی از مفاهیم و غیره را پوشش می‌دهد. ساختار مجموعه دقیقاً شبیه قرن نوزدهم به بعد در یک سلسله‌مراتب شرح داده می‌شود. همه این در یک فایل الکترونیکی خیلی بزرگ ضبط می‌شده است.

فایل‌های ای‌ای‌دی به یکپارچگی تمایل دارند و کمتر تحت تأثیر فهرست مشترک بزرگ دنیای مارک شکل می‌گیرند.

در عصر دیجیتال فهرست‌نویسان سعی کردند برای آنانی که تمایل داشتند به دنبال رویکرد ای‌ای‌دی و تکرار فرمت و محتوای رکوردهای تولیدشده توسط اجداد خود بروند، یک استاندارد ماشین‌خوان خلق کنند. قابل توجه‌ترین این فراداده‌ها افزونه رمزگذاری متن (تی‌ای‌آی)^۲ بود، استاندارد که به طور گسترده برای رمزگذاری متون الکترونیکی مورد استفاده قرار گرفت. این استاندارد تلاش می‌کند که بیشتر اجزای داده، شامل توصیف نشر قیاسی موجود در فهرست نسخه‌های خطی سنتی یک رکورد را بیرون بکشد. در نتیجه، هیچ فهرست خطی واحدی قابل استفاده برای جهانی از کتاب‌های چاپ‌شده نیست، فقط تعداد کمی از فهرست‌های محلی برای مجموعه‌های منحصر به فرد وجود دارد.

فراداده ساخته و پرداخته انسان است و از شیوه و قالب بنیان‌گذاران خود پیروی می‌کند. فراداده یک جهان‌بینی است که سعی دارد در توصیف هر آنچه انتخاب می‌کند به شکل ذهنی و انتخابی عمل کند. فراداده عینی نیست، حتی برخی از فیلسوفان آن را به واژه آلمانی Weltanschauung (به معنای واقعی کلمه چشم‌انداز از جهان) تعبیر کرده‌اند. بنابراین بلافاصله وارد حیطه ایدئولوژی شد. چنین اصطلاحی با معانی متفاوت به طور طبیعی انواع مختلفی از تعاریف را دارد که بسیاری از آن‌ها بدون شک و ناخودآگاه ایدئولوژی پدیدآورندگان این اصطلاح را نشان می‌دهند. یکی از موزترین و مرتبط‌ترین تعاریف که با هدف این بحث مرتبط است توسط یک موسیقی‌شناس به نام نیکلاس کوک^۳ بیان شده است: نظامی از باورهای شفاف که خود را به‌عنوان «هر آنچه هستند» معرفی می‌کند.

^۱ Encoded Archival Description (EAD)

^۲ Text Encoding Initiation (TEI)

^۳ Nicholas Cook

به عبارت دیگر: ایدئولوژی مجموعه‌ای از باورهایی است که ادعا دارند تصویری شفاف و بی‌طرفانه از جهان ارائه می‌کنند.

فرداده چیزی است که خطوطی از دانش به هم پیوسته را ترسیم می‌کند، بنابراین در تقابل با عدم قطعیت است و مانند پیشگامان بشری این حوزه، بعید است که هر شکستی در این خطوط با بررسی دقیق همه افشا شود.

به‌طور کلی فرداده تصویری از جهان را ارائه می‌دهد و بازتابی شفاف از «هر آنچه هست». در بیشتر ادوار تاریخ، هر نقشه از جهان فرداده‌ای است که می‌تواند از بیان ناقص علم تا بهترین آن یعنی یک باور ایدئولوژیک باشد. یک باور، همانند ایدئولوژی، مدعی تصویر یک پنجره روشن به واقعیت است.

اصطلاح‌شناسی فرداده راهی آشکار است که در آن می‌توان فرداده را به لحاظ ایدئولوژیک به کار برد. در اغلب موارد، این ناخودآگاه است. اصطلاحات مورد استفاده در اصطلاح‌نامه معمولاً به روش فیلترنشده، روش‌های غالب بیان اصطلاح را در زمان تدوین نشان می‌دهند. از آنجا که این حالات پایه‌های ایدئولوژیک دارد، بنابراین انعکاس ظاهراً شفافی از جهان را ارائه می‌دهند و فرداده خود، رنگ‌های ایدئولوژیک را به دست می‌آورد. جهان فرداده توصیفی، به‌ویژه استفاده از اصطلاحات برای توصیف موضوع، به ناچار گره در گره ایدئولوژیک دارد.

همان‌طور که قبلاً گفته شد، یکی از راه‌های اصلی تلاش انسان برای درک جهان، تقسیم‌بندی آن به دسته‌ها و ترتیب آن در سلسله‌مراتب است. این روش سازماندهی دانش در کتابخانه‌ها از زمان آثار اولیه کالیماخوس در فهرست بودلیان در قرن ۱۷ شروع و تا طرح طبقه‌بندی دیویی ادامه پیدا کرده است.

آنچه توسط لایه‌های یک سلسله‌مراتب بیان می‌شود، فقط آن است که سطوح پایین‌تر بخشی از لایه بالاتر خود هستند و لایه بالاتر موضوعات رده پایین‌تر را نیز دربرمی‌گیرد. آنچه دقیقاً «بخشی از» یا «اخص» معنی می‌شود، اغلب در صورت عدم تعریف مبهم است. لازم نیست مفهومی از برتر یا پایین‌تر بیان شود که سطوح فوقانی مهم‌تر و قابل توجه‌تر از سطوح پایینی است، در واقع، بسیاری از طرح‌های طبقه‌بندی بر دقت تأکید دارند. بنابراین استفاده از موضوعاتی را توصیه می‌کنند که تا حد امکان در پایین درخت موضوعات هستند.

طرح‌های رده‌بندی ممکن است اثر خفیفی از ایدئولوژی در زمان اشاره به رتبه نسبی مجاز نشان دهند. در این موارد، ممکن است سطوح در سلسله‌مراتب به‌عنوان ردیف‌هایی با برتری مفاهیم تحتانی و توان درونه‌ای شدن عمل کنند. همچنین ممکن است این تمایزها خودشان را در ترتیب مفاهیم در

یک سطح تکامل نشان دهند. اغلب این طور به نظر می‌رسد که موضوعات اول ذکر شده در هر لایه دارای وزن یا اهمیت بیشتر نسبت به موضوعات در لایه‌های پایین تر هستند. این امر می‌تواند یک سلسله‌مراتب ثانویه در داخل یک سطح یکنواخت ایجاد کند.

نمونه‌های فراوان از فراداده به عنوان یک ابزار ایدئولوژی وجود دارد، برخی آشکار و بسیاری از آن‌ها هم به همان اندازه برای پدید آورندگانشان و افرادی که می‌خواهند از آن‌ها استفاده کنند، نامرئی هستند. می‌توانیم به فراداده به عنوان یک امر بدیهی ایدئولوژیک نگاه موشکافانه‌ای بیاندازیم که منعکس کننده مجموعه‌ای از باورها و بازتاب شفاف از واقعیت هستند، یا می‌توانیم محدودتر باشیم و تنها زمانی آن را به عنوان ایدئولوژی در نظر بگیریم که حاوی بخشی از دستورالعمل‌های صریح برای ترویج این دیدگاه‌هاست، به شرط این که واقعیت داشته باشند.

نوعی از فراداده که توسط ادوارد اسنودن^۱ برجسته شد و روزنامه گاردین آن را به عنوان «وقتی از فناوری استفاده کنید، اطلاعات تولید می‌شوند» توصیف کرد، نشان داد که آژانس امنیت ملی آمریکا^۲ مقادیر زیادی از فعالیت‌های آنلاین ما را جمع‌آوری می‌کند. بخشی از دلایل منطقی برای کاهش نگرانی‌های عموم در این بود که به دلیل ماهیت فراداده، فقط یک ثبت عینی از معاملات آنلاین ما اتفاق می‌افتد. این عینیت دیگر خیلی آزاردهنده نبود (ریچارد گاردنر، ۱۳۹۶).

۲-۳. استانداردهای فراداده

«مجموعه‌ای از عناصر فراداده و قوانین که استفاده از آن‌ها برای یک هدف خاص طراحی شده است». محتویات یک استاندارد به ناچار آرایش فراداده، اجزای اصلی آن و روشی که در آن با یکدیگر تعامل دارند را نشان می‌دهد. فراداده معمولاً سه جزء اساسی دارد که در یک استاندارد تعریف خواهد شد:

- معناشناسی^۳: معانی فیلدها یا عناصری که فراداده در آن قرار داده شده است؛
- نحو^۴: روشی که در آن فراداده در یک صفحه گسترده، جدول پایگاه داده و یا یک فرمت کلی‌تر مانند زبان نشانه‌گذاری توسعه‌پذیر^۵ کدگذاری شده است.
- قوانین محتوایی: قوانین در صورت وجود، حاکم بر موارد ذیل است:

محتوای فراداده، آنچه ثبت شده، چه شکلی باید داشته باشد و چه چیزی کنار گذاشته شود.

^۱ Edward Snowden

^۲ National Security Agency

^۳ Semantics

^۴ Syntax

^۵ Extensible Markup Language (XML)

۲-۳-۱. معناشناسی در فراداده

در زبان‌شناسی معناشناسی، یعنی مطالعه معنا. به‌طور خاص مطالعه رابطه بین یک دال (نماد، کلمه، عبارت یا تصویر) و مدلول آن (آنچه به آن اشاره دارد) است. چنانچه سوسور^۱ در یک قرن پیش نشان داد، این رابطه اساساً اختیاری است. هیچ دلیل عینی وجود ندارد که چرا دنباله‌ای از حروف واژه «خالق» لزوماً باید به کسی اشاره داشته باشد که چیزی را به‌وجود آورده است. معنای مفاهیم به‌وسیله بافتی که در آن قرار دارند و ارتباط آن با سایر سلسله‌ای از حروف به‌دست می‌آید.

معناشناسی در فراداده، یک اصطلاح عمومی برای توصیف رابطه بین فیلدها و عناصر یک استاندارد و محتوایی که آن‌ها را پر می‌کند، است. استاندارد فراداده معمولاً مجموعه‌ای از این فیلدها را تعریف می‌کند. به‌عنوان مثال، هسته دوبلین نام عنصرش را «پوشش» می‌گذارد. آن را به‌عنوان «دامنه مکانی یا زمانی از منابع، کاربرد مکانی از منابع، و یا منابع مربوط به آن را معین می‌کند» و نشان می‌دهد که اختیاری است.

نظر سوسور در مورد ماهیت اختیاری نشانه، برای نام این فیلدها به‌کار برده می‌شود. در مارک حدود ۵۰۰ اصطلاح هم‌معنی پوشش هست، اما بار معانی آن را نمی‌رساند. مانند هر علامتی، معنای برجسب فیلد «پوشش» از بافت دریافت می‌شود (گاردنر، ریچارد، ۱۳۹۶، صص ۹۶-۹۸).

دانستن بافت^۲ برای دریافت معناشناسی یک فیلد بسیار حیاتی است. حتی یک چیز اساسی مانند «عنوان» می‌تواند به روش‌های ماهرانه مختلفی تعریف شود. این تعاریف به طرز محسوسی در آنچه که آن‌ها به‌عنوان «عنوان» در نظر می‌گیرند، متفاوت است. در فهرست‌نویسی انگلو آمریکن به‌طور خاص در «سرعنوان» قرار می‌گیرد.

هسته منابع دیداری، استاندارد برای فهرست‌نویسی تصاویر دیداری، از «عنوان»^۳ استفاده می‌کند. هسته رسانه عمومی، که یک استاندارد کلیدی در صفت رسانه است حتی مبهم‌تر آن را به نام «مربوط» به منبع می‌خواند و هر آنچه که به معنی «مربوط» است. به‌دلیل اهمیت منشأ نام فیلد به‌منظور نمایش معناشناختی آن جهان فراداده به‌جای استفاده از برجسب قابل فهم برای انسان مانند «عنوان» یا «پوشش»، یک راه دقیق‌تر برای شناسایی آن پیدا کرده است. این کار با استفاده از یک رشته از حروف، اعداد و علائم نقطه‌گذاری شناخته شده با نام شناسگر همگانی منبع^۴ (یو.آر.آی) انجام می‌شود.

نکته مهم در مورد یوآرآی این است که باید در هر نقطه از اینترنت منحصر به فرد باشد. چنین

^۱ Saussure

^۲ context

^۳ title

^۴ Uniform Resource Identifier (URI)

چیزی امکان‌پذیر است، زیرا خود یوآرای منحصر به فرد است.

□ استفاده از یک یوآرای به جای برچسب «ایجادکننده» به این معنی است که ما می‌دانیم در حال صحبت کردن در مورد ایجادکننده‌ای هستیم که توسط هسته دوبلین تعریف شده است و نه هر نوع دیگری. نیاز یک یوآرای به معناشناسی یک عنصر فراداده محدود نمی‌شود. همچنین می‌تواند برای محتوایی که درون عنصر قرار دارد و بیان روابط خود با دیگران استفاده شود. همان‌طور که بعداً خواهیم دید، یوآرای‌ها ستون فقرات وب معنایی را شکل می‌دهند، زیرا آن‌ها معانی را به‌طور منحصربه‌فرد و دقیق پیوست می‌کنند.

۲-۳-۲. نحو

جزء دوم فراداده که از یک استعاره زبانی قرض گرفته شده، نحو است. در زبان‌شناسی، نحو به قوانینی گفته می‌شود که بر چگونگی تشکیل اجزای زبان با هم در ساختارهای مربوطه، مانند عبارات یا جملات اشاره دارد. در فراداده، از قیاس برای توصیف راه‌هایی برای کدگذاری فراداده، به‌ویژه در شکل ماشین‌خوان استفاده می‌شود. استانداردهای فراداده معمولاً طراحی می‌شوند تا به مطلب خود اجازه دهند که قابل تعویض باشند به‌طوری که بتوان آن‌ها را بین سیستم‌ها منتقل کرد. آن‌ها از نحو برای امکان‌پذیر کردن این مبادلات استفاده می‌کنند. نیاز به این تعویض یا قابلیت همکاری با آن، برای فنی‌تر کردن اصطلاح‌ها از نظر مکانی و زمانی است.

ساختارهای نحوی فراداده، قواعد حاکم بر چگونگی ایجاد اجزای معنایی آن و روابط آن‌ها با یکدیگر، همان چیزی است که به آن‌ها اجازه می‌دهد تا کارهایی بیش از معرفی خود انجام دهند، و این چیزی است که باعث توانایی آن در برقراری ارتباط و نه صرفاً ایجاد رکورد می‌شود. یک فرمت برای پشتیبانی فراداده، درجه‌هایی از برتری را به‌دست می‌آورد. این خود توسط آنپری پس‌نیک^۱ از زبان توسعه‌نشانه‌گذاری و یا ایکس‌ام‌ال می‌آید.

ایکس‌ام‌ال به‌عنوان راهی برای مارک کردن متون الکترونیکی است که توان آن‌ها را تا حد زیادی برای ساده‌خواندن محتویاتشان برای رایانه‌ها بالا می‌برد.

آن‌ها به‌ویژه در حمایت از کاربرد متون پیچیده در پروژه‌های دانشگاهی تبحر دارند و پیکره بزرگی از زبان گفتاری می‌تواند برای تجزیه و تحلیل زبانی وارد شود.

بسیاری از استانداردها به‌عنوان طرح‌واره ایکس‌ام‌ال منتشر شده‌اند و فراداده‌ای که در آن ذکر می‌شود، باید در این قالب کدگذاری شود. ای‌ای‌دی، طرحی برای توصیف منابع آرشیوی، یک نمونه از این استانداردهاست که بر نحو، اصرار دارد.

¹ Unpre Possessing

پرمیس^۱ یک استاندارد معمول و مورد استفاده برای نگهداری مواد دیجیتالی و نمونه‌ای از این موارد است. هسته دوبلین، شاید زبان میانجی فراداده، به‌عنوان یک لیست از عناصر در ایکس‌ام‌ال منتشر شده است، اما همان‌طور که در مثال بالا دیده می‌شود، به آسانی در این قالب بیان شده است. این که چرا ایکس‌ام‌ال برای رمزگذاری فراداده بسیار پرطرفدار است، چند دلیل قانع‌کننده دارد: به خاطر این که به هر بسته نرم‌افزار، داده‌هایی گره خورده است و به همین دلیل زمانی که یک برنامه در دنیای دیجیتال به فراموشی سپرده می‌شود، منسوخ نخواهد شد و حرکت در اطراف سیستم‌ها برایش آسان است، زیرا یک متن کدگذاری شده است که فقط با ساده‌ترین فرمت در دسترس است. این احتمالاً برای اطمینان بیشتر در مورد ذخیره‌سازی داده‌ها به دلیل این سادگی است. در پشتیبانی سلسله‌مراتب که یکی از راه‌های مورد نظر برای سازماندهی فراداده در بسیاری از استانداردهاست، این می‌تواند خوب عمل کند. برای همه این دلایل، این استاندارد نزدیک‌ترین راه به فراداده برای نحو مشترک است.

۲-۳-۳. قوانین محتوا

سومین بخش از فراداده، بخش وسیعی از عناصر تعریف شده در استاندارد را مدیریت و کدگذاری نمی‌کند، اما ویژگی آن محتوایی است که آن‌ها در زمان استفاده از استاندارد جمع‌آوری می‌کنند و این به‌عنوان قوانین محتوا شناخته شده است. قوانین اغلب می‌توانند پاسخی به سؤالات خاص فرموله شده باشند. اگر یک مقاله از مجله ۵۰۰ نویسنده دارد (رایج در علوم) همه آن‌ها و یا تنها چند مورد از آن‌ها باید ذکر شود؟ یک فرد باید تحت نام واقعی و یا نام مستعار خود ذکر شود؟

ای‌ای‌سی آر^۲

حجم سنگین از قوانین است که ابتدا در سال ۱۹۶۷ (به‌عنوان ای‌ای‌سی آر ساده) ظاهر شد و با تطابق با کنوانسیون فهرست‌نویسی در جهان کتابخانه باقی مانده است. این مجموعه قوانین، هرچند بسیار گسترده هستند، اما نمی‌توانند هرگونه احتمالی را پوشش دهند. حتی با وجود وظیفه محدود، تهیه کنوانسیون برای فهرست‌بندی آثار در مجموعه کتابخانه‌ها که برای آن گردآوری شده بود.

روی اینترنت اعمال کل قوانین تقریباً غیرممکن است. به این دلیل که هیچ قدرت مرکزی برای اعمال آن‌ها وجود ندارد. این به آن معنا نیست که آن‌ها در زمینه‌هایی مانند رسانه‌ها و یا تجارت

^۱ استاندارد ابرداده‌ای حفاظت پرمیس (PREMIS (Preservation Metadata Maintenance Activity)

^۲ Anglo- American Cataloguing Rules

دیجیتال اعمال نمی‌شود. برعکس، روشن است که سایتی مانند آمازون از کنوانسیون قرارداد سازگار پیروی می‌کند تا عملیات پیچیده و با سرعت متغیر خود را انجام دهد. اما این‌ها تمایل خاصی به یک سرویس خاص دارند.

۲-۳-۴. واژگان کنترل‌شده

راه دیگر اجرای بعضی موارد در سازگاری و هماهنگی محتوای قرارداد این است که به مجموعه‌های محدود از اصطلاحات مجاز محدود شود. به جای صدور مجموعه‌ای از قوانین پیچیده برای فهرست‌نویسان، فهرستی از نام‌ها، موضوعات، اماکن و سایر چیزهایی که به آن‌ها اجازه استفاده از رکوردهای قرارداد را بدهد، ارائه کنند. اصطلاح معمول این فهرست، فهرست واژگان کنترل‌شده است.

نام‌های کنترل‌شده، یک بخش حیاتی از ساخت قرارداد قابل استفاده در مقیاس بزرگ هستند. اما به همان اندازه، کنترل موضوعات هم مهم است. آن‌ها باید به اندازه کافی سازگار شوند، به‌طوری‌که جستجوی موضوعی بتواند یک عمل منطقی دقیق باشد.

اگرچه ممکن است پرداختن به موضوعات با قراردادن آن‌ها در فهرست الفبایی طولانی، همانند، نام، امکان‌پذیر باشد، اما زمانی که آن‌ها به هر اندازه معنی‌داری رشد کنند، اجرای آن به سرعت دشوار می‌شود. به همین دلیل فهرست‌های موضوعی، معمولاً با گروه‌بندی مفاهیم مرتبط و تعیین روابط بین آن‌ها فهرست می‌شوند. این نیاز به طبقه‌بندی، رده‌بندی و ترتیب سلسله‌مراتبی دارد.

۲-۳-۵. طبقه‌بندی^۱

معمولاً دو مرحله برای ساخت اظهارات از «ربط موضوعی»^۲ وجود دارد. ۱. نخستین مرحله متشکل از کار نسبتاً ساده انتخاب یک برچسب برای توصیف آن است. ۲. مرحله بعدی برای ارجاع‌دادن این برچسب‌ها و مفاهیم به برخی از ساختارهایی که مفهوم بیشتری توسط موقعیت و روابط آن‌ها با یکدیگر ایجاد می‌کنند. ایجاد و شکل‌گیری این ساختارها معمولاً به‌عنوان طبقه‌بندی و علم طبقه‌بندی شناخته شده است.

اصطلاح طبقه‌بندی از دو واژه یونانی (نظم یا موقعیت) و (قانون یا اصل) مشتق شده است. با تأکید بر نام آن که پیروی از قوانین یا اصول است و صرفاً ایده‌ای فراتر از قراردادن مفاهیم یا اشیاء درون دسته‌بندی‌هاست که شامل قوانین فراگیر حاکم بر این دسته‌بندی‌ها و راه‌های ارتباط آن‌ها با یکدیگر می‌شود. این روش تلاش برای در دست گرفتن یک رشته علمی در علوم، به‌ویژه زیست‌شناسی است.

^۱ categorization

^۲ Topic relation

اساس مفهوم طبقه‌بندی از اینجاست که انسان یک ماده از چند مشاخره است، اما بسیاری از چهره‌های سرشناس در رشته انسان‌شناسی، تأکید دارند که چگونگی تعبیه عمیق فرهنگ انسانی، نیاز به قرارداددن جهان در دسته‌ها داشته است.

امیل دورکیم^۱ به طرح‌های طبقه‌بندی که بر اساس دین، سلسله‌مراتب شده‌اند اشاره می‌کند: «اعضاء غالب و دیگران که وابسته به گذشته هستند، گونه‌های مختلف یک کلاس هستند که به‌عنوان سطوح مشابه در ارتباط با یکدیگر وجود دارند. او روی یکی از ویژگی‌های اصلی طبقه‌بندی تأکید می‌کند. آن‌ها در لایه‌ها مرتب شده‌اند، اغلب با این تصور که سطوح بالاتر، دربرگیرنده اعضای کمتر، اما برتر از سطوح زیرین خود هستند.

در عمل سلسله‌مراتب برای طبقه‌بندی ضروری نیست و برخی از طبقه‌بندی‌ها بدون آن‌ها به‌کار گرفته می‌شوند. لازمه سطوح سلسله‌مراتبی برای دلالت بر مفاهیم، برتری یا حقارت نیست. اما یک ویژگی مشترک از راهی است که انسان، جهان را با در نظر گرفتن سلسله‌مراتب که یکی از ویژگی‌های اساسی شناخت و حتی زبان است، طبقه‌بندی می‌کند. ما قطعاً می‌توانیم در مفهوم توام چامسکی^۲، ساختار یک جمله ساده را در ساختارهای نحوی بسیار بانفوذ او مشاهده کنیم، تا ببینیم که چگونه در آرایش زبانی ما ظاهر می‌شوند.

اگر سلسله‌مراتب در تفکر انسان فراگیر است، تعجبی ندارد که آن‌ها آشکارا در طبقه‌بندی جهان قرار گیرند. به‌خصوص در علوم زیستی با اثر پیشگام گیاه‌شناس سوئدی قرن هجدهم و جانورشناس «کارل لینه»^۳ که به‌عنوان یکی از انسان‌های بزرگ تلاشگر در طبقه‌بندی است. لینه اولین پیشنهاد ترتیب سلسله‌مراتبی از طبیعت را داد که هنوز هم مورد استفاده برای طبقه‌بندی تمام موجودات زنده است.

۲-۳-۵-۱. طبقه‌بندی به‌عنوان قرارداد: جهان در سلسله‌مراتب

با توجه به حضور آن در همه‌جا، سلسله‌مراتب، چهره‌ای آشکار در تاریخ قراردادده است. قبلاً دیدیم که برخی از اولین تلاش‌ها در طبقه‌بندی به‌صورت سلسله‌مراتبی بود. پینکس کالیماکوس^۴ طبقه‌بندی کتابخانه بودلیان در سال ۱۶۲۰ و طبقه‌بندی دهدهی دیویی که همه از ساختار به‌طور فزاینده پیچیده و بسیار فلسفی تشکیل شدند، از این نوع بودند. در جهان آرشیو، آن‌ها برای کمک به یافتن ابزاری برای سازماندهی مجموعه استفاده شده‌اند و عملی است که تا به امروز در توصیف کدهای آرشیوی (ای‌ای‌دی) همچنان ادامه دارد.

¹ E. Durkheim

² N. Chomsky

³ K. Lineaus

⁴ Pinakes Callimachus

طبقه‌بندی ددهی دیویی، نخستین نمونه از یک سلسله‌مراتب طبقه‌بندی در یک مقیاس بزرگ است. هرچند ممکن است هنگام مرور قفسه‌ها آشکار نباشد. شماره‌های دیویی در واقع زنجیره‌ای از موضوعات از گسترده تخصصی است که همه با هم در یک نظم تعریف شده است که توسط دیویی و کسانی که شاهکار او را در سال‌های بعد تجدیدنظر کردند می‌باشد. طرح او به‌عنوان یک هرم مسطح، شبیه به هرم معروف در ستاره مصر است. در بالا ده بخش از دانش است که مباحث گسترده مانند: دین، علوم اجتماعی، زبان ادبیات است. هر کدام از این‌ها به‌طور دقیق‌تر و باریک‌تر به زیر موضوعات تقسیم شده‌اند. هر موضوع در هرم، شماره منحصر به فردی دارد که منعکس‌کننده سفر خود به محل اختصاص است. هر رقم در اعداد، نشان‌دهنده یک گام پایین‌تر از هرم با جزئیات بیشتر و دقت بیشتر است. به دلیل این که طرح بر اساس سیستم ددهی است، هر یک از این مراحل تنها محدود به ده تقسیم‌بندی می‌شود.

اصطلاح فنی برای این نوع طبقه‌بندی، که در آن به هر مورد یک محل واحد در یک سلسله‌مراتب داده شده، طرح شمارشی است. در اصل، اگر هر اصطلاح بالقوه که ممکن است مایل به استفاده از آن باشیم، در جایی از سلسله‌مراتب آن ذکر شده باشد، و یک مکان تغییرناپذیر نیز به آن داده شده باشد، طرح شمارشی است.

شماره دیویی مکانی دقیق برای یک موضوع توسط زنجیری شدن یک سری از موضوعات پیدا می‌کند، از گسترده به خاص، که در بالای آن در هرم قرار می‌گیرد. مفاهیمی که ترکیب می‌شوند و نظمی دارند، توسط دیویی و جانشینان ارائه شده است. ترکیبی از مفاهیم به فرد در انجام یک جست‌وجو کمک می‌کند (گاردنر، ریچارد، ۱۳۹۶، ص ۱۲۰).

بسیاری از فهرست‌نویسان کتابخانه در مجموعه‌هایشان از طرح طبقه‌بندی برای توصیف موضوعات پیچیده یک اثر استفاده می‌کنند. مانند کتابدار بزرگ هند «رانگاتان»^۱ ممکن است یک سرعنوان، موضوعی مثل: کتابداران - هند - کتابشناس^۲ داشته باشد که در آن سه مفهوم بسیار متفاوت به هم پیوسته است. در ایجاد این زنجیره، فهرست‌نویس باید آن‌چه که کاربر به دنبال آن هست را حدس بزند. آن‌ها در حال مونتاژ موضوعات پیچیده هستند و برخی از مفاهیم را برای جست‌وجوگر، قضاوت خواهند کرد. این روش معمولاً با نام فنی «نمایه‌سازی پیش‌همارا»^۳ خوانده می‌شود. یک تعریف مختصر از انجمن آرشیویست‌های آمریکایی آن را به‌عنوان روشی برای نمایه‌سازی مواد اطلاعاتی که مفاهیم جداگانه در یک مورد را در یک عنوان واحد ترکیب می‌کند، توصیف می‌کنند.

^۱ Ranganatan

^۲ Bibliography

^۳ Pre-Ordinate Indexing

یکی از قانع‌کننده‌ترین دلایل برای استفاده از رویکرد پیش‌همارایی این است که اجازه می‌دهد تا ما یک موضوع خاص را در چارچوب گسترده‌تر آن قرار دهیم. قبلاً دیدیم که می‌توانیم به دانش به‌عنوان مجموعه‌ای از پیوندهای معنایی با شروع آن‌ها بین داده‌ها و اطلاعات ایجاد شده، نگاه کنیم. پیش‌همارایی یک واحد معنایی کوچک است که درون یک چارچوب گسترده‌تر قرار می‌دهد. به‌طور بالقوه تمام سلسله‌مراتب یک طبقه‌بندی را می‌توان استدلال کرد که این راه بیانگر کدگذاری نردبان‌هایی از مفاهیم و بنابراین رسیدن به انتقال اطلاعات به دانش است.

۲-۳-۵. اصطلاح‌نامه: انعطاف‌پذیری در سلسله‌مراتب

یک طرح به‌شدت شمارشی که رویکردی دیگر برای ساخت طبقه‌بندی و انعطاف‌پذیری بیشتری نسبت به سلسله‌مراتب به ما می‌دهد، مربوط به فرهنگ‌نویس انگلیسی، پیتر مارک راجت^۱ است که ادعا کرد: تدوین یک «سیستم طبقه‌بندی کلامی» برای بیش از ۵۰ سال قبل، از کارهای او بوده است. «فهرست‌های دسته‌بندی‌شده کلمات» که هنوز هم عنوان اصطلاح‌نامه راجت را دارد. نام تزاروس از کلمه یونانی گنجینه مشتق شده است، یک دیدگاه نسبتاً مناسب، حداقل از دیدگاه یک فرهنگ‌نویس و تلاش برای تقسیم جهان دانش بر اساس کلماتی که در توصیف آن مورد استفاده هستند. این اثر برای اولین بار در سال ۱۸۵۲ منتشر شده در تلاش برای پارتیشن‌بندی (تقسیم‌بندی) دانش در یک سلسله‌مراتب بزرگ است که شکل فلسفی آن مدیون نام‌های بزرگی مانند ارسطو^۲ و لایبنیتس^۳ است. شش دسته اصلی و عام قرار دارد: «روابط انتزاعی»، «فضا»، «ماده»، «استعداد ذهنی»، «قدرت داوطلبانه و شعور» و «قدرت اخلاقی». بیش از یک هزار شاخه زیر کلاس‌ها از آن پیروی می‌کنند تا یک درخت با افزایش تراکم تا ۱۵۰۰۰ اصطلاح را در پایین این ساختار توانا سازد.

در اواخر دهه ۱۹۵۰، دنیای اطلاعات عنوان روت^۴ را هنگامی پیشنهاد داد که برخی وسایل برای هماهنگ کردن زبان مورد نیاز بود تا اسناد را برای بازیابی بیشتر نمایه^۵ کند. در قیاس مدل او با فهرستی از کلمات، اصطلاح اصطلاح‌نامه آمد تا راه جدیدی از تعریف واژگان کنترل‌شده در یک شکل خاصی از دانش را معنا کند. این رویکرد جدید، سلسله‌مراتبی بود. اما انعطاف‌پذیرتر از یک طرح شمارشی است که در اصطلاح‌نامه مدرن، اصطلاحات خاصی به‌عنوان «مرجع» مشخص شده‌اند. این‌ها از مواردی هستند که باید در هنگام طبقه‌بندی یک سند، استفاده شوند. آن‌ها از طریق پیوند به اصطلاحات گسترده‌تر در بالای سلسله‌مراتب، اصطلاحات پایین‌تر از خود و اصطلاحات مرتبط در

¹ Roget

² Aristotel

³ Leibnitz

⁴ Roots

⁵ Indexing

یک سطح قرار دارند.

اگرچه یک اصطلاح‌نامه بیان یک سلسله‌مراتب است، اما انعطاف‌پذیرتر از یک طرح شمارشی مانند دیویی است و این کاملاً برای یک اصطلاح موضوعی که بیش از یک محل در ساختار پیدا می‌کند، امکان‌پذیر است.

یکی از شکل‌های اصطلاح‌نامه استاندارد این است که «گسترده‌تر»، «خاص‌تر» و «مربوط به» به ما درباره روابط معنایی بین اصطلاحات، بلکه مفاهیم را مشخص می‌کنند.

یکی از مزیت‌های بزرگ این اصطلاح‌نامه‌ها نسبت به اصطلاح‌نامه‌های شمارشی غیرقابل انعطاف این است که در فراداده‌های دیجیتال، تغییر راه برای موضوعات پیچیده امکان‌پذیر است. هنگام جست‌وجو در یک پایگاه داده رایانه‌ای این امکان را به وجود می‌آورد که موضوعات پیچیده، توسط فهرست‌نویس تعریف شوند. این فناوری‌ها به کاربران اجازه می‌دهد، همان‌طور که جستجو می‌کنند اصطلاحات دلخواه را با هم ترکیب کنند و دقیقاً موضوعاتی که به آن‌ها علاقه دارند را تعریف کنند. این رویکرد که در آن موضوعات مرکب توسط کاربر در زمان انجام جستجو ایجاد می‌شود، اصطلاحی فنی یعنی پس‌همارایی است. یکی از راه‌هایی که در آن جستجوگران اجازه ترکیب اصطلاح/ها را دارند، ارتباط آن‌ها با کلمه «اند»^۱ برقرار می‌شود و تنها مواردی که شامل می‌شود و بازایی‌کننده پایگاه به این روش جستجوی بولی گفته می‌شود. یک راه بسیار متداول برای یافتن اطلاعات به‌ویژه در فهرست‌های کتابخانه یا پایگاه داده است جستجوی پس‌همارا در حال حاضر در همه‌جا هست، اما این بدان معنا نیست که سلسله‌مراتب از بین رفته است. آن‌ها ممکن است دیگر تنها راه درست برای یافتن موضوع نباشند، اما آن‌ها یکی از رایج‌ترین راه‌ها برای سازماندهی دیدگاه ما از جهان باقی می‌مانند. آن‌ها در حال حاضر به جای این که یک مسیر تجویزی ارائه دهند، تمایل دارند به‌عنوان یک راهنما باشند که به ناوبری دانش ما کمک می‌کنند (گاردنر، ریچارد، ۱۳۹۶، ص ۱۲۶).

به دلیل این که طبقه‌بندی مدیون آثار لینه است، جای تعجب نیست که شکوفایی آن در حوزه زیست‌شناسی باشد. در این جا مدل سلسله‌مراتبی هنوز هم غالب است. به‌ویژه در زیست‌شناسی تکاملی که شاخه‌ای از نقشه‌برداری از اصل و نسب مشترک موجودات مربوط است. ما استعاره‌ای مشابه در پشت طبقه‌بندی مدرن می‌یابیم که تکامل نقشه تکامل در سراسر گونه‌هاست.

طبقه‌بندی سلسله‌مراتبی در علوم پزشکی نیز رایج است. طرح طبقه‌بندی برای بیماری‌هایی که توسط سازمان بهداشت جهانی وارد شده و نگهداری شده است. طبقه‌بندی بین‌المللی آماری بیماری‌ها

¹ AND

و مشکلات مرتبط با سلامت (ای‌سی‌دی)^۱. این طبقه‌بندی شامل بیش از ۱۴۰۰۰ کد است و بیماری، علائم و علل را پوشش می‌دهد. این یک ابزار تشخیصی ضروری است که متکی به سلسله‌مراتب برای جستجو در توده عظیمی از اطلاعات موجود در آن است.

همچنین بسیاری از شرکت‌های تجاری در حال حاضر در سیستم‌های مدیریت دانش سرمایه‌گذاری می‌کنند. روش‌های حفظ دانش در شرکت‌های بزرگ و کوچک، تجمع می‌یابند که آن‌ها را به عنوان تلاش خود دنبال می‌کنند.

کلیت اغلب سیستم‌های بسیار پیچیده که نیاز به سازماندهی دقیق از دانش دارند، این است که حفظ شوند، به اشتراک گذاشته شده و دوباره استفاده شوند. طرح‌های طبقه‌بندی پیچیده، معمولاً توسط شرکت‌های بزرگ عمومی طبقه‌بندی، اغلب بر زمینه معماری اطلاعات تأکید می‌کنند. برخلاف ساختار بسته طرح‌های شمارشی، این طرح‌ها به سرعت متحول شده و مرتبط با حجم زیادی از اطلاعاتی هستند که با سرعت در حال رشدند. آن‌ها بسیار انعطاف‌پذیرتر هستند، اگرچه بسیاری هنوز هم سلسله‌مراتب را در مرکز خود دارند.

نظام سلسله‌مراتبی فناپذیر است و یک اثر ماندگار در تفکر انسان باقی مانده است و فراداده برای ثمرات کارهای فکری انسان قابل استفاده و قابل کنترل است.

۲-۳-۵. سلسله‌مراتب مسطح: رده‌بندی چهارم‌ای

یکی از نگرش‌های که از جهان کتابخانه می‌آید، کنار گذاشتن تمام سلسله‌مراتب نیست، بلکه مسطح کردن آن به حداقل مطلق است. در این جا یک‌بار دیگر کتابدار بزرگ هندی اس. آر رانگاتانان را به یاد می‌آوریم. او در ابتدا به عنوان یک ریاضیدان آموزش دید، فقط در حال تغییر ذهن منطقی خود برای نظریه‌های اساسی علم اطلاعات در دهه ۳۰، هنگامی که دو دهه از عمر خود را به عنوان کتابدار در دانشگاه مدرس (چنای) گذرانده بود، عدم انعطاف در طبقه‌بندی شمارشی یکی از اولین مسائلی بود که نگاه نقادانه او را فعال کرد.

او ادعا کرد سیستم‌های شمارشی ذاتاً سطحی و خارج از تاریخی هستند که در آن ظاهر می‌شوند. نسبت آن‌ها تنها پس از ظاهر شدنشان می‌تواند تعیین شود.

او یک رویکرد جدید از سازماندهی دانش را عنوان کرد که با نام رده‌بندی چهارم‌ای خوانده می‌شود. چهارم‌ای^۲ کلمه‌ای فرانسوی است فیس^۳ که در انگلیسی جنبه مشتق‌شده از واژه کوچک "Small face" است. در ابتدا برشی از یک الماس نامیده می‌شود. رانگاتانان برای نخستین بار آن را

^۱ International Statistical Classification of Diseases and Related Health Problems

^۲ Facette

^۳ Face

به عنوان حوزه‌ای از علم اطلاعات، اعتبار داد. او در سخنرانی‌اش با عنوان «مقدمه‌ای بر رده‌بندی کتابخانه»^۱ از آن به عنوان «یک اصطلاح عمومی مورد استفاده برای دلالت به هر یک از اجزاء - یک موضوع اساسی و یا جداشده - از یک اصطلاح مرکب» نام برد.

به عبارت دیگر مفهوم یک اتم که پیچیده‌تر از آن را می‌توان با ترکیب، ایجاد کرد. مهم‌تر از همه، هیچ سلسله‌مراتب ضمنی بین حقایق در زمان استفاده از آن‌ها توسط فهرست‌نویس برای تعریف یک موضوع مرکب وجود ندارد. آن‌ها از این منظر برابر و مساوی در نظر گرفته شده‌اند.

استفاده از چهریزه ممکن است سلسله‌مراتب سیستم شمارشی را جابه‌جا کند، اما آن‌ها لزوماً به طور کامل از بین نمی‌روند. چهریزه‌ای ترین طرح‌ها آن‌ها را در گروهی با نام «طبقات» قرار می‌دهند. رانگاتان خود ۳۱ مورد از آن‌ها را پیشنهاد داد. همه دسته‌های علم مانند: (تاریخ، پزشکی، ادبیات) که در بالای طرح دیویی ظاهر می‌شوند، از این نوع هستند. برخلاف ساختار فشرده درختی مانند دیویی، این سلسله‌مراتب فقط در دو سطح پایان نمی‌یابد: کلاس‌ها و چهریزه‌هایی که زیر کلاس‌ها قرار دارند. یکی از ویژگی‌های بسیار مهم این است که هر وجه باید منحصر به فرد و یا همسایگان خود تداخل نداشته باشد. وضوح، مرزهای معنایی روشن و پوشش کامل تمام جنبه‌های کلاس، از جمله اهدافی هستند که در هنگام ساخت یک طبقه‌بندی چهریزه‌ای مد نظر است.

رده‌بندی چهریزه‌ای تاکنون فقط در راه از بین بردن محدودیت‌های مدل سلسله‌مراتبی برای موضوعات حرکت کرده است. این رده‌بندی، سلسله‌مراتب را برای سطوح پایین تر مسطح کرده است و انعطاف‌پذیری بسیار بیشتری نسبت به منطق یک طرح شمارشی ایجاد می‌کند. اما معمولاً شامل برخی از لایه‌بندی‌ها هستند، چه در روابط بین چهریزه‌ها و کلاس و یا در مورد نظم استنادی. مرتب‌سازی مجدد، سازماندهی دانش در جهان علم اطلاعات از اوایل دهه ۱۹۹۰ رخ داد. این روش جدید به نام هستی‌شناس^۲ نام‌گذاری شده است.

فلسفه هستی‌شناسی به مطالعه ذات یا وجود توجه دارد. وجود خود به پیش از دوونیم هزار سال قبل به یونان باستان برمی‌گردد که در آن ذهن‌های بزرگی همچون افلاطون، ارسطو و پارمیندس همه افکار خود را وارد کردند تا سؤالاتی همچون «وجود چه است؟» «چه زمانی می‌گوییم چیزی هست؟» «آن چه چیز است؟» جواب دهند.

مدت‌ها بعد، سنت توماس آلویناس مفهوم «بودن چنانچه هست»^۳ را معرفی کرد، که فیلسوفان بعدی مانند کریستین ولف و گوتفرید لایبنیس تمرکز اصلی تحقیقات فلسفی خود را بر آن قرار

^۱ Prolegomena to Library Classification

^۲ Ontology

^۳ ENS gua ENS

دادند. امانوئل کانت، ذهن خود را در قرن هجدهم به آن مشغول ساخت که ما را مجبور به اذعان محدودیت ظرفیت‌مان در درک هستی‌شناسی به دلیل محدودیت شناخت‌مان کرد. در قرن بیستم مارتین هایدگر تفاوت هستی‌شناسانه ظرفیت بین بودن و نبودن^۱ در کتاب شاهکار خود «بودن و زمان»^۱ بیان کرد.

مفهوم هستی‌شناسی اولین بار توسط دانشمند رایانه، تام گروبر^۲ در یک مقاله در سال ۱۹۹۰ پیشنهاد شد که بعدها به عنوان خالق سیری^۳ به شهرت رسید.

گروبر هستی‌شناسی را به عنوان «مشخصات یک بافت‌بندی» و بافت‌بندی را به عنوان «یک چکیده، مشاهده ساده از جهان که ما مایل به نشان دادن آن برای برخی از اهداف هستیم» تعریف کرد. یکی از ویژگی‌های طرح هستی‌شناسی این است که هر جزء یا مفهوم را می‌توان به هر تعداد از اجرای خود متصل کرد. یعنی نیاز برای داشتن یک شکاف، فقط در کل درخت شمارشی است. به این معنی که حرکت در بین مفاهیم به صورت بسیار انعطاف‌پذیر، امکان‌پذیر خواهد شد.

۲-۴. فراداده‌ها در کتابخانه دیجیتالی

۲-۴-۱. تاریخچه

اصطلاح فراداده نخستین بار توسط جک ای. مایرز و اس. کی. کول جیان در سال ۱۹۶۹ به عنوان نام شرکتی تجاری (آمریکایی) برای ایجاد و توسعه محصولات مربوط به «فراالگوها» به کار رفت و در یک جزوه معرفی محصول در سال ۱۹۷۲ چاپ شد. چند سال بعد، یعنی در سال ۱۹۸۶، علامت تجاری متادیتا^۴ برای شرکت فراداده ثبت گردید. اصطلاح فراداده بعدها به وسیله متخصصان علوم رایانه، آمار، پایگاه‌های اطلاعاتی، و جامعه کتابداری و اطلاع‌رسانی به شکل‌های «فراداده»، «فرا داده» و «فرا- داده» استفاده شد. اما کاربرد مکرر این اصطلاح بیشتر به دهه ۱۹۹۰ و شکل‌گیری شبکه جهانی وب در سال ۱۹۹۳ مربوط می‌شود. انواع قالب‌های مارکی که در دهه‌های ۱۹۶۰، ۱۹۷۰، و ۱۹۸۰ شکل گرفته را می‌توان نخستین قالب‌های فراداده‌ای دانست. تاریخچه شکل‌گیری سایر قالب‌های فراداده‌ای به دهه ۱۹۹۰، به ویژه ۱۹۹۳-۱۹۹۶ بازمی‌گردد. از شناخته‌شده‌ترین قالب‌های فراداده‌ای در دهه ۱۹۹۰ می‌توان به طرح کدگذاری متن، خدمات مکان‌یابی اطلاعات دولتی، تبادل رایانه‌ای اطلاعات موزه‌ها، چارچوب توصیف منبع، توصیف کدگذاری شده آرشیوی، و طرح فراداده‌ای هسته دابلین اشاره کرد.

^۱ Magnum Opus Being and Time

^۲ Heidegger

^۳ Siri

^۴ MEADATA®

۲-۴-۲. تعریف

فراداده مکرراً «داده درباره داده» یا «اطلاعات درباره اطلاعات» تعریف شده است. به عبارت دیگر، فراداده، داده‌ای است که منابع اطلاعاتی را توصیف می‌کند. این تعریف گسترده، سطوح گوناگونی از توصیف (ساده تا پیچیده) را دربرمی‌گیرد: یادداشت توصیفی کوتاهی در مورد یک کتاب، توصیف غیررسمی را که موتورهای کاوش در مورد پیشینه‌های بازیابی شده ارائه می‌دهند، یک پیشینه کتابشناختی چاپی (فهرست‌برگه) یا الکترونیکی (پیشینه مارک)، چکیده‌ها و اصطلاحات نمایه‌ای، و حتی یک استناد کتابشناختی. اما به طور دقیق‌تر می‌توان فراداده را، «داده‌های ساختارمند درباره دیگر داده‌ها» دانست. بنابراین، فراداده عبارت است از مجموعه ساختارمندی از عناصر که منابع اطلاعاتی را به منظور شناسایی، کشف و استفاده، توصیف می‌نماید.

در تعریفی جامع‌تر ابرداده را اطلاعات ساختاریافته‌ای می‌دانند که توسط افراد و بنا به هدف با نقش خاص تهیه می‌شود و برای اشیای دیجیتالی می‌توان آن را روش نظام‌مندی دانست که منابع الکترونیکی روی شبکه اینترنت را شناسایی و توصیف و مکان‌یابی نماید.

□ داده درباره داده^۱

□ اطلاعات درباره اطلاعات^۲

□ داده‌های ساختاریافته درباره داده^۳

۲-۴-۳. ضرورت فراداده

اساس کاربرد فراداده، کاوش، بازیابی، دسترسی، کشف، مستندسازی، ارزیابی و انتخاب منابع الکترونیکی به‌ویژه شبکه‌ای است که باعث افزایش دقت بازیابی می‌گردد. به تعبیر دقیق‌تر می‌توان مهم‌ترین کاربردهای فراداده را این‌چنین توصیف کرد:

^۱ Data about data

^۲ Information about information

^۳ Structured data about data

۲-۴-۴. کارکردهای فراداده

جدول ۷-۱. کارکردهای فراداده

انواع فراداده	کارکردها	عناصر نمونه
فراداده توصیف / شناسایی (فراداده توصیفی)	کشف منبع / بازیابی اطلاعات	پدیدآورنده، عنوان، موضوع
فراداده مدیریتی	مدیریت منبع	بها، وضعیت
فراداده وضعیت / شرایط (نگهداری و کاربرد)	کاربرد منبع	حقوق، محدودیت‌های بازتولید
فراداده درجه‌بندی محتوا	استفاده از منبع توسط مخاطبان مناسب	مخاطب
فراداده منشأ	تأیید منبع و دیگر فعالیت‌های مرتبط ریشه‌ای	پدیدآورنده، منبع
فراداده رابطه / پیوند	پیوند منبع یا منابع مرتبط	ارتباط، منبع
فراداده ساختاری	نیازهای سخت‌افزاری و نرم‌افزاری منبع	نسبت تراکم

۲-۴-۵. مزایای فراداده

۱. فراهم‌نمودن امکان نمایه‌سازی طیف گسترده‌ای از انواع اطلاعات شبکه‌ای بدون نیاز به باند شبکه؛
۲. کمک به کشف و بازیابی مؤثر اطلاعات؛
۳. امکان‌پذیر ساختن تشخیص مطابقت منابع بازیابی شده با نیاز کاربران؛
۴. کنترل اطلاعات محدود از لحاظ دسترسی، و جلوگیری از دسترسی برخی کاربران به داده‌ها (مانند کودکان)؛
۵. ارائه اطلاعاتی که بر کاربر داده‌ها تأثیر می‌گذارد نظیر: شرایط قانونی، سن، اندازه و ...؛
۶. ارائه تاریخ داده‌ها مانند: منبع اصلی و هرگونه تغییر شکل بعدی؛
۷. ارائه اطلاعاتی در مورد مالک یا پدیدآورنده منبع (برای برقراری ارتباط) مانند آدرس پست الکترونیکی؛
۸. نشان‌دادن روابط با دیگر منابع از قبیل: پیوندهایی به نسخه‌های قبلی و بعدی؛
۹. فراهم‌نمودن اطلاعاتی برای فهرست‌های داده‌ای؛
۱۰. کمک به تصمیم‌گیری در مورد این که چه شکل و قالبی از داده‌ها باید بازیابی شوند (در صورت وجود قالب‌های متعدد)؛
۱۱. تسهیل چگونگی تفسیر داده‌ها (رمزگذاری و رمزنویسی)؛

۱۲. تطبیق، اشتراک و یکپارچه‌سازی منابع اطلاعاتی ناهمگن شبکه‌ای؛

۱۳. تهیه اطلاعاتی برای انتقال داده‌ها؛

۱۴. کنترل مدیریت داده‌ها.

۲-۴-۶. کاربردهای فراداده

۱. تسهیل جستجو و بازیابی منابع اطلاعاتی در وب؛

۲. نظم‌بخشیدن به منابع اطلاعاتی موجود در اینترنت؛

۳. توصیف پایگاه‌های اطلاعاتی، تصاویر دیجیتالی، فایل‌های صوتی و سایر منابع غیرمتنی شبکه؛

۴. تحلیل محتوا و نمایه‌سازی و سازماندهی انواع منابع اطلاعاتی شبکه؛

۵. تطبیق، اشتراک و یکپارچه‌سازی منابع اطلاعاتی ناهمگن شبکه اینترنت؛

۶. فراهم آوردن زمینه استفاده مجدد از انواع اطلاعات توزیع شده در محیط شبکه از طریق مستندسازی محتوای اطلاعات؛

۷. ایجاد امکان برای دسترسی کاربران شبکه به اطلاعات دقیق و مرتبط؛

۸. فراداده می‌تواند به تشخیص درباره اعتبار داده کمک کند. فراداده امکان بررسی مالکیت و اعتبار را برای یک شیء رقومی مثل یک مدرک یا تصویر الکترونیکی فراهم می‌کند. اهمیت این مسأله برای بسیاری از سازمان‌هایی که بیشتر دارای رکوردهای الکترونیکی هستند در حال افزایش است؛

۹. فراداده، کلیدی برای میان‌کنش‌پذیری است. فراداده به‌عنوان عامل انتقال اطلاعات و داده‌ها بین سیستم‌ها عمل می‌کند و به‌عنوان ترکیبی کلیدی در میان‌کنش‌پذیری محسوب می‌شود. تجارت الکترونیک یک نمونه از میان‌کنش‌پذیری می‌باشد که در آن چندین سیستم مختلف ممکن است به مبادله داده نیاز داشته باشند.

۲-۴-۷. عناصر و اجزای تشکیل‌دهنده فراداده

اجزاء و عناصر تشکیل‌دهنده فراداده به سه دسته کلی تقسیم می‌شوند که عبارتند از:

الف) عناصری که اساساً مربوط به محتوای منبع یا سایت می‌باشند. مثل عنوان، زبان، موضوع، توصیف، ارتباط، پوشش؛

ب) عناصری که با مالکیت منبع در ارتباط می‌باشند. نظیر مؤلف یا پدیدآورنده، همکار، حقوق؛

ج) عناصری که در اصل به ظاهر فیزیکی منبع مربوط هستند، مثل تاریخ، شکل یا قالب، مشخص‌کننده یا شناسه‌گر.

۲-۴-۷-۱. نمونه یک رکورد فراداده‌ای

یک رکورد فراداده‌ای شامل عناصر از پیش تعریف‌شده‌ای است که نمایانگر صفات و ویژگی‌های یک منبع اطلاعاتی هستند و به نوبه خود ممکن است یک یا چند ارزش داشته باشد. برخی از این عناصر در ارتباط با محتوای منبع، بعضی در ارتباط با مالکیت معنوی اثر و برخی دیگر با ظاهر فیزیکی منبع در ارتباط هستند.

۲-۴-۷-۲. دامنه بین‌المللی

مجموعه عناصر هسته دابلین در آغاز به زبان انگلیسی گسترش یافت، اما هم‌اکنون نسخه‌هایی در بسیاری از زبان‌های دیگر مانند فنلاندی، فرانسوی، ژاپنی، نروژی، تایلندی، پرتغالی، آلمانی، یونانی، اندونزیایی، و اسپانیایی نیز تهیه شده است. گروه علاقه‌مند به محلی کردن و بین‌المللی نمودن طرح فراداده‌ای هسته دابلین در حال ایجاد پایگاهی توزیعی برای نسخه‌های زبانی مختلف است.

۲-۴-۷-۳. گسترش‌پذیری

تولیدکنندگان دابلین کور در هنگام ایجاد توازن میان سادگی عناصر و امکان بازیابی دقیق منابع از طریق این عناصر به اهمیت فراهم‌نمودن سازوکاری برای توسعه مجموعه عناصر دابلین کور به‌منظور پاسخگویی به نیازهای بیشتر برای کشف منابع پی بردند.

فدراسیون کتابخانه‌های دیجیتالی انواع فراداده را به‌صورت زیر تقسیم کرده است:

الف) فراداده توصیفی: که اطلاعاتی درباره محتوای فکری (شبهه پیشینه‌های فهرست‌نویسی استاندارد)؛

ب) فراداده مدیریتی: اطلاعاتی برای کنترل کردن، نگهداری و آرشیو اشیاء؛

ج) فراداده ساختاری: توصیف ساختار درونی اشیاء.

۲-۵. قالب داده‌های رده‌بندی

این قالب برای کدگذاری عناصر داده‌ای مرتبط با شماره‌های رده‌بندی، عنوان، جدول مرتبط با آن‌ها، سلسله‌مراتب آن‌ها و سرعنوان‌های موضوعی که با آن‌ها همبسته‌اند به کار می‌رود.

قالب مارک ۲۱ همراه با تحولات حوزه فناوری اطلاعات، در پاسخگویی به نیازهای جدید، گام‌های مؤثری برداشته و همواره به موازات این دگرگونی‌ها رشد یافته است. تمهیدات مختلف این قالب فراداده‌ای بیانگر این مطلب است. در ادامه برخی از مهم‌ترین آن‌ها خلاصه‌وار بیان شده است.

۲-۵-۱. استانداردهای ساختار توصیفی منابع

- ◆ مارک^۱؛
- ◆ مارک^۲؛
- ◆ استاندارد توصیفی فراداده شیء گرا^۳؛
- ◆ استاندارد توصیفی فراداده مستند^۴؛
- ◆ استاندارد کدگذاری توصیفی آرشیو^۵؛
- ◆ هسته دابلین کور^۶؛
- ◆ هسته وی آرای^۷.

۲-۵-۲. استانداردهای کتابخانه دیجیتال

- ◆ استاندارد انتقال و کدگذاری فراداده^۸؛
- ◆ استاندارد فراداده سازمان ملی استاندارد اطلاعات برای تصاویر در ایکس.ام.ال.^۹؛
- ◆ استاندارد حفاظت فراداده^{۱۰}؛
- ◆ استاندارد فنی فراداده برای متن^{۱۱}؛
- ◆ ایزو/دی آی اس ۲۵۵۷۷^{۱۲}؛
- ◆ ایزو ۲۰۷۷۵^{۱۳}؛
- ◆ استاندارد فنی فراداده برای تشخیص نوری کرکترها^{۱۴}؛
- ◆ زد ۳۹/۵۰^{۱۵}؛
- ◆ استاندارد جستجو و بازیابی نشانی اینترنتی / خدمات وب^{۱۶}؛
- ◆ سی کیوال^{۱۷}.

^۱ MARC 21

^۲ XML MARCXML

^۳ Metadata Object Description Standard (MODS)

^۴ Metadata Authority Description Standard (MADS)

^۵ Encoded Archival Description (EAD)

^۶ Dublin core

^۷ VRA core

^۸ Metadata Encoding & Transmission Standard (METS)

^۹ NISO Metadata for Images in XML (MIX)

^{۱۰} Preservation Metadata (PREMIS)

^{۱۱} Technical Metadata for Text (TextMD)

^{۱۲} ISO/DIS 25577

^{۱۳} ISO 20775

^{۱۴} ALTO -Technical Metadata for Optical Character Recognition

^{۱۵} Z39.50

^{۱۶} Search and Retrieve URL/ Web Service (SRU/ SRW)

^{۱۷} Contextual Query Language (CQL)

۶-۲. پردازش هسته دوبلین چیست؟

استاندارد فراداده‌ای هسته دوبلین، طرحی بین‌المللی و میان‌رشته‌ای است که مجموعه عناصری ساده و کارآمد برای توصیف طیف گسترده‌ای از منابع اطلاعاتی شبکه‌ای ارائه می‌دهد. پدیدآمدن این قالب فراداده‌ای مانند دیگر قالب‌های فراداده‌ای برخاسته از اهداف و نیازهایی بوده که توسعه‌دهندگان و پدیدآورندگان آن در نظر داشته‌اند. در نهایت، استاندارد دابلین کور مورد پذیرش بسیاری از کتابخانه‌ها و مراکز فهرست‌نویسی قرار گرفته و مبنای کار سازماندهی منابع اینترنتی واقع شده است.

این طرح استاندارد دو سطح دارد:

□ ساده و ویژه (مقید به توضیح‌گرها). سطح ساده شامل پانزده عنصر، و سطح ویژه شامل هفت عنصر (مخاطب، منشأ (ریشه)، نگهدارنده حقوق) بیشتر، یعنی بیست و دو عنصر می‌باشد. هسته دوبلین، همچنین دارای گروهی از توضیح‌گرها (پالایش‌های عنصر و طرح‌های کدگذاری عنصر) است که عناصر را از لحاظ معناشناختی به‌منظور فرایند کشف منبع پالایش می‌کنند.

مباحث مربوط به معناشناختی عناصرهسته دوبلین توسط گروهی از متخصصان رشته‌های کتابداری و اطلاع‌رسانی، علم رایانه، کدگذاری متن، جامعه موزه‌ها، و دیگر حوزه‌های مرتبط پرداخته شده است.

۶-۲-۱. اهداف هسته دوبلین

به‌طور کلی اهداف طرح فراداده‌ای هسته دوبلین عبارتند از:

۶-۲-۱-۱. سادگی ایجاد و نگهداری

طرح فراداده‌ای هسته دوبلین بر آن است که عناصر خود را تا حد ممکن ساده و محدود حفظ کند تا یک غیرمتخصص نیز بتواند به سادگی و به‌طور ارزان‌قیمت پیشینه‌های ساده توصیفی - که در عین سادگی موجب بازیابی مؤثر منابع اطلاعاتی شبکه‌ای نیز می‌شوند - ایجاد کند.

۶-۲-۱-۲. قابل فهم‌بودن برای همه (از لحاظ معناشناختی)

کشف اطلاعات شبکه اینترنت به لحاظ تفاوت‌های موجود در روش‌های توصیفی و اصطلاح‌شناختی حوزه‌های دانش بشری دشوار گردیده است. دابلین کور به کاوشگران غیرمتخصص کمک می‌کند تا مجموعه عناصری عمومی و قابل فهم از لحاظ معناشناختی، برای بازیابی منابع اطلاعاتی خود بیابند.

۲-۶-۲. مجموعه عناصر فراداده‌ای هسته دابلین

۲-۶-۲-۱. عناصر سطح ساده

جدول ۲-۱. عناصر سطح ساده

نام عنصر		شرح عنصر
عنوان	Title	نامی که به منبع یا اثر داده می‌شود و اثر به وسیله آن شناخته می‌گردد
پدیدآورنده	Creator	شخص یا سازمانی که مسئولیت اصلی خلق محتوای اثر را بر عهده دارد
موضوع	Subject	موضوع محتوای منبع که معمولاً از طریق کلیدواژه‌ها، عبارات کلیدی یا نشانه‌های رده‌بندی بیان می‌شود
توصیف	Description	شرح محتوای منبع که معمولاً شامل مواردی چون چکیده، فهرست مطالب و ارجاع به نمایش گرافیکی محتوا می‌باشد
ناشر	Publisher	شخص حقیقی یا حقوقی (سازمان) که مسئولیت پشتیبانی (امور فنی چاپ) و توزیع منبع یا اثر را بر عهده دارد
همکار	Contributor	شخص حقیقی یا حقوقی که در تهیه محتوای منبع مشارکت دارد و سهم قابل ملاحظه‌ای نسبت به منبع دارد. نظیر: ویراستار
تاریخ	Date	عنصر تاریخ معمولاً در مورد زمان تدوین، ایجاد، انتشار یا توزیع منبع، مورد استفاده قرار می‌گیرد
نوع منبع	Type	عنصر نوع، معمولاً در مورد ماهیت یا گونه محتوای منبع به کار می‌رود
قالب	Format	شکل فیزیکی یا دیجیتالی منبع که معمولاً شامل نوع رسانه یا ابعاد منبع است
شناسه گر	Identifier	یک رشته حرفی [یا دیجیتالی] که منحصراً منبع را مشخص کرده و ارجاع روشن و بدون ابهام به آن دارد. مثل «نشانی سایت» ^۱
زبان	Language	زبان محتوای فکری منبع، بهترین روش برای این منظور استفاده از عنصر زبانی دو حرفی برای کد زبان است. مثل en برای انگلیسی
ارتباط	Relation	ارجاع به متن مرتبط با اثر یا منبع
منبع اصلی	Source	ارجاع به مأخذی که اثر حاضر از آن اخذ شده است
پوشش	Coverage	پوشش، معمولاً شامل موقعیت مکانی و فضایی (نام مکان یا نام جغرافیایی) و دوره زمانی (نشانه دوره یا دامنه زمانی) است
حقوق	Rights	اطلاعات مربوط به حقوق منبع یا اثر، این اطلاعات شامل حقوق مالکیت معنوی، حق چاپ و نشر و سایر حقوق مالکیت است

¹ URL

۲-۲-۶-۲. عناصر سطح ویژه

جدول ۲-۲. عناصر سطح ویژه

نام عنصر	شرح عنصر
مخاطب	Audience
منشأ (ریشه)	Provenance
مالک حقوق	Right Shoulder
شیوه آموزشی	Instructional Method
شیوه گسترش	Accrual Method
تناوب گسترش	Accrual Periodicity
خط‌مشی گسترش	Accrual Policy

۲-۲-۶-۳. مارک در زبان نشانه‌گذاری گسترش‌پذیر^۱

- مارک ایکس‌ام‌ال همان اطلاعات تنظیم‌شده در قالب مارک ۲۱ است که به زبان ایکس‌ام‌ال تبدیل شده که هر کاربری توانایی استفاده از آن را دارد؛
- مارک ایکس‌ام‌ال از استانداردهای قالب داده‌ای به‌شمار می‌رود و توسط مرورگرهای وب قابل خواندن است و این قابلیت، نیاز کاربر را به فهرست‌های عمومی پیوسته (اپک) مرتفع می‌سازد. همچنین توسط موتورهای جستجوی وب مانند گوگل قابل درک است. از دیگر مزایای ایکس‌ام‌ال این است که طرح تمامی داده‌های کدگذاری‌شده مارک را صرف‌نظر از قالب آن‌ها پشتیبانی می‌کند؛
- ایکس‌ام‌ال امکانات نیرومند و انعطاف‌پذیری برای تغییر شکل زبانی دارد؛
- ایکس‌ام‌ال خواناتر از مارک است و برای توصیف آن نیازی به کمک گرفتن از مستندات اضافی نیست؛
- و از همه مهم‌تر سادگی و انعطاف‌پذیری و گسترش‌پذیری و امکان تبدیل پیشینه‌های مارک به مارک ایکس‌ام‌ال و برعکس، از دیگر مزایای آن است.

^۱ MARCXML

۲-۶-۲-۴. طرح فراداده‌ای توصیف شیء^۱

مدز یا مودس یک استاندارد فراداده‌ای شیء گراست که برای توصیف اشیای دیجیتالی توسط شبکه کتابخانه ملی و اداره استاندارد مارک ایجاد شده است. هدف این استاندارد اساساً منطبق کردن هدف دوگانه سازگاری^۲ و دقت^۳ است که ناکارآمدی استانداردهای دیگر را برطرف کند. این استاندارد از پرکاربردترین قالب‌های فراداده‌ای در کتابخانه‌های دیجیتالی محسوب می‌شود که بر مبنای ایکس‌ام‌ال و قالب مارک طراحی شده و سعی دارد تا با ایجاد اجزای گسترده نسبت به دابلین کور، توصیف کتابشناختی برای اکثریت فیلدها به راحتی انجام گیرد تا فرایند دسترسی به منابع هم آسان گردد.

به‌طور کل، کاربرد این استاندارد انتقال داده‌های انتخاب‌شده از رکوردهای مارک ۲۱ و ایجاد رکوردهای توصیفی از منابع اصلی است و جامع‌تر و غنی‌تر از دابلین کور است اما پیچیدگی‌های مارک را ندارد و از مارک ایکس‌ام‌ال نیز قابل فهم‌تر است.

1. Language-based tags

۲-۶-۲-۵. مهم‌ترین مزایای این طرح عبارت است از:

- مجموعه عناصری غنی‌تر از هسته دابلین را ارائه می‌کند؛
- مجموعه عناصری سازگارتر از اونیکس^۴ را برای داده‌های کتابخانه‌ای ارائه می‌کند؛
- نسبت به مارک ایکس‌ام‌ال کاربرمدارتر است؛
- ایجاد فراداده‌های مبتنی بر آن از قالب کامل مارک آسان‌تر است؛
- برترین ویژگی آن، تک‌کاربردی بودن آن است و می‌تواند جهت توصیف یکی از اشیاء به‌عنوان یک کل یا هر بخش فرعی مورد استفاده قرار گیرد.

ضرورت استفاده از این استاندارد و سایر استانداردهای فراداده‌ای در راستای نیاز به استانداردهای فراداده‌ای همگام با توسعه طرح کتابخانه دیجیتالی بسیار تشدید شده و امروزه بسیاری از کتابخانه‌های دیجیتالی از مودس، جهت تسهیل دسترسی به محتوای خود استفاده می‌کنند علاوه بر این فراداده به‌وسیله مودس توسط ابزارهای جستجوی وب نیز قابل دسترسی هستند.

در مجموع مودس را می‌توان به‌عنوان یک پیشرفت مهم تلقی کرد چرا که این استاندارد مشکل فقدان استاندارد کاربردی جهت فراداده توصیفی را حل نموده و به همراه متس (استاندارد کدگذاری و انتقال فراداده) محیط استاندارد را برای حفظ ابرداده‌های کتابخانه دیجیتالی فراهم

¹ Metadata Object Description Standard (MODS)

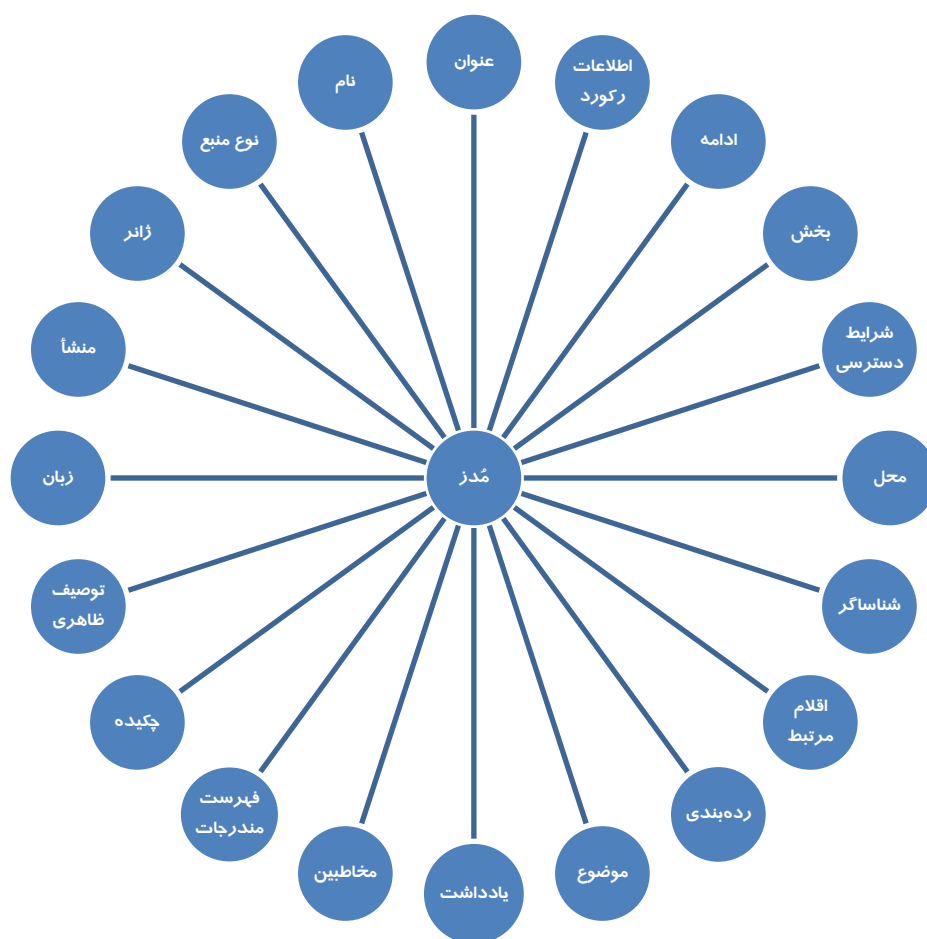
² interoperability

³ precision

⁴ ONIX

نموده است.

مدز حاوی گزیده فیلدهای مارک است و به‌جای تگ‌های عددی، تگ‌های زبانی را به‌کار می‌گیرد. عناصر اصلی مدز عبارتند از:



عنوان	title info
نام	name
نوع منبع	Type Of Resource
ژانر	genre
منشأ	Origin Info
زبان	Language
توصیف ظاهری	Physical Description
چکیده	Abstract
فهرست مندرجات	Table Of Contents
مخاطبین	Target Audience
یادداشت	Note
موضوع	Subject
رده‌بندی	Classification
اقلام مرتبط	Related Item
شناساگر	Identifier
محل	Location
شرایط دسترسی	Access Condition
بخش	Part
ادامه	Extension
اطلاعات رکورد	Record info

شکل ۲-۳. طرح فراداده‌ای توصیف مستند MADS (Metadata Authority Description Schema)

از جمله استانداردهای مرتبط با مستندسازی، استاندارد فراداده‌ای توصیفی مستند مدز است که محصول دفتر استانداردهای مارک و توسعه شبکه کتابخانه کنگره است و نخستین بار در سال ۲۰۰۴ توسط کتابخانه کنگره منتشر شد.

□ این استاندارد برای تهیه پیشینه‌های فراداده‌ای به‌منظور مستندسازی و توصیف نام افراد حقیقی و حقوقی، سرعنوان‌های موضوعی، نام‌های جغرافیایی، رویدادهای مهم و غیره به‌کار می‌رود؛

- هدف از ارائه این طرح، پیاده‌سازی قالب مستندات مارک ۲۱ در بستر زبان نشانه‌گذاری گسترش‌پذیر و استفاده از مزایای این محیط توسعه‌یافته است.

۲-۶-۳. تعریف مستندسازی

- به عمل تولید و استفاده از سیاهه واژگان پذیرفته‌شده در نظام بازیابی اطلاعات مستندسازی می‌گویند؛
- کار مستندسازی به دلیل نیاز به بررسی دقیق منابع و ثبت تصمیم‌های اتخاذشده، فعالیت وقت‌گیر و پرهزینه‌ای در امر فهرست‌نویسی است به همین دلیل در کتابخانه‌کنگره استانداردهای مربوط به پردازش انواع متفاوت پیشینه‌ها شامل داده‌های کتابشناختی و داده‌های مستند تهیه شده است؛
- یکی از استانداردهای مرتبط با مستندسازی، استاندارد فراداده‌ای توصیفی مستند (مدز) است.

۲-۶-۴. مزایای مدز

- با این استاندارد ارتباط بین عناصر کتابشناختی و پیشینه‌های مرتبط در قالب مستند به آسانی و در محیطی منعطف امکان‌پذیر می‌شود؛
- این طرح کنترل، نرمال‌سازی و مدیریت انواع داده‌های استفاده‌شده در منبع و همچنین مبادله و مدیریت توصیفات کدگذاری‌شده را پشتیبانی می‌کند؛
- امکان جاسازی پیشینه‌های ایجادشده بر مبنای این استاندارد نیز در متن منابع الکترونیکی وجود دارد؛
- ویژگی مذکور می‌تواند برخی از ارزش‌های پایگاه‌های اطلاعاتی ارجاعی مانند «کنترل مستندات» را در پی داشته باشد.

۲-۶-۵. اهداف مدز

- حمایت از نیازهای بومی و سفارشی؛
- امکان استفاده از عناصر مدز/مودز^۱ توسط سایر استانداردها؛
- تداوم نگهداری ساختار و محتوا؛
- حفظ یک روش برای رمزگذاری یک قطعه اطلاعات؛
- ساخت نام‌های عنصر ویژگی به گونه‌ای که برای عموم مخاطبان قابل فهم باشد؛
- حمایت از رمزگذاری یک توصیف مستند برای نام، عنوان، نام/عنوان، موجودیت، نوع/موجودیت جغرافیایی، موجودیت جغرافیایی سلسله‌مراتبی، حرفه.

¹ MADS/MODS

۲-۷. استاندارد انتقال و کدگذاری فراداده‌ها^۱

پروژه ام.ا.ی.^۲، در زمینه تهیه یک قالب کدگذاری برای فراداده‌های توصیفی، مدیریتی، و ساختاری برای منابع متنی و تصویری به‌منظور پاسخگویی به نیازهای سازماندهی کتابخانه‌های دیجیتال فعالیت می‌کند. در همین راستا، «استاندارد انتقال و کدگذاری فراداده‌ها» که محصول «فدراسیون کتابخانه‌های دیجیتال» است برای تحقق اهداف پروژه ام.ا.ی.^۲، با ارائه یک قالب مدرک در بستر نحوی «زبان نشانه‌گذاری گسترش‌پذیر» به‌منظور ایجاد فراداده‌های ضروری برای مدیریت اشیای محتوایی کتابخانه دیجیتال یک سازمان و تبادل این اشیای محتوایی میان چند سازمان و کاربران آن‌ها، تلاش می‌کند. بسته به نوع استفاده، این استاندارد می‌تواند در بسته‌های گوناگونی به کار گرفته شود.

پروژه ام.ا.ی.^۳

فراداده‌های ضروری برای مدیریت موفق و استفاده از شیء‌های دیجیتال، گسترده‌تر و متفاوت‌تر از ابرداده‌های مورد استفاده در مدیریت مجموعه آثار چاپی و سایر منابع فیزیکی می‌باشد.

□ در فهرست‌های کتابخانه‌ها ابرداده توصیفی هر کتاب ثبت می‌شود. اگر کتابخانه در ثبت دقیق ابرداده ساختاری سازمان انتشاردهنده کتاب ناموفق باشد، به‌هم پیوستگی صفحه‌های کتاب مذکور از بین نخواهد رفت؛

□ همچنین اگر کتابخانه در بیان نحوه چاپ کتاب و استفاده از ابزارهای ویژه دقت نکرده باشد، در بازیابی کتاب توسط محققان آثار منفی در پی نخواهد داشت، اما این مسائل درخصوص نسخه دیجیتال همان کتاب صحیح نیست. بدون فراداده ساختاری، تصویر صفحه یا فایل متنی به‌وجود آورنده اثر دیجیتال کمتر مورد استفاده قرار خواهد گرفت؛

□ برای اهداف مدیریت داخلی، کتابخانه باید به فراداده فنی مناسب جهت روزآمدسازی و انتقال متناوب داده و اطمینان از ماندگاری منابع باارزش دسترسی داشته باشد؛

□ این قالب استاندارد با فراهم کردن یک قالب کدگذاری برای فراداده توصیفی، مدیریتی و ساختاری جهت آثار متنی و تصویری به حل این مسئله کمک کند؛

□ نتیجه و حاصل این کار، مدارک با قالب ایکس‌ام‌ال می‌باشد که از این طریق کدگذاری فراداده‌های ضروری جهت مدیریت شیء‌های دیجیتال در مخزن و نیز تبادل چنین شیء‌هایی بین مخزن‌ها (یا بین مخزن‌ها و کاربران آن‌ها) فراهم می‌شود.

¹ Metadata Encoding Transmission Standard

² MOA2

³ Making America 2 project

فصل سوم

فرا اصطلاح نامه (مناظراروس)

۳-۱. فرا اصطلاح نامه (متا تزاروس)

با گسترش شبکه جهانی وب نقش اصطلاح نامه ها در حال تغییر است. این تغییر بیشتر حاکی از گسترش فایده اصطلاح نامه است تا کاهش آن. این گسترش ممکن است نه فقط نقش اصطلاح نامه ها در تقویت عملکرد سامانه های متن کامل را شامل شود، بلکه نقش اصطلاح نامه به عنوان ابزاری برای استفاده در وبگاه ها و اینترنت ها، یا برای استفاده در نمایه سازی، گسترش عبارت کاوش و سازماندهی بصری را دربرگیرد.

ظهور وب، همراه با گسترش کاربرد اصطلاح نامه در بازیابی اطلاعات بیش از ابزارهای سازماندهی و نیز رواج پایگاه های اطلاعاتی متن کامل، سبب پدید آمدن نسلی تازه از اصطلاح نامه ها شده است که نقش جدیدی - چه در زمینه نمایه سازی و چه در زمینه جستجو - ایفا می کنند. در واقع، در آینده نزدیک، اصطلاح نامه ها بیشتر به عنوان یک ابزار بازیابی به شمار می آیند تا ابزار ورود و ذخیره اطلاعات. حرکت به سوی افزایش استفاده از اصطلاح نامه به عنوان ابزاری برای بازیابی باعث شده تا ظرفیت و اندازه عملکرد اصطلاح نامه ها در نظام های بازیابی اطلاعات گسترش یابد.

۳-۱-۱. اصطلاح نامه چیست؟

اصطلاح نامه، از دیدگاه کارشناسان اطلاع رسانی، مجموعه ای از واژگان کنترل شده ای است که برای نمایه سازی اطلاعات در یک پایگاه اطلاعاتی به کار می رود و بنابراین می توان از آن برای جستجوی اطلاعات در پایگاه اطلاعاتی استفاده کرد و مفاهیم مشابه توسط کلمه های مشابه را از این طریق نمایش داد.

سالیان سال، اصطلاح نامه بیشتر به عنوان فهرست الفبایی شده از واژگان کلیدی عرضه می شده که از مدارک و اسناد استخراج می شده و برای نمایه کردن همراه با مراجع آن ها و دیگر واژگان مورد نیاز استفاده می شده است. این عملکرد سنتی در سال های اخیر تغییر پیدا کرده و بیشتر بر اساس ساختار تکنیکی - تحلیلی پایه گذاری شده است.

طبق تعریف واژه‌نامه مریام وبستر^۱، اصطلاح‌نامه عبارت است از:

الف. کتاب‌واژه‌ها یا کتاب اطلاعات در مورد حوزه‌های خاص یا مجموعه مفاهیم، به‌ویژه واژه‌ها و مترادف‌های آن‌ها.

ب. فهرستی از سرعنوان‌های موضوعی یا توصیفگرها که معمولاً به همراه یک نظام ارجاع متقابل، برای استفاده در سازماندهی مجموعه مدارک به‌منظور ارجاع و بازیابی به کار می‌روند.

با استفاده از اصطلاح‌نامه می‌توان به جستجوی مفهومی پرداخت و روابط مفهومی بین واژگان را به صورت کارآمد مورد شناسایی و استنباط قرار داد. اصطلاح‌نامه در نظام‌های بازیابی اطلاعات، از فهرست‌نویسی در مراکز اطلاع‌رسانی و کتابخانه‌ها تا اینترنت و موتورهای کاوش مورد استفاده قرار می‌گیرد. هدف اصلی اصطلاح‌نامه در درجه اول کمک به بازیابی اطلاعات و نیز در مراحل بعدی کمک به درک کلی یک حوزه موضوعی از طریق ترسیم، نظام مفاهیم، کمک به تعاریف اصطلاح‌ها و همچنین فراهم کردن امکان تولید فهرست‌های کلیدواژه‌ای و مانند آن می‌باشد.

اصطلاح‌نامه در نمایه‌های چاپی و رایانه‌ای، برای کنترل مترادف‌ها و ایجاد مبنای ارجاع‌ها کاربرد دارد. امروزه در این نظام‌ها، اصطلاح‌نامه برای نمایه‌سازی واژگان کنترل‌شده و برای تقویت نظام‌های بازیابی هوشمند تمام‌متن با فنون زبانی و اطلاعات معناشناختی و آماری مورد استفاده قرار می‌گیرد.

¹ Webster's Third New International Dictionary of the English Language. Vol. 3, S.V. "Thesaurus"



UNESCO Thesaurus

8 records found for: **Psychology**

Click on the [number] to display the records indexed with that descriptor in unesdoc/unesbib.

Term: **Attitudes** [407]

MT **4.10 Psychology**
NT Attitude change [299]
NT Intolerance [80]
NT Prejudice [64]
NT Tolerance [363]

Term: **Behaviour** [407]

MT **4.10 Psychology**
UF Human behaviour
NT Emotions [85]
NT Lateral dominance [5]
NT Psychological effects [276]
....NT2 Mental stress [68]
NT Sensorimotor activities [13]
....NT2 Perception [147]
....NT2 Visualization [28]
NT Sexual behaviour [319]
NT Unconscious [9]
....NT2 Sleep [17]

Term: **Cognition** [385]

MT **4.10 Psychology**
UF Cognitive processes

شکل ۳-۱. نمونه‌ای از نمایش اینترنتی سلسله‌مراتبی اصطلاح‌نامه یونسکو

۳-۲. فرااصطلاح‌نامه چیست؟

از بسیاری نقطه‌نظرات، تکنولوژی فرااصطلاح‌نامه برای رسمی ساختن روابط میان ساختارهای دانش - به بیان صریح روابط مفهومی در اصطلاح‌نامه‌ها - و قادر ساختن جست‌وجوی متقابل و جست‌وجوهای منظم چندگانه می‌باشد. فرااصطلاح‌نامه مجموعه‌ای از مفاهیم کلیدی از یک حوزه، روابط معنایی میان مفاهیم مانند مترادف‌ها، هاپوینیم‌ها و نیز طرح و نقشه معنایی از منابع مختلف اصطلاح‌نامه برای مفاهیم کلیدی را پشتیبانی می‌کند.

فرااصطلاح‌نامه یک پایگاه اطلاعاتی واژگان چندزبانه، دارای اهداف چندگانه و بسیار بزرگ است که شامل اطلاعاتی در مورد مفاهیم موضوعی خاص در یک زمینه، با اسامی مختلف در آن رشته موضوعی، و رابطه‌های میان آن‌ها می‌باشد. فرااصطلاح‌نامه از نسخه‌های الکترونیکی اصطلاح‌نامه‌های مختلف، رده‌بندی‌ها، مجموعه رمزها و فهرستی از واژگان کنترل‌شده که در زمینه‌های موضوعی همان رشته وجود دارد و نیز نمایه‌ها و فهرست‌های متون آن رشته تخصصی ساخته می‌شوند.

فرااصطلاح‌نامه به خاطر پیشوند «فرا» به معنای جامعیت بیشتر، بر اصطلاح‌نامه برتری یافته است. از سویی برتر بودن فرااصطلاح‌نامه به خاطر وجود اصطلاح‌نامه‌های خاص، فرهنگ لغات و رده‌بندی‌هایی

که در آن وجود دارند، می‌باشد. فرااصطلاح‌نامه به‌خاطر وجود منابع منحصر به فردش که دامنه آن‌ها را دربرگرفته است، و نیز به‌خاطر سازماندهی مجموعه نتایج از طریق معانی و افزودن اطلاعات مفید و روابط بین معنایی که در هیچ‌یک از فرهنگ لغات منابع یافت نمی‌شوند بر اصطلاح‌نامه‌ها برتری دارند.

پیشوند متا (فرا) پیشوندی لاتین است که برای نشان‌دادن مفهومی انتزاعی از مفاهیم دیگر به‌کار می‌رود. برای نمونه متافیزیک مربوط به مواردی در فیزیک می‌شود که ماورای آن و برتر از آن است. شناخت‌شناسی پیشوند فرا (متا)، نشان می‌دهد که برای پیرامون همان موضوع استفاده می‌شود. برای نمونه، فراداده، داده درباره داده است (چه کسی که آن را تهیه کرده، چه فرمتی از آن وجود دارد، چه زمانی تهیه‌شده و مانند این‌ها)

در واژه‌نامه هایپر کامپیوتینگ در مورد این پیشوند آمده که:

«معنی پیشوند فرا این است که لایه‌ای از اطلاعات انتقالی شده است». فرض کنیم که ایکس مفهومی است در نتیجه فرا (متا) ایکس داده‌ای درباره آن یا فرایندهای عملیاتی درباره ایکس است. فرااصطلاح‌نامه نوعی از اصطلاح‌نامه است که سعی در یکپارچه‌سازی اصطلاح‌نامه و لغات و واژگان موجود دارد. فرااصطلاح‌نامه نرم‌افزار سیستمی هوشمند است. در واقع می‌توان گفت که تا حدی فرااصطلاح‌نامه، اصطلاح‌نامه اصطلاح‌نامه‌ها می‌باشد. هر رکورد در فرااصطلاح‌نامه شامل واژگان اصلی مستند شده است که به‌عنوان مرجعی برای واژگان مبهم به‌کار می‌رود و همراه با ترتیب واژگان رابطه‌ای، ذخیره و طراحی شده است.

۳-۳. تفاوت‌های اصطلاح‌نامه و فرااصطلاح‌نامه (متا تراوس)

فرااصطلاح‌نامه ترکیبی است از زبان‌های کنترل‌شده موجود که از طریق برقراری پیوند، ادغام و تلفیق آن‌ها به‌دست آمده است:

الف. اصطلاح‌نامه‌ها با یکدیگر مطابقت می‌یابند و از هر مفهوم در یک اصطلاح‌نامه، نشانگرهایی به نزدیک‌ترین مفهوم معادل در اصطلاح‌نامه‌های دیگر ایجاد می‌شود. این امر امکان ترجمه پرسش کاربر به کاوشی برای یک پایگاه اطلاعاتی مشخص را که با اصطلاح‌نامه به‌خصوصی نمایه‌سازی شده است، فراهم می‌کند؛

ب. واژگان نیز ادغام می‌شوند و یک مبنای دانش غنی‌تر و جامع‌تر ایجاد می‌گردد. در هر یک از واژگان متشکله، نامی به‌عنوان نام مرجح برای مفهوم مربوطه انتخاب می‌شود و ارتباطات معنایی چندجانبه بین واژگان را فراهم می‌کند؛

ج. فرااصطلاح‌نامه مجموعه واژگان کنترل‌شده نیست، بلکه ابزاری برای افزایش سودمندی و

بهره‌وری از فرهنگ لغات و اصطلاح‌نامه‌های موجود است.

د. بنابراین فرااصطلاح‌نامه، همان اصطلاح‌نامه است، اما با این تفاوت که در فرااصطلاح‌نامه چندین اصطلاح‌نامه با هم تلفیق می‌شوند و برای جست‌وجوی هرچه بهتر و کارآمدتر، رابطه‌های مفهومی میان واژگان را ایجاد می‌نمایند.

اصطلاح‌نامه عمومی چندزبانه محیط زیست

شکل ۳-۲. نمایش کامل مدرک مربوط به "employment" در GEMET

۳-۴. فلسفه وجودی فرااصطلاح‌نامه

فرااصطلاح‌نامه‌ها به منظور ایجاد جهانی نزدیک‌تر و پیوسته‌تر از نمایش مفاهیم و معانی ضمنی و تلویحی در منابعی که از آن‌ها ساخته شده‌اند به وجود آمدند. این روش همزمان یک روش جامع و مانع است. از این رو مانع است که تعدادی از معانی به تعداد معانی بیان شده در فرهنگ لغات ترکیبی محدود می‌شود و از این رو جامع است که تمام معانی بیان شده در فرهنگ‌های لغات ترکیبی در نهایت در فرااصطلاح‌نامه به نمایش درمی‌آیند. منظور از ایجاد فرااصطلاح‌نامه ایجاد طرح و شمای رمزگذاری منفرد، یک سیستم رده‌بندی پوشاننده و یا سیستم واژگان کنترل شده نیست بلکه به این دلیل ایجاد شده که سازوکاری را برای ترجمه، تفسیر و برطرف نمودن ابهامات در مورد معانی و

اسامی خاص یک رشته - مثلاً زیست‌شناسی - گسترش دهد. هدف از برتر بودن فرااصطلاح‌نامه چگونگی بیان اسامی، ماهیت و تعیین مکان اطلاعاتی که باید برای استفاده دسترس‌پذیر شوند، می‌باشد.

۳-۵. ضرورت و فایده فرااصطلاح‌نامه‌ها

در ۳۰ سال گذشته یا حتی بیشتر، بسیاری از زبان‌های کنترل‌شده انتشار یافته‌اند که مستقل از یکدیگرند و اگرچه در یک حوزه موضوعی واحد می‌باشند، اما از لحاظ ساختار، دیدگاه و تخصصی‌شدن با هم تفاوت‌هایی دارند. این ناسازگاری، انتقال رکورد در بین فیلدهای پایگاه‌های اطلاعاتی و نیز جست‌وجو در آن‌ها را با مشکل روبه‌رو ساخته است. در نتیجه انطباق و یا تلفیق اصطلاح‌نامه‌ها، راهی برای غلبه کردن بر ناسازگاری می‌باشد.

از جمله شیوه‌های انطباق زبان‌های کنترل‌شده ناسازگار مطابقت‌دادن، ادغام و تلفیق است که از ویژگی‌های فرااصطلاح‌نامه نیز شمرده می‌شود. مطابقت‌دادن، اصطلاحات را به‌طور مستقیم از یک زبان کنترل‌شده، به زبانی دیگر یا به واژگان‌های چندگانه پیوند می‌دهد. اصطلاحات یک زبان کنترل‌شده، به اصطلاحات معادل در زبان دیگر اشاره می‌کنند.

مشهورترین نمونه از فرااصطلاح‌نامه یا اصطلاح‌نامه‌های ادغام‌شده، اصطلاح‌نامه نظام زبان پزشکی واحد، که یک اصطلاح‌نامه مفهومی است می‌باشد. «یو.ام.ال.اس»^۱ توسط کتابخانه ملی پزشکی آمریکا از سال ۱۹۸۶ با همکاری نهادهایی از داخل و خارج ایالات متحده اجرا شد. هدف آن یکپارچه‌سازی تمام نام‌ها همراه با معانی تقریباً یکسان در یک مفهوم می‌باشد. نام‌هایی که از جزء اصلی یا منابعی مشتق می‌شوند که شامل تعداد بسیاری از واژگان کنترل‌شده و طبقه‌بندی شده، تمام متن و سیستم‌های خبره است.

«یو.ام.ال.اس» شامل چهار منبع دانش می‌شود:

۱. فرااصطلاح‌نامه؛
۲. شبکه معناساختی؛
۳. نقشه منابع اطلاعاتی؛
۴. و فرهنگ اصطلاحات تخصصی.

۳-۶. زبان‌های فرااصطلاح‌نامه

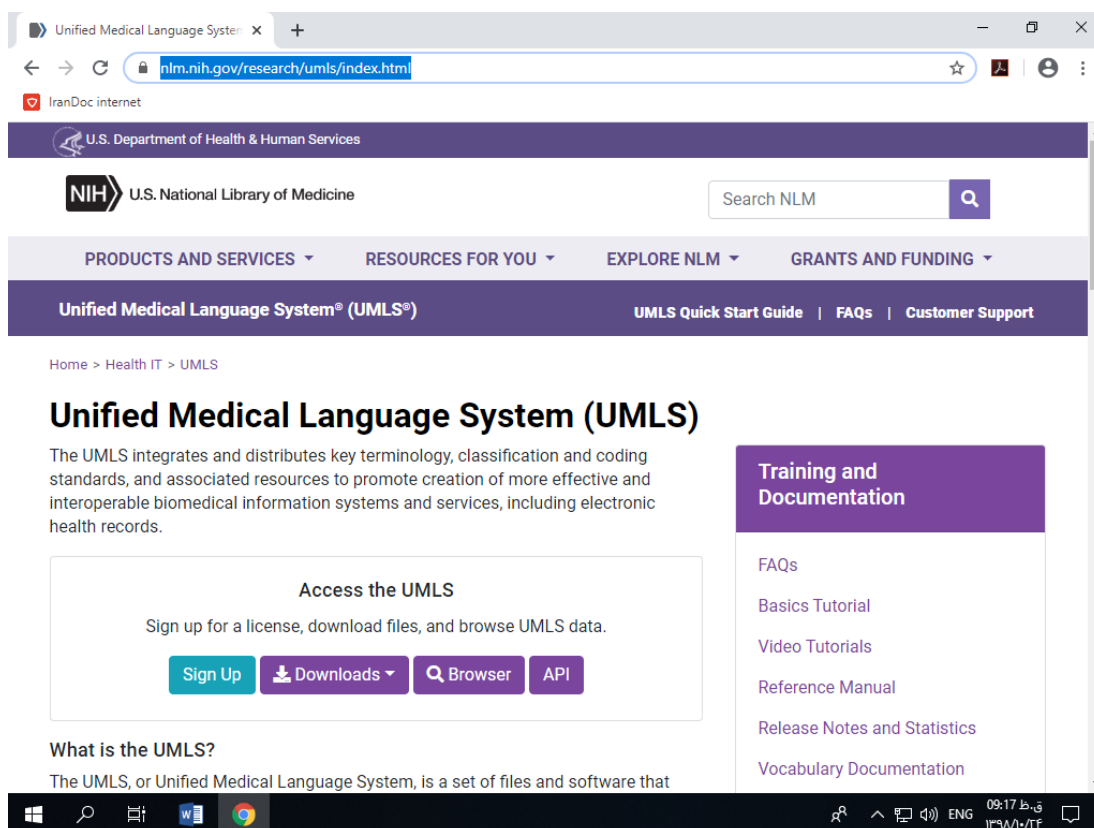
فرااصطلاح‌نامه ترکیبی از زبان‌های کنترل‌شده موجود است که از طریق برقراری پیوند، ادغام و

¹ Unified Medical Language System (UMLS)

تلفیق آن‌ها به دست آمده است. با توجه به این که در هر یک از واژگان متشکله، نامی به عنوان نام مرجح برای مفهوم انتخاب می‌شود؛ این امر امکان ایجاد ارتباطات متقابل معنایی بین واژگان‌ها را فراهم می‌کند. در فرااصطلاح‌نامه «یو.ام.ال.اس» چنانچه مفهومی در مش وجود داشته باشد، شکل مرجح آن مفهوم، صورت سرعنوان اصلی در مش است؛ در غیر این صورت اصطلاحات از دیگر واژگان‌ها به ترتیبی مشخص گزینش می‌شوند. تمام اصطلاحات، عبارات و زنجیره‌های دیگری که برای بیان مفاهیم مورد استفاده قرار می‌گیرند، با نام مندرج در فرااصطلاح‌نامه مطابقت داده می‌شوند.

۳-۷. شبکه معناشناختی

این شبکه دسته‌بندی دائمی از تمام مفاهیم موجود در فرااصطلاح‌نامه بر طبق دسته‌ها و یا انواع معناشناختی و رابطه‌های بین این دسته‌ها ارائه می‌کند. شیوه تنظیم انواع معناشناختی، سلسله‌مراتبی است. انواع معناشناختی به تشخیص معانی گوناگون مرتبط با یک اصطلاح، کمک می‌کنند. رابطه‌ها غیرسلسله‌مراتبی هستند و به پنج گروه تقسیم می‌شوند: رابطه‌های فیزیکی، رابطه‌های فضایی، رابطه‌های زمانی، رابطه‌های کارکردی و رابطه‌های مفهومی.



شکل ۳-۳. UMLS

۳-۸. نقشه منابع اطلاعاتی

نقشه منابع اطلاعاتی نشان می‌دهد که کدام پایگاه اطلاعاتی به سؤالی خاص مربوط می‌شود، اطلاعاتی درباره منابع خاص را برای کاربران فراهم می‌کند، به‌طور خودکار به منابع متصل می‌شود و عمل بازیابی را بر روی یک یا چند منبع اطلاعاتی انجام می‌دهد.

۳-۹. فرهنگ اصطلاحات تخصصی

این فرهنگ برای فراهم آوردن اطلاعات واژگانی مورد نیاز در نظام پردازش زبان طبیعی اصطلاحات تخصصی ایجاد شده است. هدف از این منبع آن است که، یک فرهنگ زبان انگلیسی با بسیاری از اصطلاحات زیست‌پزشکی باشد. با بررسی نظام «یو.ام.ال.اس» می‌توان دریافت که در فرااصطلاح‌نامه، چگونه ادغام زبان‌های کنترل‌شده، مزایای یک محصول نهایی را به‌دست می‌دهد، محصولی که به لحاظ اصطلاحات و رابطه‌ها از هر یک از واژگان‌های متشکله به تنهایی، غنی‌تر است؛ بازیابی اطلاعات را با استفاده از اصطلاحات متعارف زبان کنترل‌شده تسهیل می‌کند؛ و همچنین به‌عنوان ابزاری برای تأمین دستیابی به متون زبان طبیعی در پایگاه‌های اطلاعاتی مختلف عمل می‌کند که تمام این موارد می‌تواند از ویژگی‌های یک فرااصطلاح‌نامه باشد.

۳-۱۰. ویژگی‌های فرااصطلاح‌نامه

۳-۱۰-۱. دامنه فرااصطلاح‌نامه‌ها

دامنه فرااصطلاح‌نامه توسط دامنه ترکیبی از فرهنگ لغات منبع تعیین می‌شوند. بسیاری از روابط (متراذفات اولیه)، صفات مفهومی و چند اسم مفهومی و ... که از یک منبع خاص موجود در آن رشته اخذ می‌شود دامنه آن را تعیین می‌کنند. اما به‌طور کل تمام مفاهیم، خودشان از یک یا چند منبع مختلف استخراج خواهند شد. بدیهی است، چنانچه یک مفهوم در واژگان منبعی ظاهر نشود در فرااصطلاح‌نامه نیز ظاهر نخواهد شد.

۳-۱۰-۲. حفظ مفاهیم و معانی از واژگان منبع

فرااصطلاح‌نامه، معانی، اسامی مفهومی و روابط را از فرهنگ لغات منبع منعکس و حفظ می‌کند. زمانی که دو فرهنگ لغات منبع متفاوت از یک اسم مشابه و یکسانی برای مفاهیم مختلف استفاده می‌کند، فرااصطلاح‌نامه، هم معنی را نشان می‌دهد و هم معین می‌کند که در حال حاضر کدام معنی در کدام فرهنگ لغات منبع استفاده می‌گردد.

۳-۱۰-۳. لزوم کاربرپسند کردن فرااصطلاح‌نامه

به‌دلیل این که فرااصطلاح‌نامه یک منبع چندمنظوره است که شامل مفاهیم و واژگان مختلف از

منابع مختلف می‌باشد و برای اهداف مختلفی گسترده شده است، لذا فرااصطلاح‌نامه باید برای استفاده مؤثر کاربرپسند شود. فرااصطلاح‌نامه شامل فرهنگ لغات منابعی است که توسط تهیه‌کنندگان مختلف دارای حق مؤلف تولید می‌شود.

اطلاعات در فرااصطلاح‌نامه به خود منابع برچسب زده شده است، بنابراین ممکن است اسامی مفهومی، ویژگی‌ها و روابطی از واژگان منبع برآید و ویژگی‌ها و روابطی که در هنگام ساخته شدن فرااصطلاح‌نامه اضافه شده است را تعیین کند. برچسب‌هایی که به شما اجازه می‌دهد زیرمجموعه‌ای از فرااصطلاح‌نامه را - که برای استفاده از آن‌ها نیازی به اجازه و مجوز ندارید - به استثنای اطلاعات از واژگان منبع خاص ایجاد کنید. همچنین مستثنی کردن تمام فرهنگ لغات منبعی که سطوح محدودیتی خاص دارند یا تمام اطلاعات در زبان‌های خاص را تسهیل می‌نماید.

۳-۱۰-۴. فرهنگ واژگان منبع

فرااصطلاح‌نامه‌ها شامل مفاهیم، اسامی مفهومی، و دیگر ویژگی‌ها و مشتمل بر بیش از صد اصطلاح، رده‌بندی و اصطلاح‌نامه در چند ویرایش می‌باشند. در فرااصطلاح‌نامه برای هر فرهنگ لغات منبع مفهومی وجود دارد، که با نوع معنایی «محصول هوشمند» مشخص می‌شوند و شامل فایل خاصی است که هر نسخه فرهنگ واژگان منبع کنونی در ویرایش خاص از فرااصطلاح‌نامه را ذخیره می‌کند. فایل‌های دیگر فرااصطلاح‌نامه که واژگان منبع مرجع هستند، از ریشه و بدون نگارش اختصارات استفاده می‌کنند (به عنوان مثال به جای ^۱ICD9CM2003 از ^۲ICD9CM استفاده می‌کنند). بنابراین باید از روزآمدسازی روزمره و عادی برای منعکس ساختن نسخه‌های تازه پرهیز شود.

۳-۱۰-۵. شمول مجموعه رموز استانداردهای بین‌الملل و ترمینولوژی

فرااصطلاح‌نامه‌ها شامل مجموعه‌ای از کدهای اختیاری برای استفاده در تراکنش‌های اجرایی الکترونیکی است که فرااصطلاح‌نامه تمام مفاهیم و واژگانی از این مجموعه کدها را شامل می‌شود.

۳-۱۰-۶. شمول زبان‌ها به غیر از زبان انگلیسی

ساختار فرااصطلاح‌نامه‌ها باید به این گونه باشد که بتوانند ترجمه‌های فرهنگ واژگان منبع را به دیگر زبان‌ها به غیر از زبان انگلیسی در خود جای دهند و بسیاری از ترجمه‌ها در بسیاری از زبان‌های مختلف در فرااصطلاح‌نامه نمایش داده شود مانند: فرااصطلاح‌نامه یو.ام.ال.اس.

۳-۱۰-۷. مفاهیم، اسامی مفهومی و شناسه‌های آن‌ها

فرااصطلاح‌نامه‌ها با مفاهیم سازماندهی می‌شوند. یکی از نخستین اهداف فرااصطلاح‌نامه‌ها

^۱ International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)

^۲ ICD-9-CM 2003: Diagnostic Coding in Obstetrics and Gynecology

برقراری ارتباط بین اسامی مختلف برای مفاهیم مشابه از واژگان گوناگون می‌باشد. فرااصطلاح‌نامه‌ها شناسه‌های ثابت چندین نوع خاص را برای مفاهیم در نظر می‌گیرند و علاوه بر حفظ تمام شناسه‌هایی که در فرهنگ لغات منبع موجود است، اسامی مفهومی را نیز دربردارند. ساختار مفهومی فرااصطلاح‌نامه‌ها شامل اسامی مفهومی، شناسه‌های آن‌ها، و ویژگی‌های کلیدی این اسامی مفهومی می‌باشد مانند (زبان، واژگان منبع، انواع اسم).

۳-۱۰-۸. مفاهیم و شناسه‌های مفهومی^۱

یک مفهوم دارای یک معنی است. یک معنی می‌تواند اسم‌های مختلفی داشته باشد. هدف اصلی از ایجاد فرااصطلاح‌نامه‌ها درک انتخاب معنی هر اسمی است که در هر فرهنگ لغات منبع وجود دارد و به تمام اسامی از فرهنگ لغات منبع که معنی مشابهی دارند (مترادف‌ها) متصل است. ساختن فرااصطلاح‌نامه‌ها بر مبنای این فرض است که کارشناسان موضوعی آموزش دیده، می‌توانند مترادف‌ها را همراه با درجه بالایی از صحت و دقت تعیین کنند و این ویرایشگرهای فرااصطلاح‌نامه‌ها هستند که تصمیم می‌گیرند چه دیدگاهی از مترادف در ساختار مفهومی فرااصطلاح‌نامه کنونی ارائه شود. هر مفهوم یا هر معنی در فرااصطلاح‌نامه یک شناسه مفهومی ثابت و خاصی دارد. زمانی یک شناسه از فرااصطلاح‌نامه خارج خواهد شد که دو شناسه اسمی، برای مفاهیم مشابه به کار رود، به عبارت دیگر، تا زمانی که مترادف‌ها کشف نشده‌اند شناسه مفهومی ثابت استخراج می‌شود. در این گونه موارد یکی از آن دو شناسه باقی خواهد ماند، تمام اطلاعات مناسب در فرااصطلاح‌نامه به آن پیوند داده شده و شناسه دیگر حذف خواهد شد و دیگر از شناسه‌های حذف شده استفاده نمی‌کنند. گفتمانی است که هر فرااصطلاح‌نامه باید شامل فایل‌هایی باشد که جزئیات هر گونه تغییر از ویرایش قبلی را داشته باشد.

۳-۱۰-۹. اسامی مفهومی و شناسه‌های زنجیره‌ای

هر اسم مفهومی خاص یا شناسه‌های زنجیره‌ای در هر زبانی در فرااصطلاح‌نامه یک شناسه زنجیره‌ای ثابت و منحصر به فردی دارد. هر تغییری در نظم ویژگی‌ها، یا نقطه‌گذاری، زنجیره جداگانه‌ای است، با یک شناسه زنجیره.

۳-۱۰-۱۰. اتم‌ها و شناسه‌های اتمی (شناسه‌های عنصری و عناصر)

ساختن واحدها یا عناصر از فرااصطلاح‌نامه‌ای که ساخته شده است در واقع اسامی مفهومی یا زنجیره‌ها از فرهنگ واژگان منبع هستند.

¹ Conceptual identifiers

۳-۱۰-۱۱. واژه‌ها و شناسه‌های لغوی

در فرااصطلاح‌نامه‌ها برای مدخل زبانی، هر رشته، به تمام رمزهای جایگزین لغوی یا تغییرات جزئی توسط شناسه لغوی منحصر به فرد پیوند داده خواهد شد.

۳-۱۰-۱۲. شناسه‌های رابطه‌ای و روابط در فرااصطلاح‌نامه

فرااصطلاح‌نامه‌ها شامل روابط بسیاری بین مفاهیم متفاوت هستند (علاوه بر رابطه مترادفی، ساختار مفهومی در فرااصطلاح‌نامه‌ها نیز مطرح است). بسیاری از این روابط از فرهنگ واژگان منبع تخصصی ایجاد می‌شوند. برای نمونه، چنانچه فرااصطلاح‌نامه ما در زمینه پزشکی باشد، از ساختار آن، رده‌بندی کتابخانه ملی پزشکی ایجاد می‌شود. بعضی دیگر از فرااصطلاح‌نامه توسط کاربران برای حمایت از انواع معینی از کاربردها توزیع می‌شوند. به‌طور کل، فرااصطلاح‌نامه‌ها هر رابطه‌ای را نشان می‌دهند، بعضی از روابط که در سال‌های نخستین گسترش فرااصطلاح‌نامه اضافه شده است به فرااصطلاح‌نامه‌ها منسوب می‌شود، اما حقیقت این است که این روابط از واژگان منابع خاص استخراج شده‌اند.

۳-۱۰-۱۳. اساس انواع روابط غیر مترادف‌ها

فرااصطلاح‌نامه، روابط غیر مترادفی ممکن بین مفاهیم را شامل نمی‌شود، بلکه تمام روابط حاضر در خود فرهنگ واژگان منبع را فرا می‌گیرد و بعضی از روابط افزوده برای ایجاد ارتباط میان مفاهیم مرتبط در نظر گرفته شده است.

۳-۱۰-۱۴. برچسب‌های روابط

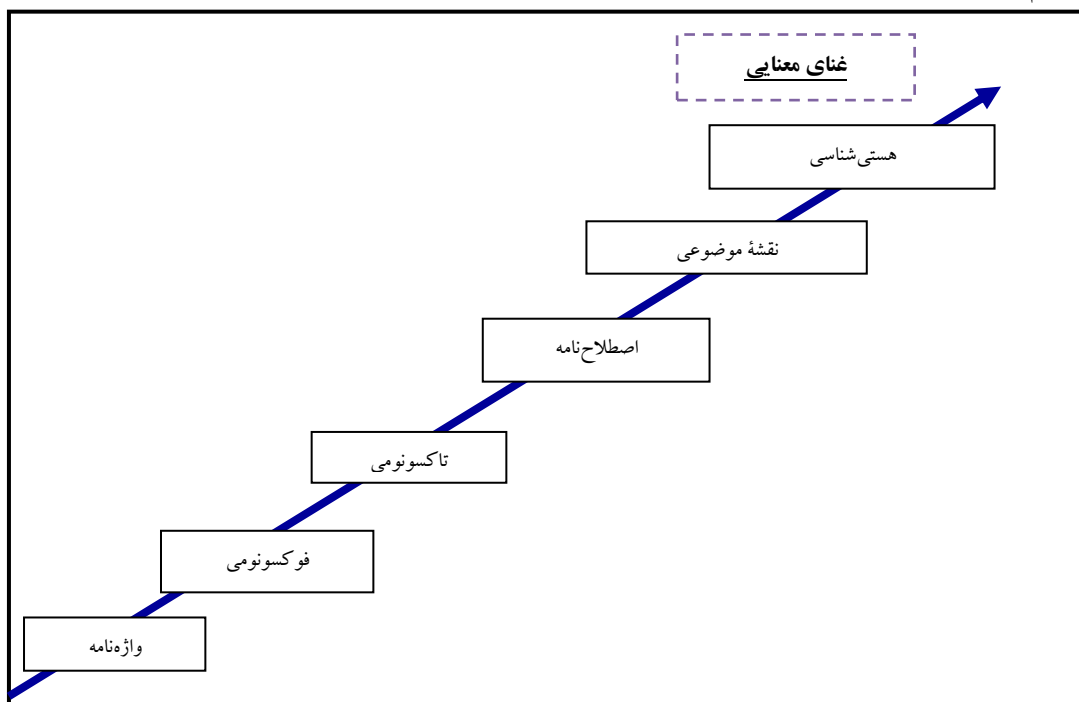
تمام روابط (خارج از ساختار مفهومی پایه) در فرااصطلاح‌نامه یک برچسب عمومی را حمل می‌کند، که اساس ماهیت‌شان را توضیح می‌دهد، مانند اعم‌تر، اخص‌تر، ارجاعات، توصیفگر. این گونه‌ها توسط منبع‌شان مشخص می‌گردند. بیشتر این روابط هم به‌صورت مستقیم در فرهنگ واژگان منبع و هم از طریق ساختار فرهنگ واژگان منبع اثبات می‌شوند.

یک چهارم روابط در فرااصطلاح‌نامه برچسب ارتباطی افزوده را حمل می‌کنند، که از یک فرهنگ واژگان منبع به‌دست آمده است و ماهیت روابط را بیشتر از آنچه که هست توضیح می‌دهد، مانند این که این لغت شاخه یا جزئی از کدام واژه است. فرهنگ لغات علم کالبدشناسی دیجیتالی و RxNorm نمونه‌هایی از فرهنگ واژگان منبعی است که شامل برچسب‌های این گونه روابط است.

۳-۱۱. تدوین فرااصطلاح‌نامه

به‌طور کلی استفاده از فرااصطلاح‌نامه‌ها برای رسیدن به یک مفهوم صریح و آشکار در علوم مختلف از جمله پزشکی امری تردیدناپذیر است. اما این به‌معنای حذف منابع موجود در آن رشته

موضوعی خاص نمی‌باشد، زیرا همان‌طور که پیش‌تر اشاره شد، این منابع به مفاهیم کلی‌تری پرداخته‌اند که مفاهیم جزئی را دربرنمی‌گیرد، ولی فرااصطلاح‌نامه به تجزیه و ترکیب مفاهیم، اصطلاحات و مترادف‌های آن‌ها در گستره‌های تخصصی پرداخته که ما را قادر به رسیدن به معانی و مفاهیم ویژه در آن حوزه موضوعی خاص می‌کند.



شکل ۳-۴. جایگاه هستی‌شناسی در پلکان معنایی^۱ (بلومر و پلگرینی، ۲۰۰۶)

در دهه‌های اخیر بازنمایی و بازیابی دانش در حوزه مستندسازی از فناوری‌های مهمی چون اینترنت و فناوری‌های وابسته به آن تأثیرات فراوانی یافته است. این فناوری جدید و تغییر و تحولات مهم در حوزه بازنمایی اطلاعات دیجیتال و بازیابی آن و تأثیرات عوامل مختلف در حوزه اطلاع‌رسانی و ابزارهای جدید بازنمایی و بازیابی، بر کاربران اطلاعاتی نیز اثر گذاشته است. در سایه این رشد روزافزون متون دیجیتال، ابزارهای جدید بازنمایی اطلاعات نیز تغییر یافته، به شکلی که رشته‌های دیگر علمی نیز در این عرصه سهیم شده‌اند و در نتیجه رشته‌هایی مانند زبان‌شناسی، هوش مصنوعی، مستندسازی، مهندس زبان و غیره نیز باید مورد مطالعه قرار گیرند. بر این اساس باید موضوعات خاص مانند تاکسونومی‌ها یا سیستم‌های رده‌بندی، واژگان رایانه‌ای، پایگاه‌های واژگانی، اصطلاح‌نامه‌ها، فرااصطلاح‌نامه، نقشه مفهومی، شبکه معنایی واژگان و هستی‌شناسی‌ها به عنوان ابزارهای بازنمایی دانش مورد توجه قرار گیرند (شکل ۱۴-۱). در این میان دامنه موضوعی گسترده

¹ Semantic staircase

«بازنمایی دانش»^۱، فرااصطلاح‌نامه‌ها و هستی‌شناسی‌ها جزء مهم‌ترین ابزار سازماندهی اطلاعات محسوب می‌شوند، هرچند تعلق به گرایش‌های علمی کاملاً متفاوت دارند.

واژه اصطلاح‌نامه^۲ به معنای «گنجینه‌ای از واژگان (گنجواره)» است. در سال ۱۸۵۲ «پیرمارک» روزه نتایج پژوهشی خود را با عنوان «اصطلاح‌نامه روزه»^۳ که در آن هدفش کمک به متخصصین و کارشناسان فنی برای بیان ایده‌ها و تعاریف مفاهیم علمی و انتقال دانش فنی بود، منتشر کرد. «اصطلاح‌نامه روزه» ساختاری «مفهوم بنیان» داشت که در آن، مفاهیم بیانگر مدخل‌های واژگانی و زیربنای ساختار ارتباطات و گروه‌بندی واژه‌ها بود و نشانگر نامگذاری مفاهیم مورد نظر در متون گوناگون، معانی و کاربردهای «اصطلاحات» بود.

فرااصطلاح‌نامه ساختار معناشناختی را ارائه می‌نماید که عمدتاً با روابط حاکم میان مفاهیم و در نهایت اصطلاحات (که بازنمونی از مفاهیم هستند) بیان می‌شوند و در حوزه مستندسازی و طبقه‌بندی اطلاعات، ابزاری است برای کنترل اصطلاحات به کار گرفته شده، زیرا ساختار فرااصطلاح‌نامه بر اساس مفاهیم استوار است.

مفاهیم به وسیله اصطلاحات گزینش شده ارائه می‌شوند که خود نوعی کنترل واژگان است و «ترادف» و «چندمعنایی» که هر دو از ویژگی‌های زبان‌های طبیعی هستند را ارزیابی می‌نماید.

همان گونه که گفته شد، نتایج و عواقب کارکرد در محیط الکترونیکی و دیجیتالی ابزارها و تجهیزات و حتی رفتار کاربران را در رابطه با بازنمایی و بازیابی اطلاعات تحت تأثیر قرار داده و امکانات جدیدی در طراحی، گسترش و مدیریت کاربرد ابزارها در حوزه بازنمایی و بازیابی اطلاعات ایجاد نموده است. این عوامل تأثیرگذار عبارتند از:

۱. ساختار فرااصطلاح‌نامه: گسترش کارکرد آن‌ها بر اساس محیط فرامتنی است که منجر به ایجاد و ساخت پیوندهای فرامتنی در میان تمامی عناصر ساختاری (توصیفگرها، غیرتوصیفگرها، یادداشت دامنه‌ها و ...) و نیز در میان بخش‌های مختلف فرااصطلاح‌نامه می‌گردد.

۲. کاهش هزینه‌های توسعه و روزآمدسازی اصطلاح‌نامه‌ها: با دیجیتالی کردن فرایند ساخت فرااصطلاح‌نامه‌ها و صرف نظر نمودن از فرمت‌های کاغذی به عنوان ابزار انتشاراتی، کاهش هزینه‌ها تا حدود زیادی مقدور می‌باشد.

۳. مشارکت کاربران: سهم نمودن نظرات کاربران در فرایند تدوین، مدیریت و روزآمدسازی

¹ knowledge representation

² thesaurus

³ Roget thesaurus

اصطلاح‌نامه‌ها و فرااصطلاح‌نامه که به‌وسیله آزمون‌های کاربردی، استفاده از تکنیک‌های الگوسازی، کاربران و ... امکان‌پذیر شده است. این روش امکان تولید ابزارهایی را فراهم می‌سازد که کاربران و نظرات آنان نیز مورد توجه قرار می‌گیرد و ایجاد ساختارهای کاملاً نظری و غیر کاربرمدار را منتفی می‌سازد.

۴. مهندسی مجدد^۱: امکان به‌کارگیری شیوه‌های کاربردی مهندسی مجدد و بازطراحی و صرفه‌جویی، در زمان طراحی و ایجاد فرااصطلاح‌نامه است. این بازطراحی، کاربرد و استفاده از اطلاعات واژگانی و مفهومی موجود در دیگر پایگاه‌ها و سیستم‌های اطلاعاتی را امکان‌پذیر می‌سازد، و همین عامل کلیدی، در ایجاد نسل‌های جدید فرااصطلاح‌نامه دیجیتالی مؤثر است.

۵. هستی‌شناسی: کاربرد اطلاعات مفهومی و زبان‌شناختی ذخیره شده در دیگر انواع ماشین‌های جستجو (مانند هستی‌شناسی‌ها) ما را قادر به پیشبرد عناصر ساختاری فرااصطلاح‌نامه و نیز ماهیت کاربرپسند نمودن و توسعه این ابزارها خواهد نمود (سوا، ۲۰۱۰).

۳-۱۱-۱. تلفیق اصطلاح‌نامه‌ها

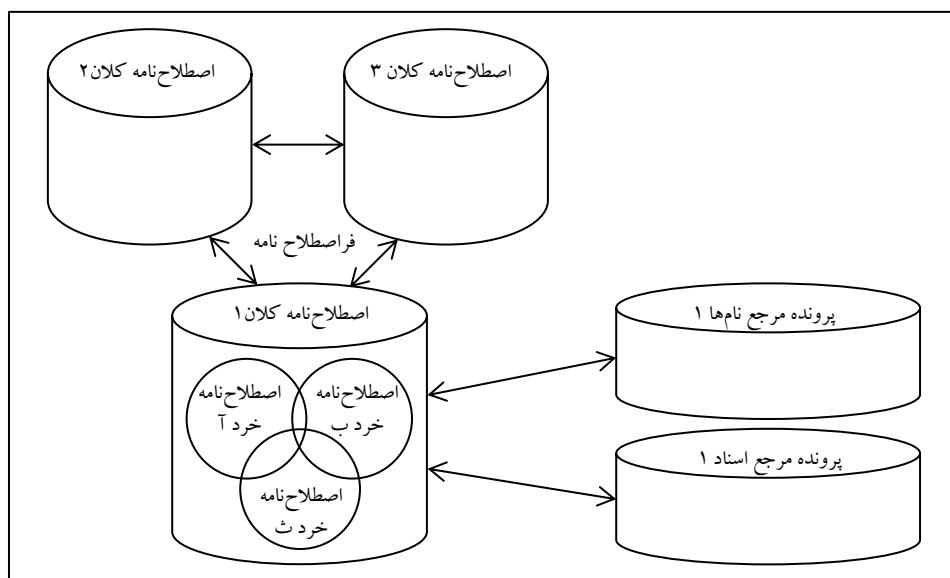
در سال‌های گذشته بسیاری از زبان‌های کنترل شده - چه به‌صورت اصطلاح‌نامه‌ها و چه طرح‌های رده‌بندی یا فهرست سرعنوان‌های موضوعی - انتشار یافته‌اند که کلاً مستقل از یکدیگر هستند و حتی بعضاً هرچند در یک حوزه موضوعی واحد به‌صورت اصطلاح‌نامه‌های خرد^۲ می‌باشند، به لحاظ ساختار، دیدگاه و تخصیصی‌بودن با یکدیگر تفاوت دارند. این فقدان سازگاری، انتقال رکورد در بین سامانه‌ها و کاوش در پایگاه‌های اطلاعاتی مختلف را با مشکل مواجه می‌سازد. انطباق یا تلفیق اصطلاح‌نامه‌ها، راهی برای غلبه بر ناسازگاری آن‌ها می‌باشد.

وجود استانداردهایی برای تدوین اصطلاح‌نامه‌های خرد و تخصیصی و نیز ادغام آن‌ها به‌صورت اصطلاح‌نامه کلان^۳ فقط به‌طور محدودی به کاهش ناسازگاری‌های بین زبان‌های کنترل‌شده کمک می‌کند. تفاوت‌های عمده در زبان‌های کنترل‌شده در یک فیلد واحد شامل مواردی مانند: میزان اخصیت، فراگیری و پوشش موضوعی، اصطلاحات مرکب، مترادف‌ها، متضادها و چندمعنایی میان واژگان می‌باشد (شکل ۱۴-۲).

^۱ Reengineering

^۲ Microthesaurus

^۳ Macrothesaurus



شکل ۳-۵. تلفیق اصطلاح‌نامه‌ها (Yancey, T. & Carson, J., 2004)

شیوه‌های انطباق زبان‌های کنترل‌شده ناسازگار شامل مطابقت‌دادن^۱، جهت‌گرانی^۲، ادغام^۳ و تلفیق^۴ می‌شود.

مطابقت‌دادن، اصطلاحات را مستقیماً از یک زبان کنترل‌شده به زبانی دیگر یا به مجموعه واژگان چندگانه پیوند می‌دهد، اصطلاحات یک زبان کنترل‌شده، به اصطلاحات معادل در زبان دیگر اشاره می‌کنند. اطلاعات موجود بیان می‌کند که آیا اصطلاح مشابه است، مترادف است یا دارای رابطه‌ای دیگر با آن اصطلاح در زبان‌های دیگر می‌باشد.

در جهت‌گردانی خودکار^۵، اصطلاحات یک واژگان، به‌طور خودکار به اصطلاحات واژگان دیگری تغییر شکل می‌یابد. جهت‌گردانی بر مبنای انطباق‌ها، زبان‌های میانجی، با الگوریتمی برای کنترل اصطلاح‌های دروندادی استوار است (ایچیسن و همکاران، ۱۳۸۵، ص ۲۴۴).

۳-۱۱-۲. توسعه و روزآمدسازی اصطلاح‌نامه‌های تخصصی

یک زبان نمایه‌سازی به محض انتشار باید به‌طور منظم روزآمد شود. اصطلاح‌نامه‌هایی دارای واژگان تخصصی که با استفاده از نمایه‌سازی عمیق در سامانه‌ها به کار برده می‌شوند، به دفعات بیشتر از اصطلاح‌نامه‌های محدود و عام‌تر نیازمند روزآمدسازی و توسعه می‌باشند. همان‌گونه که انتظار می‌رود با افزایش تعداد اسناد نمایه‌ای، صرف نظر از ماهیت واژگان یا عمق نمایه‌سازی، نیاز به اصلاح و تغییر

^۱ mapping

^۲ switching

^۳ merging

^۴ integration

^۵ automated switching

نیز افزایش می‌یابد. اصطلاح‌نامه در مراحل اولیه ایجاد پایگاه اطلاعاتی، دارای بیشترین رشد می‌باشد و در صورت گسترش پوشش موضوعی پایگاه اطلاعاتی رشد آن افزوده می‌شود.

روزآمدسازی اصطلاح‌نامه نیازمند مدیریتی روشمند است، در غیر این صورت با عدم انسجام و بی‌دقتی اصطلاح‌نامه و دشواری فزاینده در استفاده و نگهداری و کاوش با آن دچار اختلال خواهد شد. مسئولیت کنترل و توسعه اصطلاح‌نامه باید بر عهده ویراستاری باشد که در صورت امکان، گروه کوچکی از دستیاران و متخصصین موضوعی مشاور به وی کمک می‌کنند. گردآوری و به‌کارگیری اصطلاحاتی که از کارهای در دست اقدام نمایه‌سازان و جستجوگران شکل می‌گیرند، امری رایج است. این اصطلاحات گزینش شده را می‌توان در جایی که نرم‌افزار امکان می‌دهد، به‌عنوان اصطلاحات قابل کاوش، در فیلد کلیدواژه مجزایی وارد پایگاه اطلاعاتی کرد. به محض اتخاذ تصمیمی از سوی ویراستار در خصوص جایگاه این اصطلاحات، می‌توان آن‌ها را به فیلد اصطلاح نمایه‌ای ارتقاء داد، حذف نمود یا اصلاح کرد و به آن جایگاهی جدید تخصیص داد. ارزیابی هر پیشنهاد از وظایف ویراستار اصطلاح‌نامه است که با در نظر گرفتن اعتبار پیشنهاد، اجرای آن و تأثیراتی که از اعمال آن در طرح ناشی می‌شوند، این کار را انجام می‌دهد.

در صورتی که ویراستار در رسیدن به تصمیمات دچار مشکل شود، آنگاه می‌تواند یا به‌صورت غیررسمی و یا در جلسات منظم با متخصصین شورای اصطلاح‌نامه به رایزنی و اتخاذ تصمیم نهایی بپردازد (ایچ‌سن و همکاران، ۱۳۸۵، ص ۲۳۳).

۳-۱۱-۳. تدوین فرااصطلاح‌نامه

برای استفاده از چند اصطلاح‌نامه خرد و کلان به‌صورت یک فرااصطلاح‌نامه و تبدیل آن به هستی‌شناسی، در اولین قدم باید کلیه مفاهیم مشترک در اصطلاح‌نامه‌ها همسان شوند، در یک مجموعه اصطلاح‌نامه با ده‌ها هزار واژه که در زمان‌های مختلف و توسط متخصصان موضوعی مختلف تولید شده‌اند، همسان‌سازی بدون استفاده از توانایی‌هایی که سیستم‌های رایانه‌ای در اختیارمان می‌گذارد، تقریباً ناشدنی است و نیاز به زمان و کار بسیار زیاد نیروی انسانی متخصص خواهد داشت تا بتوان فرایند همسان‌سازی را تا حدی به نتیجه مطلوب نزدیک سازند.

وقتی زبان‌های کنترل‌شده ادغام می‌شوند، اصطلاحات مرجح و اصطلاحات مدخل، در کنار رابطه‌های اصطلاح‌نامه‌ای همراه، در یک فهرست الفبایی واحد در کنار هم آورده می‌شوند و به زبان‌های کنترل‌شده‌ای که اصطلاحات خاص، یادداشت‌های دامنه و رابطه‌ها را دربرمی‌گیرند نیز اشاره‌ای می‌شود. تطابق‌ها و مغایرت‌های موجود در رابطه‌های همابندی و سلسله‌مراتبی در اصطلاح‌نامه‌های گوناگون و معادل‌های به زبان خارجی برای اصطلاحات ادغام‌شده را نیز نشان

می‌دهد. در هنگام ادغام اصطلاح‌نامه‌ها، از رابطه‌های بین همین دو اصطلاح در اصطلاح‌نامه دیگر نیز وجود دارد.

هدف از ادغام زبان‌های کنترل‌شده، رها کردن زبان‌های کنترل‌شده ساختاریافته قبلی و جذب آن‌ها در ابزار یگانه و جدید برای نمایه‌سازی و جستجوی اطلاعات است. راه دیگر آن است که زبان‌های کنترل‌شده قبلی استقلال خود را حفظ کنند و همچنان برای نمایه‌سازی هر یک از پایگاه‌های اطلاعاتی مورد استفاده قرار گیرند. در این شیوه، در هنگام کاوش در پایگاه‌های اطلاعاتی گوناگون برای جابه‌جایی بین اصطلاحات و واژگان‌های ادغام‌شده، می‌توان از جهت‌گردانی استفاده کرد. در صورت نیاز، با افزودن رابطه‌های جدید بین اصطلاحات می‌توان هر یک از زبان‌ها را ارتقاء بخشید. مشهورترین نمونه‌های اخیر از اصطلاح‌نامه ادغام‌شده «نظام زبان پزشکی واحد» (یو.ام.ال.اس) است که روی وب‌سایت نظام پزشکی واحد تشریح شده است و پروژه‌ای بلندمدت تحقیق و توسعه است که توسط «کتابخانه ملی پزشکی» آمریکا از سال ۱۹۸۶ با همکاری نهادهایی از داخل و خارج ایالات متحده اجرا شده است. هدف از این طرح فراهم آوردن امکان دستیابی یکسان به تعدادی از پایگاه‌های اطلاعاتی توزیع‌شده و ناهمخوان است که از مجموعه اصطلاحات متفاوتی استفاده می‌کند.

«یو.ام.ال.اس» شامل چهار «منبع دانش»^۱ می‌شود:

«فرااصطلاح‌نامه»، «شبکه معناشناختی»، «نقشه منابع اطلاعاتی» و «واژه‌نامه‌های تخصصی»

الف- «فرااصطلاح‌نامه» ترکیبی است از زبان‌های کنترل‌شده موجود که از طریق برقراری پیوند، ادغام و تلفیق آن‌ها به دست آمده است. اصطلاح‌نامه‌ها با یکدیگر مطابقت می‌یابند و از هر مفهوم در یک اصطلاح‌نامه، نشانگرهایی به نزدیک‌ترین مفهوم معادل در اصطلاح‌نامه‌های دیگر ایجاد می‌گردد. این امر امکان ترجمه پرسش کاربر به کاوش برای یک پایگاه اطلاعاتی مشخص را که با اصطلاح‌نامه مخصوصی نمایه‌سازی شده، فراهم می‌کند. واژگان‌ها نیز ادغام می‌شوند و یک مبنای دانش غنی‌تر و جامع‌تر ایجاد می‌گردد. در هر یک از واژگان‌های ساختاریافته نامی به عنوان نام مرجع برای مفهوم مربوطه انتخاب می‌شود (نمودار ۲، ص ۲۵۳، ایچ‌سی‌سن، ۱۳۸۵).

¹ Knowledge source

The screenshot shows the 'Metathesaurus Retrieval System' interface. At the top, there is a search bar with the text 'Query Term: aids'. Below the search bar are several buttons: 'Clear', 'Help', 'Exit', 'Abort', and 'Terms'. A 'Concept Info' panel is open on the right side, displaying a list of concepts related to 'aids'. The main area shows the search results for 'aids'.

Concept Name	
Query Term	aids
Term Matched:	AIDS
Concept Name:	Acquired Immunodeficiency Syndrome
UI:	MC0001175
MSH/ MH:	Acquired Immunodeficiency Syndrome
ICD/ PT:	Acquired Immunodeficiency Syndrome
MSH/ ET:	AIDS
ICD/ AB:	UNSPECIFIED AIDS

The 'Concept Info' panel on the right lists the following items:

- Concept Name
- Concept Definition
- Concept Variants
- Synonyms
- Related Concepts
- Associated Expressions
- Semantic Types
- Context
- Term Data

شکل ۳-۶. فرااصطلاح‌نامه «یوام ال اس» اطلاعات مربوط به هر یک از مفاهیم

این امر امکان جهت‌گردانی بین واژگان‌ها را فراهم می‌کند در صورتی که آن مفهوم در «مش»^۱ وجود داشته باشد، شکل مرجع آن مفهوم، صورت سرعنوان اصلی در «مش» است. در غیر این صورت اصطلاحات از دیگر واژگان‌ها به ترتیبی مشخص گزینش می‌شوند. تمام اصطلاحات، عبارات و زنجیره‌های دیگر که برای بیان مفهوم مورد استفاده قرار می‌گیرند، یا نام مندرج در «فرااصطلاح‌نامه» مطابقت داده می‌شوند.

ب- شبکه معناشناختی: دومین «منبع دانش» یعنی «شبکه معناشناختی یوام ال اس» یک دسته‌بندی دائمی از تمام مفاهیم موجود در «فرااصطلاح‌نامه» بر طبق دسته‌های نوع‌های معناشناختی و رابطه‌های بین این دسته‌ها ارائه می‌کند.

¹ Mesh thesaurus

گروه‌بندی عمده درون «موجودیت‌ها» شامل «شیء فیزیکی»، «جانداران»، «بافت‌های کالبدی»، «مواد شیمیایی» و «مفاهیم یا ایده‌ها» می‌شوند.

در رخدادهای پدیده‌یافرایند و «کارکرد زیست‌شناختی» هستند. نحوه تنظیم «نوع‌های معناشناختی» سلسله‌مراتبی است.

رابطه‌های غیرسلسله‌مراتبی به پنج گروه تقسیم می‌شوند: رابطه‌های فیزیکی (مانند: بخشی از، مشتمل بر ...) رابطه‌های فضایی، رابطه‌های زمانی (مانند: قبل از، همزمان با ...) رابطه‌های کارکردی (مانند: موجب می‌شود، تولید می‌کند، تحت تأثیر قرار می‌دهد)، و رابطه‌های مفهومی (مانند: اندازه‌گیری‌ها، ارزیابی‌ها). این رابطه‌ها بین گروه‌های سطح بالای دسته‌های معناشناختی بیان می‌شوند، مانند «فرایند» بین «کارکرد زیست‌شناختی» و «جانداران». از آن‌جا که یک رابطه عبارت است از یک (isa) بین «نوع‌های معناشناختی» وجود دارد، این حالت بین «کارکرد اندامی و بافتی» (یک نوع از «کارکرد زیست‌شناختی») و «حیوانات» (یک نوع از «جانداران») نیز مصداق دارد. رابطه‌های ارائه‌شده نمایانگر تمام رابطه‌های ممکن بین «انواع معناشناختی» تعریف‌شده نیستند، بلکه مرحله‌ای آغازین در ایجاد مجموعه مفیدی از رابطه‌ها می‌باشند. کاربردهای عمده «شبکه معناشناختی»، در تدوین پرسش و در پالایش تعاملی پرسش جستجو است.

ج- نقشه منابع اطلاعاتی و واژه‌نامه‌های تخصصی

سومین «منبع دانش یووال‌اس» «نقشه منابع اطلاعاتی» است که نشان می‌دهد کدام پایگاه اطلاعاتی به سؤالی خاص مربوط می‌شود. اطلاعات درباره منابع تخصصی را برای کاربران فراهم می‌کند، به‌طور خودکار به منابع متصل می‌شود و عمل بازیابی را بر روی یک یا چند منبع انجام می‌دهد.

د- چهارمین «منبع دانش»، واژه‌نامه‌های تخصصی است که برای فراهم آوردن اطلاعات واژگانی مورد نیاز در «نظام پردازش زبان طبیعی تخصصی» ایجاد شده است. هدف از این منبع آن است که یک واژه‌نامه با بسیاری از اصطلاحات زیست‌پزشکی فراهم گردد. واژه‌نامه‌های تخصصی برای هر اصطلاح، دارای اطلاعات نحوی، واژه‌شناختی و املائی نیز می‌باشد.

نظام «یووال‌اس» نشان می‌دهد که چگونه ادغام زبان‌های کنترل‌شده، مزایای یک محصول نهایی را به‌دست می‌دهد، محصولی که به لحاظ اصطلاحات و رابطه‌ها از هر یک از واژگان‌های ساختاریافته به تنهایی، غنی‌تر است. جهت‌گردانی بین واژگان‌های مستقل را میسر می‌سازد، بازیابی اطلاعات را با استفاده از اصطلاحات متعارف زبان کنترل‌شده تسهیل می‌کند و نیز به‌عنوان ابزاری برای تأمین دستیابی به متون زبان طبیعی در پایگاه‌های اطلاعاتی زیست‌شناسی عمل می‌کند. این نظام «مهم‌ترین گنجینه تجارب زیست‌پزشکی» به اشتراک گذاشته شده نامیده شده است.

امر انطباق اصطلاح‌نامه‌ها و دیگر زبان‌های کنترل‌شده با مشکلات بسیاری مواجه شده است. یکی از این مشکلات آن است که اصطلاح‌نامه‌هایی که قرار است تلفیق شوند، اغلب به لحاظ ساختاری دارای تفاوت اساسی هستند و بسیاری از آن‌ها از استانداردهای ایزو و یا انسی پیروی نمی‌کنند. در صورتی که بین اصطلاح‌نامه‌ها زمینه مشترکی وجود داشته باشد، احتمال موفقیت بیشتر خواهد بود. تجربه‌ای که مکرراً رخ می‌دهد آن است که مسئله بسیار رایج فقدان تناظر یک به یک بین اصطلاحات اصطلاح‌نامه‌های گوناگون، فرایند انطباق را با ناکامی مواجه می‌کند. با وجود تمام دشواری‌ها، این تلاش در صورتی که نتیجه نهایی انطباق، اصطلاح‌نامه‌های تلفیق شده باشد، ارزشمند است. فرااصطلاح‌نامه‌ای که ذخیره‌ای غنی از گزینه‌های جایگزین را برای مجموعه اصطلاحات ارائه می‌کند تا کاربر بتواند این گزینه‌ها را در کاوش متن آزاد و نیز در کاوش به زبان کنترل‌شده اعمال کند (ایچسن و همکاران، ۱۳۸۵، صص ۲۵۶-۲۵۲).

در مرحله تلفیق، موارد اختلاف بین اصطلاح‌نامه‌ها بر اساس استاندارد «ایزو ۲۵۹۶۴» که جدیدترین استاندارد ایزو و مجموعه‌ای از استانداردهای تک‌زبانه و چندزبانه انسی و ایزو است بررسی می‌شوند. اگر بیش از دو اصطلاح‌نامه نیاز به همسان‌سازی^۱ داشته باشند، با همسان‌سازی همه اصطلاح‌نامه‌ها به صورت دوه‌دو، یک مجموعه اصطلاح‌نامه همسان خواهیم داشته که شامل موارد ذیل می‌شوند:

۱. اختلاف نوشتاری بین یک مفهوم در دو اصطلاح‌نامه؛
۲. اختلاف در واژه‌های اخص برای یک مفهوم در دو اصطلاح‌نامه؛
۳. اختلاف در واژه‌های نامرجح برای یک مفهوم در دو اصطلاح‌نامه؛
۴. اختلاف بین ترجمه‌های یک مفهوم در یک زبان خاص در دو اصطلاح‌نامه؛
۵. اختلاف در واژه‌های وابسته برای یک مفهوم در دو اصطلاح‌نامه؛
۶. کمبود یک واژه وابسته که معادل مفهوم مورد نظر در اصطلاح‌نامه دیگر آمده؛
۷. استفاده از ترجمه یکسان برای دو مفهوم مختلف در یک زبان خاص؛
۸. اختلاف در انتخاب یک واژه مرجح برای یک مفهوم؛
۹. اختلاف در نوع رابطه بین دو مفهوم؛
۱۰. ایجاد حلقه نامحدود بیان مفاهیم (شبکه مفاهیم)؛
۱۱. عدم یگانگی یک مفهوم در برابر یک اصطلاح (چندمعنایی) (حسینی بهشتی، ۱۳۹۲، ص ۲۵۰).

¹ Synchronizing

۳-۱۱-۴. مراحل تلفیق اصطلاح‌نامه‌ها

نظر به این که تلفیق اصطلاح‌نامه‌ها و تولید یک فرااصطلاح‌نامه، مقدمه ایجاد هستی‌شناسی است، زمان شروع آن بلافاصله پس از یکسان‌سازی کامل و صحیح و دقیق اصطلاح‌نامه‌ها خواهد بود که توسط متخصصان موضوعی انجام می‌شود.

فرایند تبدیل از واژه‌بنیان به مفهوم‌بنیان، کاملاً ماشینی خواهد بود و نیاز آن به متخصصان موضوعی کمتر است. اما قبل از آن، باید مراحل زیر را به‌طور کامل و دقیق طی کرد:

۱. تصمیم‌گیری برای ادغام اصطلاح‌نامه‌ها یا ایجاد نقشه معنایی بین آن‌ها و پیاده‌سازی آن توسط متخصصین؛

۲. بررسی مفاهیم و تعریف اجتماع انواع روابط سلسله‌مراتبی، روابط هم‌ارزی و روابط وابسته توسط ماشین؛

۳. تبدیل هر یک از روابط موجود (سلسله‌مراتبی، هم‌ارز و وابسته) در اصطلاح‌نامه، به روابط تعریف شده جدید بین مفاهیم توسط متخصص؛

۴. تعیین میزان ارتباط (وابسته و سلسله‌مراتبی) بین مفاهیم توسط متخصص؛

۵. ایجاد پیوند بین مفاهیم و منابع مرتبط توسط متخصص؛

۶. بررسی مفاهیم و تعریف انواع روابط سلسله‌مراتبی، وابسته و هم‌ارز توسط ماشین؛

۷. بررسی مفاهیم و تخصیص «یوآرای»^۱ یکسان توسط متخصص به کمک ماشین؛

۸. یکسان‌سازی مفاهیم مشترک بین اصطلاح‌نامه‌ها توسط متخصص به کمک ماشین؛

۹. بررسی روابط هم‌ارزی و انتخاب واژه مرجع مشترک توسط متخصص؛

۱۰. بررسی روابط سلسله‌مراتبی و وابسته و مشخص کردن روابط نامتجانس بین مفاهیم یکسان توسط ماشین؛

۱۱. تصمیم‌گیری برای هر یک از روابط نامتجانس توسط متخصص؛

۱۲. بررسی روابط سلسله‌مراتبی و وابسته و یکسان‌سازی آن‌ها بر اساس روابط تعریف شده جدید توسط متخصص به کمک ماشین؛

۱۳. انتقال اصطلاح‌نامه‌ها از استاندارد ۵۹۶۴ به استاندارد «ایزو ۲۵۹۶۴» و اعمال تغییرات لازم توسط متخصص؛

۱۴. تبدیل از حالت واژه‌بنیان به مفهوم‌بنیان توسط ماشین (حسینی بهشتی، ۱۳۹۲، ص ۲۵۱).

بنابراین «فرااصطلاح‌نامه» نوعی اصطلاح‌نامه است که در نتیجه ادغام و تلفیق اصطلاح‌نامه‌های

¹ URI

متعدد به‌دست می‌آید و میان مجموعه واژگان ملی و بین‌المللی، ارجاعات متقابل ایجاد می‌کند. به‌منظور بازیابی در پایگاه‌های متعدد و چندگانه که به واسطه واژگان کنترل‌شده متعدد نمایه شده‌اند، نوعی فرااصطلاح‌نامه ایجاد می‌شود. هر اصطلاح در فرااصطلاح‌نامه با برابر نهاده‌ها و اصطلاحات مرتبط از تمامی اصطلاح‌نامه‌های تلفیق‌شده در آن، به نمایش درمی‌آیند. فرایند خودکار در جستجو و بازیابی، توسعه می‌یابد و کاربر را قادر می‌سازد بازیابی را در نمایه‌های منفرد یا چندگانه و تلفیق‌شده و با استفاده از اصطلاحات هر یک از منابع واژگانی، انجام دهد. فرااصطلاح‌نامه مفهوم بنیان است و هدف آن ایجاد وحدت میان نامگذاری معانی مختلف یک مفهوم است. این نامگذاری‌ها از منابع واژگانی متعدد و رده‌بندی‌ها و اطلاعات پیرامون آن به‌دست آمده است.

گسترش سریع میزان واژگان یک فرااصطلاح‌نامه باعث شده است که به‌کارگیری رایانه برای سهولت در ایجاد آن مؤثر واقع شود. روزآمدی واژگان، بهره‌مندی از مجموعه واژه‌های متعدد، الگوریتم‌گزینی و بازنگری متخصصین موضوعی را می‌طلبد. اصطلاح‌نامه‌های مختلف را می‌توان تحت شرایط زیر در یک اصطلاح‌نامه تلفیق نمود.

۱. از ابتدا اصطلاح‌نامه‌های مختلف الزاماً بر اساس پیش‌زمینه نظری واحدی ساختاربندی شده

باشند. این امر گام نخست برای تلفیق و کاملاً ضروری است؛

۲. در گام دوم، شیوه تخصصی شناسه‌های یکسان به‌کار گرفته شود و زمینه‌های بازنگری و رفع

مشکلات شناسایی شده، فراهم گردد؛

۳. در گام سوم، توسعه کنترل کیفیت دیدگاه‌ها و توافقات به عمل آمده صورت پذیرد؛

۴. در گام چهارم، شناسایی زمینه‌های جابه‌جایی و دگرگونی ظریف و نامحسوس معانی و

مفاهیم در هر یک از گرایش‌ها ضروری است.

فرااصطلاح‌نامه تنها یک مجموعه واژگان استاندارد شده نیست، بلکه همچنین ابزاری برای افزایش نهایی واژگان موجود و مفید است و منبعی غنی برای متخصصین اطلاعات علمی می‌باشد. نظارت بر تناسب اجزای فرااصطلاح‌نامه در هر ویرایش جدید، سبب توسعه و افزایش میزان مفاهیم یک فرااصطلاح‌نامه می‌گردد. فناوری فرااصطلاح‌نامه برای فرمول‌بندی روابط معنایی میان ساختارهای دانش فنی (به‌صورت مفاهیم و روابط سلسله‌مراتبی آن‌ها) از دیدگاه‌های گوناگون به‌کار گرفته می‌شود و کاربران را قادر به تورق، جستجو و بازیابی میان رشته‌ای در گرایش‌های گوناگون علمی می‌سازد. فرااصطلاح‌نامه دربرگیرنده مجموعه‌ای از مفاهیم کلیدی در رأس، و روابط معنایی میان مفاهیم (همانند مترادف‌ها، همنام‌ها و ...) و نیز ترسیم نقشه معنایی از منابع گوناگون اصطلاح‌نامه‌ها و مفاهیم کلیدی است که به نسبت تعداد کمی از این‌ها تألیف و ابداع شده‌اند (وود، ۲۰۰۴).

۳-۱۱-۵. نحوه پیاده‌سازی فرااصطلاح‌نامه

با وجود این که سازمان‌های بسیاری در حوزه‌های گوناگون تخصصی ضرورت ایجاد اصطلاح‌نامه‌های تخصصی را دریافته و اقدام به تدوین اصطلاح‌نامه متعددی نموده‌اند، تعجب‌آور است که در زمینه تلفیق اصطلاح‌نامه‌های گوناگون و ایجاد فرااصطلاح‌نامه کمتر تلاش شده است. به‌ویژه این مسئله معطوف به انتشاردهندگان اصطلاح‌نامه‌هاست که زمینه‌های موضوعی مشابهی را کار کرده‌اند (فاسکت، ۱۹۸۰، ص ۴۵۶).

هدف از همسان‌سازی واژگان اصطلاح‌نامه‌ها و تلفیق به شکل فرااصطلاح‌نامه مبتنی بر مفهوم یک‌پارچه‌سازی علوم است، بدین معنا که تمامی رشته‌های علمی متعلق به یک ایده واحد است. مثلاً در مورد فیزیک و روانشناسی، که گرایش‌های متمایز علوم محسوب می‌شوند، نظریه وحدت وجود بر این باور است که به‌صورت سلسله‌مراتبی با یکدیگر مرتبط و طبقه‌بندی شده‌اند. بدین شکل که علم فیزیک در زیربنایی‌ترین سطح و شیمی در سطح بالاتر از آن، زیست‌شناسی در رده بالاتر از آن و روانشناسی بالاتر از زیست‌شناسی و سرانجام علوم اجتماعی بالاتر از روانشناسی قرار گرفته است. تلاش‌های فراوان علمی عامل مؤثری در ایجاد و مدیریت اصطلاح‌نامه‌ها به شکل خرد، کلان و فرااصطلاح‌نامه می‌باشد. تفکر وحدت در علوم نشانگر تفکر چندبعدی و چندمنظوره است که منجر به تلفیق علوم و ایجاد ارتباط میان رشته‌های گوناگون و سرانجام تحقیقات میان‌رشته‌ای می‌گردد. (امیرحسینی، مازیار، ۲۰۰۸).

به دلایل مختلف، هنوز در همه زمینه‌ها همسان‌سازی و یکدست کردن تزاروس‌ها به وقوع نپیوسته است. در سال‌های اخیر روش‌های یکسان و استانداردها مورد توجه قرار گرفته‌اند. بسیاری از سازمان‌ها از وجود سرعنوان‌های موضوعی و اصطلاح‌نامه‌های مشابه و یا مرتبط موجود آگاهی ندارند. در صورت آگاهی از این مسئله گرایش به همکاری سازمانی تحقق خواهد داشت، زیرا با رشد روزافزون تکنولوژی نیاز به این ابزار نیز بیشتر شده است. اگر روش تدوین و ساخت اصطلاح‌نامه‌ها یکسان باشد، همکاری میان سازمان‌های بین‌المللی مانند یونسکو (سازمان آموزشی، علمی و فرهنگی ملل متحد)^۱ و سازمان همکاری‌های توسعه اقتصادی منجر به ایجاد ماکروتزاروس مانند: ا.ای.سی.دی.^۲ می‌شود.^۳

همان‌گونه که گفته شد «فرااصطلاح‌نامه» نوعی اصطلاح‌نامه برای اصطلاح‌نامه‌هاست که از این طریق می‌توان تعدادی از مجموعه واژگان کنترل شده را با یکدیگر تطابق داد و در جستجوی متقابل

^۱ United Nations Educational, Scientific and Cultural Organization (UNESCO)

^۲ Organisation for Economic Co-operation and Development (OECD)

^۳ <http://bibliotechavirtual.clasco.org.ar/ar/oecd-macroth/en/index.htm>

سازگاری ایجاد نمود (کمیت‌فرعی فراداده و تحلیل موضوعی ۱۹۹۹). فرااصطلاح‌نامه دربرگیرنده مفاهیم کلیدی یک دامنه موضوعی خاص و روابط معنایی میان مفاهیم مانند مترادف‌ها و متضادها، و نیز ایجاد ارتباطات معنایی میان تزاروس‌های متفاوت هستند. فرااصطلاح‌نامه‌ها توسط مهندسين سیستم‌های رایانه‌ای و نیز کتابداران و اطلاع‌رسانان مورد استفاده قرار می‌گیرد (یو.ام.ال.اس، ۲۰۰۴). بر اساس آنچه که گفته شد، شکل ۱۴-۲ مدلی است که در آن رابطه یک سری از مجموعه واژگان‌های کنترل شده، نشان داده می‌شود. این مدل شامل سه ماکروتزاروس می‌شود که هر یک از آن‌ها سه میکروتزاروس را دربرمی‌گیرد. در این مجموعه سه اصطلاح‌نامه کلان به شکلی با هم ادغام شده‌اند که یک متاتزاروس یا فرااصطلاح‌نامه را به وجود آورده‌اند. سرانجام یکی از ماکروتزاروس‌ها برای نمایه مستقل نام‌ها (اشخاص و اسامی جغرافیایی) و نیز پایگاه اسناد و مدارک موجود اختصاص یافتند (یانسی و کارسون، ۲۰۰۴).

تلفیق اصطلاح‌نامه‌های تخصصی متعدد ضرورت ایجاد میکروتزاروس را ایجاد می‌کند. رابطه میان تزاروس‌های تخصصی و اصطلاح‌نامه خرد و کلان به واسطه واژگان واسط و یا کدهای انتقالی (مقولات طبقه‌بندی) صورت می‌پذیرد و سرانجام اصطلاح‌نامه‌های کلان تلفیق و به صورت فرااصطلاح‌نامه (متاتزاروس) ایجاد می‌گردد.

نگاه تک‌بعدی به علوم در تقابل با تفکر چندبعدی و ایجاد وحدت میان علوم است و منتج به ایجاد میکرو، ماکرو و متاتزاروس می‌گردد. دیدگاه وحدت و همبستگی علوم سبب تولید تزاروس‌های مدرن و تلفیق و ادغام علوم و دسترسی آسان به اطلاعات علمی را در پی خواهد داشت. بازتاب تحول و ارتقاء در فرااصطلاح‌نامه‌ها در ساختار مفاهیم و عملکرد آن‌ها در ذخیره و بازیابی اطلاعات به وقوع می‌پیوندد.

فصل چهارم

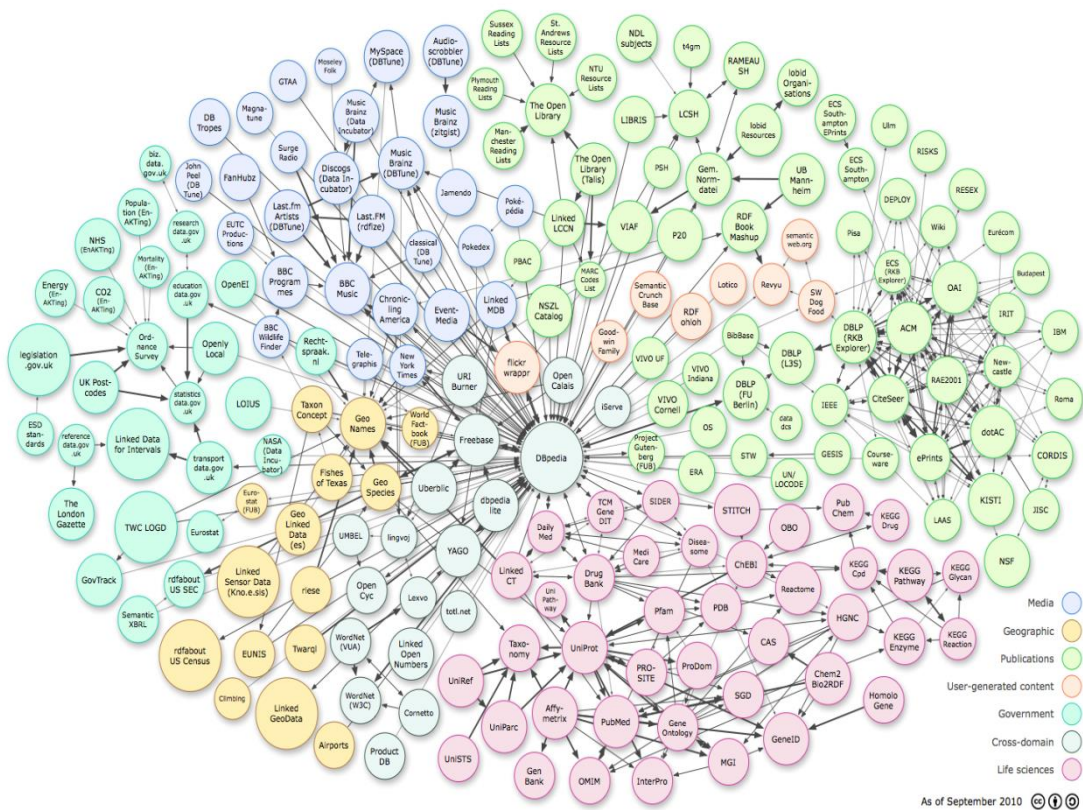
تأثیر آنتولوژی‌ها بر اصطلاح‌نامه و فرااصطلاح‌نامه

۴-۱. تأثیر آنتولوژی‌ها بر اصطلاح‌نامه و فرااصطلاح‌نامه

آنتولوژی به معنای هستی‌شناسی است، که در حوزه فلسفه مطرح است و به موضوع موجودیت، مطالعه هستی و ساختار واقعیت می‌پردازد. به عبارت دیگر، «هستی‌شناسی یک علم در فلسفه است که به مطالعه آنچه موجود است و آنچه باید فرض شود که وجود دارد؛ به منظور دستیابی به یک توصیف متقاعدکننده از واقعیت» می‌پردازد.

سال‌هاست که هستی‌شناسی‌ها به اشکال گوناگونی مانند طرح‌های رده‌بندی، فرااصطلاح‌نامه‌ها، واژگان کنترل‌شده، اصطلاح‌شناسی‌ها و واژه‌نامه‌ها معنا شده‌اند. اگرچه ممکن است هستی‌شناسی‌ها دارای چنین ویژگی‌هایی باشند، اما قطعاً نمی‌توان آن‌ها را مساوی با هر یک از این ساختارها دانست. تعریف «سوا»^۱ از هستی‌شناسی به این شرح است: «هستی‌شناسی دسته‌بندی از موضوعات یا عناصر موجود در یک حوزه خاص را بررسی می‌کند و بر اساس آن، فهرستی از موضوعات را ارائه می‌دهد. این فهرست که به تفصیل، انواع موضوعات و روابط میان آن‌ها را در حوزه مورد بررسی بیان می‌کند، هستی‌شناسی نامیده می‌شود».

¹ J.F. Sowa



شکل ۴-۱. هستی‌شناسی علوم پایه

لازم به ذکر است که یک هستی‌شناسی الزاماً شامل یک مجموعه واژه مرکب از اصطلاحات است و بعضی از مشخصات و معانی آن‌ها می‌باشد. بنابراین می‌توان هستی‌شناسی را - مانند اصطلاح‌نامه‌ها - مجموعه‌ای از لغات کنترل‌شده دانست، اما با این تفاوت که مفاهیم یک حوزه خاص با روابط معنایی خاص با توجه به نوع حوزه موضوعی، از موضوعی به موضوعی دیگر به شکل خاصی طبقه‌بندی می‌شوند که در قالب زبان طبیعی به نمایش درمی‌آید.

از هستی‌شناسی در حوزه‌های مختلفی از جمله هوش مصنوعی، مهندسی نرم‌افزار و علوم اطلاعات استفاده شده است. در حوزه علوم اطلاعات تعاریف زیادی از هستی‌شناسی ارائه شده است اما کمتر می‌توان تعریف یکسانی از آن به‌دست آورد. از هستی‌شناسی‌ها بر روی وب نیز استفاده می‌شود که به بازیابی متون بر روی وب به کاربران کمک می‌کند. «سورگل» با توجه به این نکته هستی‌شناسی را یک ابزار سازماندهی منابع وب که به بازیابی متون با زبان طبیعی کمک می‌کند، توصیف کرده است.

محیط وب دستیابی گسترده به اطلاعات را فراهم کرده است. این محیط اطلاعاتی جدید و تغییر توانایی‌های کاربران و نیازهای اطلاعاتی متفاوت، نیاز به ابزارهای یاری‌دهنده در امر بازیابی و سازماندهی اطلاعات رو به رشد را تشدید کرده است. یکی از این ابزارها هسته‌شناسی‌ها هستند.

در محیط وب هستی‌شناسی ابزار فشرده و منظمی برای تعریف معانی منابع وب است. هستی‌شناسی، مفاهیم حوزه مرتبط، خصیصه‌های آن‌ها، و روابط ممکن میان مفاهیم و خصیصه‌ها را مشخص می‌کند.

یک اصطلاح‌نامه نشان می‌دهد که چه اصطلاحاتی در یک پایگاه داده خاص وجود دارد، اما هستی‌شناسی‌ها، روابط موجود در اصطلاح‌نامه‌ها را حمایت می‌کنند و با فراهم‌آوری روابط مفهومی میان موجودیت‌ها به‌طور دقیق‌تر ارتباطات معنایی را نیز ارائه می‌کند و به کاربر در این شناخت کمک می‌کنند. بنابراین هستی‌شناسی‌ها ترکیبی از لغات کنترل‌شده محدود و در عین حال قابل گسترش، تفسیر بدون ابهام از رده‌ها و روابط میان اصطلاحات، روابط فرعی سلسله‌مراتبی صریح و وجود یک سری اصول منطقی برای ایجاد روابط توسط ماشین می‌باشند.

فرااصطلاح‌نامه و هستی‌شناسی، هر دو شبکه‌ای از اصطلاحات و نمایشی از روابط میان آن‌ها جهت ارائه مفاهیم می‌باشند که بر پایه مفاهیم برای نمایش دانش پیچیده موضوعی بنا می‌شوند و در سازماندهی منابع اطلاعاتی و بازیابی اطلاعات کاربرد مؤثری دارند. اما می‌توان تأثیر هستی‌شناسی‌ها را در اصطلاح‌نامه‌ها و در پی آن در فرااصطلاح‌نامه این‌گونه دانست که:

۱. اصطلاح‌نامه به ارائه واژگان و روابط میان آن‌ها برای کمک به کاربران در درک مفهوم یک اصطلاح می‌پردازند، اما نوع اصطلاحات را مشخص نمی‌کند. اما با کمک هستی‌شناسی به‌طور واضح، دانش مفهومی با استفاده از زبانی قراردادی براساس یک حوزه خاص همراه با معانی دقیق آن‌ها ارائه می‌شود؛
۲. اصطلاح‌نامه‌ها سیاهه‌ای از واژگان برای منابع موجود است که بیشتر درباره مدارک موجود و برای استفاده در مورد نوع خاصی از مدارک است. اما در هستی‌شناسی‌ها به لحاظ الگوهای وضع‌شده، اعمال یک کنترل واژگان یکنواخت به آسانی امکان‌پذیر است؛
۳. اصطلاح‌نامه برای کاربران انسانی معنادار و قابل درک است، در حالی که استفاده از هستی‌شناسی برای پردازش اطلاعات توسط انسان و ماشین کاربرد دارد؛
۴. ابهام‌های معنایی موجود در اصطلاح‌نامه‌ها که ناشی از روابط تعریف‌شده محدود هستند با استفاده از هستی‌شناسی‌ها برطرف شده و گسترش می‌یابد؛
۵. با کمک هستی‌شناسی‌ها می‌توان اصطلاح‌نامه‌ها و فرااصطلاح‌نامه‌ها را بر روی وب منتقل کرد؛ زیرا زبان هستی‌شناسی برای پردازش دانش قابل درک است و از زبان تعریف‌شده در

قالب استاندارد های آر.دی.اف، اُ.دبلیو.ال^۱، اُ.ال.آی^۲ و دی.ای.ام^۳ استفاده می‌کند که امکان ویرایش برای روزآمدسازی و استفاده مجدد از دانش گردآمده را به آسانی امکان‌پذیر می‌کند.

۴-۲. وب معنایی

۴-۲-۱. مقدمه

پیش از اختراع وب اطلاعات روی رایانه‌های مختلف ذخیره می‌شد. برای بازیابی اطلاعات نیاز به یک واسط انسانی و مستلزم آشنایی کار با تک‌تک رایانه‌ها بود. با ایجاد وب رایانه‌های مختلف به هم پیوند خوردند و دنیایی به هم متصل را به وجود آوردند. که هر فرد در هر زمان می‌توانست به اطلاعات دسترسی داشته باشد. یکی از ابزارهای اصلی بازیابی اطلاعات در بین مردم جست‌وجوی کلیدواژه‌ای در موتورهای جست‌وجو هست. مشکلات عمده موتورهای جست‌وجوی مبتنی بر کلیدواژه غیرقابل فهم و پردازش بودن ماشین‌هاست. به طوری که تشخیص اختلاف معنای دو جمله «من دانشجو هستم» یا «شما ممکن است فکر کنید من دانشجو هستم» در موتورهای جست‌وجوی کلیدواژه‌ای ممکن نیست.

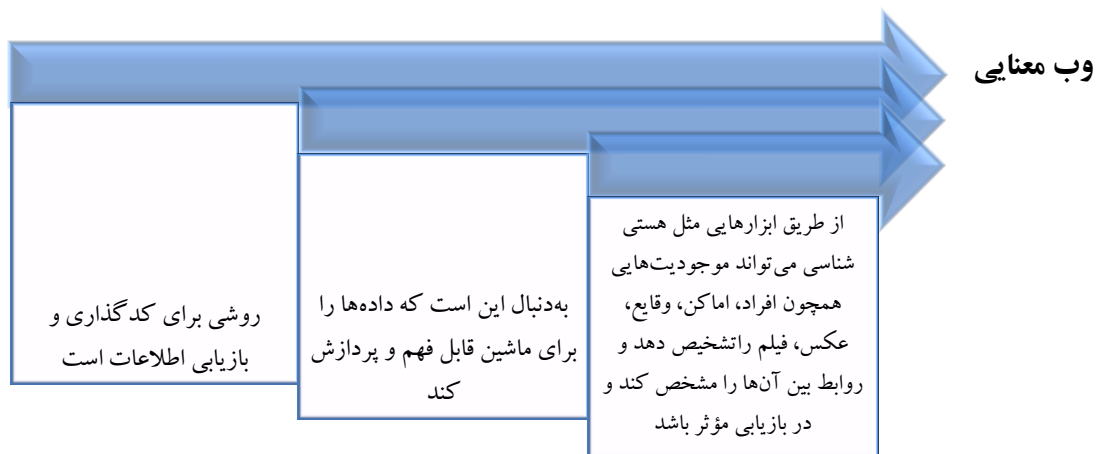


¹ Online Writing Lab (OWL)

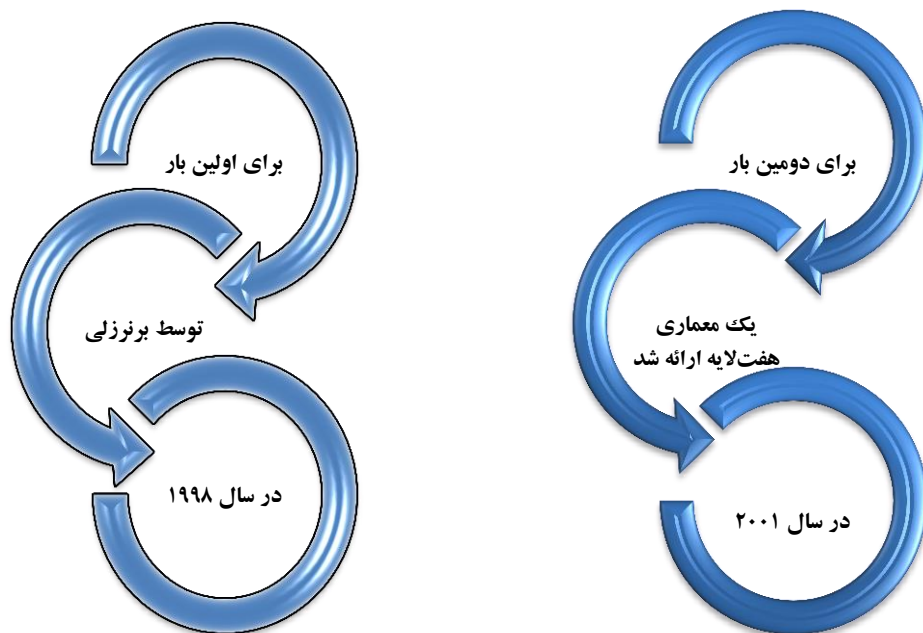
² OLI

³ Digital Asset Management (DAM)

۴-۲-۲. تعریف



۴-۲-۲-۱. معرفی وب معنایی



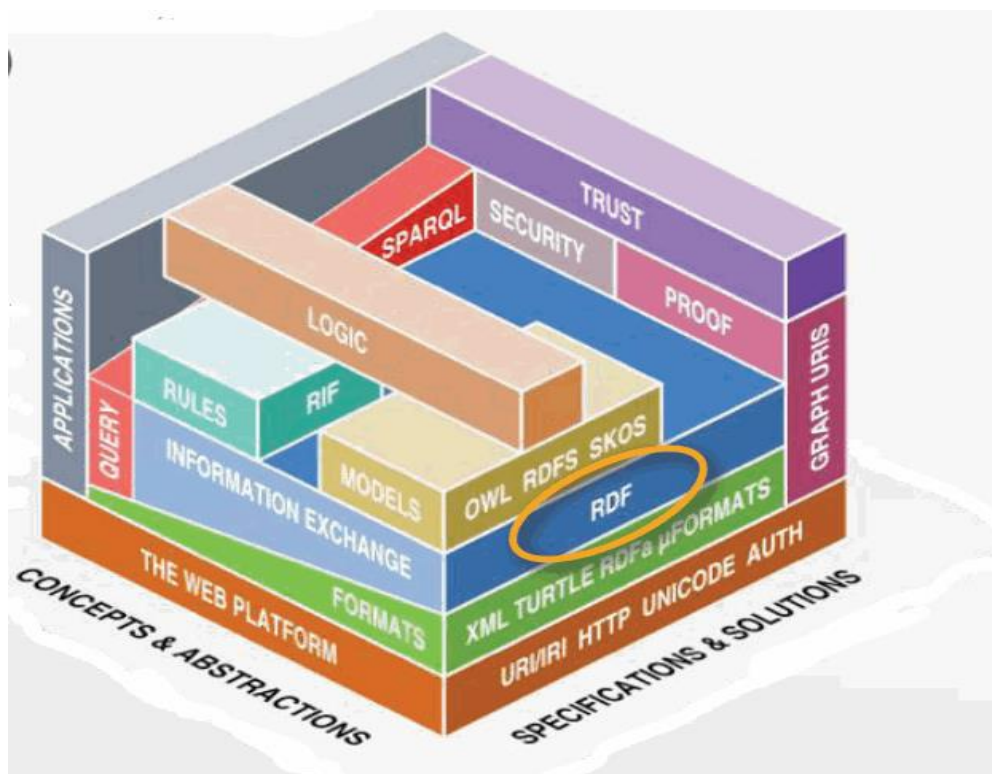
۴-۲-۳. وب معنایی فراتر از کلمات

در وب معنایی جست‌وجو متکی به کلمات نخواهد بود مثلاً می‌توان به یاری صوت به دنبال صدایی مشابه گشت

۴-۲-۴. اهداف وب معنایی

- هدف از توسعه وب معنایی ساختارمند کردن داده‌ها اضافه کردن معانی آن‌ها و در نهایت بازنمایی دانش به کمک ماشین با استفاده از فناوری‌ها و استانداردهایی است که به وسیله کنسرسیوم وب جهان گستر در حال تدوین و تکمیل هستند؛
- به عبارت دیگر در وب معنایی ماشین‌ها با درک داده‌ها در فرایند تحلیل و استخراج اطلاعات مورد نظر کاربر مشارکت کرده و در این زمینه هوشمندانه‌تر عمل می‌کنند؛
- در وب معنایی قرار است ماشین‌ها (روبات‌ها، سرورها و کامپیوترها) توانایی فهم محتویات اینترنت را داشته باشند. در این ساختار ماشین‌ها باید بتوانند با هم ارتباط برقرار کنند نه فقط با انسان؛
- واضح‌تر این که وقتی یک ربات وب‌کاو جمله‌ای را بخواند باید بتواند از نظر دستور زبان آن جمله را بفهمد یعنی این که بتواند فاعل، مفعول و گزاره‌ای که در آن جمله آمده را درک کند و اگر یک پاراگراف را بخواند، باید بتواند نوع وابستگی جمله‌ها را با یکدیگر را تشخیص دهد. وب معنایی در حقیقت وبی است که در بخش طبقه‌بندی و بازیابی محتوای اینترنت به تکامل رسیده باشد؛
- در ساختار معنایی وب وقتی کاربر یک جمله محاوره‌ای را جستجو می‌کند، موتور جستجو باید بتواند مفهوم جمله را تشخیص دهد، این که سؤالی است یا خبری، طنز است یا جدی، علمی است یا از نوع سرگرمی. بنابراین وب جدید باید چنان در زبان‌شناسی و فلسفه تجهیز شود که تمام محتویات وب را بتواند آنالیز کرده و میزان وابستگی بخش‌ها را از طریق رسم گراف‌ها تعریف کند.

۴-۲-۵. معماری وب معنایی^۱

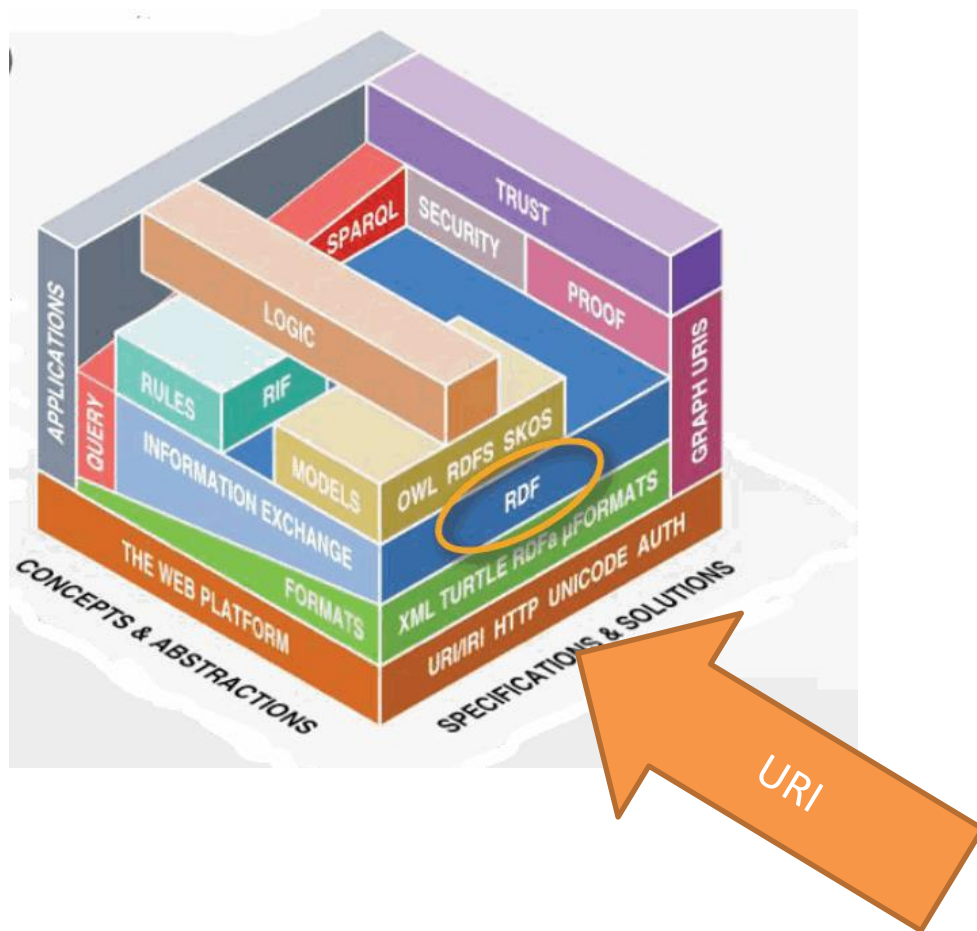


لایه اول:

- این لایه یونی‌کد و یوآرای همین صفحات معمولی وب است؛
- با زبان اچ.تی.ام.ال^۲ نوشته شده است؛
- بنابراین وب معنایی بر روی وب کنونی پیاده می‌شود و چیزی مستقل از آن نیست؛
- برای این که شیء در وب معنایی قابل شناسایی باشد باید دارای یوآرای منحصر به فرد باشد.

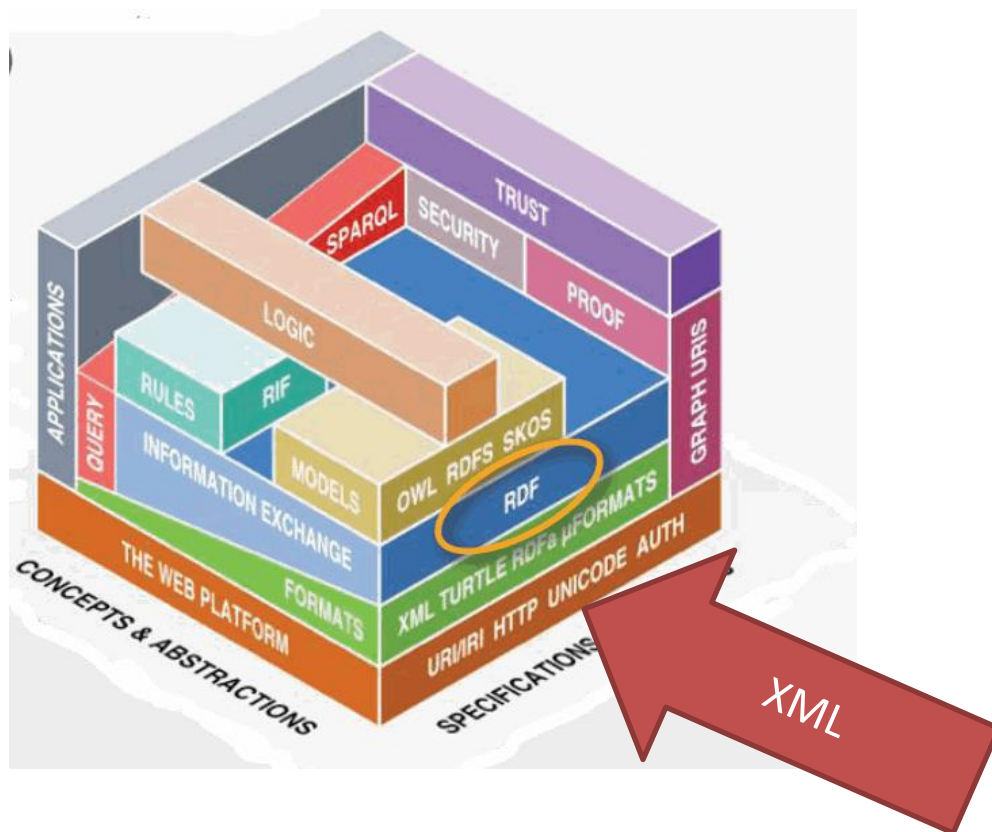
^۱ Semantic Web Architecture

^۲ HTML



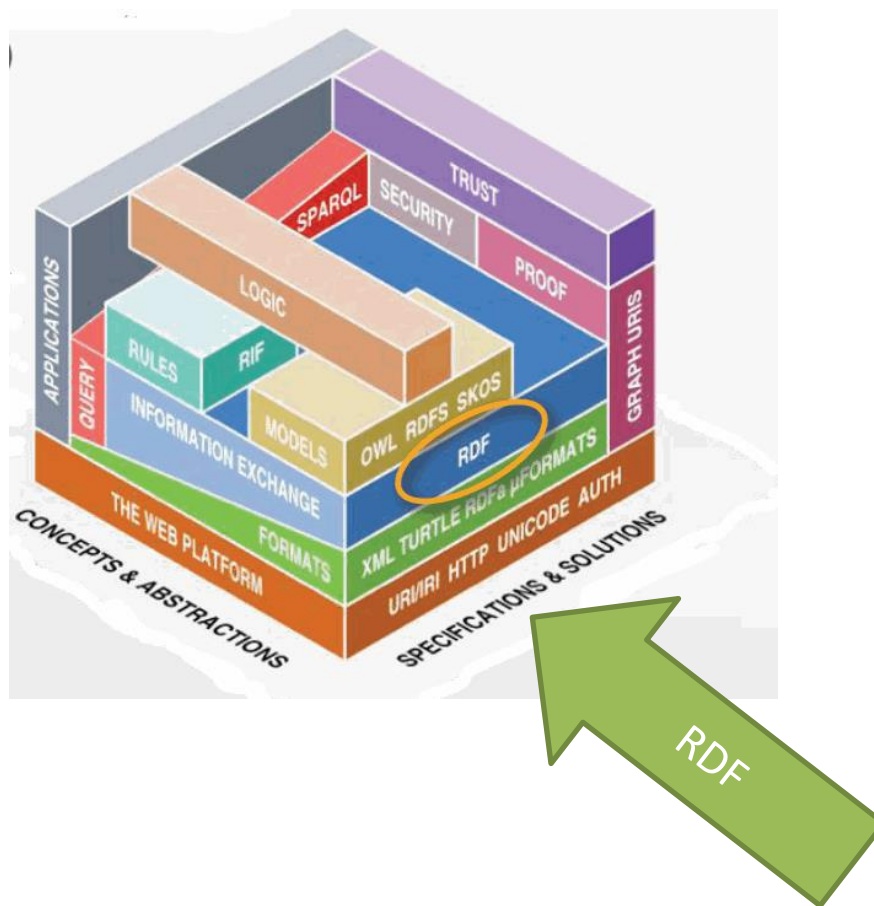
لایه دوم:

- زبان XML است؛
- در قالب نحو به گونه‌ای طراحی شده که به راحتی قابل دسترس باشد؛
- به وسیله این زبان می‌توان اطلاعاتی در رابطه با داده‌ها به آن‌ها ضمیمه کرد.



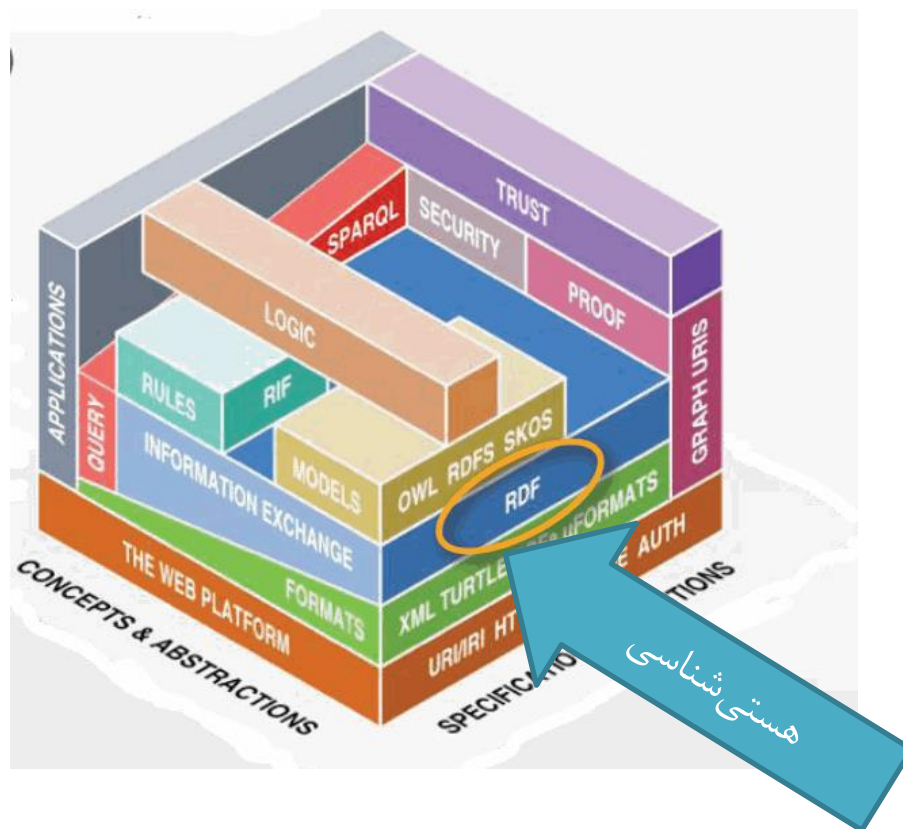
لایه سوم:

- آردی اف یا چارچوب توصیف منبع؛
- زبان علامت‌گذاری شده فرامتنی که در وب جهان‌گستر استفاده می‌شود؛
- RDF نمونه عمومی چارچوبی است که برای توصیف منبع به‌ویژه توسط ابرداده‌ها به کار می‌رود.



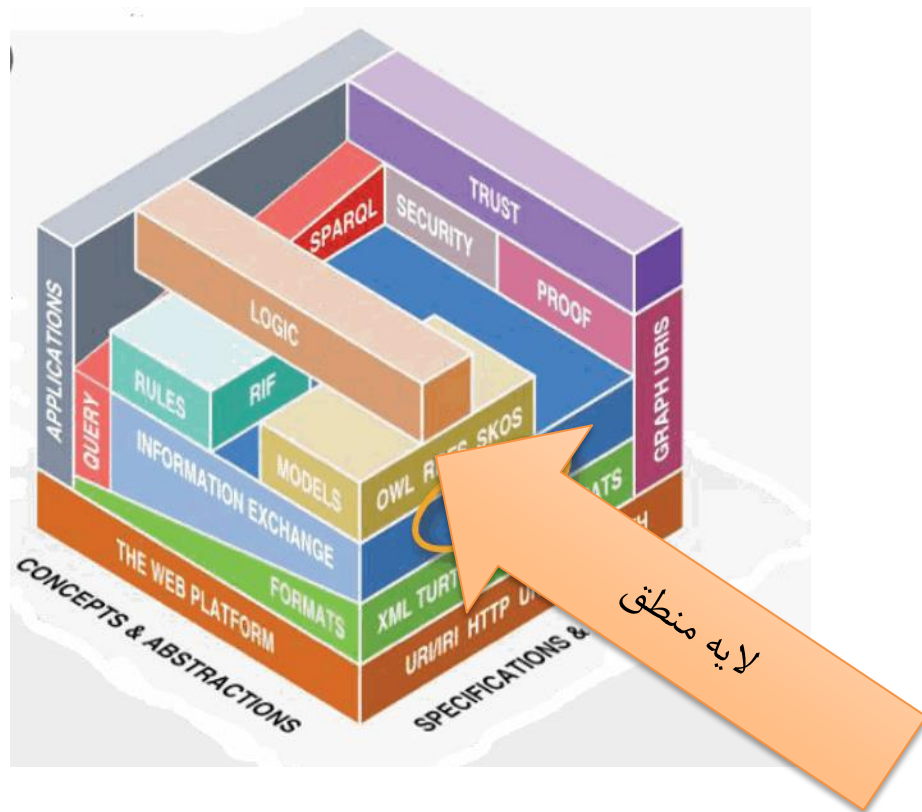
لایه چهارم:

- لایه هستی‌شناسی یا لایه ارتباط بین معانی است؛
- ارتباط بین معانی توسط زبان OWL صورت می‌گیرد و امکان استنتاج برقرار می‌شود؛
- به عبارت دیگر هستی‌شناسی، هم‌زمانی مشترک بین صفحات وب ایجاد می‌کند.



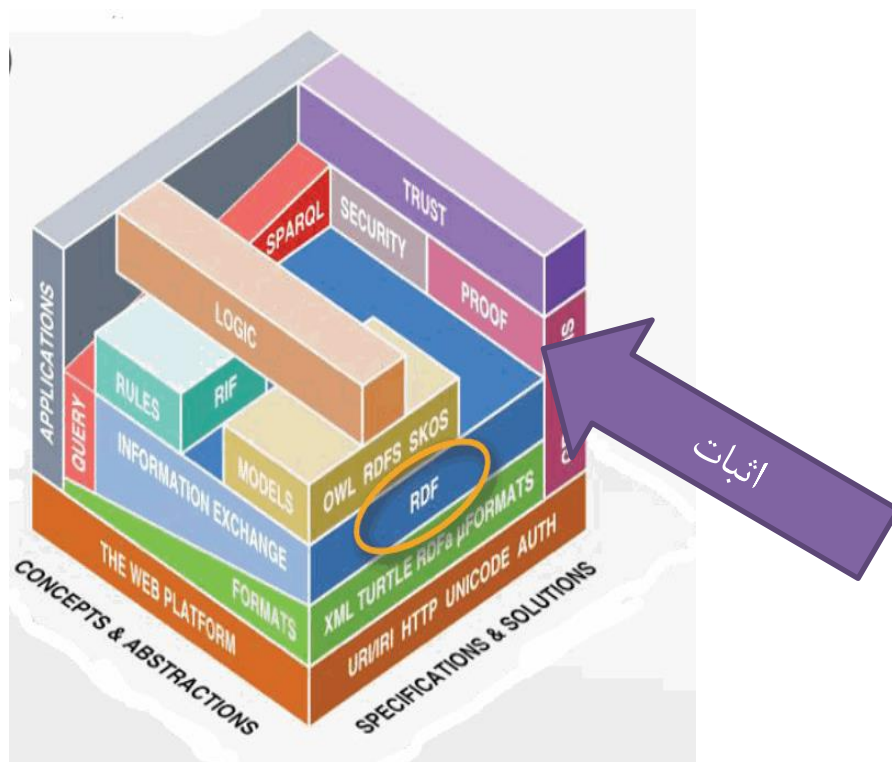
لایه پنجم:

- لایه منطق برای تقویت بیشتر زبان‌های هستی‌شناسی به کار می‌رود؛
- هدف از کاربرد این لایه ایجاد محیط و چارچوب به‌منظور نوشتن اصول و قواعد سیستم‌ها؛
- بر مبنای قانون با منطقی یکنواخت است.



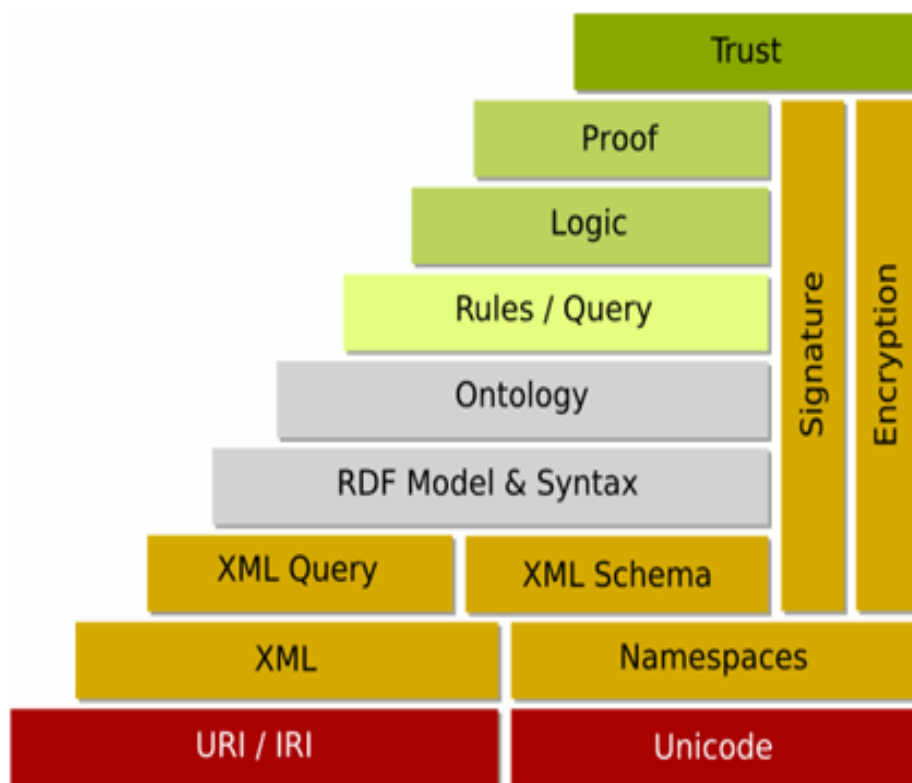
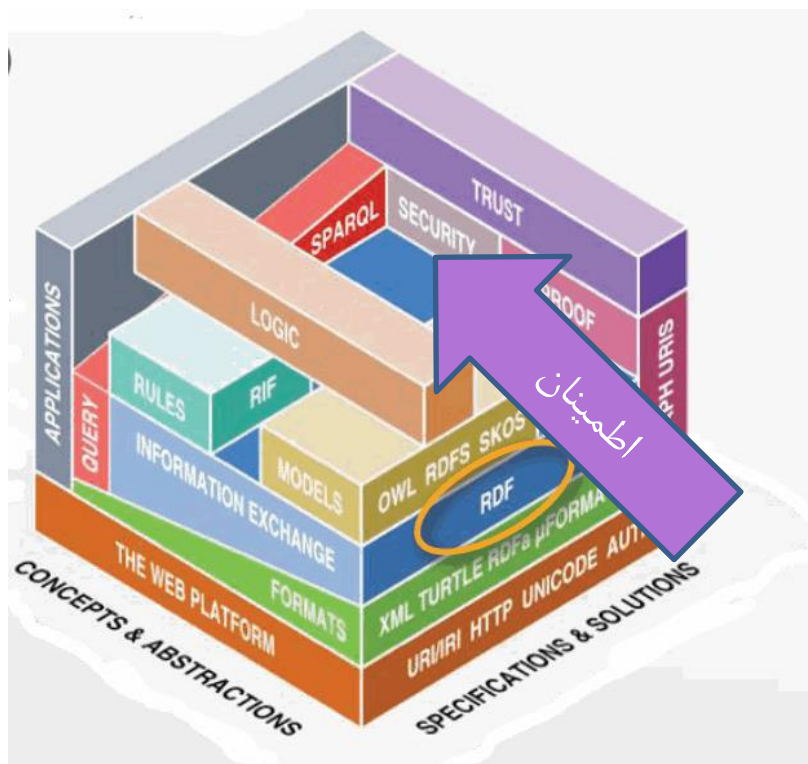
لایه ششم:

- لایه اثبات است؛
- مسئول فرایند واقعی استنتاج و ارائه اثبات به زبان‌های وب از لایه‌های پایین‌تر و اثبات اعتبار است.

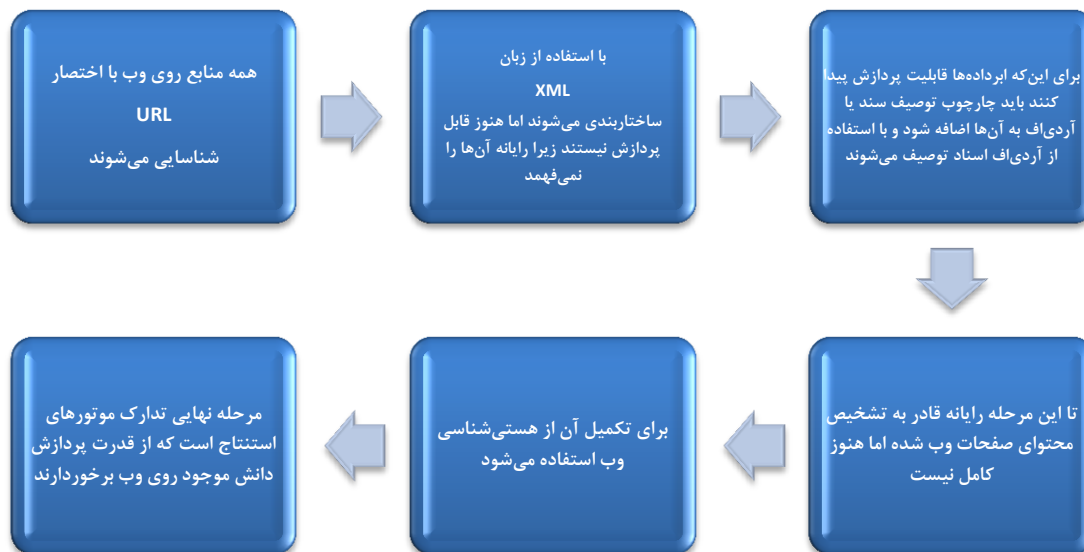


لایه هفتم:

- لایه اطمینان، در حقیقت برای کسب اطمینان از صحت اطلاعات وب معنایی و تکمیل‌کننده؛
- ارکان اصلی این ساختار است؛
- لایه اطمینان هدف نهایی وب معنایی است؛
- هر عبارت یا مفهومی در وب معنایی مرتبط به زمینه و موضوع خاص حوزه مفهومی است.



۴-۲-۶. مروری مختصر بر عملکرد وب معنایی



۴-۲-۷. نقش هستی‌شناسی در وب معنایی

□ هستی‌شناسی‌ها در حقیقت شالوده‌ی وب معنایی هستند و وب معنایی فقط با تعریف و استفاده از آن‌هاست که می‌تواند به سرانجام مناسب برسد. باید در وب معنایی تدابیری اتخاذ گردد که معنای اطلاعات پراکنده در وب توسط ماشین‌ها نیز قابل فهم باشد. یکی از مهم‌ترین تدابیر، تعریف و استفاده مناسب از هستی‌شناسی‌هاست. یک هستی‌شناسی با مشخص کردن مفاهیم و موجودات یک دامنه و توصیف روابط بین آن‌ها به صورت رسمی و قابل فهم توسط ماشین و با به خدمت گرفتن مجموعه‌ای از قواعد، در حقیقت دانش موجود در آن دامنه را به صورت رسمی بیان می‌کند. با توصیف رسمی دانش به کمک هستی‌شناسی‌ها، عامل‌های نرم‌افزاری یا به طور کلی ماشین‌ها در وب معنایی قادر می‌شوند از آن دانش به نحو مطلوب استفاده کنند.

می‌توان نقش هستی‌شناسی‌ها را در وب معنایی چنین بیان کرد:

- هستی‌شناسی‌ها توانایی پردازش دانش، به اشتراک گذاری آن بین عامل‌های مختلف و استفاده مجدد از آن را در محیط وب ایجاد می‌کنند. هستی‌شناسی‌ها چنین توانایی‌هایی را مدیون تعریف و توصیف مفاهیم و موجودات و بیان ارتباط بین آن‌ها به صورت رسمی، می‌باشند؛
- با تجزیه و تحلیل محتوای صفحات وب، می‌توان داده‌های موجود در صفحات را به مفاهیم موجود در هستی‌شناسی‌ها نگاشت کرد. چنین عملی باعث بالا بردن درجه همکاری بین انسان و ماشین شده به گونه‌ای که ماشین‌ها مستعد به عهده گرفتن بسیاری از وظایف انسان می‌شوند؛

□ هستی‌شناسی‌ها برای فعالیت سرویس‌های هوشمند، بستر مناسبی فراهم می‌کنند. به عنوان مثال در سرویس‌هایی مانند جستجو هوشمند در وب، فیلترهای اطلاعاتی، یکپارچه‌سازی هوشمند اطلاعات و مدیریت دانش، استفاده از هستی‌شناسی‌ها می‌تواند بسیار کارساز باشد.

۴-۲-۸. اُ.دبلیو.ال^۱ زبان بازنمون هستی‌شناسی

□ پیش‌نیاز کاربرد هستی‌شناسی‌ها در وب معنایی، توسعه استاندارد برای تعریف و مبادله هستی‌شناسی یا به عبارتی زبان‌های بازنمون هستی‌شناسی است. در این راستا کنسرسیوم وب جهانی، زبان چارچوب توصیف منابع یا آر.دی.اف^۲ را توسعه داد. آر.دی.اف، زبانی برای کدگذاری دانش موجود در صفحات وب، به منظور قابل فهم کردن این دانش برای عامل‌های الکترونیکی جستجوگر اطلاعات بود؛

□ دی.ای.آر.پی.ای^۳ نیز با همکاری کنسرسیوم وب جهانی، زبان دی.ای.ام.ال^۴ را ایجاد کرد. زبان دی.ای.ام.ال گسترشی از آر.دی.اف^۵ بود که در آن ساخت‌های توصیفی بیشتری استفاده می‌شد. هدف این ساخت‌ها تسهیل تعامل عامل‌ها در وب بود. سرانجام زبان هستی‌شناسی وب یا اُ.دبلیو.ال بر مبنای زبان دی.ای.ام.ال ایجاد شد. از میان زبان‌های مطرح‌شده، اُ.دبلیو.ال جدیدترین استاندارد زبانی توصیه شده است و قابلیت‌های بیشتری نیز نسبت به سایر زبان‌ها دارد. به کارگیری زبان‌های بازنمون دانش، امکان تعریف، برقراری ارتباط میان هستی‌شناسی‌ها و تحقق میان‌کنش‌پذیری و سازگاری را فراهم می‌کند.

□ وب معنایی بسط و توسعه وب جاری است که در آن معنی و منظور اطلاعات به خوبی تعریف و بیان شده است به گونه‌ای که کامپیوتر و انسان را قادر می‌سازد تا بهتر با هم مشارکت و تعامل داشته و ارتباط برقرار کنند.

اساس وب معنایی تأکید بر این نکته است که داده‌های موجود در وب قابل تفسیر و ترجمه به وسیله ماشین باشد به گونه‌ای که فراداده‌ها بهتر قادر به خودکارسازی، ترکیب، یکپارچه‌سازی، کشف و استفاده مجدد برای کاربردهای گوناگون خود و کاربران باشد.

¹ OWL

² RDF

³ Defense Advanced Research Projects Agency (DARPA)

⁴ DAML

⁵ RDF

فصل پنجم

جمع بندی

۵-۱. جمع‌بندی

در دهه‌های اخیر بازنمایی و بازیابی دانش در حوزه مستندسازی از فناوری‌های مهمی چون اینترنت و فناوری‌های وابسته به آن تأثیر فراوانی یافته است. این فناوری‌های جدید و تغییر و تحولات مهم در حوزه بازنمایی اطلاعات دیجیتالی و بازیابی آن‌ها و تأثیرات عوامل مختلف در حوزه اطلاع‌رسانی و ابزارهای جدید بازنمایی و بازیابی، بر کاربران اطلاعاتی نیز اثر گذاشته است.

در سایه این رشد روزافزون متون دیجیتالی، ابزارهای جدید بازنمایی دانش نیز تغییر یافته، به شکلی که رشته‌های دیگر علمی نیز در این عرصه سهیم شده‌اند و در نتیجه رشته‌هایی مانند زبان‌شناسی، هوش مصنوعی، مستندسازی، مهندسی زبان و غیره نیز باید مورد مطالعه قرار گیرند. موضوعات خاص مانند تاکسونومی‌ها یا سیستم‌های رده‌بندی، مجموعه واژگان رایانه‌ای، پایگاه‌های واژگانی، اصطلاح‌نامه‌ها، فرااصطلاح‌نامه، نقشه مفهومی، شبکه معنایی واژگان و هستی‌شناسی‌ها به عنوان ابزارهای بازیابی و بازنمایی دانش به‌ویژه فرااصطلاح‌نامه و هستی‌شناسی که از مهم‌ترین ابزارهای سازماندهی اطلاعات در محیط وب معنایی هستند، مورد توجه قرار گرفته‌اند.

فرااصطلاح‌نامه ساختار معناشناختی را ارائه می‌نماید که عمدتاً با روابط حاکم میان مفاهیم و در نهایت اصطلاحات که برچسب مفاهیم هستند، بیان می‌شوند و در حوزه مستندسازی و طبقه‌بندی اطلاعات ابزاری است برای نظارت و ارزیابی اصطلاحات به کار گرفته شده، زیرا ساختار فرااصطلاح‌نامه بر اساس مفاهیم استوار است. از نظر ساختاری، گسترش کارکرد آن بر اساس محیط فرامتنی است که منجر به ایجاد و ساخت پیوندهای فرامتنی در میان تمامی عناصر ساختاری مانند توصیفگرها، غیرتوصیفگرها، یادداشت دامنه‌ها و ارجاعات می‌گردد.

در سال‌های گذشته بسیاری از مجموعه واژگان‌ها و اصطلاح‌نامه‌های تخصصی انتشار یافته‌اند که مستقل از یکدیگر بوده‌اند و بعضاً در موضوعی واحد به صورت اصطلاح‌نامه‌های مختلف می‌باشند، به لحاظ ساختار، دیدگاه و تخصصی بودن با یکدیگر تفاوت دارند. این فقدان سازگاری انتقال رکورد بین سامانه‌ها و کاوش در پایگاه‌های اطلاعاتی مختلف را با مشکل مواجه می‌سازد. انطباق یا تلفیق

اصطلاح‌نامه‌ها راهی برای غلبه بر ناسازگاری آن‌ها می‌باشد. مطابقت دادن، اصطلاحات را مستقیماً از یک زبان کنترل‌شده به زبانی دیگر یا به مجموعه واژگان چندگانه پیوند می‌دهد و اصطلاحات یک زبان کنترل‌شده به اصطلاحات معادل در زبان دیگر اشاره می‌کنند و این اطلاعات بیان می‌کند که آیا اصطلاح مشابه، مترادف و یا دارای رابطه‌ای دیگر با آن اصطلاح در زبان‌های دیگر است. در جهت گردانی خودکار، اصطلاحات یک واژگان به طور خودکار به اصطلاحات واژگان دیگری تغییر شکل می‌یابد که بر مبنای انطباق زبان‌های میانجی با الگوریتمی برای کنترل اصطلاح‌های دروندادی استوار است.

در مرحله تدوین فرااصطلاح‌نامه، برای استفاده از چند اصطلاح‌نامه خرد و کلان به صورت یک فرااصطلاح‌نامه و تبدیل آن به هستی‌شناسی، ابتدا کلیه مفاهیم مشترک در اصطلاح‌نامه‌ها همسان می‌شوند. در یک مجموعه اصطلاح‌نامه با ده‌ها هزار واژه و مفهوم که در زمان‌های مختلف و توسط متخصصین موضوعی مختلف تولید شده‌اند، برای همسان‌سازی باید از توانایی‌هایی که سیستم‌های رایانه‌ای در اختیارمان می‌گذارد استفاده نماییم تا در صرف زمان و نیروی متخصص و نیز بودجه صرفه‌جویی شده و فرایند همسان‌سازی نیز در حد مطلوب انجام پذیرد.

به منظور بازیابی در پایگاه‌های متعدد و چندگانه که به واسطه واژگان کنترل‌شده متعدد نمایه شده‌اند، فرااصطلاح‌نامه ایجاد می‌شود. هر اصطلاح در فرااصطلاح‌نامه با برابرنهاده‌ها و اصطلاحات مرتبط از تمامی اصطلاح‌نامه‌های تلفیق شده در آن به نمایش در می‌آیند. فرایند خودکار در جستجو و بازیابی، توسعه می‌یابد و کاربر را قادر می‌سازد بازیابی را در نمایه‌های منفرد یا چندگانه و تلفیق شده و با استفاده از اصطلاحات هر یک از منابع واژگانی، انجام دهد. فرااصطلاح‌نامه مفهوم بنیان است و هدف آن ایجاد وحدت میان نامگذاری‌هایی که از منابع مختلف و رده‌بندی‌ها و اطلاعات پیرامون آن به دست آمده است. گسترش سریع میزان واژگان یک فرااصطلاح‌نامه باعث شده است که به کارگیری رایانه برای سهول در ایجاد آن مؤثر واقع شود. روزآمدی واژگان، بهره‌مندی از مجموعه واژه‌های متعدد، الگوریتم گزینش و بازنگری متخصصین موضوعی را می‌طلبد.

بر این اساس فرااصطلاح‌نامه تنها یک مجموعه واژگان استاندارد نیست، بلکه همچنین ابزاری برای افزایش نهایی واژگان موجود و مفید است و منبعی غنی برای متخصصین اطلاعات علمی می‌باشد. نظارت بر تناسب اجزای فرااصطلاح‌نامه در هر ویرایش جدید، سبب توسعه و افزایش میزان مفاهیم یک فرااصطلاح‌نامه می‌گردد. فناوری این ابزار برای فرمول‌بندی روابط معنایی میان ساختارهای دانش فنی به صورت روابط سلسله‌مراتبی با دیدگاه‌های گوناگون، کاربران را قادر به تورق، جستجو و بازیابی میان‌رشته‌ای در گرایش‌های گوناگون علمی می‌سازد.

با وجود این که سازمان‌های بسیاری در حوزه‌های گوناگون تخصصی ضرورت ایجاد اصطلاح‌نامه‌های تخصصی را دریافته و اقدام به تدوین اصطلاح‌نامه‌های متعددی نموده‌اند، لکن در زمینه تلفیق اصطلاح‌نامه‌های گوناگون و ایجاد فرااصطلاح‌نامه تلاشی صورت نگرفته است. هدف از همسان‌سازی واژگان اصطلاح‌نامه‌ها و تلفیق به شکل فرااصطلاح‌نامه مبتنی بر مفهوم، یک پارچه‌سازی علوم است، بدین معنا که تمامی رشته‌های علمی متعلق به یک ایده واحد است. تاکنون تلاش‌های فراوان علمی عامل مؤثری در ایجاد و مدیریت اصطلاح‌نامه به شکل خرد، کلان و فرااصطلاح‌نامه شده است. تفکر وحدت در علوم و رابطه سلسله‌مراتبی آن‌ها نشانگر تفکر چندبعدی و چندمنظوره است که منجر به تلفیق و ایجاد ارتباط میان رشته‌های گوناگون و سرانجام تحقیقات میان‌رشته‌ای می‌گردد. به دلایل مختلف یکسان‌سازی تزاروس‌ها هنوز به وقوع نپیوسته است. اما در سال‌های اخیر روش‌های یکسان و رعایت و به‌کارگیری استانداردها مورد توجه قرار گرفته‌اند. بسیاری از سازمان‌ها که از وجود اصطلاح‌نامه‌های مشابه و یا مرتبط موجود آگاهی نداشته‌اند. اما آگاهی از این امر، گرایش همکاری بین سازمانی افزایش می‌دهد، زیرا با رشد روزافزون تکنولوژی نیاز به این ابزار نیز بیشتر خواهد شد. چنانچه روش تدوین اصطلاح‌نامه‌ها یکسان شود و فرااصطلاح‌نامه ایجاد گردد، همکاری سازمان‌های تحقیقاتی داخل کشور با سازمان‌های بین‌المللی مانند یونسکو و «سازمان توسعه همکاریهای اقتصادی» و غیره، ایجاد فرااصطلاح‌نامه به زبان فارسی را نیز میسر خواهد نمود.

فهرست منابع

- ابراهیمی جعفری، پیروز، ۱۳۸۶، وب ۳، کتابداری، آرشیو و نسخه‌پژوهی: اطلاع‌یابی و اطلاع‌رسانی: بهمن ۱۳۸۶ - شماره ۴.
- ایچینسن، جین و همکاران، ۱۳۸۵، «تدوین و کاربرد اصطلاح‌نامه» ترجمه: محسن عزیزی، انتشارات پژوهشگاه اطلاعات و مدارک علمی.
- جمالی مهموئی، حمیدرضا، ۱۳۸۲، وب معنایی: شیوه‌ای رو به تکامل برای ذخیره و بازیابی کارآمدتر اطلاعات روی اینترنت، کتابداری، آرشیو و نسخه‌پژوهی: اطلاع‌شناسی: زمستان ۱۳۸۲ - شماره ۲.
- حسینی بهشتی، ملوک‌السادات، ۱۳۹۲، «ساختواره، اصطلاح‌شناسی و مهندسی دانش» تهران، پژوهشگاه علوم و فناوری اطلاعات ایران، انتشارات چاپار.
- صفری، مهدی، ۱۳۸۷، ابر داده و کاربرد آن در سازماندهی منابع وب با تأکید بر عناصر ابر داده‌ای دوبلین کور در بستر نحوی HTML. فصلنامه کتاب ۵۵.
- صنعت‌جو، اعظم، ۱۳۸۴، «ضرورت بازنگری در ساختار اصطلاح‌نامه‌ها: بررسی عدم کارایی اصطلاح‌نامه‌ها در محیط اطلاعاتی جدید و قابلیت‌های هستی‌شناسی‌ها در مقایسه با آن» فصلنامه کتاب. ۶۴. زمستان ۱۳۸۴: صص ۸۰-۹۲.
- کفاشان، مجتبی؛ فتاحی، رحمت‌الله، ۱۳۹۰، نظام‌های نوین سازماندهی دانش وب معنایی هستی‌شناسی و ابزارهای سازماندهی دانش عینی، کتابداری، آرشیو و نسخه‌پژوهی: کتابداری و اطلاع‌رسانی: تابستان ۱۳۹۰ - شماره ۵۴ (علمی-پژوهشی / ISC).
- گاردنر، ریچارد، ۱۳۹۶، «فراداده، شکل‌دادن به دانش از عهد باستان تا وب معنایی» مترجمان: صفیه طهماسبی لیمونی، فتنه ترکاشوند، محمد کاردان نشاطی. انتشارات چاپار.
- محمداسماعیل، صدیقه، لاله صمدی، ۱۳۸۵، «فرااصطلاح‌نامه‌ها (متاتزاروس‌ها): معرفی، ضرورت و فلسفه وجودی». کتاب ماه کلیات.
- محمدی استانی، مرتضی، ۱۳۹۵، وب معنایی و کاربرد آن در مدیریت دانش، کتابداری، آرشیو و نسخه‌پژوهی: مدیریت اطلاعات و دانش‌شناسی: پاییز ۱۳۹۵، سال سوم - شماره ۳.
- محمدی، فخرالسادات، سیروس علیدوستی، ۱۳۸۶، «فرایند تدوین اصطلاح‌نامه» پژوهشگاه اطلاعات و مدارک علمی ایران.
- مرادی، خدیجه، ۱۳۹۵، «فضای مفهومی» رویکردی جهت توسعه وب معنایی، کتابداری، آرشیو و نسخه‌پژوهی:

مطالعات دانش‌شناسی: زمستان ۱۳۹۵ – شماره ۹.

- موسوی چلک، افشین، ۱۳۸۶، استانداردهای کتابخانه دیجیتال. کتاب ماه کلیات، مرداد و شهریور.
- مهراد، جعفر، محمدرضا فلاحتی فومنی، ۱۳۸۴، «معناشناسی و بازیابی اطلاعات، هفت گفتار» انتشارات کتابخانه زاویه.
- میدو، چارلز تی. و همکاران، ۱۳۹۰، «نظام‌های بازیابی اطلاعات متنی» ترجمه: نجلا حریری، انتشارات چاپار.
- Agogino, A. (2004) Developing a Learner-Centered Metathesaurus for Science, Mathematics, Engineering and Technology Education. http://best.me.berkeley.edu/pubs/04_1202_P.pdf
- Amirhosseini, Maziar (2008) Dialectic Schemes in Thesaurus Creation, Library Philosophy and Practice (e-journal). 163. <http://digitalcommons.unl.edu/libphilprac/163>.
- Galinski, Christian (2001) Terminology and Knowledge representation (There is no knowledge without terminology) 2001-03-12 CG, Infoterm, Vienna.
- Hjorland, Birger (2015) Theories are Knowledge Organizing Systems (KOS), Knowl. Org, 42 (2015) No.2.
- Hjorland, Birger (2017) Reviews of Concepts in Knowledge Organization, In: Knowledge Organization, 45 (2018) No. 3
- Hjorland, Birger (2018) Indexing: Concepts and theory, Encyclopedia of Knowledge Organization [Http://bibliothecavirtual.clasco.org.ar/ar/oecd-macroth/en/index.htm](http://bibliothecavirtual.clasco.org.ar/ar/oecd-macroth/en/index.htm)
- [Http://faculty.ung.edu/kmelton/Documents/DataWisdom.pdf](http://faculty.ung.edu/kmelton/Documents/DataWisdom.pdf)
- [Http://journals.pnu.ac.ir/article_4410_504a5ba8896eb4962ae3ad6e9d4606cb.pdf](http://journals.pnu.ac.ir/article_4410_504a5ba8896eb4962ae3ad6e9d4606cb.pdf)
- [Http://lis.aqr-libjournal.ir/article_43734.html](http://lis.aqr-libjournal.ir/article_43734.html)
- [Http://nastinfo.nlai.ir/article_748_e8c027d9da2fc45f27a74802f12296e4.pdf](http://nastinfo.nlai.ir/article_748_e8c027d9da2fc45f27a74802f12296e4.pdf)
- [Http://payavard.tums.ac.ir/article-1-143-en.pdf](http://payavard.tums.ac.ir/article-1-143-en.pdf)
- [Http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B1%D8%AF%D9%87%20%D8%A8%D9%86%D8%AF%D9%8A%20%D9%83%D9%88%D9%84%D9%86.aspx](http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B1%D8%AF%D9%87%20%D8%A8%D9%86%D8%AF%D9%8A%20%D9%83%D9%88%D9%84%D9%86.aspx)
- [Http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B7%D8%B1%D8%AD%20%D9%81%D8%B1%D8%A7%D8%AF%D8%A7%D8%AF%D9%87%20%D8%AA%D9%88%D8%B5%DB%8C%D9%81%20%D8%A7%D8%B4%DB%8C%D8%A7%20\(%D9%85%D9%88%D8%AF%D8%B2\).aspx](http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B7%D8%B1%D8%AD%20%D9%81%D8%B1%D8%A7%D8%AF%D8%A7%D8%AF%D9%87%20%D8%AA%D9%88%D8%B5%DB%8C%D9%81%20%D8%A7%D8%B4%DB%8C%D8%A7%20(%D9%85%D9%88%D8%AF%D8%B2).aspx)
- <https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/bul2.2017.1720430211>
- https://en.wikipedia.org/wiki/Dublin_Core
- [https://en.wikipedia.org/wiki/Ontology_\(information_science\)](https://en.wikipedia.org/wiki/Ontology_(information_science))
- <https://en.wikipedia.org/wiki/Semantics>
- <https://fa.wikipedia.org/wiki/%D9%85%D8%B9%D9%86%D8%A7%D8%B4%D9%86%D8%A7%D8%B3%DB%8C>
- <https://www.loc.gov/standards/mods/design-principles-mods-mads.html>
- <https://www.loc.gov/standards/premis/v3/index.html>
- https://www.researchgate.net/publication/220854050_A_Functional_Semantic_Web_Architecture
- https://www.researchgate.net/publication/220854050_A_Functional_Semantic_Web_Architecture
- Pisanelli, D.M., Gangem, A. and Steve G. (2006) An Ontological Analysis of the UMLS Metathesaurus, Roma: Institio di technology Biomediche-CNR.
- Soergel, D. et al (2004) Reengineering Thesauri for New Application: The AGROVOC Example. Journal of Digital Information, Volume 4, Issue 4, Article No. 254.

- Roget, p. (2000). Roget's International Thesaurus of English Words and Phrases. C.O. Sylvester Mawson. <http://Bartleby.com>
- To increase usability of application axiomatizations, they might be specified in multiple specification languages, such as DAML+OIL, OWL, RuleML, EER, UML, etc. Figure 10 shows the above Bookstore axiomatization expressed in OWL.
https://www.researchgate.net/publication/251320331_Ontology_engineering_-_The_DOGMA_approach
- Smiraglia, R.P. (2014) The Elements of Knowledge Organization, Springer International Publishing Switzerland 2014, DOI 10.1007/978-3-09357-4_2.
- Sowa, J.F. (2000) Knowledge Representation: Logical, Philosophical, and Computational Foundations. Oxford: Black well.
- Sowa, J.F. (2010) The Role of Logic and Ontology in Language and Reasoning, Theory and Applications of ontology. Ink. Poli and J.seibt (Eds.), Philosophical Perspectives, PP. 231- 263. Berlin. Springer.
- Webster's Third New International Dictionary of the English Language .Vol. 3, S.V. "Thesaurus."
- Wood, W-H. (2004) Developing Learner- Centered Meta thesaurus for Science, Mathematics Engineering and Technology Education. Uni of California Berkeley.
- UMLS Metathesaurus (2004). <http://www.nlm.nih.gov/pubs/factsheets/factsbj.html>
- Yancey, T. & Carson, J. (2004) Controlled Vocabulary Component Model.
- Yancey, T., & Carson, J. (2004) Lexicography without Limits: A Web-Based Solution. <http://www.udc.es/lexicografia/x-zw.htm>

Abstract

Metathesauri is a combination of the existing structured languages which has been obtained by linking, merging and integrating them.

Thesauri corresponds with each other and there are semantic relation made for each concept in a thesaurus on the closest equivalent concept in other thesauri. This enables a certain database, which has been indexed with a specialized thesaurus, to translate a user's question into a search item. The lexicon is also merged and becomes the basis for a richer and more comprehensive source of knowledge. In each structured lexicon, a word is chosen as the "preferred term" for a concept.

Metathesaurus is not only a structured collection of words, but it is a tool to increase efficiency and usefulness of the existing dictionaries and thesauri in order to coordinate between the structured languages through corresponding, merging and integrating the thesauri. This is one of metathesaurus's traits. This coordination facilitates the transfer record between different fields of database and searching within them.

Therefore, metathesaurus is a large comprehensive lexicon database that is expanding, multi-purpose and multilingual which results from coordinating, linking and merging the thesauri and the existing structured languages and it is originally created for its developers to use. It can also be a useful source for creators of databases, librarians, information specialists and other interested users. Also, it can provide the needed information to create the standard data for computer programs, interpret the users' questions and communicate with them through refining the questions and convert the users' terms into the lexicon used in the related sources. Metathesaurus is usually formed in a specific field of science.

In 1380s, specialized thesauri have been created in IRANDOC, which were updated in 1396-97. In addition to these thesauri, a set of dictionaries and specialized lexicon have been compiled and printed.

It is necessary for a more precise search and analysis that these lexicon and thesauri are merged as a metathesaurus in order to get more precision in searching of

information in the Ganj database and a fast access to specialized information is provided.

In this report we first study the documents with the topic “metathesaurus” , the related concepts, the usage and the process of its compilation.

Keywords: Thesaurus, Metathesaurus, Ontology, Semantic Web, IranDoc



**Iranian Research Institute
for Information Science and Technology
(IranDoc)**

Research Projects:

Study of Concepts and Method of the Metathesaurus Building

By:

Molouk Sadat Hosseini Beheshti

2020