

بررسی هستی‌شناختی فرا اصطلاح‌نامه نظام زبان پزشکی واحد* صدیقه محمداسماعیل^۱، لاله صمدی^۲

چکیده

مقدمه: سالیان متمادی است که هستی‌شناسی به اشکال گوناگونی مانند طرح‌های رده‌بندی، تزاروس‌ها، واژگان کنترل شده، اصطلاح‌شناسی‌ها و واژه‌نامه‌ها معنی شده است. از آن‌جا که اصطلاح‌ها و منابع دانش، اجزای مهمی از تعامل میان نظام‌های نام‌تجانس است، هدف این مقاله، بررسی هستی‌شناختی فرا اصطلاح‌نامه‌ی نظام زبان پزشکی واحد (که در یک سیستم منفرد اصطلاحات از علوم زیست‌شناسی ادغام شده است و توسط کتابخانه‌ی ملی پزشکی آمریکا تهیه گردید که بخشی از زبان واحد پزشکی است) بوده است.

روش بررسی: در این پژوهش با استفاده از روش مطالعه‌ی سندی و کتابخانه‌ای، به مطالعه‌ی متون و منابع موجود در زمینه‌ی مقایسه‌ی فرا اصطلاح‌نامه Unified Medical Language System یا UMLS و هستی‌شناسی‌های دیگر (مانند WordNet و UCO و کارهایی که در این خصوص توسط افرادی چون Burgun و همکار، در نتیجه‌ی بررسی و مرور منابع معتبر این حوزه و تهیه‌ی فهرستی از لغات مرتبط با علم زیست‌شناسی، دال بر وجود اسامی، روابط مورد نیاز در علم اطلاع‌رسانی زیست‌شناسی، اقدام به مطابقت آن‌ها بر اساس مفاهیم فرا اصطلاح‌نامه UMLS و در نهایت تعیین وجود روابط در منابع اصلی این فرا اصطلاح‌نامه صورت گرفت) پرداخته شده است. علاوه بر این، به منظور بررسی وجود سازگاری میان نقشه‌ی معنایی UMLS و دو هستی‌شناسی دیگر، مطابقت دادن در انواع معنایی UMLS و هستی‌شناسی سلسله‌مراتبی Cyc و WordNet انجام گرفت و برای ایجاد ساختاری یک‌پارچه در UMLS و سازگاری آن با Cyc، مفاهیم در Cyc به سه گروه تقسیم‌بندی گردید.

یافته‌ها: قریب به ۵۰ مقوله از Cyc به طور دقیق برای پوشش دادن شبکه‌ی معنایی UMLS استفاده شده است. نتایج نشان داد که نزدیک به نیمی از آن‌ها در هر دو سیستم شبیه به هم هستند و میان انواع Cyc و شبکه معنایی UMLS هم‌پوشانی وجود دارد. همچنین مشخص شده است که برای بعضی از گونه‌های معنایی UMLS مقوله‌های همسان و مشابهی در نسخه‌ی عمومی هستی‌شناسی سلسله‌مراتبی Cyc وجود ندارد. بنابراین، مقوله‌هایی در Cyc وجود دارد که دارای هیچ گونه معادلی در UMLS نمی‌باشد. در زمینه‌ی مقایسه‌ی رده‌های معنایی UMLS و WordNet نتایج نشان داد که ۲ درصد از مفاهیم رده‌ی اختلال سلامتی UMLS در WordNet وجود دارد که مطابقت و سازگاری میان رده‌ها ۴۸ درصد بوده است.

نتیجه‌گیری: هم‌اکنون هیچ استاندارد Upper-Level-Ontologies در دسترس نیست. ممکن است در یک رده، مفاهیم به طور متفاوتی دسته‌بندی شوند، حتی زمانی که دسته‌بندی‌ها شبیه به هم به نظر برسند.

واژه‌های کلیدی: اصطلاح‌شناسی؛ اصطلاح‌نامه؛ واژه‌نامه‌ها؛ ذخیره و بازیابی اطلاعات پزشکی.

نوع مقاله: تحقیقی

پدیرش مقاله: ۱۸/۶/۵

اصلاح نهایی: ۱۷/۱۱/۳

دریافت مقاله: ۱۷/۵/۱۲

ارجاع: محمداسماعیل صدیقه، صمدی لاله. بررسی هستی‌شناختی فرا اصطلاح‌نامه نظام زبان پزشکی واحد. مدیریت اطلاعات سلامت ۱۳۸۹؛ ۷ (۲): ۱۶۰-۱۴۷.

* این مقاله حاصل تحقیقی مستقل است که بدون حمایت مالی سازمان انجام شده است.

۱. استادیار، کتابداری و اطلاع‌رسانی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران، تهران، ایران. (نویسنده‌ی مسؤول).

Email: m.esmaeili@sri.au.ac.ir

۲. کارشناس ارشد، کتابداری و اطلاع‌رسانی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، تهران، ایران.

مقدمه

سالیان متمادی است که هستی‌شناسی‌ها به اشکال گوناگونی مانند طرح‌های رده‌بندی، تزاروس‌ها، واژگان کنترل شده، اصطلاح‌شناسی‌ها و واژه‌نامه‌ها معنی شده است. اگر چه ممکن

فرا اصطلاح‌نامه UMLS:

یکی از مشهورترین نمونه‌های فرا اصطلاح‌نامه یا اصطلاح‌نامه‌های ادغام شده، اصطلاح‌نامه‌ی نظام زبان پزشکی واحد (UMLS) است که یک اصطلاح‌نامه‌ی مفهومی می‌باشد. "UMLS" توسط کتابخانه‌ی ملی پزشکی آمریکا از سال ۱۹۸۶ با همکاری نهادهایی از داخل و خارج ایالات متحده اجرا شده است. هدف آن یک پارچه‌سازی تمام اسامی همراه با معانی تقریباً یکسان در یک مفهوم می‌باشد. اسامی که از جز اصلی مشتق می‌شوند یا منابع تنوع گسترده‌ای از ترمینولوژی‌های بیوشیمی است یا شامل تعداد بسیاری از واژگان کنترل شده و طبقه‌بندی شده‌ی تمام متن و سیستم‌های خبره است. در واقع هدف اصلی از ایجاد "UMLS" فراهم آوردن امکان دستیابی یکسان به تعدادی از پایگاه‌های اطلاعاتی توزیع شده و ناهمخوان است که از مجموعه‌ی اصطلاحات متفاوتی استفاده می‌کنند. "UMLS" شامل چهار منبع دانش (Knowledge source) است: «فرا اصطلاح‌نامه»، «شبکه‌ی معنی شناختی»، «نقشه‌ی منابع اطلاعاتی» و «واژه فرهنگ تخصصی (Specialist lexicon)» (۴) که در ادامه به توضیح هر کدام از آن‌ها پرداخته می‌شود:

فرا اصطلاح‌نامه: همان طور که در ابتدا نیز به این مسأله اشاره شد فرا اصطلاح‌نامه‌ها ترکیبی از زبان‌های کنترل شده موجود است که از طریق برقراری پیوند، ادغام و تلفیق آن‌ها به وجود آمده است. با توجه به این که در هر یک از واژگان‌های متشکله، نامی به عنوان نام مرجح برای مفهوم انتخاب می‌شود این امر امکان جهت‌گردانی بین واژگان‌ها را فراهم می‌کند. در فرا اصطلاح‌نامه "UMLS" چنانچه مفهوم در Mesh وجود داشته باشد، شکل مرجح آن مفهوم، صورت سرعنوان اصلی در Mesh است؛ در غیر این صورت، اصطلاحات از دیگر واژگان‌ها به ترتیبی مشخص‌گزینه می‌شوند. تمام اصطلاحات، عبارات و زنجیره‌های دیگری که برای بیان مفاهیم مورد استفاده قرار می‌گیرند با نام مندرج در فرا اصطلاح‌نامه مطابقت داده می‌شوند.

است هستی‌شناسی‌ها دارای چنین ویژگی‌هایی باشند اما به طور قطع نمی‌توان آن‌ها را یکسان با هر یک از این ساختارها دانست (۱). هستی‌شناسی و فرا اصطلاح‌نامه‌ها فقط به جهت ایجاد طرح و شمای رمزگذاری منفرد نظام رده‌بندی و یا نظام واژگان کنترل شده به وجود نیامده است. بلکه هدف از ایجاد آن‌ها عبارت است از مکانیزمی برای ترجمه، تفسیر، برطرف نمودن ابهامات، معانی و اسامی خاص یک رشته. فرا اصطلاح‌نامه و هستی‌شناسی، هر دو شبکه‌ای از اصطلاحات و نمایشی از روابط میان آن‌ها جهت آرایه مفاهیم می‌باشد که بر پایه‌ی مفاهیم، در راستای آرایه دانش موضوعی پیچیده بنا می‌شوند و در سازماندهی منابع اطلاعاتی و بازیابی اطلاعات کاربرد مؤثری دارند. با این حال می‌توان تأثیر هستی‌شناسی‌ها را بر روی فرا اصطلاح‌نامه‌ها بررسی کرد (۲). بخشی از تحقیقات در علم پزشکی بر استفاده‌ی مجدد دانش در کاربردهای جدید و طراحی هستی‌شناسی‌های قابل اشتراک متمرکز است. مومن منابع اصلی اشتراک دانش و استفاده مجدد از آن را شامل مشکلاتی از قبیل عدم اتفاق آرا در خصوص معنی اصلی دانش، برشمردن ویژگی‌های مفهومی و دانش زمینه‌ای مورد نیاز برای توضیح یک حوزه‌ی موضوعی خاص می‌داند. با این حال تلاش‌های متعددی در این زمینه صورت گرفته است از جمله ترکیب هستی‌شناسی‌های عمومی، استفاده‌ی مجدد هستی‌شناسی‌های تخصصی برای کاربردهای تخصصی و یک پارچه‌سازی هستی‌شناسی‌های تخصصی با کتابخانه‌های هستی‌شناسی بزرگ مقیاس (۳). از این رو هدف اصلی این مقاله توسعه‌ی روش‌هایی است که به موجب آن هستی‌شناسی‌ها به توانند از منابع موجود (فرا اصطلاح‌نامه UMLS) تهیه شوند، به طوری که همانند سایر منابع دانش و در مقایسه‌ی با آن‌ها دارای اعتبار باشند. بنابراین، در این بررسی بر روی شبکه معنایی UMLS که می‌تواند زمینه‌ی بالقوه هستی‌شناسی در علوم زیستی باشد متمرکز شده است و هدف بررسی سازگاری مقوله‌های UMLS با مقوله‌هایی از هستی‌شناسی‌های کلی یا سلسله مراتبی است.

شناختن ارگانسیم‌ها، هستی‌شناسی ژن و توضیح محصولات ژنی در تمام مدل‌های مختلف ارگانسیم. همچنین فرا اصطلاح‌نامه متون زیست‌شناسی، سرعنوان موضوعی Mesh را پوشش می‌دهد که از واژگان کنترل شده‌ی آن برای نمایه‌سازی در Medline استفاده می‌شود. علاوه بر آن، حوزه‌های موضوعی اصلی مانند آناتومی، در فرا اصطلاح‌نامه ارایه می‌شود. در نهایت می‌توان گفت که بهترین حوزه‌ی موضوعی ارایه شده، اجزا بالینی زیست‌شناسی است همراه با اصطلاح‌نامه عام آن، مانند SNOMED یا Systematized Nomenclature of human Medicine. شکل ۱ نشان می‌دهد که فرا اصطلاح‌نامه‌ی UMLS چگونه با یک پارچه‌سازی این اصطلاحات مختلف می‌تواند نه تنها به عنوان پیوندی میان واژگان مختلف باشد، بلکه توانایی نشان دادن حوزه‌های موضوعی آن‌ها را نیز دارد (۵).

با بررسی «نظام UMLS» می‌توان دریافت که چگونه با ادغام زبان‌های کنترل شده، محصولی به وجود می‌آید که به لحاظ اصطلاحات و رابطه‌ها از هر یک از واژگان‌های متشکله به تنهایی، غنی‌تر است، جهت‌گردانی بین واژگان‌های مستقل میسر می‌باشد، بازیابی اطلاعات با استفاده از اصطلاحات متعارف زبان کنترل شده تسهیل می‌گردد و همچنین می‌تواند به عنوان ابزاری برای تأمین دستیابی به متون زبان طبیعی در پایگاه‌های اطلاعاتی مختلف عمل نماید (۴).

هدف عمده‌ی فرا اصطلاح‌نامه‌ی UMLS، تسهیل کار کارشناسان سلامت پزشکی و محققان برای بازیابی و یک‌پارچه‌سازی اطلاعات مناسب از منابع خودکار نامتجانس است مانند: رکوردهای بیماری مبتنی بر کامپیوتر، بانک‌های اطلاعاتی واقعی، پایگاه‌های کتاب‌شناختی و تمام متن و سیستم‌های خبره.

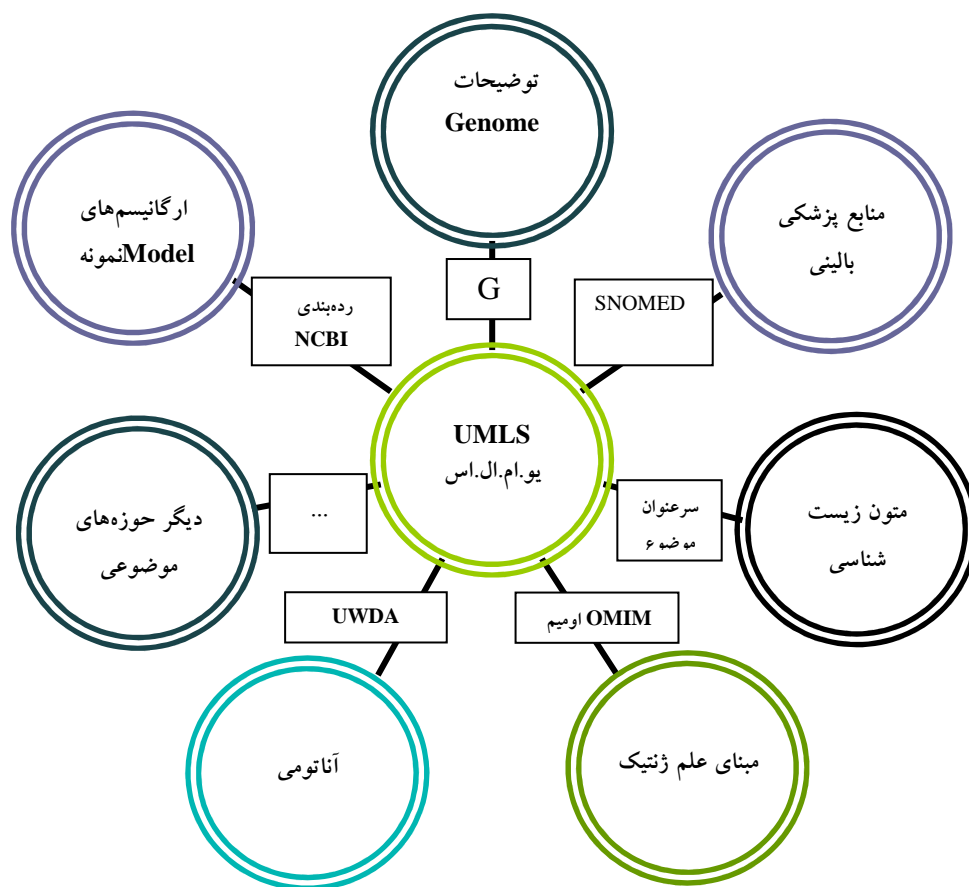
همچنین این فرا اصطلاح‌نامه بر ارتباط مفهومی تأکید دارد و منابع دانشی را ایجاد می‌کند که می‌تواند با استفاده از برنامه‌های هوشمند بر مشکلات زیر فایق آید:

شبکه‌ی معنی‌شناختی: این شبکه دسته‌بندی دایمی از تمام مفاهیم موجود در فرا اصطلاح‌نامه را بر طبق دسته‌ها، انواع معنی‌شناختی و رابطه‌های بین این دسته‌ها ارایه می‌کند. نحوه‌ی تنظیم انواع معنی‌شناختی، سلسله‌مراتبی است. انواع معنی‌شناختی به تشخیص معانی گوناگون مرتبط با یک اصطلاح کمک می‌کنند. رابطه‌ها غیر سلسله‌مراتبی هستند و به پنج گروه تقسیم می‌شوند: رابطه‌های فیزیکی (Physical relationship)، رابطه‌های فضایی، رابطه‌های زمانی (Chronological relationship)، رابطه‌های کارکردی (Functional relationship) و رابطه‌های مفهومی (Conceptual Relationship). این رابطه‌ها بین گروه‌های سطح بالای دسته‌های معنی‌شناختی بیان می‌شوند. رابطه‌های ارایه شده نمایانگر تمام رابطه‌های ممکن بین انواع معنی‌شناختی‌های تعریف شده نیستند، بلکه مرحله‌ای آغازین در ایجاد مجموعه‌ی مفیدی از رابطه‌ها می‌باشند. قابل ذکر است که کاربردهای عمده‌ی شبکه‌ی معنی‌شناختی در تدوین پرسش و در پالایش تعاملی آن است.

نقشه‌ی منابع اطلاعاتی: نقشه‌ی منابع اطلاعاتی نشان می‌دهد که کدام پایگاه اطلاعاتی به چه سؤالی خاص مربوط می‌شود، اطلاعات درباره‌ی منابع مخصوص را برای کاربران فراهم می‌کند که به طور خودکار به منابع متصل می‌شود و عمل بازیابی را بر روی یک یا چند منبع انجام می‌دهد.

واژه‌رنگ تخصصی: این واژه‌رنگ برای فراهم آوردن اطلاعات واژگان مورد نیاز در نظام پردازش زبان طبیعی اسپشیالیست ایجاد شده است. هدف از این منبع آن است که یک واژه‌رنگ زبان انگلیسی با بسیاری از اصطلاحات زیست‌پزشکی باشد.

UMLS نه تنها برای نیازهای اطلاعاتی زیست‌شناسان گسترش یافته است، بلکه اصطلاحات استفاده شده در زیست‌شناسی را نیز شامل می‌شود. به عنوان مثال، اصطلاح‌نامه‌ی جامع اخیر مشتمل است بر رده‌بندی National (NCBI) Center for Biotechnology Information برای



شکل ۱: حوزه‌های موضوعی مختلف مجتمع در UMLS (۵)

به هم مرتبط می‌شوند. در نتیجه نموداری غنی از اصطلاحات در UMLS ایجاد می‌گردد که روابط مفهومی- درونی (Inter- concept) آن هم موروث ساختاری از لغات منبع است که توسط سردبیران فرا اصطلاح‌نامه ایجاد گردیده است. روابط سمبلیک (isa یا is a kind of یا isa و یا Part of) یا روابط همبسته (location of یا Caused by) می‌توانند به صورت سلسله مراتبی وجود داشته باشند. همچنین در آن رابطه‌ی آماری بین مفاهیم و لغات Mesh ایجاد شده است که از نمایه‌سازی لغات Mesh در استادهای Medline به طور هم‌زمان مشتق می‌شود. در نهایت، هر مفهوم در متاتزاروس، با اهداف شبکه‌ی معنایی، توسط سردبیران فرا اصطلاح‌نامه دسته‌بندی می‌شوند (۵).

۱. عدم توافق (ناهماهنگی) در زبان که توسط کاربران و منابع اطلاعاتی مختلف استفاده خواهد شد.
 ۲. مشکلات در مورد شناسایی بسیاری از منابع اطلاعاتی مرتبط. که برای تحقق این امر، از ابزارهای چند کاره یا نرم‌افزارهای هوشمند برای گسترش نظام‌های خود استفاده می‌کند؛ مانند فرا اصطلاح‌نامه، برنامه‌های لغوی، لغت‌نامه‌های تخصصی و شبکه‌ی معنایی (۶).

اصول یک پارچه‌سازی اصطلاحات در UMLS:

در UMLS، دانش بر اساس مفهوم، سازماندهی (بر اساس معنی لغت) می‌شود و واژه‌های مترادف در قالب یک مفهوم در یک مقوله جای می‌گیرند که بر اساس انواع مختلف رابطه‌های موجود

تعریف هستی‌شناسی یا آنتولوژی:

آنتولوژی (Ontology) (با حرف O بزرگ) به معنای هستی‌شناسی است، که درحوزه‌ی فلسفه مطرح است و به موضوع موجودیت، مطالعه‌ی هستی و ساختار واقعیت می‌پردازد. به عبارت دیگر، هستی‌شناسی یک علم در فلسفه است که «به مطالعه آن چه موجود است، آن چه باید فرض شود که وجود دارد و به منظور نایل آمدن به یک توصیف متقاعد کننده از واقعیت» می‌پردازد (۱). John Sowa از هستی‌شناسی تعریفی ارائه می‌دهد: «هستی‌شناسی، دسته‌بندی از موضوعات یا عناصر موجود در یک حوزه‌ی خاص را بررسی می‌کند و بر اساس آن بررسی، فهرستی از موضوعات را ارائه می‌دهد. این فهرست که به تفصیل، انواع موضوعات و روابط میان آن‌ها را در حوزه‌ی مورد بررسی بیان می‌کند، هستی‌شناسی نامیده می‌شود (با حرف O کوچک)». لازم به ذکر است که یک هستی‌شناسی الزاماً شامل یک مجموعه واژه‌ی مرکب از اصطلاحات و بعضی از مشخصات و معانی آن‌ها می‌باشد. بنابراین، می‌توان هستی‌شناسی را مانند اصطلاح‌نامه‌ها، مجموعه‌ای از لغات کنترل شده دانست، اما با این تفاوت که مفاهیم یک حوزه‌ی خاص با روابط معنایی خاص با توجه به نوع حوزه‌ی موضوعی، از موضوعی به موضوعی دیگر به شکل خاصی طبقه‌بندی می‌شوند که در قالب زبان طبیعی به نمایش در می‌آید Tom Gruber یکی از پیشگامان در معرفی هستی‌شناسی است، بر اساس نظر او، هستی‌شناسی «ذکر خصوصیات صریح از مفاهیم به اشتراک گذاشته شده است.» (۱).

بررسی هستی‌شناختی (آنتولوژیکی) فرا اصطلاح‌نامه‌ی UMLS:

بر طبق تحقیقات به عمل آمده از نظام‌های مبتنی بر به کارگیری اصطلاحات (Paper-based)، این نظام‌ها دیگر قادر به پاسخگویی به خواسته‌های جدید نظام‌های اطلاع‌رسانی، مثل تقاضای مراکز بهداشتی برای استفاده‌ی مجدد از اطلاعات بیمار (بایگانی)، انتقال و دسته‌بندی آن‌ها نمی‌باشند.

آنتولوژی‌ها در این میان، با فراهم آوردن اطلاعات مؤثر، کارآمد و در واقع بسته‌بندی دوباره، این اطلاعات را در تمامی زمینه‌ها به خصوص علم پزشکی فراهم کرده است که شامل

امکان استفاده مجدد از این اطلاعات می‌باشد. رسیدن به چنین ساختاری احتیاج به تحلیلی عمیق و مفهومی از اطلاعات پزشکی دارد، که در صورت استفاده‌ی مفهومی از این اصطلاحات امکان‌پذیر می‌باشد. صرف نظر از معنای فلسفی آنتولوژی «هستی‌شناختی»، این واژه در متون پزشکی به معنای مفهوم‌سازی نسبی به کار می‌رود. همان طور که در بالا اشاره شد تعریف "Sowa" از آنتولوژی، مطالعه‌ی طبقه‌بندی نسبی مفاهیم موجود در حوزه‌ی خاص است که محصول چنین برداشتی «آنتولوژی» نامیده می‌شود. در واقع نظری واحد برای تعریف آنتولوژی و این که آنتولوژی چه مفهومی دارد، وجود ندارد. عده‌ای از آن به عنوان یک توصیف معمولی مبتنی بر «ساختار سلسله مراتبی» که در بعضی موارد استفاده از زبان‌شناسی را سازمان می‌دهد نام برده‌اند و برخی دیگر آن را به عنوان تئوری متشکل از لغات رسمی با معانی ثابتی از آن لغات، برای توصیف‌های منطقی لغات می‌دانند. اما آنتولوژی به بیان مفهوم نسبی یک واژه در حوزه‌ی مربوط به خود می‌پردازد. از آن جا که به طور معمول واژه‌ها با معانی مختلفی (چند معنایی Polysemous) در یک حوزه استفاده می‌شوند، آنتولوژی به بررسی جداگانه‌ی هر مفهوم با استفاده از ارائه متنی برای تفهیم بهتر آن می‌پردازد. در اصطلاح به چنین فرایندی تحلیل آنتولوژیکی (Ontological analysis) یا تحلیل مفاهیم گفته می‌شود. در این جا به برخی از سیستم‌های کاربردی مهم اصطلاحات پزشکی که توسط این فرایند فراهم شده است، اشاره می‌شود.

تجربیات گذشته حاکی از این است که برای رسیدن به مفهومی روشن و صریح از کاربرد اصطلاحات، نیاز مبرمی به مفاهیم فلسفی و زبان‌شناسی می‌باشد، به عنوان مثال، مرلوژی (Mereology) تئوری نقش کلمات، توپولوژی (Topology) تئوری جایگاه شناسی کلمات، مرفولوژی (Morphology) نظریه‌ی صرف و معادل‌سازی و مواردی از این قبیل. گفتنی است که این گونه تئوری‌ها برای توسعه‌ی آنتولوژی مؤثر می‌باشند. این تئوری‌ها از مفاهیم رسمی و تشریفاتی تشکیل نشده‌اند بلکه می‌توانند حاوی مفاهیم و

۲. IS_A و تفاوت بین مفاهیم موجود در CUI،
 UMLS (بیماری عضلات IS_A Fibromyolgia)
 ۳. IS_A ارتباط موجود بین CUI و معانی مختلف آن.
 (مثال بیماری - سندرم IS_A Fibromyolgia)
 ۴. معنی CUI در متون ساده مانند متون ذکر شده،
 فرهنگ‌نامه‌های پزشکی است.

این اطلاعات بعد از سازماندهی شدن مجدد در برنامه‌ی
 MS Access توسط کتابخانه‌ی ملی پزشکی آمریکا به
 صورت جدول زیر ارائه شد:

۱. ۳۳۱۷۵۶ اطلاعات ثبت شده با استفاده از CUI و زبان
 انتخابی آن.

۲. مقایسه‌ی ۵۱۸۴ اطلاعات ثبت شده با استفاده از کودک
 CUI و والد CUI (Parent CUI) و مفاهیم به کار رفته
 ۳. ۴۴۳۷۷۰ اطلاعات ثبت شده در رابطه با انواع CUI
 ۴. ۲۱۸۹۵ اطلاعات ثبت شده و به دست آمده از CUI
 لازم به ذکر است که در حدود ۴۴۳۰۰۰ مفهوم CUI به
 صورت زوج ارائه شده است که در این میان سهم اندکی از آن
 به CUI والد تعلق دارد.

پایگاه داده‌های موجود در CUI به صورت رایانه‌ای است که
 علاوه بر اطلاعات ارائه شده، دیگر اطلاعات نیز به طور متناوب
 و در جای خود ذکر شده‌اند. با داشتن ۳۳۱۷۵۶ CUI موجود، از
 ابزاری به نام Loom برای طبقه‌بندی آن‌ها به صورت مفهومی
 با ارائه مثال‌ها و معانی مربوطه استفاده شده است. در این
 طبقه‌بندی بعضی مفاهیم قدیمی به منظور رسیدن به مفهومی
 کلی‌تر و دقیق‌تر حذف شده است که امکان تسلط بیشتر در
 جهت رسیدن به مفهومی صریح‌تر را برای کاربر فراهم می‌کند.
 مسأله دیگر عدم رعایت در ساختار طبقه‌بندی و سلسله
 مراتبی، به کار بردن اصطلاحات در جای مناسب خود می‌باشد
 که در این میان استفاده از روش CUI باعث قرار گرفتن
 مفاهیم در جای مناسب خود می‌شود که این خود باعث یافتن
 مفاهیم به نحو بهتری خواهد شد.

شناسه‌های مفهومی خاص CUI با پرداختن به بیش از
 یک معنی از هر واژه، الگوی بسیار متداولی در متاتزاروس

مستندات عادی و روزمره نیز باشند. به تازگی، رایج نظام‌های
 کامل در این زمینه با رایج خدماتی نظیر طبقه‌بندی مفاهیم و
 متون رسمی باعث توسعه‌ی گستره این تئوری‌ها به صورت
 تخصصی شده‌اند.

در این راستا "ONIONS" تأمین کننده اصطلاحات
 پزشکی در پروژه‌ی Gallen، یک روش شناسایی ترکیبات و
 گستره‌های کاربردی اصطلاحات مبتنی بر تئوری‌های عمومی
 زبان شناسی است که در قالب یک کتابخانه به بهره‌برداری
 رسید. ONIONS ترکیبی مفهومی است که شامل جملات،
 الگوها و اصول متعارف کاربردی می‌شود که در تجزیه و
 ترکیب سیستم‌های مهم کاربرد اصطلاحات پزشکی، از جمله
 در شبکه معنایی UMLS، SNOMED III، GMN، ۱۰
 ICD مدل هسته‌ای Gallen به کار برده شده است.

در کاربردهای اصطلاحات پزشکی به هر دو جنبه‌ی مهم و
 جزئی اصطلاحات پرداخته شده است. این پروژه توسط
 کتابخانه‌ی ملی پزشکی آمریکا، مبتنی بر جمع‌آوری اطلاعات
 چشمگیر استخراج شده در آمریکا و سایر کشورها، به انجام
 رسیده است. این استخراج باعث به وجود آمدن منبع عظیم
 اطلاعاتی در حوزه‌ی اصطلاحات پزشکی شده است، اما این
 بدان معنی نمی‌باشد که این استخراج در همین گستره محدود
 شود بلکه، مفاهیم علوم دیگر را نیز در بر می‌گیرد (۶).

حال به بررسی تحلیل هستی شناختی فرا اصطلاحنامه
 UMLS پرداخته می‌شود:

اولین فرا اصطلاحنامه در سال ۱۹۹۸ در نسخه‌ای مبنی بر
 CUI که بعدها تکمیل شد ارائه گردید. این نسخه شامل
 ۳۳۰۰۰۰ مفهوم، مترادف و واژه‌های استخراج شده از زبان‌های
 انگلیسی، اسپانیایی، فرانسوی، آلمانی، روسی و پرتغالی می‌باشد.
 بدین معنی که در مورد واژگانی که دارای چندین معنی در یک
 گستره‌ی مفهومی بودند با استفاده از CUI به بیان این مفاهیم
 پرداخته شده است و محدود کردن این کاربرد، منجر به محدود
 شدن صراحت این مفاهیم نشده است.

در فرا اصطلاحنامه این اطلاعات به صورت زیر ارائه می‌شود:
 ۱. اسامی مرجع در CUI (مثال: Fibromyolgia)

و روابط استنتاجی تعریف شوند که استدلال را برای کارهای منتخب قادر سازند.

انواع معنایی (Semantic Types یا STs): همانند مقوله‌ها به نظر می‌رسند. نمایش STs باید در یک هستی‌شناسی عمومی که عملکرد آن ارایه دانش است مقایسه شود. انتخاب Cyc زمینه‌ی مناسبی را که ممکن است شامل نظرات خاصی باشد فراهم می‌نماید (نظریه خرد). STs به عنوان رده‌ها به نظر می‌آیند. UMLS دو سطح نمایشی را فراهم می‌کند، جایی که STs مقوله‌هایی تحت مفاهیم متاتزاروس باشند آن را رده یا گسترش ST می‌نامیم که مجموعه‌ای از مفاهیم متاتزاروس است و به یک ST داده می‌شود. نمایش گسترش ST با WordNet مقایسه می‌شود. بر خلاف Cyc، ساختار WordNet بسیار شبیه متاتزاروس است (واژه‌ها، مفاهیم و ساختار سلسله مراتبی). رده‌ها در Wordnet می‌تواند از روابط هایپونامیکی (Hyponymic) مشتق شود. بنابراین، Wordnet یک ترمینولوژی عمومی را در بسیاری از زمینه‌های معنایی ایجاد می‌کند که شامل حوزه‌ی زیست پزشکی است.

هستی‌شناسی سلسله مراتبی (سطح بالایی) Cyc (۳): در سال ۱۹۹۷ منتشر شده است مجموعه‌ای قریب به ۳ هزار مفهوم عمومی و جهانی است که بر ۲ بخش انفرادی و گروهی (مجموعه) تقسیم می‌شود. بخش‌های Cyc به صورت سلسله مراتبی سازمان یافته است که بر مبنای دو رابطه‌ی ساختاری می‌باشد.

۱. #isa: رده‌بندی کلاسیکی را نمایش می‌دهد. به عنوان مثال (#isa EI col) یعنی EI عضوی از مجموعه col است.
۲. #genls: روابط میان یک مجموعه و رده‌ی بالایی آن است. (#genls col sup) به این معنی که sup مقوله‌ای است که رده‌ی بالایی آن col می‌باشد.

WordNet: اطلاعات فرهنگ لغات در واژه‌هایی که دارای روابط معنایی و مفهومی هستند را سازماندهی می‌کند. در Wordnet مترادف‌ها بر اساس معنی دسته‌بندی می‌شوند که مجموعه‌ای از مترادف‌ها را در Wordnet، Synset می‌نامند.

می‌باشد. یک سوم از ۹۰۰۰۰ مفهوم به دست آمده بر اساس این طرح مقوله‌بندی شده‌اند، که با پرداختن به بیش از ۲ الی ۵ معنی، CUI کاربر را به کاربرد صحیح هر مفهوم در متن مناسب خود رهنمون می‌کند. این طبقه‌بندی باعث تسهیل تحلیل مفاهیم در فرا اصطلاح‌نامه می‌شود (۷).

تعاریف مفهومی و عملیاتی:

شبکه‌ی معنایی UMLS و هستی‌شناسی‌ها (۳): شبکه‌ی معنایی در UMLS نماینده‌ی سطح پیشرفته حوزه‌ی زیست پزشکی مبتنی بر انواع معنایی هر گونه مفاهیم فرااصطلاح‌نامه‌ای است. ۴ دسته‌ی اصلی هستی‌شناسی‌ها بر اساس سطح پوشش و وظایفی که به خاطر آن‌ها طراحی شده‌اند عبارتند از:

هستی‌شناسی عمومی (General ontologies): دانش عمومی را نشان می‌دهد که مستقل از حوزه‌ها و وظایف خاص، همراه با نسبیّت در سطح متوسط است.

هستی‌شناسی موضوعی (Domain ontologies): هستی‌شناسی‌هایی که جنبه‌ی خاصی از یک موضوع هستند اما مستقل از یک وظیفه‌ی خاص عمل می‌کنند، به طور کلی، هستی‌شناسی موضوعی نامیده می‌شود که باید منعکس کننده‌ی هسته‌ی اصولی و نظریه‌ی آن حوزه باشد.

هستی‌شناسی سطح بالایی (Upper-Level-Ontologies یا ULOS): در این هستی‌شناسی مفاهیمی که به فضا یا زمان مربوط می‌شوند نمونه‌ای از هستی‌شناسی سطح بالا هستند که در تمام حوزه‌ها کاربرد دارند. این هستی‌شناسی باید جهانی باشد به این معنی که نباید به یک حوزه‌ی خاص محدود شود. هر مفهوم به یک حوزه‌ی خاص نیاز دارد که به تواند به ULOS آن مرتبط گردد به همین دلیل باید چند منظوره باشند.

هستی‌شناسی‌های کاربردی (Application ontologies): این گونه هستی‌شناسی‌ها دامنه را محدود می‌کنند و از کارهای خاص مشتق می‌شوند. هستی‌شناسی‌های کاربردی کما بیش در قالب کاربرد خاصی تعبیه شده است و شامل تعداد به نسبت کوچکی از مفاهیمی هستند که باید بسیار مفصل همراه با قوانین

UMLS شامل مقوله‌هایی است که متعلق به حوزه‌ی زیست پزشکی خاصی نیستند مانند Physical Object یا Animal. در حالی که هستی شناسی ULOS شامل مقوله‌های زیست پزشکی است.

از آن جا که این مطالعه به روش سندی و کتابخانه‌ای صورت گرفته است و به فرا اصطلاحنامه UMLS از دیدگاه هستی شناختی نگریسته است. لذا، جامعه‌ی آن را دو هستی شناسی Cyc و WordNet که به تفکیک (و به تفصیل) در بخش تعاریف مفهومی و عملیاتی تعریف شده‌اند، تشکیل می‌دهد.

مطابقت دادن شبکه‌های معنایی به هستی شناسی سطح بالایی Cyc:

برای حفظ ظاهر Cyc، ویژگی‌های STs با استفاده از روابط $cys\#\$isa$ و $\#\$genls$ به صورت دستی انجام شده است. از مقوله‌های اضافی Cyc جهت کسب اطمینان از سازگاری و هماهنگی استفاده گردید. رابطه‌ی بین ST، T و نزدیکترین مفهوم Cyc، U مشابهت نامیده می‌شود به شرط این که T معادل U باشد حتی اگر نقش اسم را داشته باشد. اگر میان بخشی از T و U همپوشانی وجود داشته باشد در این مورد T و U با هم سازگارند و دارای یک رده‌ی مشترک هستند که این همپوشانی نامیده می‌شود مثالی از همپوشانی دو مقوله، سگ و حیوان (Dog و Pet) است (۳).

تطبیق گونه‌های معنایی UMLS در WordNet: تطابق رده‌های UMLS و WordNet بر اساس مقایسه‌ی مجموعه مفاهیمی است که بر اساس STs در UMLS رده‌بندی شدند و مجموعه‌ی هایپونیم‌هایی است که در Synset و WordNet رده‌بندی شدند. بنابراین، تمرکز بر روی دو رده است: حیوان (Animal) که یک رده‌ی عمومی است و فرض می‌شود که به همین شکل در سیستم نشان داده شده است، رده‌ی دیگر مختل شدن سلامتی (Health disorder) است که نوعی از رده‌ی پزشکی است. جزئیات در مورد ساختار این رده‌ها در جدول ۱ آمده است.

روابط هایپونیمی میان synset در Wordnet، بر اساس مشخصه‌ی یک مفهوم توسط Synset این گونه نمایش داده می‌شود $\{x,x',\dots\}$ که می‌گوید "هایپونیمی از مفهوم باش که توسط مترادف Synset نشان داده می‌شود $\{y,y',\dots\}$ ". در سال ۲۰۰۱ شبکه‌ی معنایی UMLS که ۱۳۴ انواع معنایی مربوط به حوزه زیست پزشکی را نشان می‌دهد و ۸۰۰ هزار مفهوم فرا اصطلاحنامه‌ای را مقوله‌بندی می‌کند منتشر شد. پیوند isa در آن اجازه داد که STs جانشین خواص گروه‌های سطوح بالاتر شود.

روش بررسی

بخشی از پژوهش در انفورماتیک پزشکی بر روی کاربردپذیری مجدد دانش در کاربردهای تازه و طراحی هستی شناسی‌های قابل اشتراک متمرکز است. موانع اصلی بر سر راه اشتراک دانش و استفاده‌ی دوباره‌ی آن از دیدگاه Musen سختی‌هایی است که در کسب اتفاق آرا در مورد آن چه که دانش معانی را نشان می‌دهد و ویژگی‌های موضوعی مورد نیاز دانش برای نسبت دادن معنی به دانش ارایه شده در فرایندهای خاص است. بنابراین، تلاش‌های متعددی در این زمینه انجام شده است:

۱. ترکیب هستی شناسی‌های عمومی
۲. استفاده دوباره آنتولوژی‌های حوزه‌ای برای کاربردهای خاص

۳. یک پارچه‌سازی هستی شناسی حوزه‌ای در مقیاس وسیع کتابخانه‌های هستی شناسی.

هدف اصلی در این بخش گسترش روش‌هایی است که به وسیله‌ی آن هستی شناسی‌ها به توانند از منابع موجود (شامل UMLS) ایجاد شوند به طوری که در مقابل دیگر منابع دانش معتبر باشند. بنابراین، این مقاله بر روی UMLS به عنوان یک هستی شناسی موضوعی فعال در حوزه‌ی زیست پزشکی متمرکز می‌گردد.

در ابتدا قابلیت‌های موجود در مقوله‌های UMLS با مقوله‌های هستی شناسی‌های ULOS سنجیده می‌شود.

تعریف می‌کند. بنابراین، تب در هر دو سیستم متعلق به رده‌ی اختلال سلامتی است.

این پژوهش با استفاده از روش سندی و کتابخانه‌ای به بررسی و مقایسه‌ی جریان موجود در گروه‌های دسته‌بندی شده در UMLS و WordNet پرداخته است.

یافته‌ها

هستی‌شناسی سطح بالای Cyc در UMLS:

قریب به ۵۰ مقوله‌ی Cyc به طور دقیق برای پوشش دادن شبکه معنایی UMLS استفاده شده است که نزدیک به نیمی از آن‌ها در هر دو سیستم به یکدیگر شبیه بودند. به عنوان مثال: ماهی (Fish) شبیه به Cyc و UMLS بود. در دیگر موارد، میان انواع Cyc و شبکه معنایی UMLS همپوشانی وجود داشت. به عنوان مثال شرایط ژنتیکی در Cyc شرایط نابهنجاری را نشان می‌دهد که در یک ارگانسیم خاص به علت وضعیت ژنتیکی گسترش یافته است، بنابراین،

فهرستی از واژگان و مفاهیم متعلق به رده‌های UMLS و واژه‌های مطابقت داده شده در WordNet به عنوان ابزار پژوهش در نظر گرفته شده است.

به منظور تجزیه و تحلیل داده‌ها، کار با فهرستی از واژگان و مفاهیمی که به رده‌های UMLS تعلق داشت و واژه‌هایی که در WordNet مطابقت داده شده بود آغاز شد. عمل تطابق بر اساس عملکردهای استاندارد WordNet بود. برای هر مفهوم این نکته که آیا واژه تطبیق داده شده WordNet به رده‌ی مشابه آن در WordNet تعلق دارد یا نه ثبت می‌شد. به عنوان مثال تب (Fever) در UMLS به عنوان یک نشانه یا علامت (Sign or symptom) دسته‌بندی شده بود اما در WordNet هاپیونیمی از نشانه است. همان گونه که در جدول ۱ مشاهده می‌شود نشانه یا علامت (Sing or symptom) یکی از STs است که رده‌ی اختلال سلامتی را در UMLS تعریف می‌کند. نشانه (Symptom) یکی از Synsetها است که رده‌ی اختلال سلامتی را در WordNet

جدول ۱: رده‌های حیوان و اختلال سلامتی (۳)

UMLS	WordNet	رده
تمامی مفاهیم فرا اصطلاح‌نامه که به انواع معنایی حیوان یا هرگونه زیرمجموعه آن اختصاص دارد.	مترادف Animal و تمامی هاپیونیم‌هایش	حیوان (Animal)
مفاهیم اصطلاح‌نامه که به هرگونه انواع معنایی اختصاص دارد: Anatomical abnormality Congenital abnormality Acquired abnormality Finding Sing or symptom Pathologic function Disease or syndrome Mental or behavioral dysfunct Neoplastic process Cell or molecular dysfunct Experimental model of disease Injury or poisoning	مترادف‌های یک پارچه و تمامی هاپیونیم‌هایش در زیر آمده است: Symptom Ill health Disorder (sense 1) Mental retardation Mental illness Defect (sense 1) Abnormalcy	اختلال سلامتی (Health disorder)

۴۸ درصد) متعلق به WordNet در رده‌ی اختلال سلامتی است.

ممکن است در یک رده، مفاهیم به طور متفاوتی دسته‌بندی شوند، حتی زمانی که دسته‌بندی‌ها شبیه به هم به نظر برسند. به عنوان مثال: Symptom دارای معانی مشابهی در WordNet است، جایی که "هر گونه احساس یا تغییر در عملکرد بدنی وجود دارد که توسط یک بیمار تجربه شده است و همراه با یک بیماری خاص می‌باشد". در UMLS جایی که Sing or Symptom است که «مشاهده‌ی ظهور یک بیماری، یا موقعیتی که بر اساس قضاوت پزشکی است یا جلوه‌ای از یک بیماری قابل مشاهده است» این مشابهت معنایی منجر به تناسب بالایی از مفاهیم دسته‌بندی شده است که در هر دو سیستم به هم شبیه هستند به عنوان مثال یرقان Cyanosis، تب.

بحث

نتایج حاصل از تحقیق در بخش گوناگون هستی‌شناسی‌ها نشان داد که در بخش تنوع در ساختار، با وجود تلاش‌های جاری گروه کاری استاندارد IEEE، هم اکنون هیچ استاندارد ULO در دسترس نیست. به طور مثال در نظریه John Sowa ۵ مقوله با هم ترکیب می‌شوند: چکیده، جسمی، مستقل، نسبت، میانجی‌گری. چنین تنوع ULOS نشان می‌دهد که مشکلات بالقوه‌ای (پنهانی) در تعمیم بخشیدن یافته‌ها وجود دارد. تنوع در ظاهر، زمانی که UMLS میان مقوله‌های سطح بالا در شبکه‌ی معنایی و مفاهیم سطح پایین در فرا اصطلاح‌نامه تمایز ایجاد کند، در این صورت هم Cyc و هم WordNet از ظاهر منحصر به فردی استفاده کرده است. هر چند که سطوح مفاهیم در آن‌ها سلسله‌مراتبی است. ظاهر نمایش مقوله‌های سطح بالا، ویژگی‌ها و نقش‌ها را تحت تأثیر قرار می‌دهد، ویژگی‌هایی مانند غیر انسانی، در بعضی از ULOS‌ها دسته‌بندی می‌شوند در حالی که آن‌ها ویژگی‌های STs در UMLS هستند.

نه عملکرد ژنتیک و نه سلول یا عدم کارایی مولوکولی در UMLS، بلکه همه‌ی آن‌ها به طور کلی مطابقت داده شدند. اجزای مقوله‌ی Cyc ناحیه‌ی بدن حیوانات، مناطق ناسالم بدن است مانند تاول‌ها و زخم‌های ترک خورده، که در UMLS در زیر مجموعه‌های نابهنجاری قرار گرفتند. Cyc، بخش بدن حیوانات زیرمجموعه‌ای از منطقه‌ی بدن حیوان است که هم شامل ارگان‌ها و هم شامل سیستم‌های بدنی است.

برای بعضی از گونه‌های معنایی UMLS (مانند شیمیایی‌ها) مقوله‌های همسان و مشابهی در نسخه‌ی عمومی هستی‌شناسی سطح بالایی Cyc وجود ندارد. علاوه بر این، مقوله‌هایی در Cyc وجود دارد که دارای هیچ گونه معادلی در UMLS نیست، اما در پژوهش حاضر، در UMLS یک پارچه شدند. برای این که ساختاری ایجاد شود که با Cyc سازگار باشد موارد زیر ایجاد شدند:

۱. گروه‌های میانجی، مانند پستانداران (پستانداران انسان (#\$genls)، (پستاندار پستانداران (\$genls))
 ۲. مفاهیم کلی مانند درمان ساده که زیرمجموعه‌ای از عمل معالجه‌ی پزشکی است.
 ۳. دانش افزوده، مانند زیست‌شناسی گیاهی که اطلاعاتی را در مورد مقوله‌های زیست‌شناسی بر اساس رده‌بندی جانداران عمومی از نحوه‌ی زندگی آن‌ها ارایه می‌کند. مقایسه‌ی رده‌های معنایی UMLS و WordNet:
- مجموعه‌ای از ۱۱۶۳۴ مفهوم در UMLS از رده‌ی حیوان (Animal) در WordNet مطابقت داده شده است که رده‌ی Animal شامل ۳۹۸۴ Synset می‌باشد. تعداد ۲۱۵۴ مفهوم UMLS (معادل ۱۹ درصد) در WordNet پیدا شده است. ۷۳ درصد از آن‌ها در رده‌ی Animal در WordNet بودند. رده‌ی اختلال سلامتی در UMLS شامل بیش از ۱۴۰ هزار مفهوم است که در WordNet مطابقت داده شده است. تعداد ۲۶۳۹ مفهوم در UMLS (معادل ۲ درصد) در WordNet پیدا گردید و در بین آن‌ها ۱،۲۵۷ مفهوم (معادل

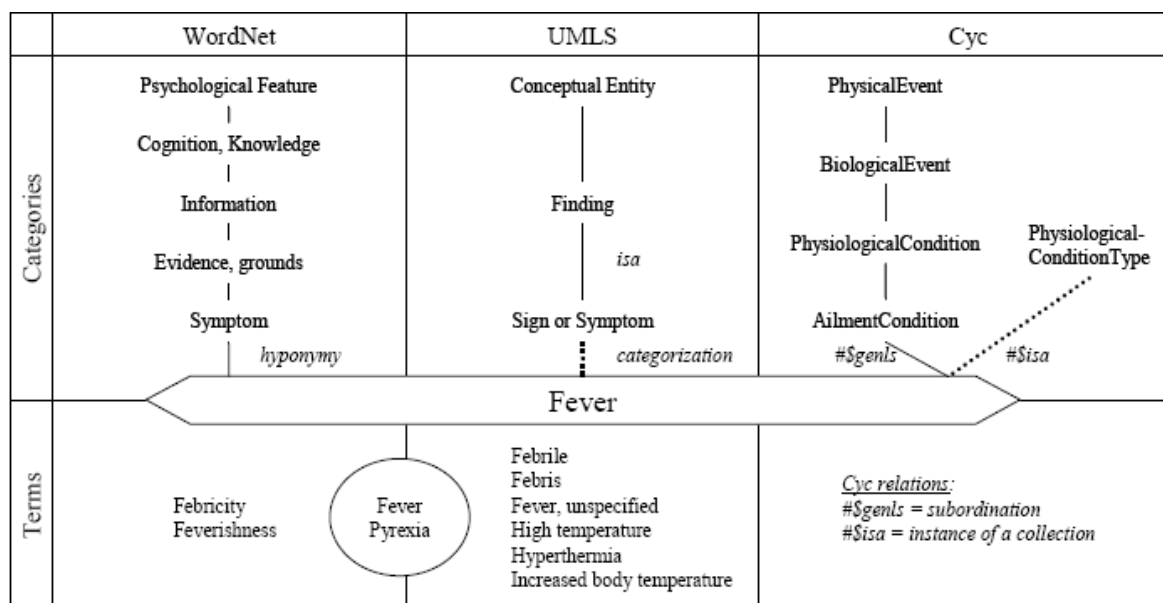
یک حوزه منعکس می‌شود بعضی از آنتولوژی‌ها بر اساس آن‌ها نمایش داده می‌شود. به عنوان مثال در WordNet، صرع (Epilepsy) تخریب سیستم عصبی مرکزی است که به عنوان فقدان آگاهی و تشنج شناخته شده است. برای متخصصین سلامتی، اگر چه این تعریف تنها به شکلی از پزشکی به نام صرع مربوط می‌شود. اما جنبه‌ی دیگر آن دانش، منتج از عقل سلیم است که توسط ارایه مفاهیم تهیه شده از آنتولوژی‌های زبان عمومی نمایش داده می‌شود.

ابر دانش:

آنتولوژی‌ها اغلب استفاده از ابر دانش را برای نشان دادن نه تنها جهان خود بلکه برای ارایه مدلی از آن جهان تهیه می‌کنند. رده‌بندی گیاهی زیست پزشکی که رده‌ای از ارگانیسم‌های زنده در سلسله‌ها، گونه‌های بشری و غیره است مدل سازمانی را فراهم می‌کند، اما دنیای آن را تعریف نمی‌کند. چنین ابر دانشی ممکن است توسط مقوله‌ها نمایش داده شود. به عنوان مثال زمانی که Cyc و WordNet، مقوله‌هایی را برای رده‌های گیاهی زیست پزشکی دارند، UMLS آن مقوله را دارا نیست. در Cyc رده‌ی گیاهی زیست پزشکی گونه‌ای از

هر آنتولوژی نه تنها چشم اندازی از دنیای خویش را نشان می‌دهد بلکه بخش‌های مختلف یک دانش را نیز به تصویر می‌کشد (شکل ۲). در حوزه‌ی زیست شناسی، نقشه برداری میان آنتولوژی‌های حوزه‌ای و عمومی، تطابق میان معنای عمومی و پزشکی خاص را شامل می‌شود.

مفاهیم عمومی: مفاهیم عمومی به مفاهیم کلی مانند Path یا Simple repairing مربوط می‌شود. ULOs به مفهوم‌سازی که یکسان به نظر می‌آید استناد می‌کند. Path (مسیر) برای سفر به کار می‌رود اما می‌تواند برای دستگاه گردش خون نیز کاربردپذیر باشد. Simple repairing به طور مجازی برای تعمیرات در فرایندهای هر حوزه‌ای کاربرد دارد (مکانیزم، روش‌های پزشکی و ...) به همین نحو دلیلی وجود ندارد که آناتومی از مقوله‌های عمومی که نشان دهنده‌ی اهداف فضایی است استفاده نکند (خط، فضا، ...). بنابر روش این مقاله در تطابق میان آنتولوژی‌های حوزه و عمومی ممکن است که توجه به بررسی اعتبار آنتولوژی‌های عمومی پایه باشد. دانش منتج از عقل سلیم: در قالب نمایش گروهی حوزه‌ی زیست پزشکی بعضی اوقات دانش به کار رفته در آن متفاوت از دانش تخصصی است و زمانی که نظریه‌های دیگری از علوم از



شکل ۲: تب در WordNet، Cyc و UMLS: مقوله‌بندی و همبستگی پدیده زبان‌شناختی (۳)

نشان داد که در بخش تنوع در ساختار، با وجود تلاش‌های جاری گروه کاری استاندارد IEEE، هم اکنون هیچ استاندارد ULO در دسترس نیست. در بخش مفاهیم عمومی مشخص شد که مفاهیم عمومی به مفاهیم کلی مربوط است. علاوه بر این، آنتولوژی‌ها اغلب استفاده از ابر دانش را برای نشان دادن نه تنها جهان خود بلکه برای ارایه مدلی از آن جهان تهیه می‌کنند. همچنین در قالب نمایش گروهی حوزه‌ی زیست پزشکی، بعضی اوقات دانش به کار رفته در آن متفاوت از دانش تخصصی است و زمانی که نظریه‌های دیگری از علوم از یک حوزه منعکس می‌شود بعضی از آنتولوژی‌ها بر اساس آن‌ها نمایش داده می‌شود.

پیشنهادها

پیشنهاد می‌شود که برای خودکارسازی تطابق میان آنتولوژی‌ها، الگوریتم‌های منطقی اختصاص داده شود که دو موضوع معنایی اصلی را مخاطب قرار دهد تا به توان هستی‌شناسی‌های تطبیقی گوناگون را منتشر کرد، نظایر:

۱. مقوله‌هایی که در آنتولوژی‌های مختلف دارای اسامی مشابه هستند، اما از لحاظ معنایی تفاوت دارند. به عنوان مثال: هویت، بخش بدنی، منطقه‌ی بدنی، معنی مشابهی در Cyc و UMLS ندارد.
۲. دو مقوله که به شدت به هم شبیه باشند، اما نه رده‌های ترتیبی آن‌ها مساوی و یکسان باشد و نه این که به طور کلی رده‌ی دیگری را شامل شود، همان گونه که توسط گسترش‌های Symptom در WordNet و UMLS نشان داده شده است.

رده‌ی زیست پزشکی است که به فرزند گونه‌ی رده‌بندی مرسوم است که دیگر فرزندان او ابر دانش را نشان می‌دهد. انتخاب ما برای پیشبرد نمایش ابر دانش، به شدت تحت تأثیر توانایی سیستمی است که از رابطه‌ی رده‌بندی استنتاج می‌شود. به عنوان مثال همان طور که در شکل ۲ نشان داده شده است می‌توان استنتاج کرد که Cyc به رابطه‌ی \$genls# محدود شده است در حالی که Fever یک رویداد زیست پزشکی است. طبق رابطه‌های رده‌بندی گیاهی، Fever هدف غایی در مقوله‌ی Sychological feature در WordNet است و به عنوان یک هویت مفهومی در UMLS به کار می‌رود.

نتیجه‌گیری

نتایج مطالعات سندی نشان داد که از ۵۰ مقوله‌ای که در Cyc به طور دقیق برای پوشش دادن شبکه‌ی معنایی UMLS استفاده شده است، قریب به نیمی از آن‌ها در هر دو سیستم شبیه به هم هستند و میان انواع Cyc و شبکه معنایی UMLS همپوشانی وجود دارد. همچنین مشخص شده است که برای بعضی از گونه‌های معنایی UMLS، مقوله‌های همسان و مشابهی در نسخه‌ی ULOS هستی‌شناسی سلسله مراتبی Cyc وجود ندارد. علاوه بر این، مقوله‌هایی در Cyc موجود است که دارای هیچ گونه معادلی در UMLS نیست. مقایسه‌ی رده‌های معنایی UMLS و WordNet نتایج نشان داد که ۲ درصد از مفاهیم UMLS از رده‌ی اختلال سلامتی در WordNet وجود داشت و مطابقت و سازگاری میان رده‌ها ۴۸ درصد بود. هم چنین نتایج حاصل از تحقیق در بخش گوناگونی هستی‌شناسی‌ها

Reference

1. Jacob EK. Ontologies and the Semantic Web. Trans. Sheykh Shoaee F, Tehran: Quarterly Ketab 2005; 63(4): 189-94.
2. Sanatjoo A. Necessity of the revision in thesauri structures: a review of thesauri inefficiency in the new information environment and ontology abilities in comparison with them. Quarterly ketab 2005; 16(4): 79-92.
3. A, Bodenreider O. Mapping the UMLS Semantic Network into general ontologies. Proc AMIA Symp 2001; 81-85.
4. Aitchison J, Gillchrist A, Bawden D. Thesaurus construction and use: a practical manual. Trans. Azizi M, Tehran: IRAN DOC 2006. p. 252-6.

5. Bodenreider. "The unified medical language system (UMLS): integrating biomedical terminology". *Nucleic Acids research* 2004; 32(database issue): D267-D70.
6. Mohammedsmaeil S, Samadi L. "Metathesaurus: Introduction, necessity and existence philosophy". *Kolliyat-Mah* 2006; 10(1-2): 31-42
7. Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: representing different views of biomedical concepts. *Bull Med Libr Assoc* 1993; 81(2): 217-22.

Ontology of Unified Medical Language System (UMLS) Metathesaurus* *Sedigheh Mohammadesmaeil, PhD¹; Laleh Samadi²*

Abstract

Introduction: For many years, Ontologies have been available in many different forms, for example: classification schema, thesauruses, Authorized vocabularies, Terminologies and dictionaries. Terminology and knowledge resources are essential components of interoperability among disparate systems. So, this paper considered the ontology of UMLS (which was prepared by National Library Medicine of American and it is a part of Unified Medical Language System). In addition to, we evaluate whether names and relationships needed in biomedical informatics are present in the UMLS, and analyzed the compatibility between an ontology of the biomedical domain (the UMLS semantic network) and two other Ontologies : the Upper Cyc Ontology (UCO) and WordNet.

Methods: In this study which has been done by the documentary and library methods (print & digital existing documentatins) about comparison between the UMLS and the other Ontologies (such as WordNet and Cyc), we try to consider the works 's like Burgun and Bodenerreider, who surveyed on this field (and extracted the words which are related to the Biomedical informatics), in order to show the existence of relationships between them, as well as to determine and map them to the concept's of UMLS, and it's resources of it.

Cyc categories that had no equivalent in the UMLS were integrated in our representation of the UMLS in order to build structure that was consistent with Cyc. They represent:

Intermediate nodes, such as Primate (#\$ genls Person Primate), (#\$genls Primate Mammal)

Generic concepts, such as simple repairing which is a super type of medical treatment event.

Additional knowledge, such as biological taxon which provides information about biological categories, according to the general taxonomy of living beings.

Analyzing in WordNet was held in two categories of Animal and Health disorder.

Results: Roughly 50 Cyc categories were used for strictly covering the UMLS Semantic Network field. Approximately half of them were similar in both systems, for example Fish is similar in Cyc and in the UMLS. For the others, there was overlap between the Cyc and semantic network of UMLS. Thus it maps totally neither genetic function nor cell or molecular dysfunction in UMLS. Representation of anatomy differs, since elements of Cyc category Animal Body Region may be unhealthy body regions such as blisters, puncture wounds, which are abnormalities in the UMLS. In Cyc, Animal body part is a subtype of animal body region that includes both organs and body systems. For several UMLS semantic network (e.g. chemicals), there was no equivalent category in the public version of Upper Cyc Ontology. Additional

The set of 11,634 UMLS concepts from the ANIMAL class was mapped to WordNet, whose ANIMAL class contains 3,984 synsets. 2,154 concepts (19%) were found in WordNet, 73% of them in the WordNet ANIMAL class. The UMLS Health disorder class contains more than 140000 concepts, which were mapped to WordNet, 2,639 UMLS concepts (2%) were found in WordNet, and among them, 1,257 concepts (48%) belonged to the WordNet health disorder class.

Conclusion: There is no standard ULO available yet. within a class, concepts may be categorized differently, even when the categories look similar.

Keywords: Terminology; Thesaurus; Dictionaries; Medical Information Storage & Retrieval.

Type of article: Original Article

Received: 3 Aug, 2008

Accepted: 27 Aug, 2009

Citation: Mohammadesmaeil S, Samadi L. **Ontology of Unified Medical Language System (UMLS) Metathesaurus.** Health Information Management 2010; 7(2): 160.

* This article resulted from independent research.

1. Assistant Professor, Library & Information Sciences, Research and Science Branch, Azad University, Tehran, Iran. (Corresponding Author) Email: m.esmaeili@sri.au.ac.ir

2. MSc, Library & Information Sciences, Research and Science Branch, Azad University, Tehran, Iran.