

بررسی هستان‌شناسی‌های توسعه یافته مبتنی بر اصول هستان‌شناسی‌های منبع باز زیست پزشکی

نادر عالیپشان کرمی^۱، محسن حاجی زین العابدینی^۲، ایرج رداد^۳،

سید جواد قاضی میرسعید^۴

چکیده

زمینه و هدف: هستان‌شناسی‌ها در زیست‌پزشکی کار ادغام، مبادله، جستجو و پرسش از داده‌ها را آسان می‌کنند. مخزن هستان‌شناسی‌های باز زیست‌پزشکی (مخزن) یک راه‌کار برای ایجاد هستان‌شناسی‌های مرجع می‌باشد که در آن طراحی هستان‌شناسی‌ها مبتنی بر اصولی است که تعامل آنها را به صورت یک نظام منفرد تبدیل نماید. هدف این مقاله تعیین وضعیت هستان‌شناسی‌های توسعه یافته طبق اصول مزبور می‌باشد.

روش بررسی: این پیمایش توصیفی-مقطعی در سال ۱۳۹۶ انجام شد. اطلاعات اجزای اصلی، موضوع، زبان و ابزار ساخت هستان‌شناسی‌ها از وب سایت مخزن هستان‌شناسی باز زیست پزشکی و OntoBee (سرور پیش فرض داده‌های پیوندی هستان‌شناسی‌های مخزن) بازیابی گردید. فراوانی کلاس‌ها، نمونه‌ها، روابط شیئی، روابط نوع-داده و روابط ابرداده‌ای حاشیه‌نویسی با استفاده از آمار توصیفی، دسته بندی و آرایه گردید.

یافته‌ها: موضوعات اصلی ۱۵۱ هستان‌شناسی موجود آناتومی، سلامت و فنوتیپ بود. بیشترین زبان و ابزار ساخت به ترتیب نسخه‌های OBO و OBO-Edit بودند. ۵ درصد هستان‌شناسی‌ها بیش از ۷۸۹۷۷ کلاس، ۱۵۰ رابطه شیئی، ۳۹ رابطه‌ی نوع-داده، ۱۱۰ رابطه‌ی ابرداده‌ای حاشیه‌نویسی و ۳۵۶ نمونه داشتند.

نتیجه‌گیری: افزایش تعداد هستان‌شناسی‌های مخزن و استفاده‌ی بیشتر از زبان OBO و ابزار OBO-Edit حاکی از پذیرش توسعه دهندگان هستان‌شناسی برای استفاده از این ابزار و زبان برای ساخت هستان‌شناسی مبتنی بر اصول مخزن است.

واژه‌های کلیدی: هستان‌شناسی، هستان‌شناسی زیست پزشکی، مخزن هستان‌شناسی‌های منبع باز زیست پزشکی

دریافت مقاله: اردیبهشت ۱۳۹۷

پذیرش مقاله: شهریور ۱۳۹۷

* نویسنده مسئول:

محسن حاجی زین العابدینی؛

دانشکده علوم تربیتی و روانشناسی دانشگاه

شهید بهشتی

Email:

m_zabedini@sbu.ac.ir

^۱ دانشجوی دکتری علم اطلاعات و دانش‌شناسی، دانشکده علوم انسانی، دانشگاه بین‌المللی امام رضاع، مشهد، ایران

^۲ استادیار گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید بهشتی، تهران، ایران

^۳ استادیار گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم انسانی، دانشگاه بین‌المللی امام رضاع، مشهد، ایران

^۴ دانشیار گروه کتابداری و اطلاع‌رسانی پزشکی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران

مقدمه

روند افزایشی اطلاعات زیست‌پزشکی نه تنها مدیریت دستی اینگونه اطلاعات را غیرممکن می‌سازد، بلکه مدیریت آن به وسیله رایانه‌ها را نیز با مشکلاتی روبرو می‌نماید. این مشکلات؛ از یک طرف، ناشی از ابهامات موجود در زبان طبیعی از قبیل استفاده از کلمات مختلف برای بیان یک مفهوم و یا استفاده از یک کلمه در متون مختلف برای مفاهیم متفاوت بوده؛ و از طرف دیگر، به علت گوناگونی شکل‌های داده‌ای و ساختارهای داده می‌باشد که یکپارچه‌سازی اطلاعات در نظام‌های اطلاعاتی را با مشکل روبرو می‌کند (۱). این امر، به نوبه‌ی خود استفاده‌ی مجدد از اطلاعات در نظام‌های دیگر را بسیار دشوار و یا عملاً غیرممکن می‌سازد. چنانچه بخواهیم داده‌های زیست‌پزشکی مفید واقع شوند، باید با دقت تمام، تعریف و توضیح داده شوند. به منظور نیل به این مهم، در علوم رایانه و هوش مصنوعی از ساختاری به نام هستان‌شناسی استفاده می‌شود. هستان‌شناسی؛ در حقیقت، واژگانی است که به خوبی تعریف شده‌است، و توصیف و شرح دقیق از مفاهیم در یک حیطه‌ی موضوعی با قابلیت فهم ماشینی را فراهم می‌آورد (۲ و ۱).

به طور کلی، دانش موجود در هستان‌شناسی‌ها با استفاده از هفت جزء رسمی‌سازی می‌شود؛ یا به عبارت دیگر، دارای قابلیت فهم ماشینی می‌گردد. این هفت جزء عبارتند از: نمونه‌ها (instances)، کلاس‌ها (classes)، ویژگی‌ها (attributes)، روابط (relations)، توابع (functions)، اصول بدیهی (axioms) و قواعد طرح‌ریزی (planning rules). "نمونه‌ها" در اغلب موارد اسامی ذات هستند؛ یعنی اینکه قائم به خود بوده و با حواس درک می‌شوند و در سایر موارد اسامی معنی می‌باشند. "کلاس‌ها" که نمونه‌ها و یا کلاس‌های دیگر را در خود جای می‌دهند، در تاکسونومی‌ها سازماندهی می‌شوند. یک مفهوم که در یک کلاس قرار می‌گیرد، می‌تواند هر آن چیزی باشد که در مورد آن در یک حیطه‌ی موضوعی، اطلاعاتی رد و بدل می‌شود، از جمله شرح یا توضیحی از یک کار، وظیفه، اقدام، استراتژی، فرایند و نظیر آنها. "ویژگی‌ها" جنبه‌ها، خواص، خصوصیات، مشخصات یا پارامترهایی هستند که نمونه‌ها و کلاس‌ها از آنها برخوردارند. "روابط" نوعی تعامل بین مفاهیم یک حیطه‌ی موضوعی را بازنمایی می‌کنند و نشان می‌دهند که کلاس‌ها و نمونه‌ها چگونه به یکدیگر مرتبط می‌شوند. "روابط" معمولاً بین کلاس‌ها تعریف می‌شود، و نمونه‌های روابط بین نمونه‌ها نیز برقرار می‌گردد. از "اصول بدیهی" برای مدل‌سازی جملات و گزاره‌هایی که در یک حیطه‌ی موضوعی همواره درست هستند

استفاده می‌شود. آنها همچنین بیان می‌دارند که چگونه می‌شود حقایق جدید را از حقایق موجود در هستان‌شناسی استخراج نمود. بخش آخر هستان‌شناسی نیز "قواعد طرح‌ریزی" می‌باشد که تشریح می‌کنند که حیطه‌ی موضوعی علمی که هستان‌شناسی برای آن ساخته شده است، چگونه ممکن است تغییر نماید (۳).

در زیست‌پزشکی، پرکاربردترین هستان‌شناسی، هستان‌شناسی ژن (۴) است. موفقیت حاصل از به کارگیری این هستان‌شناسی و هستان‌شناسی‌های مشابه در مدیریت اطلاعات این حوزه‌ی علمی، موجب افزایش کمی انواع این گونه هستان‌شناسی‌ها شده است (۵ و ۶). این افزایش تعداد هستان‌شناسی‌ها، به دلیل عدم استفاده از روش‌شناسی‌های مشترک در طراحی، ساخت، نگهداری و توسعه‌ی آنها، خود به مانعی در راه یکپارچه‌سازی اطلاعات تبدیل شده‌است (۷). در راستای سازماندهی هستان‌شناسی‌های زیست‌پزشکی دو منبع اصلی اینترنتی آنها را به اشتراک گذاشته‌اند: (۱) بایوپورتال (۸) که تا شهریور ۱۳۹۶ دارای بیش از ۶۶۰ هستان‌شناسی بود و نگهداری آن توسط مرکز ملی هستان‌شناسی‌های زیست‌پزشکی وابسته به موسسه ملی سلامت ایالات متحد آمریکا صورت می‌گیرد؛ و (۲) نظام جستجوی هستان‌شناسی (۹) با برخورداری از بیش از ۲۰۰ هستان‌شناسی متعلق به موسسه بیوانفورماتیک اروپاست.

در تلاشی برای مدیریت و غلبه بر مشکلات ناشی از روش‌شناسی‌های مختلف تولید هستان‌شناسی‌های زیست‌پزشکی، کنسرسیوم هستان‌شناسی‌های منبع باز زیست‌پزشکی (۷) یک راهبرد را دنبال می‌کند. این راهبرد عبارت است از ایجاد هستان‌شناسی‌های جدید بر اساس یک مجموعه از اصول -شامل: (۱) منبع باز بودن، (۲) فرمت یا شکل مشترک، (۳) فضای شناساگر، (۴) نسخه‌بندی، (۵) دامنه، (۶) تعاریف متنی، (۷) روابط، (۸) مستندسازی، (۹) مستندسازی کاربران هستان‌شناسی، (۱۰) تعهد به مشارکت، (۱۱) منبع قدرت، (۱۲) قوانین نام‌گذاری و (۱۳) نگهداری- به اشتراک گذاشته شده که خود نیز در حال رشد و تکامل هستند. در نتیجه‌ی توسعه‌ی هستان‌شناسی‌ها بر مبنای این اصول، فرض بر این است تا خانواده‌ای از هستان‌شناسی‌ها ایجاد شود که از آغاز برای میان‌کنش‌پذیری، نامتناقض بودن، منطقی بودن و بازنمایی درست و دقیق موجودیت‌های زیست‌پزشکی طراحی شده باشند. در این نوآوری، طراحی هستان‌شناسی‌ها به گونه‌ای است که بتوانند با یکدیگر به نحوی کار کنند که یک نظام منفرد و بدون افزونگی ایجاد شود. هستان‌شناسی‌های توسعه‌یافته بر اساس اصول مزبور پس

وبسایت کنسرسیوم قابل دسترسی می‌باشند. در این وب سایت اطلاعات اولیه نظیر PURL یا آدرس دائمی هستان‌شناسی که کاربر را به فایل هستان‌شناسی راهنمایی می‌کند، نوع حق چاپ و نشر، آدرس وب سایت هستان‌شناسی که توسط توسعه دهندگان هستان‌شناسی نگهداری می‌شود، آدرس پست الکترونیک توسعه دهنده، آدرس ردیابی تغییرات هستان‌شناسی GitHub و حیطه‌ی موضوعی یا Domain (دامنه) ارائه می‌شود (۱۴). اطلاعات تکمیلی در مورد موضوع، زبان ساخت هستان‌شناسی، مشخصات اجزای اصلی هستان‌شناسی که از ویژگی‌های مهم در ساخت هستان‌شناسی‌ها محسوب می‌شوند در یک سرور و مرورگر جداگانه داده‌های پیوندی مبتنی بر وب به نام OntoBee (۱۶ و ۱۷) سازماندهی می‌شود و در دسترس همگان قرار می‌گیرد.

این پژوهش با توجه به روند توسعه‌ی هستان‌شناسی‌های زیست‌پزشکی مبتنی بر اصول کنسرسیوم هستان‌شناسی‌های منبع باز زیست‌پزشکی بر آن است تا تصویری از وضعیت زبان‌ها و ابزارهای ساخت هستان‌شناسی و همچنین اجزای هستان‌شناسی‌های زیست‌پزشکی که تاکنون در مخزن مزبور ساخته شده‌اند، ارائه نماید.

روش بررسی

جامعه‌ی پژوهش این پیمایش توصیفی-مقطعی را که در شهر یور ۱۳۹۶ انجام گردید، کلیه‌ی هستان‌شناسی‌های منبع باز زیست‌پزشکی به اشتراک گذاشته شده در مخزن هستان‌شناسی‌های منبع باز زیست‌پزشکی (<http://www.obofoundry.org>) تشکیل دادند که اطلاعات آنها در وب سایت مزبور قابل دسترسی بود.

در این پژوهش، اطلاعات توصیفی هستان‌شناسی‌ها از قبیل: نام، پیشوند، حیطه موضوعی، ابزار ساخت، زبان‌های ساخت هستان‌شناسی، تعداد کلاس‌ها، تعداد نمونه‌ها، تعداد روابط بین کلاس‌ها (روابط شیئی (Object Property))، تعداد روابط نوع-داده (Data Type Property) یا روابط بین نمونه‌ها و مقادیر یا ارزش‌ها، و تعداد روابط ابرداده‌ای حاشیه‌نویسی (Annotation Property)، که از آنها برای افزودن اطلاعات به کلاس‌ها، نمونه‌ها و انواع روابط در هستان‌شناسی استفاده می‌شود از تمام هستان‌شناسی‌های موجود و فعال در وب سایت کنسرسیوم با روش زیر استخراج شد:

ابتدا سایت مخزن (<http://www.obofoundry.org>) باز شد. لیست هستان‌شناسی‌ها در قسمت پایین صفحه اول سایت قرار داشت. هستان‌شناسی‌ها در یک جدول قرار داشتند. هستان‌شناسی‌های منسوخ

از تایید متخصصان همتا، در مخزن این کنسرسیوم منبع باز در دسترس همگان قرار می‌گیرند (۱۰) درست همان‌گونه که مقالات ارسالی به مجلات علمی پس از داوری توسط همتایان در دسترس جوامع علمی قرار می‌گیرند. توسعه‌دهندگان (سازندگان) هستان‌شناسی‌هایی که مایلند هستان‌شناسی‌شان عضو این مخزن باشد، باید از این اصول پیروی کنند. این افراد باید با متخصصان مخزن تعامل داشته باشند، تا این اطمینان حاصل شود که محصولشان طبق اصول مزبور توسعه یافته است (۷).

سیزده اصل کنسرسیوم مزبور، الزامات توسعه‌ی هستان‌شناسی‌ها را تعیین می‌کنند. بر همین اساس مجموعه‌ای از هستان‌شناسی‌های توسعه یافته مبتنی بر اصول ذکر شده در وب سایت کنسرسیوم گردآوری شده و قابل دسترسی هستند. مطالعات نشان می‌دهند هستان‌شناسی‌های موجود در مخزن از این اصول پیروی می‌کنند (۱۱) و به واسطه‌ی افزایش استفاده‌ی مجدد (۱۲) از واژگان بین هستان‌شناسی‌های این مخزن، میزان تعامد (orthogonality) هستان‌شناسی‌ها افزایش یافته است (۱۳). مطالعه‌ای که به منظور تجزیه و تحلیل اولیه هستان‌شناسی‌های مخزن و همچنین رشد نسخه‌های مختلف هر یک از این هستان‌شناسی‌ها در سال ۲۰۱۷ با استفاده از اکوآر (OQuARE) گزارش شد نشان داد تمامی هشت هستان‌شناسی مورد مطالعه از اصول مخزن پیروی می‌کنند و رشد و توسعه نسخه‌های جدید هستان‌شناسی‌ها دارای امتیازات کیفی بالاتر اکوآر هستند (۱۱).

نکته‌ی قابل تأمل این است که اصول مخزن محدودیتی برای نوع زبان و ابزار ساخت هستان‌شناسی، و همچنین کمیت اجزای سازنده‌ی هستان‌شناسی اعمال نمی‌کنند. اطلاعات فراهم آمده در وب سایت مخزن برای توسعه دهندگان هستان‌شناسی‌ها که مایلند محصول آنها طبق ۱۳ اصل کنسرسیوم توسعه یابد، به‌ویژه در گام‌های آغازین مهندسی هستان‌شناسی که همانا شناسایی و استفاده‌ی مجدد از هستان‌شناسی‌های موجود در یک حیطه‌ی موضوعی است، کافی نیست و آنها را با چالش‌هایی مواجه می‌نماید. توسعه دهندگان نیاز دارند که بدانند در حوزه‌های مختلف زیست‌پزشکی از کدام زبان‌ها و ابزار ساخت هستان‌شناسی استفاده می‌شود. بررسی‌ها همچنین نشان می‌دهد مطالعه‌ای که به طور کلی وضعیت هستان‌شناسی‌های کنسرسیوم را به لحاظ ابزارها و زبان‌های مورد استفاده، حیطه موضوعی، تعداد کلاس‌ها، نمونه‌ها و انواع روابط بین مفاهیم بررسی کرده باشد، انجام نشده است. تا زمان انجام این مطالعه، بیش از ۱۵۰ هستان‌شناسی زیست‌پزشکی طبق اصول مزبور توسعه یافته است که از طریق



بر اساس اصول سیزده‌گانه کنسرسیوم توسعه یافته‌اند برای علاقمندان حوزه هستان‌شناسی‌های زیست‌پزشکی فراهم آید.

Ontobee که از طریق آدرس <http://www.ontobee.org> قابل دسترسی است، یک سرور و مرورگر داده پیوندی مبتنی بر وب می‌باشد که به طور خاص برای واژه‌های هستان‌شناسی تهیه شده است. هدف از Ontobee تسهیل به اشتراک‌گذاری داده‌های هستان‌شناسی، تجسم یا تصویرسازی از داده‌های هستان‌شناسی، پرسش از هستان‌شناسی‌ها، یکپارچه‌سازی و تجزیه و تحلیل هستان‌شناسی‌ها می‌باشد. این مرورگر یک رابط کاربری کاربرپسند برای نمایش جزئیات و سلسله مراتب یک واژه‌ی خاص در هستان‌شناسی ارائه می‌کند. این ابزار یک روش کارآمد با قابلیت دسترسی برای عموم برای ارتقای فرآیندهای به اشتراک گذاشتن، میان‌کنش پذیری و یکپارچه‌سازی داده‌های هستان‌شناسی فراهم می‌آورد (۱۷).

یافته‌ها

بررسی وب سایت مخزن هستان‌شناسی‌های منبع باز زیست‌پزشکی در زمان انجام پژوهش (شهریور ۱۳۹۶) نشان داد که ۱۵۱ هستان‌شناسی از طریق این وب سایت قابل دسترسی بود. این هستان‌شناسی‌ها طیف وسیعی از موضوعات زیست‌پزشکی را در بر می‌گرفتند.

شده که با رنگ زمینه‌ی خاکستری مشخص شده بودند، از مطالعه خارج شدند.

در گام بعد، با کلیک بر روی لینک هستان‌شناسی جزئیات بیشتری از هستان‌شناسی در صفحه جدید که در حقیقت پروفایل هستان‌شناسی در مخزن بود باز شد. اطلاعات موجود در این صفحه عبارت بودند: از معرفی انواع فایل هستان‌شناسی (نظیر owl و obo)، لینک وبسایت اصلی هستان‌شناسی، نوع حق نشر و چاپ، فضای شناسایی (ID Space)، PURL یا آدرس دائمی هستان‌شناسی، مسئول تماس، آدرس ردیابی مسایل (Trackers) در GitHub (۱۴) و حیطه‌ی موضوعی. لینک‌هایی نیز به سایر منابع وجود داشت از قبیل AberOWL، OntoBee (۱۷)، OLS (۱۸) و BioPortal (۲۱ و ۲۰ و ۷). از میان این موارد به منظور استخراج اطلاعات در مورد تعداد کلاس، تعداد نمونه، تعداد رابطه شیئی، رابطه نوع-داده و رابطه ابرداده‌ی حاشیه‌نویسی و همچنین فرمت‌های هستان‌شناسی از OntoBee استفاده شد. زیرا OntoBee سرور پیش‌فرض داده پیوندی هستان‌شناسی برای کتابخانه هستان‌شناسی‌های مخزن هستان‌شناسی‌های زیست‌پزشکی است. فراوانی کلاس‌ها، نمونه‌ها، روابط شیئی، روابط نوع-داده و رابطه ابرداده‌ای حاشیه‌نویسی با استفاده از آمار توصیفی دسته بندی و ارائه گردید تا به طور کلی تصویر روشنی از وضعیت کنونی هستان‌شناسی‌های زیست‌پزشکی که تاکنون

جدول ۱: فراوانی موضوعی هستان‌شناسی‌های به اشتراک گذاشته شده در مخزن هستان‌شناسی‌های منبع باز زیست‌پزشکی

ردیف	حیطه‌ی موضوعی	فراوانی	درصد فراوانی	ردیف	حیطه‌ی موضوعی	فراوانی	درصد فراوانی
۱	Not available	۳۸	۲۵/۱۷	۲۱	Food	۱	۰/۶۶
۲	adverse events	۱	۰/۶۶	۲۲	Fungal	۱	۰/۶۶
۳	Agronomy	۱	۰/۶۶	۲۳	genotype-to-phenotype-associations	۱	۰/۶۶
۴	Algorithms	۱	۰/۶۶	۲۴	Health	۱۸	۱۱/۹۲
۵	Anatomy	۲۷	۱۷/۸۸	۲۵	information	۱	۰/۶۶
۶	anatomy and development	۱	۰/۶۶	۲۶	Medicine	۲	۱/۳۲
۷	Behavior	۲	۱/۳۲	۲۷	molecular structure	۱	۰/۶۶
۸	Biochemistry	۳	۱/۹۹	۲۸	MS experiments	۱	۰/۶۶
۹	biological process	۱	۰/۶۶	۲۹	phenotype	۱۳	۸/۶۱
۱۰	biological sequence	۱	۰/۶۶	۳۰	plant phenotypes	۱	۰/۶۶
۱۱	Biology	۱	۰/۶۶	۳۱	Proteins	۲	۱/۳۲
۱۲	biomedical health	۱	۰/۶۶	۳۲	provenance	۱	۰/۶۶

۰/۶۶	۱	Relations	۳۳	۰/۶۶	۱	Cells	۱۳
۰/۶۶	۱	Resources	۳۴	۱/۹۹	۳	Clinical	۱۴
۰/۶۶	۱	scholarly contribution role	۳۵	۰/۶۶	۱	collections of organisms	۱۵
۰/۶۶	۱	Scientific claims· evidence	۳۶	۰/۶۶	۱	Development	۱۶
۱/۳۲	۲	Statistics	۳۷	۲/۶۵	۴	Disease	۱۷
۱/۹۹	۳	Taxonomy	۳۸	۰/۶۶	۱	Ecological functions and ecological interactions	۱۸
۰/۶۶	۱	Upper	۳۹	۲/۶۵	۴	Environment	۱۹
*	*	****	*	۳/۳۱	۵	Experiments	۲۰

شده است. بیشترین فراوانی موضوعی هستان‌شناسی‌ها به ترتیب در آناتومی، سلامت و فنوتیپ بود. به استثنای ۲۵/۱۷ درصد از موارد که موضوع آنها قابل بازیابی نبود، سایر موارد، موضوع کلی حیطه موضوعی خود را مشخص کرده بودند.

چنانچه جدول ۱ نشان می‌دهد هستان‌شناسی‌هایی را که حیطه موضوعی آنها قابل بازیابی بوده است، می‌توان به ۳۹ دسته موضوع طبقه بندی نمود. در این جدول موضوعات به ترتیب الفبایی ارایه شده‌اند و فراوانی هستان‌شناسی‌ها در هر یک از موضوعات نمایش داده

جدول ۲: فراوانی انواع ابزارهای سافت هستان‌شناسی‌های در مخزن هستان‌شناسی‌های منبع باز زیست‌پزشکی

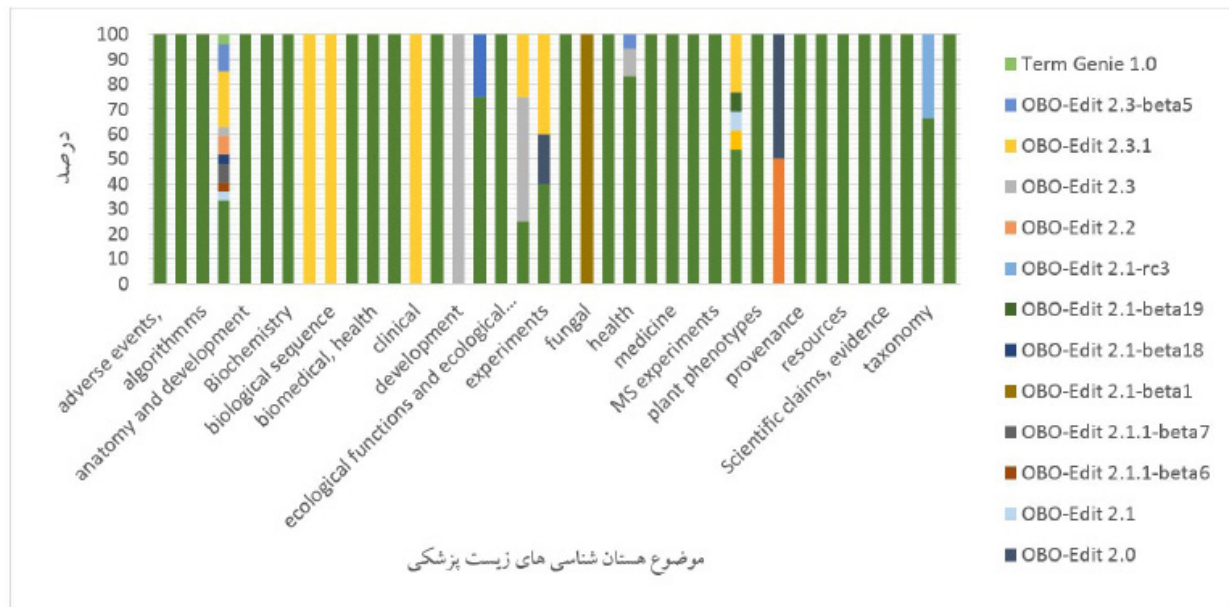
جمع کل	عدم ثبت اطلاعات	Term Genie 1.0	BO-Edit 2.3-beta5	OBO-Edit 2.3.1	OBO-Edit 2.3	OBO-Edit 2.2	OBO-Edit 2.1-rc3	OBO-Edit 2.1-beta19	OBO-Edit 2.1-beta18	OBO-Edit 2.1-beta1	OBO-Edit 2.1.1-beta7	OBO-Edit 2.1.1-beta6	OBO-Edit 2.1	OBO-Edit 2.0	OBO-Edit 1.101	OBO-Edit 1.002	FormulaOBO.PL	DAG-Edit 1.419 rev 3	ابزارهای ساخت
۱۵۱	۹۸	۱	۴	۲۲	۶	۲	۱	۱	۱	۱	۲	۱	۳	۲	۲	۱	۱	۲	فراوانی
۱۰۰	۶۴/۹۰	۰/۶۶	۲/۶۵	۱۴/۵۷	۳/۹۷	۱/۳۲	۰/۶۶	۰/۶۶	۰/۶۶	۰/۶۶	۱/۳۲	۰/۶۶	۱/۹۹	۱/۳۲	۱/۳۲	۰/۶۶	۰/۶۶	۱/۳۲	درصد

OBO-Edit نیز نسخه OBO-Edit 2.3.1 بیشترین فراوانی را دارد. توسعه دهندگان هستان‌شناسی‌های مورد مطالعه از سایر ابزارها مثل Term Genie و FormulaOBO.PL، DAG-Edit نیز استفاده کرده‌اند.

بررسی نشان داد از ابزارهای مختلف ساخت هستان‌شناسی برای موضوعات مختلف زیست‌پزشکی استفاده می‌شود. هستان‌شناسی‌های موضوعات آناتومی و فنوتیپ که بیشترین فراوانی را دارند تنوع ابزاری بیشتری نیز دارند.

یکی از الزامات برای ساخت هستان‌شناسی‌ها استفاده از ابزار مناسب برای تولید هستان‌شناسی است. در ساخت هستان‌شناسی‌های منبع باز زیست‌پزشکی از ابزار مختلفی با رعایت اصل باز بودن هستان‌شناسی مورد استفاده قرار می‌گیرند (جدول ۲).

به طور کلی، ابزارهای مورد استفاده در ساخت هستان‌شناسی‌های مورد بررسی را می‌توان به ۴ دسته تقسیم کرد. صرف نظر از مواردی که نوع ابزار آنها در OntoBee ذکر نشده است، اکثر قریب به اتفاق هستان‌شناسی‌ها از نسخه‌های مختلف OBO-Edit برای ساخت هستان‌شناسی استفاده نموده‌اند (۵۹ مورد). بین نسخه‌های مختلف



جدول ۱: فراوانی موضوعی هستان‌شناسی‌های به اشتراک گذاشته شده در مخزن هستان‌شناسی‌های منبع باز زیست پزشکی

نکته دیگری که می‌باید در ساخت هستان‌شناسی‌ها مدنظر قرارگیرد، استفاده از زبان‌های ساخت هستان‌شناسی است. در ساخت هستان‌شناسی‌های منبع باز زیست‌پزشکی از زبان‌های باز ساخت هستان‌شناسی استفاده شد (جدول ۳).

از آنجاکه فراوانی هستان‌شناسی‌ها در برخی موضوعات در سطح حداقل بوده و یا حیطه‌ی موضوعی آنها ذکر نشده است و همچنین به علت مشخص نبودن ابزار ساخت آنها نمی‌توان به روشنی اذعان داشت که کدامیک از انواع ابزارها بیشترین کاربرد را در ساخت چه موضوعات هستان‌شناسی داشته است.

جدول ۳: فراوانی انواع زبان‌های سافت هستان‌شناسی‌های به اشتراک گذاشته شده در مخزن هستان‌شناسی‌های منبع باز زیست‌پزشکی

درصد	فراوانی	زبان هستان‌شناسی
۱/۳۲	۲	application/rdf+xml
۶/۶۲	۱۰	OBO
۱/۹۹	۳	OBO-1.1
۳۷/۰۹	۵۶	OBO-1.2
۰/۶۶	۱	OBO-1.4
۴/۶۴	۷	OWL-DL
۴۷/۶۸	۷۲	No
۱۰۰	۱۵۱	جمع کل

اجزای اصلی هستان‌شناسی‌های منبع باز زیست‌پزشکی که در این مطالعه امکان بررسی آنها وجود داشت عبارتند از: کلاس، رابطه شیئی، رابطه نوع-داده، رابطه ابرداده‌ای حاشیه‌نویسی و نمونه‌ها. جدول شماره ۴ وضعیت اجزای اصلی را طبق اطلاعات استخراج شده از OntoBee نشان می‌دهد.

بیشترین زبان‌های ساخت هستان‌شناسی در مخزن نیز متعلق به نسخه‌های مختلف OBO می‌باشد و سایر موارد نظیر application/rdf+xml و OWL-DL انگشت شمار است. این مطالعه نشان داد که اطلاعات زبان ساخت هستان‌شناسی‌های مورد بررسی در وب سایت کنسرسیوم و همچنین در OntoBee برای ۴۷/۶۸ درصد موارد ناقص است.

جدول ۴: آماره‌های مربوط به اجزای اصلی هستان‌شناسی‌های توسعه یافته بر اساس اصول کنسرسیوم هستان‌شناسی‌های منبع باز زیست‌پزشکی

اجزای هستان‌شناسی	تعداد هستان‌شناسی	تعداد مورد ثبت نشده	میانگین	انحراف معیار	حداقل	حداکثر	صدک ۱۵م	صدک ۲۵م	میانه	صدک ۱۷۵م	صدک ۱۹۵م	صدک ۱۹۹م
کلاس	۱۵۱	۸	۲۰۵۴۹/۹۸	۹۵۷۹۳/۷۷	۲۱	۱۰۰۰۰۰	۶۶	۳۷۱	۱۱۰۲	۴۴۲۷	۷۸۹۷۷	۴۳۴۶۶۳
رابطه شیئی	۱۵۱	۱۸	۵۷/۳۵	۱۰۲/۷۷	۱	۶۳۹	۱	۶	۲۵	۶۹	۱۵۰	۵۲۸
رابطه نوع-داده	۱۵۱	۱۰۹	۱۱/۱۲	۲۵/۹۲	۱	۱۵۹	۱	۲	۴	۷	۳۹	۱۵۹
رابطه ابرداده‌ای حاشیه‌نویسی	۱۵۱	۰	۴۰/۹۴	۳۵/۳۳	۰	۲۰۶	۰	۱۵	۳۱	۶۱	۱۱۰	۱۴۹
نمونه	۱۵۱	۱۰۴	۱۴۱۶۹/۱۹	۹۶۴۰۰/۸۵	۲	۶۶۰۹۹۳	۳	۱۸	۲۲	۳۵	۳۵۶	۶۶۰۹۹۳

بررسی کلاس‌ها در هستان‌شناسی‌ها نشان می‌دهد که تفاوت بسیار زیادی از نظر تعداد کلاس بین هستان‌شناسی‌های موجود در مخزن مشاهده می‌شود؛ نتایج مطالعه در صدک‌های مختلف به قرار زیر است:

• صدک پنجم برابر ۶۶ شده است که بدین معناست که ۵ درصد کل هستان‌شناسی‌ها کمتر از ۶۶ کلاس دارند.

• صدک ۱۲۵م (چارک اول) نشان می‌دهد که ۲۵ درصد هستان‌شناسی‌ها کمتر از ۳۷۱ کلاس دارند.

• صدک ۵۰م یا میانه نشان می‌دهد که ۵۰ درصد هستان‌شناسی‌ها دارای ۱۱۰۲ کلاس می‌باشند.

• صدک ۱۷۵م یا چارک سوم نشان دهنده‌ی اینست که ۲۵ درصد هستان‌شناسی‌ها بیش از ۴۴۲۷ کلاس دارند.

• صدک ۱۹۵م نشان می‌دهد که ۵ درصد هستان‌شناسی‌ها بیش از ۷۸۹۷۷ کلاس دارند.

چنانچه از بررسی تعداد کلاس‌ها بر می‌آید تفاوت زیادی بین تعداد کلاس‌های صدک‌های ۱۵م و ۱۹۵م وجود دارد. NCBI organismal classification با یک میلیون کلاس بالاترین تعداد کلاس را دارد و رتبه‌های بعدی متعلق به PRotein Ontology (۳۱۲۷۱۵)، The Drug Ontology (۴۳۴۶۶۳) و NCI Thesaurus OBO Edition (۱۳۰۰۴۳) می‌باشد.

بررسی رابطه شیئی بین کلاس‌های هستان‌شناسی‌های مورد مطالعه نتایج زیر را نشان می‌دهد:

• صدک پنجم برابر ۱ شده است که بدین معناست که ۵ درصد کل هستان‌شناسی‌ها کمتر از ۱ رابطه شیئی دارند.

• صدک ۱۲۵م (چارک اول) نشان می‌دهد که ۲۵ درصد

هستان‌شناسی‌ها کمتر از ۶ رابطه شیئی دارند.

• صدک ۱۵۰م یا میانه نشان می‌دهد که ۵۰ درصد هستان‌شناسی‌ها دارای ۲۵ رابطه شیئی می‌باشند.

• صدک ۱۷۵م نشان دهنده‌ی اینست که ۲۵ درصد هستان‌شناسی‌ها بیش از ۶۹ رابطه شیئی دارند.

• صدک ۱۹۵م نشان می‌دهد که ۵ درصد هستان‌شناسی‌ها بیش از ۱۵۰ رابطه شیئی دارند.

به لحاظ تعداد روابط شیئی نیز بین صدک‌های ۱۵م و ۱۹۵م تفاوت زیادی دیده می‌شود. بیشترین تعداد اینگونه روابط متعلق به RNA Ontology، Population and Community Ontology، Relations Ontology و Obstetric and Neonatal که به ترتیب با تعداد ۶۳۹، ۵۲۸، ۵۲۲، ۴۹۷ و ۴۴۰ رابطه شناسایی گردید.

بررسی آماری رابطه نوع-داده یا "روابط بین نمونه‌ها و مقادیر یا ارزش‌ها" در هستان‌شناسی‌های مورد مطالعه نتایج زیر را نشان می‌دهد:

• صدک پنجم برابر ۱ شده است که بدین معناست که ۵ درصد کل هستان‌شناسی‌ها کمتر از ۱ رابطه نوع-داده دارند.

• صدک ۱۲۵م (چارک اول) نشان می‌دهد که ۲۵ درصد هستان‌شناسی‌ها کمتر از ۲ رابطه نوع-داده دارند.

• صدک ۵۰م یا میانه نشان می‌دهد که ۵۰ درصد هستان‌شناسی‌ها دارای ۴ رابطه نوع-داده می‌باشند.

• صدک ۱۷۵م نشان دهنده‌ی اینست که ۲۵ درصد هستان‌شناسی‌ها بیش از ۷ رابطه نوع-داده دارند.

• صدک ۱۹۵م نشان می‌دهد که ۵ درصد هستان‌شناسی‌ها بیش از ۳۹ رابطه نوع-داده دارند.

از نظر تعداد روابط نوع-داده، بیشترین تعداد این نوع رابطه در



Non-Coding RNA Ontology, eagle-I resource ontology, Biological Collections Ontology, The Drug-Drug Interactions Ontology, و Ontology for Parasite Life Cycle به ترتیب با تعداد ۱۵۹، ۶۰، ۳۹، ۳۲ و ۱۷ رابطه مشاهده شد. از روابط ابر داده‌ای حاشیه نویسی برای افزودن اطلاعات به کلاس‌ها، نمونه‌ها و انواع روابط در هستان‌شناسی استفاده می‌شود. بررسی آماری رابطه ابر داده‌ای حاشیه‌نویسی هستان‌شناسی‌های مورد مطالعه نتایج زیر را نشان می‌دهد:

- صدک پنجم نشان می‌دهد که ۵ درصد کل هستان‌شناسی‌ها دارای رابطه ابر داده‌ای حاشیه‌نویسی نیستند.
- صدک ۱۲۵ام نشان می‌دهد که ۲۵ درصد هستان‌شناسی‌ها کمتر از ۱۵ رابطه ابر داده‌ای حاشیه‌نویسی دارند.
- صدک ۱۵۰ام یا میانه نشان می‌دهد که ۵۰ درصد هستان‌شناسی‌ها دارای ۳۱ رابطه ابر داده‌ای حاشیه‌نویسی می‌باشند.
- صدک ۱۷۵ام نشان دهنده‌ی اینست که ۲۵ درصد هستان‌شناسی‌ها بیش از ۶۱ رابطه ابر داده‌ای حاشیه‌نویسی دارند.
- صدک ۱۹۵ام نشان می‌دهد که ۵ درصد هستان‌شناسی‌ها بیش از ۱۱۰ رابطه ابر داده‌ای حاشیه‌نویسی دارند.
- از نظر تعداد روابط ابر داده‌ای حاشیه‌نویسی تفاوت نسبتاً زیادی بین صدک‌های ۱۵ام و ۱۹۵ام وجود دارد. بیشترین تعداد این نوع رابطه به ترتیب در Uberon multi-species anatomy ontology، Obstetric and Neonatal Ontology، Cell Line Ontology و Drug Interaction and Evidence Ontology و MFO Mental Disease Ontology با تعداد ۲۰۶، ۱۴۹، ۱۴۸، ۱۱۴ و ۱۱۳ رابطه می‌باشد.

بررسی آماری تعداد نمونه‌های موجود در هستان‌شناسی‌های مورد مطالعه به قرار زیر است:

- صدک پنجم برابر ۳ شده است که بدین معناست که ۵ درصد کل هستان‌شناسی‌ها دارای کمتر از ۳ نمونه هستند.
- صدک ۱۲۵ام برابر ۱۸ شده است که بدین معناست که ۲۵ درصد هستان‌شناسی‌ها کمتر از ۱۸ نمونه دارند.
- صدک ۱۵۰ام یا میانه نشان می‌دهد که ۵۰ درصد هستان‌شناسی‌ها دارای ۲۲ نمونه می‌باشند.
- صدک ۱۷۵ام نشان دهنده‌ی اینست که ۲۵ درصد هستان‌شناسی‌ها بیش از ۳۵ نمونه دارند.
- صدک ۱۹۵ام نشان می‌دهد که ۵ درصد هستان‌شناسی‌ها بیش

از ۳۵۶ نمونه دارند.

از نظر تعداد نمونه هم تفاوت نسبتاً زیادی بین صدک‌های ۱۵ام و ۱۹۵ام وجود دارد. بیشترین تعداد نمونه به ترتیب در Gazetteer، FOODON، Hymenoptera Anatomy Ontology و Ontology of Arthropod Circulatory Systems با تعداد ۶۶۰۹۹۳، ۲۷۵۷، ۳۵۶، ۳۲۳ و ۲۵۴ نمونه می‌باشد.

بحث

مخزن هستان‌شناسی‌های باز زیست‌پزشکی مجمعی است از توسعه دهندگان هستان‌شناسی که به پیروی از اصول به اشتراک گذاشته‌ای متعهد شده‌اند. ماموریت مخزن عبارت است از توسعه یک خانواده از هستان‌شناسی‌های میان‌کنش پذیر که هم به لحاظ منطقی غیرمبهم و هم به لحاظ علمی دقیق و درست باشند. در این راستا، شرکت کنندگان در مخزن از یک سری اصول در حال توسعه، بر اساس مدل‌های موفق هستان‌شناسی نظیر هستان‌شناسی ژن پیروی می‌کنند و سپس به توسعه‌ی بیشتر آن اصول کمک می‌کنند. این مخزن توسط یک کمیته عملیاتی مشتمل بر گروه‌های کاری تحریریه، فنی و توسعه نظارت می‌شود. فرایندهای گروه کاری تحریریه بر اساس فرایند داوری مجلات مدل سازی می‌شوند. در وب سایت این مخزن جدول هستان‌شناسی‌های موجود در اشکال مختلف به همراه جزئیات هر یک از هستان‌شناسی‌ها و مستندات بر اساس اصول هستان‌شناسی‌های زیست پزشکی ارائه می‌گردد (۲۲).

تعداد هستان‌شناسی‌های به اشتراک گذاشته شده در مخزن از ۶۰ هستان‌شناسی در سال ۲۰۰۷ (۷) به ۱۵۱ هستان‌شناسی در سال ۲۰۱۷ در حدود ۴۰ حوزه موضوعی زیست‌پزشکی افزایش یافته است که همگی آنها از اصول هستان‌شناسی‌های باز زیست‌پزشکی پیروی نموده‌اند. در حال حاضر، بیشترین هستان‌شناسی‌ها در موضوعات آناتومی، سلامت و فنوتایپ ساخته شده‌اند. بنابراین می‌توان گفت توسعه دهندگان هستان‌شناسی در زمینه موضوع‌های ذکر شده فعال‌تر بوده‌اند. این تنوع موضوعی زیست‌پزشکی می‌تواند دلیلی بر پذیرفته شدن اصول مزبور در حیطه‌های موضوعی زیست‌پزشکی باشد. لذا می‌توان ادعا نمود که جامعه علمی زیست‌پزشکی نیز در راستای ایجاد امکان میان‌کنش پذیری بیشتر بین نظام‌های مختلف اطلاعاتی، پیروی از اصول کنسرسیوم را ترجیح می‌دهد. این امر موجب می‌شود

که توسعه‌دهندگان لازم دانسته‌اند، بسط داده شده باشد. محدودیت عمده‌ی پژوهش حاضر عبارت است از به روز نبودن اطلاعات منابع اینترنتی نظیر OntoBee که توسط مخزن به عنوان ارایه‌کنندگان اطلاعات معرفی شده‌اند. این امر ممکن است ناشی از عدم شکل‌گیری کامل منابعی باشد که کار اطلاع‌رسانی هستان‌شناسی‌ها را انجام می‌دهند.

نتیجه‌گیری

از آنجایی که پژوهشگران ساخت و توسعه‌ی هستان‌شناسی‌ها بر سر روش‌شناسی‌های ساخت هستان‌شناسی با یکدیگر اجماع ندارند، اقدامات انجام شده در حیطه‌ی هستان‌شناسی‌های زیست‌پزشکی جایگاه ویژه‌ای در توسعه‌ی هستان‌شناسی‌ها دارد. از افزایش تعداد هستان‌شناسی‌های مخزن، تنوع موضوعات زیست‌پزشکی در هستان‌شناسی‌های موجود و همچنین استفاده از ابزارها و زبان‌های مشترک ساخت هستان‌شناسی می‌توان اینگونه نتیجه‌گیری نمود که جامعه‌ی علمی زیست‌پزشکی نیز در راستای ایجاد امکان میان‌کنش‌پذیری بیشتر بین نظام‌های مختلف اطلاعاتی، پیروی از اصول کنسرسیوم را در اولویت قرار داده است.

توسعه‌دهندگان تازه کار هستان‌شناسی‌های موضوعی زیست‌پزشکی در مدیریت چالش‌های اولیه ساخت هستان‌شناسی خود می‌توانند با پیروی از اصول کنسرسیوم ساخت هستان‌شناسی‌های منبع باز زیست‌پزشکی آزادانه با استفاده‌ی مجدد از هستان‌شناسی‌های موجود، انتخاب زبان و ابزار پر استفاده در مخزن هستان‌شناسی‌های منبع باز زیست‌پزشکی (زبان OBO و ابزار OBO-Edit) مبادرت به ساخت هستان‌شناسی‌هایی نمایند که از آغاز برای میان‌کنش‌پذیری، نامتناقض بودن، منطقی بودن و بازنمایی درست و دقیق موجودیت‌های زیست‌پزشکی شکل گرفته‌اند. توسعه‌دهندگان هستان‌شناسی‌ها با به کارگیری ابزارهای منبع باز پرکاربرد، که در این مطالعه معرفی شده‌اند، می‌توانند بدون محدودیت‌های کمی، اجزای مختلف هستان‌شناسی‌های خود را در راه بازنمایی هر چه صریح‌تر جهان مورد نظر/حیطه‌ی موضوعی بسازند و مفاهیم موجود در جهان مورد نظر خود را ماشین خوان نمایند.

علاوه بر این، توسعه‌دهندگان برنامه‌های کاربردی؛ برای مثال، انواع نظام‌های مدیریت اطلاعات پزشکی، نظام‌های پشتیبانی از تصمیم‌گیری و ... می‌توانند جهت هوشمندسازی سامانه‌های خود از

که هستان‌شناسی‌های باکیفیت‌تری (۱۱) جهت استفاده‌ی مجدد در هنگام مهندسی هستان‌شناسی‌های جدید در اختیار توسعه‌دهندگان هستان‌شناسی‌ها و یا توسعه‌دهندگان انواع نظام‌های اطلاعاتی هوشمند قرارگیرد.

هستان‌شناسی‌ها معمولاً با استفاده از زبان‌های هستان‌شناسی کدگذاری می‌شوند. در علوم رایانه و هوش مصنوعی، زبان‌های هستان‌شناسی به زبان‌های رسمی اطلاق می‌گردد که برای ساخت هستان‌شناسی‌ها از آنها استفاده می‌شود. این زبان‌ها امکان کدگذاری دانش در یک حیطه‌ی موضوعی خاص را به وسیله‌ی ابزارهای ساخت هستان‌شناسی فراهم می‌آورند (۲۳). ابزار و زبان اصلی ساخت هستان‌شناسی‌های به اشتراک گذاشته شده در مخزن به ترتیب OBO-Edit (۲۴) و OBO (۲۵) می‌باشد. شواهد حاکی از پذیرش جامعه‌ی توسعه‌دهنده‌ی هستان‌شناسی‌های زیست‌پزشکی برای استفاده از این دو ابزار و زبان برای ساخت هستان‌شناسی‌هاست. هر چند از ابزارهای دیگری نیز برای ساخت هستان‌شناسی‌ها استفاده می‌شود، توسعه‌دهندگان بیشتر ترجیح داده‌اند از نسخه‌های مختلف OBO-Edit برای ساخت هستان‌شناسی موجود در مخزن استفاده کنند. OBO-Edit یک ویراستار هستان‌شناسی منبع باز و مستقل از پلتفرم است که توسط کنسرسیوم هستان‌شناسی ژن توسعه داده شده و نگهداری می‌شود. این ابزار برای ویراستاری هستان‌شناسی‌های زیست‌پزشکی بسیار ارزشمند گزارش شده است (۲۶). توسعه‌دهندگان هستان‌شناسی‌های موجود در مخزن همچنین بیشتر سعی داشته‌اند از نسخه‌های مختلف زبان OBO استفاده کنند. OBO یک زبان بازنمایی دانش‌نمودار محور است که به طور گسترده در هستان‌شناسی‌های زیستی و زیست‌پزشکی از آن استفاده می‌شود (۲۳).

هر چند مطالعات پیشین امتیازات کیفی هستان‌شناسی‌های مخزن در اکوآر را بالا گزارش می‌کنند (۱۱)، مطالعه‌ی حاضر هیچ رابطه‌ای را بین فراوانی اجزای تشکیل‌دهنده‌ی هستان‌شناسی‌ها کلاس، رابطه‌ی شیئی، رابطه‌ی نوع-داده، رابطه‌ی ابرداده‌ای حاشیه‌نویسی و تعداد نمونه‌ها نشان نداد. این امر نشان می‌دهد که هستان‌شناسی‌های مخزن با توجه به اهداف توسعه‌دهندگان صرفاً مطابق با اصول کنسرسیوم هستان‌شناسی‌های منبع باز زیست‌پزشکی برای بازنمایی یک جهان واقعی در یک موضوع خاص زیست‌پزشکی بسط یافته‌اند و کمیت آنها برای توسعه‌دهندگان اهمیتی ندارد. در این راستا، انتظار می‌رود که اجزای تشکیل‌دهنده‌ی هستان‌شناسی‌های موجود در مخزن به اندازه‌ی



هستان‌شناسی‌های موجود در مخزن استفاده‌ی مجدد نمایند.

بانک‌های اطلاعاتی مورد بررسی قرار گیرد.

در این مطالعه سعی شد وضعیت فعلی هستان‌شناسی‌های زیست پزشکی موجود در مخزن هستان‌شناسی‌های منبع باز زیست پزشکی که بر اساس سیزده اصل مخزن توسعه یافته‌اند، به لحاظ زبان و ابزار طراحی و فراوانی اجزای تشکیل دهنده‌ی هستان‌شناسی‌های مورد مطالعه تبیین گردد. هر چند مطالعات پیشین بر کیفیت این هستان‌شناسی‌ها صحت گذاشته‌اند، جا دارد کیفیت این هستان‌شناسی‌ها در برنامه‌های کاربردی؛ به عنوان نمونه، ذخیره و بازیابی اطلاعات از

تشریح و قدردانی

این مقاله حاصل بخشی از پایان‌نامه‌ی مقطع دکتری تخصصی علم اطلاعات و دانش‌شناسی با عنوان "امکان‌سنجی طراحی و ساخت هستان‌شناسی صرع و سنجش کارآمدی آن در بازیابی معنایی اطلاعات" به شماره‌ی ثبت ۱۵۸۱ در دانشگاه بین‌المللی امام رضا(ع) می‌باشد.

منابع

- Chandrasekaran B, Josephson JR & Benjamins VR. What are Ontologies, and why do we need them? IEEE Intelligent Systems 1999; 14(1): 20-6.
- Konopka BM. Biomedical ontologies-A review. Biocybernetics and Biomedical Engineering 2015; 35(2): 75-86.
- Fleuriot J. Semantic web systems: Ontological representation. Available at: <https://www.inf.ed.ac.uk/teaching/courses/sws/lectures/SWS-03-Representation.pdf>. 2018.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: Tool for the unification of biology. Nature Genetics 2000; 25(1): 25-9.
- Moore B, Fan G & Eilbeck K. SOBA: Sequence ontology bioinformatics analysis. Nucleic Acids Research 2010; 38(1): 161-4.
- Huang DW, Sherman BT & Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2009; 4(1): 44-57.
- Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 2007; 25(1): 1251-5.
- Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: Ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research 2009; 37(1): 170-3.
- Côté RG, Jones P, Apweiler R & Hermjakob H. The Ontology lookup service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics 2006; 7(1): 97.
- OBO Working Group. The OBO foundry. Available at: <http://www.obofoundry.org>. 2018.
- Quesada-Martínez M, Duque-Ramos A, Iniesta-Moreno M & Fernández-Breis JT. Preliminary analysis of the OBO foundry ontologies and their evolution using OQuaRE. Studies in Health Technology and Informatics 2017; 235(1): 426-30.
- Ochs C, Perl Y, Geller J, Arabandi S, Tudorache T & Musen M. An empirical analysis of ontology reuse in BioPortal. Journal of Biomedical Informatics 2017; 71(1): 165-77.
- Ghazvinian A, Noy NF & Musen MA. How orthogonal are the OBO foundry ontologies? Journal of Biomedical Semantics 2011; 2(S): 2.
- Dabbish L, Colleen Stuart H, Tsay J & Herbsleb J. Social coding in GitHub: Transparency and collaboration in an open software repository, Washington: The 2012 ACM Conference on Computer Supported Cooperative Work, 2012.
- Bizer C, Heath T & Berners-Lee T. Linked data-the story so far. International Journal on Semantic Web and Information Systems 2009; 5(3): 1-22.
- Xiang Z, Mungall C, Ruttenberg A & He Y. Ontobee: A linked data server and browser for ontology terms, USA: International Conference on Biomedical Ontology, 2011.

17. Ong E, Xiang Z, Zhao B, Liu Y, Lin Y, Zheng J, et al. Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic Acids Research* 2017; 45(1): 347-52.
18. Hoehndorf R, Slater L, Schofield PN & Gkoutos GV. Aber-OWL: A framework for ontology-based data access in biology. *BMC Bioinformatics* 2015; 16(1): 26.
19. Côté RG, Jones P, Martens L, Apweiler R & Hermjakob H. The ontology lookup service: More data and better tools for controlled vocabulary queries. *Nucleic Acids Research* 2008; 36(1): 372-6.
20. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, et al. BioPortal: Enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Research* 2011; 39(1): 541-5.
21. Rubin DL, Moreira DA, Kanjamala P & Musen MA. BioPortal: A web portal to biomedical ontologies. Available at: <http://www.aaai.org/Papers/Symposia/Spring/2008/SS-08-07/SS08-07-011.pdf>. 2008.
22. OBO Working Group. Principle: Overview. Available at: <http://www.obofoundry.org/principles/fp-000-summary.html>. 2018.
23. Hoehndorf R, Schofield PN & Gkoutos GV. The role of ontologies in biological and biomedical research: A functional perspective. *Briefings in Bioinformatics* 2015; 16(6): 1069-80.
24. Raffat SK, Sarim M, Iqbal S & Shahab M. OBO edit: A tool for classifying the biological data. *FUUAST Journal of Biology* 2015; 5(2): 249-55.
25. Golbreich C, Horridge M, Horrocks I, Motik B & Shearer R. OBO and OWL: Leveraging semantic web technologies for the life sciences, Korea: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, 2007.
26. Day-Richter J, Harris MA, Haendel M & Lewis S. OBO-Edit an ontology editor for biologists. *Bioinformatics* 2007; 23(16): 2198-200.



A Study on the Ontologies Developed Based on the Principles of Open Biomedical Ontologies Foundry

Nader Alishan Karami¹ (M.S.) - Mohsen Hajzeinalabedini² (Ph.D.) -
Iraj Radad³ (Ph.D.) - Seyed Javad Ghazi Mirsaeed⁴ (Ph.D.)

1 Ph.D. Candidate in Knowledge and Information Science, School of Human Sciences, Imam Reza International University, Mashhad, Iran

2 Assistant Professor, Knowledge and Information Science Department, School of Letters and Human Sciences, Shahid Beheshti University, Tehran, Iran

3 Assistant Professor, Knowledge and Information Science Department, School of Human Sciences, Imam Reza International University, Mashhad, Iran

4 Associate Professor, Medical Library and Information Science Department, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

Abstract

Received: Apr 2018

Accepted: Aug 2018

Background and Aim: Ontologies facilitate data integration, exchange, searching and querying. Open Biomedical Ontologies (OBO) Foundry is a solution for creating reference ontologies. In this foundry, the design of ontologies is based on established principles which allow for their interactions as a single system. The purpose of this study is to determine the main features of ontologies developed based on foundry principles.

Materials and Methods: This descriptive cross-sectional study (2017) retrieved data about the ontology components, subject, language and tool from the OBO foundry website, and OntoBee (the default linked data server for the foundry ontologies). The frequency of classes, instances, Object Properties, DataType Properties and Annotation Properties are classified by the use of descriptive statistics.

Results: The main subjects of 151 ontologies were Anatomy, health and Phenotype. Most languages and tools were, respectively, OBO and OBO-edit versions used for the development of ontologies. Five percent of ontologies contained more than 78977 classes, 150 object relations, 39 DataType relations, 110 annotation relations, and 356 instances.

Conclusion: Increasing the number of ontologies developed based on OBO foundry principles as well as using OBO language and OBO-Edit is indicative of the fact that ontology developers have accepted the principles and tools.

Keywords: Ontology, Biomedical Ontology, Foundry of Open Biomedical Ontologies

* Corresponding Author:
Hajzeinalabedini M
Email:
m_zabedini@sbu.ac.ir