توزیع آماری رشد موضوعات علوم پزشکی در طول زمان: تحلیل همبستگی بین توسعه‌ی MeSH و رشد Medline*

محمد توکلی زاده راوری[1]

چکیده

مقدمه: سرعنوان‌های موضوعی پزشکی (MeSH) با بیش از ۲۳ ۰۰۰ سرعنوان، منعکس کننده‌ی موضوعات حوزه‌ی علوم پزشکی است. لذا در این پژوهش با همبستگی بین توسعه‌ی MeSH و افزایش رکوردهای Medline می‌توان چگونگی گسترش آماری موضوعات حوزه‌ی پزشکی را در طول زمان مورد سنجش قرار داد.

روش بررسی: در یک پژوهش تحلیلی، حدود یک میلیون رکورد از موضوعات علوم پزشکی و انتشارات نمایه شده در دامنه‌ی Medline به طور تصادفی انتخاب شد. از طریق تعیین همبستگی بین افزایش تعداد مدارک Medline و افزودن سرعنوان‌های جدید به MeSH، به دو سؤال در خصوص گسترش موضوعات پزشکی پاسخ داده شد: ۱. موضوعات پزشکی در طول زمان از لحاظ کمّی چگونه در MeSH گسترش یافته است؟ ۲. امروزه در ازای تولید و انتشار چند مدرک علمی پزشکی، یک موضوع جدید تولید و به صورت سرعنوان به MeSH افزوده می‌شود؟

یافته‌ها: توسعه‌ی MeSH در طول زمان، به صورت لگاریتمی در سه فاز مختلف صورت گرفته است. با تلفیق این سه فاز در یک فاز واحد، فرمولی برای پیش‌بینی توسعه‌ی MeSH و موضوعات پزشکی در ازای افزایش انتشارات این حوزه به دست آمد. از طرفی روشن گردید که امروزه، به صورت خطی، در ازای انتشار بیش از ۲۵۰ مدرک در موضوعات گوناگون علوم پزشکی، یک موضوع جدید فراهم و به MeSH افزوده می‌شود.

نتیجه‌گیری: نتایج نشان داد که بر خلاف برخی از نظرها، گسترش موضوعات علمی پویاست و این پویایی به صورت خطی است و هرگز متوقف نمی‌شود.

واژه‌های کلیدی: Medline؛ سرعنوان‌های موضوعی؛ اصطلاح‌نامه.

نوع مقاله: تحقیقی

دریافت مقاله: ۸۷/۲/۳ *اصلاح نهایی: ۸۷/۵/۱۵* *پذیرش مقاله: ۸۷/۶/۱۶*

ارجاع: توکلی زاده راوری محمد. **توزیع آماری رشد موضوعات علوم پزشکی در طول زمان: تحلیل همبستگی بین توسعه‌ی MeSH و رشد Medline**. مدیریت اطلاعات سلامت ۱۳۸۶؛ ۴ (۲): ۱۸۵-۱۹۲.

مقدمه

Thesaurus را می‌توان مجموعه‌ای فشرده از موضوعاتی دانست که در یک زمینه خاص گردآوری شده و به طور عام به یک سیستم نمایه‌سازی پیوسته است. این بدان معناست که سرعنوان‌ها یا توصیفگرهای موجود در آن یک‌باره پدید نیامده‌اند، بلکه بر اثر گذشت زمان، با توجه به پدید آمدن موضوعات یا طبقه‌ای از موضوعات جدید، در مدارک نمایه شده و به بانک اطلاعاتی مربوط به آن افزوده و گسترش یافته است. بنا بر این، یکی از ویژگی‌های Thesaurus این

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

۱. استادیار کتابداری و اطلاع‌رسانی، دانشگاه یزد (نویسنده‌ی مسؤول)

E-mail: mravari@yahoo.com

Archive of SID

www.SID.ir

است که به مرور زمان با پدید آمدن آثار جدید، رشد می‌کند. هنگامی که نمایه‌سازان در مواجهه با این منابع جدید، به موضوع یا طبقه‌ای از موضوعات جدید برخورد می‌کنند، به صورت کنترل‌شده، آن را به Thesaurus مربوط می‌افزایند. به عبارتی، گسترش یک Thesaurus همواره تابع توسعه بانک اطلاعاتی پیوسته به آن است. از این رو، پژوهش‌هایی که به چگونگی رشد واژگان کنترل‌شده پرداخته‌اند و همواره رشد آن را به صورت متغیری وابسته به متغییر رشد نمایه‌نامه‌ها یا در اصطلاح جدید، بانک‌های اطلاعاتی مربوط به آن مورد بررسی قرار داده‌اند. از این میان می‌توان به پژوهش‌های Wurm، McClelland و همکار، Blagden، Umstätter، Lancaster و Tavakolizadeh اشاره کرد (۱– ۶).

این پژوهش‌ها صورت‌های گوناگونی از رشد Thesaurus را به تصویر کشیده‌اند. Lancaster معتقد است که یک واژگان کنترل‌شده، در طول عملیات نمایه‌سازی در ابتدا بسیار سریع به صورت لگاریتمی رشد می‌کند و سپس این رشد به صورت یک خط افقی ادامه می‌یابد (۵). این سخن به آن معناست که زمانی خواهد رسید که این رشد متوقف خواهد شد. وی در دهه‌ی هشتاد میلادی به آثاری اشاره می‌کند که سقفی برای رشد واژگان کنترل‌شده قائل است. او می‌گوید: هنگامی که تعداد موضوعات پزشکی به حدود بیست هزار برسد، رشد واژگان کنترل شده در این حوزه متوقف خواهد شد. او این عدد را برای هر حوزه‌ای گوناگون می‌داند و معتقد است که بزرگی یک Thesaurus به عواملی مانند زمینه‌ی موضوعی، سطح خاص بودن و نوع اصطلاحات به کار رفته در آن بستگی دارد (۵).

با توجه به این که پیش‌بینی توقف رشد واژگان کنترل‌شده، در حوزه‌ی پزشکی در سقف بیست هزار، در دو دهه‌ی گذشته صورت گرفته است، ضرورت دارد تا ببینیم که آیا رشد موضوعات پزشکی همان‌گونه که پیش از این گفته شد به نهایت رسیده است یا این رشد همچنان پویاست و متوقف نخواهد شد.

یکی از معروف‌ترین واژگان کنترل‌شده در زمینه‌ی پزشکی سرعنوان‌های موضوعی پزشکی است که در اصطلاح MeSH نامیده می‌شود. توسعه‌ی MeSH تابع نمایه‌سازی و افزودن مدارک جدید به بانک اطلاعاتی مربوط به آن یعنی Medline است و امروزه بیش از بیست و سه هزار سرعنوان را در خود جای داده است. از این رو به نظر می‌رسد که رشد یک واژگان کنترل شده مانند MeSH نباید هیچ‌گاه متوقف شود و به صورت پویا همواره ادامه خواهد یافت (۶).

بررسی این پویایی هدف اصلی پژوهش حاضر بوده است. از این رو، نحوه‌ی گسترش موضوعات پزشکی در طول زمان مورد بررسی آماری قرار گرفت. برای رسیدن به این هدف دو سؤال اصلی مورد توجه قرار گرفت: ۱. از لحاظ آماری موضوعات پزشکی در طول زمان چگونه رشد کرده است؟ و همبستگی بین رشد موضوعات و تعداد انتشارات نمایه شده این حوزه چگونه بوده است؟۲. امروزه در ازای انتشار و نمایه شدن چند مدرک یک موضوع جدید به صورت سر عنوان در حوزه ی پزشکی به وجود می آید؟

روش بررسی

پژوهش حاضر از نوع تحلیلی و جامعه آماری شامل MeSH و Medline بود که پژوهشگر به بررسی رشد موضوعات پزشکی در طول زمان پرداخت. بررسی رابطه‌ی این دو متغیر با توجه به پژوهش‌های گذشته صورت گرفت؛ برای این منظور Medline به عنوان معروف‌ترین و بزرگ‌ترین بانک اطلاعاتی حوزه‌ی پزشکی و حتی جهان با بیش از هفده میلیون رکورد نمایه شده به عنوان ماده اصلی برای استخراج داده‌های مورد نیاز مورد توجه قرار گرفت. برای این منظور، ابتدا واژه‌ی UP از طریق Pubmed در Medline جستجو شد. انتخاب این واژه برای جستجو پس از بررسی واژه‌های گوناگون صورت گرفت. یکی از محاسن آن این بود که معنای خاصی را در بر نداشت و می‌شد از طریق آن به صورت تصادفی تعداد محدودی رکورد را از تمامی موضوعات پزشکی بازیابی کرد. نتیجه‌ی این جستجو بازیابی نزدیک به یک میلیون رکورد بود. سپس این رکوردها به صورت متنی ذخیره شد و مورد تجزیه و تحلیل قرار گرفت.

از آن‌جا که بررسی این تعداد رکورد به صورت دستی امکان‌پذیر نبود، تصمیم گرفته شد تا این کار با استفاده از امکان پردازش متن در زبان‌های برنامه‌نویسی انجام شود. لذا از زبان برنامه‌نویسی دلفی برای این منظور استفاده شد.

Archive of SID

www.SID.ir

یکی از بهترین شکل‌های ممکن برای بررسی رکوردها به صورت خودکار، ذخیره‌سازی آن‌ها به شکل برچسبی بود که در اصطلاح فرمت Medline نامیده می‌شود. در این شکل متنی، به راحتی می‌توان آغاز و پایان هر رکورد را مشخص ساخت و همچنین امکان تشخیص هر یک از اجزاء رکوردها مانند عنوان و تاریخ توصیفگرها و ... به راحتی و با استفاده از نشان هر جزء ممکن می‌شد.

برای یافتن همبستگی بین رشد MeSH و Medline لازم بود مشخص شود که در ازای نمایه‌سازی و افزودن چند رکورد جدید در Medline یک توصیفگر جدید فراهم شده است. به عبارتی باید آشکار گردد که با افزودن اولین رکورد به Medline چند توصیفگر در MeSH پدید آمده است و با افزایش تعداد رکوردهای Medline به دو، مجموع تعداد سرعنوان‌های MeSH به چند رسیده است. این عمل باید ادامه می‌یافت تا معلوم گردد امروزه با افزودن آخرین رکوردها به Medline تعداد توصیفگرهای MeSH به چند رسیده است؛ از این رو، رکوردها باید به ترتیب ورودشان به Medline ذخیره شده، مورد بررسی قرار گیرند؛ یعنی قدیمی‌ترین رکورد در این مجموعه اولین باشد و جدیدترین آن در آخر بیاید.

متأسفانه از راه Pubmed تنها امکان ذخیره‌ی رکوردهای بازیابی شده به صورت برعکس وجود دارد؛ یعنی جدیدترین رکورد در هنگام نمایش و ذخیره اطلاعات اول بار می‌آید. برای حل این مشکل، همان‌گونه که رکوردها نمایش داده شده بودند، یعنی به صورت نظم جدیدترین به قدیم‌ترین ذخیره شدند و ترتیب قرار گرفتن آن‌ها با استفاده از امکانات برنامه‌نویسی به صورت دلخواه تغییر داده شد؛ یعنی قدیمی‌ترین رکورد در اولین و جدیدترین آن در آخرین جایگاه رکورد در فایل متنی قرار گرفت.

سپس، رکوردها برای بررسی آماده شدند. در این بررسی جزء MH (MeSH Headings) هر رکورد و شماره‌ی رکورد مورد توجه قرار گرفت. اجزای PMID یا Pub Med Identification Number (شماره‌ی متمایز کننده‌ی رکورد در Pubmed) و SO یا Source به ترتیب امکان تشخیص ابتدا و انتهای هر رکورد را فراهم آورد.

اساس کار بر این بود که برنامه‌ای از قبل نوشته شده یک به یک و از ابتدا رکوردها را مورد توجه قرار می‌داد و پس از رسیدن به هر رکورد، بلافاصله به دنبال جزئی از MH می‌گشت. هر رکورد می‌توانست چند جزء MH را داشته باشد زیرا یک جزء تکرارپذیر است. ضمن آن که از هر سرعنوان فقط سرعنوان اصلی مورد توجه قرار می‌گرفت و از سرعنوان‌های فرعی صرف نظر می شد. پس از بررسی هر رکورد، برنامه‌ی کامپیوتری شماره‌ی ترتیب آن را یادداشت و در برابر آن، مجموع سرعنوان‌های جدیدی را که از ابتدا برای اولین بار در Medline مورد استفاده قرار گرفته بودند، ثبت می‌کرد. به عنوان مثال اولین رکورد نشان می‌داد که در برگیرنده‌ی پنج سرعنوان جدید است که برای اولین بار در Medline به کار رفته است، بنا بر این برای آن عدد پنج ثبت شد. رکورد دوم نشان می‌داد که از میان شش سرعنوان آن، سه سرعنوان برای اولین بار در Medline مورد استفاده قرار گرفته است، لذا در برابر عدد دو (معادل شماره رکورد)، عدد هشت ثبت می‌شد؛ یعنی مجموع سرعنوان‌های جدید Medline در دو رکورد شماره یک و دو، هشت سرعنوان بود و این عمل تا آخرین رکورد ادامه پیدا کرد. در آخر نیز مجموع سرعنوان‌های به کار رفته در Medline حدود ۲۳۰۰۰ محاسبه شد.

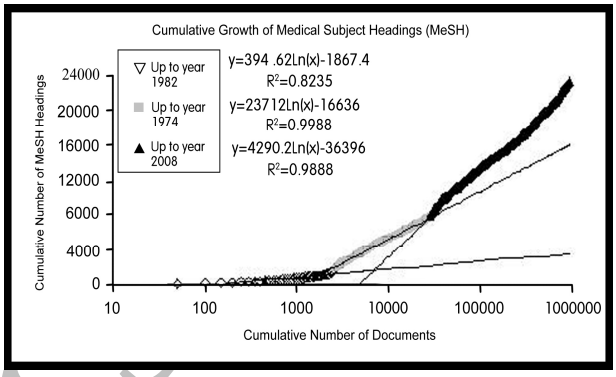
برای این که برنامه بتواند تشخیص دهد که آیا سرعنوان مورد بررسی پیش از این در Medline به کار رفته است یا نه، ابتدا در یک بانک حاوی سرعنوان‌هایی که تا رکورد مورد بررسی در Medline یافت شده بود، جستجو انجام می‌شد. اگر سرعنوان برای اولین بار مورد استفاده قرار گرفته بود، آن را به پایگاه مربوط اضافه می‌کرد تا در بررسی بعدی مشخص گردد که از این سرعنوان در رکوردهای پیشین Medline استفاده شده است تا دیگر به عنوان سرعنوان جدید شمرده نشود. در پایان، داده‌های حاصل‌شده، شامل شماره رکورد در سمت چپ و مجموع موضوعات جدید با هر شماره در سمت راست در دو ستون، به برنامه Excel منتقل شد و از طریق تعیین همبستگی بین این دو متغیر، مورد بررسی قرار گرفت.

Archive of SID

www.SID.ir

یافته‌ها

ابتدا سعی شد تا دریابیم که نحوه‌ی رشد سرعنوان‌های MeSH در برابر رشد Medline چگونه بوده است. از این طریق نحوه‌ی رشد MeSH میسر شد. نمودار ۱ این همبستگی را نشان می‌دهد:



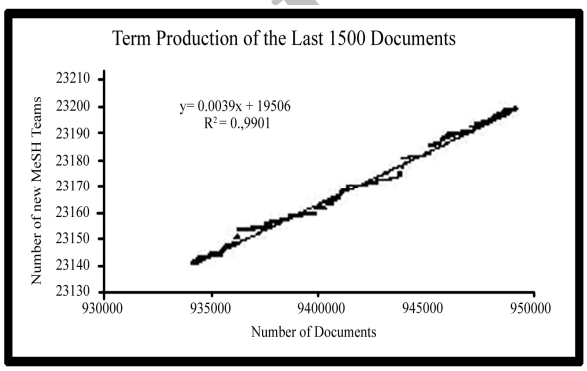
نمودار ۱. همبستگی بین مجموع تعداد سرعنوان‌های MeSH و مجموع تعداد مدارک در Medline

محور افقی در نمودار ۱ به صورت لگاریتمی در آمده است تا بتوان نتایج را به صورت خط مستقیم نمایش داد. همان‌گونه که می‌بینیم، MeSH در سه فاز مختلف رشد کرده و هر فاز به صورت لگاریتمی بوده است. در فاز اول در ازای نمایه‌سازی و افزودن ۲۶۰۰ مدرک به Medline تعداد ۱۵۶۵ سرعنوان جدید تولید شده است. لوزی‌های توخالی نمایشگر رشد در این مرحله است. فاز دوم پس از آن آغاز و تا رکورد ۲۸۸۵۰ ادامه یافته است. در این مرحله تعداد سرعنوانهای MeSH به ۷۸۵۳ می‌رسد. این مرحله با مربع‌های خاکستری و روشن‌تر قابل تشخیص است. آخرین فاز که با مثلث‌های مشکی نمایش داده شده است تا رکورد ۹۴۹۰۰۰ از مجموعه رکوردهای نمونه ادامه می‌یابد و تعداد سرعنوان‌ها به ۲۳۰۰۰ می‌رسد.

بر پایه‌ی بررسی‌های بعدی معلوم شد که رکورد شماره‌ی ۲۶۶۰ از نمونه‌ی مورد بررسی در آخرین ماه در سال ۱۹۶۲، رکورد ۲۸۸۵۰ در سال ۱۹۷۴ و رکورد ۹۴۹۰۰۰ در سال ۲۰۰۶ به Medline افزوده شده است. بنا بر این می‌توان نتیجه گرفت که تا کنون سه بار در سال‌های ۱۹۶۲، ۱۹۷۴ و ۲۰۰۶ میلادی خط مشی کتابخانه ملی پزشکی برای افزودن سرعنوانهای جدید به MeSH تغییر یافته یا به عبارتی ضریب رشد MeSH سه مرتبه بهبود یافته است.

با توجه به داده‌ها و نتایج بالا می‌توان دریافت که در واقع موضوعات پزشکی چگونه باید افزایش یابد. به عبارتی می‌توان از سه فرمول بالا، فرمول واحدی را استخراج کرد. استخراج یک فرمول واحد می‌تواند چندین فایده داشته باشد. مهم‌تر از همه، به کسانی که در پی فراهم کردن و گسترش واژگان مشابه هستند کمک خواهد کرد که رشد آن واژگان را بر اساس این فرمول کنترل کنند تا کم‌تر یا بیشتر از سطح مورد نیاز رشد نکند. این کار در عمل استانداردی را برای فراهم کردن و گسترش Thesaurusها و سرعنوان‌های مشابه به دست می‌دهد. عملیات میانی برای دستیابی به این فرمول نیز واقعیت‌هایی را نمایان می‌کند. یکی از آن‌ها نحوه‌ی پویایی رشد موضوعات پزشکی است که در این میان آشکار خواهد شد.

ابتدا باید بررسی شود که در حوزه‌ی پزشکی امروزه در ازای انتشار و نمایه‌سازی چند مدرک یک موضوع جدید تولید و به MeSH افزوده می‌شود؟ برای این کار آخرین رکوردهای موجود در نمونه مورد توجه قرار گرفت. به این منظور ۱۵۰۰ رکورد آخر، از مجموعه‌ی نمونه استخراج شد و میزان تولید سرعنوان‌های جدید برای این تعداد مورد بررسی قرار گرفت. نمودار ۲ حاصل این عملیات است.

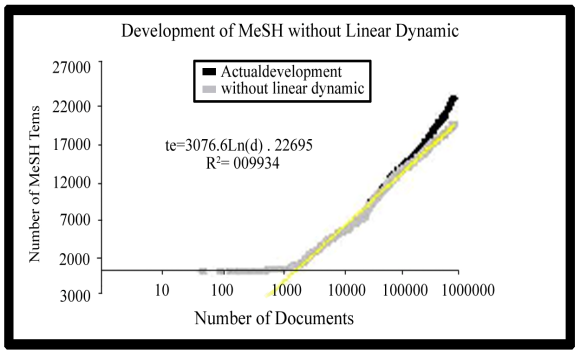


نمودار ۲. تولید موضوعات جدید توسط ۱۵۰۰ رکورد آخر در مجموعه‌ی نمونه

www.SID.ir

همان گونه که نمودار ۲ نشان می‌دهد، امروزه در ازای هر مدرک منتشره و نمایه‌سازی شده، تعداد ۰/۰۰۳۹ موضوع جدید تولید و به MeSH افزوده می‌شود. از آن جا که همبستگی به صورت خطی است، می‌توان به راحتی از قاعده‌ی تناسب استفاده کرد. در این صورت در می‌یابیم که در ازای هر ۲۵۶ مدرک، یک موضوع یا طبقه‌ی موضوعی جدید در پزشکی پدید می‌آید. فرمول بالا نشان می‌دهد که خط شاخص همبستگی در نقطه ۱۹۵۰۶، محور افقی را قطع می‌کند؛ لذا این رشد خطی به موازات رشد لگاریتمی، پس از ورود ۱۹۵۰۶ مدرک به سیستم نمایه‌سازی شروع می‌شود.

رشد خطی در نمودار ۲ و رشد لگاریتمی در نمودار ۱، نشان می‌دهد که رشد موضوعات، ترکیبی از رشد لگاریتمی و خطی است، لذا می‌توان رشد خطی را (که نشان پویایی است) از نمودار ۱ کاست تا معلوم شود که بدون رشد خطی چه حالتی به وجود می‌آید. نمودار ۳ این حالت را نشان می‌دهد.

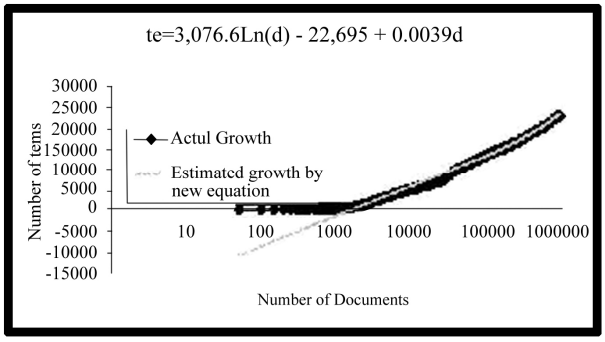


نمودار ۳. رشد MeSH بدون در نظر گرفتن بخش رشد خطی آن

بخش سیاه رنگ در نمودار ۳، رشد واقعی MeSH را از آغاز تا سال ۲۰۰۶ نشان می‌دهد و بخش خاکستری رنگ، این کار را با حذف رشد دینامیکی (پویا) یا خطی آن به نمایش می‌گذارد. محور افقی در این نمودار نیز به صورت لگاریتمی نمایش داده شده است.

خط مستقیمی که از میان نمودار خاکستری گذشته است، خط شاخص همبستگی است. اگر به این خط دقت کنیم، پس از آن که تعداد مدارک به ۱۶۰۰ رسید، بر خط خاکستری منطبق می‌شود و از منفی به سوی مثبت سیر می‌کند. این حالت نشان می‌دهد که حداقل مدرک مورد نیاز برای محاسبه‌ی رشد موضوعات و نیز پدید آمدن یک واژگان کنترل شده، وجود حداقل ۱۶۰۰ مدرک در زمینه‌های مختلف پزشکی است.

با توجه به فرمول همبستگی در نمودار ۲ و ۳ و ترکیب آنها می‌توان یک فرمول واحد به دست آورد که در عمل واقعیت رشد موضوعات پزشکی را از گذشته تا کنون نشان می‌دهد و همچنین بر اساس آن می‌توان گسترش آینده‌ی MeSH را نیز پیش بینی کرد. نمودار ۴ نتایج عملیات حاصل از ترکیب دوباره رشد خطی و لگاریتمی را نشان می‌دهد.



نمودار ۴. مقایسه‌ی رشد موضوعات در MeSH بر اساس فرمول ترکیبی و رشد واقعی

در بالا فرمول گسترش موضوعات پزشکی ذکر شده است. قابل توجه است که محور افقی در این نمودار، به صورت لگاریتمی در آمده است. سه مجهول te، Ln و d در این فرمول، به ترتیب عبارت است از تعداد تخمینی موضوعات پزشکی، لگاریتم طبیعی، و تعداد مدارک نمایه و افزوده شده به Medline.

بحث

نتایج نشان داد که تولید موضوعات یا طبقات موضوعی جدید در پزشکی و افزودن آن به MeSH تابع خط مشی سیستم نمایه‌سازی Medline بوده است. به گونه‌ای که تغییر در آن موجب پدید آمدن سه فاز مختلف در رشد موضوعات شده است. اگر به تاریخچه‌ی این سیستم نمایه‌سازی توجه کنیم، می‌بینیم که دو واقعه تأثیر به سزایی در بالا بردن شتاب تولید

Archive of SID
www.SID.ir

موضوعات پزشکی جدید در MeSH داشته است. گزارش سالانه‌ی کتابخانه‌ی ملی پزشکی (NLM Fiscal Year) در سال ۱۹۶۵ نشان داد که از سال ۱۹۶۴ و پیش از آن، اولین پیشرفت‌ها در مکانیزه کردن ایندکس مدیکوس انجام شده است (۷). اگر به فاز اول توسعه نگاه کنیم، می‌بینیم که تغییر شتاب رشد حوالی همین تاریخ (سال ۱۹۶۲) صورت گرفته است. دومین تغییر فاز متأثر از توجه Medline به چکیده‌ها و رشد تدریجی آن‌ها در این سیستم بوده است. گزارش سالانه‌ی کتابخانه‌ی ملی پزشکی در سال ۱۹۷۵ نشان می‌دهد که در سال ۱۹۷۴ افزودن چکیده‌ها جزء خط مشی Medline قرار گرفته است (۸). نکته‌ی قابل توجه این است که شتاب رشد در هر یک از این تغییر فازها افزایش یافته است.

موارد یاد شده‌ی بالا (مکانیزه کردن ایندکس مدیکوس و افزودن چکیده‌ها) نشان می‌دهد که بررسی توسعه‌ی موضوعات پزشکی از طریق MeSH با متغیرهای مزاحمی رو به رو است. چگونگی رشد دقیق موضوعات جدید در حوزه‌ی پزشکی به طور مستقیم از نمودار ۱ آشکار نیست، زیرا رشد نشان داده شده تابع متغیرهای دیگری بوده است. تنها چیزی که می‌توان از آن به طور مستقیم نتیجه گرفت این است که با افزایش امکانات و قابلیت‌ها، تولید موضوعات جدید به صورت سرعنوان سرعت بیشتری یافته است.

نتیجه‌گیری

افزایش سرعت رشد در هر فاز نمی‌تواند به این معنی باشد که تولید موضوعات جدید در مقاطع مختلف زمانی به یک‌باره شتاب گرفته است. بلکه نشان‌گر آن است که موضوعات بسیاری با تأخیر و یک‌باره وارد MeSH شده است. پس برای بررسی رشد واقعی موضوعات حوزه پزشکی باید تمهیداتی اندیشید. عملیات صورت گرفته در نمودارهای بعدی (نمودار ۲، ۳ و ۴) به این امر کمک کرد.

نمودار ۲ آشکار ساخت که هنوز گسترش MeSH ادامه دارد و امروزه MeSH به عنوان شاخص گسترش موضوعات پزشکی به صورت خطی در حال رشد است. عددی که از نمودار یاد شده استخراج شد، قابل توجه است. پدید آوردن یک موضوع یا طبقه‌ی موضوعی جدید در ازای تولید و نمایه‌سازی بیش از ۲۵۰ مدرک از موضوعات مختلف پزشکی آشکار می‌سازد که بر خلاف نمایش رشد لگاریتمی موضوعات در پژوهش حاضر و پژوهش‌های مشابه، در عمل گسترش موضوعات به صورت خطی است. دلیل این که ما در این گونه پژوهش‌ها با رشد لگاریتمی مواجه می‌شویم این است که یک واژگان کنترل‌شده از یک زمان خاص ایجاد می‌شود و توسعه می‌یابد و باید موضوعاتی را که پیش از ایجاد و گسترش به وجود آمده است، با شتاب در خود جای دهد. لذا اگر ما تصور کنیم که یک واژگان کنترل شده از ابتدای فعالیت‌های علمی بشر در زمینه‌ی پزشکی به وجود می‌آمد، می‌توانست به صورت خطی رشد کند و رشد واقعی موضوعات را نشان دهد.

نکته‌ی دیگر در یافته‌های این پژوهش عدد ۲۵۰ است. Price در سال ۱۹۶۳ نشان داد که در ازای انتشار حدود ۲۵۰ نشریه‌ی جدید، یک پایگاه اطلاعاتی جدید پدید می‌آید (۹). Umstätter و همکار در سال ۱۹۸۴بر پایه این یافته، این سؤال را مطرح کردند که در ازای انتشار چند مقاله، یک نشریه‌ی جدید منتشر می‌شود. نتایج پژوهش آنان نشان داد که در ازای حدود ۲۵۰ مقاله یک نشریه‌ی جدید منتشر شده است (۱۰). اگر یافته‌های این پژوهش‌ها را در کنار هم قرار دهیم، درمی‌یابیم که انتشار هر نشریه بر پایه‌ی ظهور یک موضوع جدید است و با ظهور ۲۵۰ موضوع جدید، می‌توان انتظار یک پایگاه اطلاعاتی جدید را داشت. با این دست‌آورد می‌توان دریافت که موضوعات هر حوزه را باید در ۲۵۰ طبقه جای داد که این ادعا نیاز به پژوهش‌های بیشتر دارد.

نمودار ۲ و ۳ نشان دادند که چگونه رشد موضوعات در یک Thesaurus، ترکیبی از یک همبستگی خطی و یک همبستگی لگاریتمی است و چگونه می‌شود این دو را از هم جدا کرد. نمودار ۲ با این فرض که توزیع مدارک نمایه‌شده بین موضوعات گوناگون پزشکی به یک اندازه است، نشان داد که اگر تعداد مدارک به ۱۹۶۰۵ برسد، رشد خطی در کنار رشد

Archive of SID
www.SID.ir

لگاریتمی شروع می‌شود. نمودار ۳ نیز نشان داد که برای ایجاد یک واژگان کنترل‌شده و هم‌چنین بررسی آغاز رشد موضوعات در یک حوزه‌ی خاص، مانند پزشکی، دست کم به ۱۶۰۰ مدرک در موضوعات گوناگون آن حوزه نیاز است؛ زیرا خط شاخص همبستگی در این نقطه، محور افقی را قطع می‌کند.

نمودار ۴ نیز به این نکته اشاره داشت که اگر دست اندرکاران MeSH از ابتدا از میزان تولید موضوعات حوزه‌ی پزشکی در آینده آگاه بودند، ایجاد موضوعات یا طبقات موضوعی جدید در پزشکی چگونه جریان پیدا می‌کرد. بر اساس این تجربه می‌توان فرمولی را استخراج کرد که بر پایه‌ی آن بتوان یک واژگان کنترل‌شده در سطح MeSH را پدید آ ورد و گسترش داد و شاخصی را در این زمینه به دست داد. بر این اساس بهتر می‌توان رشد واقعی موضوعات را در مدارک حوزه‌ی پزشکی مورد توجه قرار داد.

References

1. Wurm B R. The Relation between Number of Documents and Number of Terms and their Discriminatory Power in Information Retrieval for U.S. Pharmaceutical Patents. Fourth Annual Meeting of the Committee for International Cooperation in Information Retrieval Among Examining Patent Offices ICIREPAT 1964;:349-360.
2. MacClelland R. M., Mapleson W. W. Construction and Usage of Classified Schedules and Generic Features in Coordinated Indexing. ASLIB Proc 1966; 18: 290-299.
3. Blagden J. F. Management Information Retrieval: a New Indexing Language. London: Management Publications Ltd. for the British Institute of Management; 1971.
4. Umstätter W. Informetrische Hilfen durch das intelligente Terminal [in German]. Deutscher Dokumentartag. 1986; 556-564.
5. Lancaster F. W. Vocabulary Control for Information Retrieval(Second Edition). Washington: Information Resources Press;1986.
6. Tavakolizadeh-Ravari M. Analysis of the Long Term Dynamics in Thesaurus Developments and its Consequences, Ph.D. Thesis in Library and Information Science. Berlin: Philosophische Fakulität I., Humboldt-Universität zu Berlin; 2007.
7. NLM Fiscal Year 1965 [WWW Document] available at: http://www.nlm.nih.gov/hmd/manuscripts/nlmarchives/annualreport/1965.pdf 04.02.2008.
8. NLM Fiscal Year 1975 [WWW Document] available at: http://www.nlm.nih.gov/hmd/manuscripts/nlmarchives/annualreport/1975.pdf 04.05.2008.
9. Price D. Little Science Big Science. Colombia: Colombia University Press; 1963.
10. Umstätter W, Rehm M. Bibliothek und Evolution [in German]. Nachr. Bok. 1984; 35 (6): 237 – 249.

Archive of SID

www.SID.ir

The Growth of Medical Sciences Subjects: A Correlation Analysis between Development of MeSH and Medline*

Mohammad Tavakolizadehravari[1]

Abstract

Introduction: The subject headings of MeSH in medicine is more than 23,000 which is an index for the medical sciences subjects. This study aims to find the correlation between development of MeSH and increase of medline records to assess the statistical development of medical sciences subjects during the time.

Methods: This is an analytical study of the subjects of medical sciences and medline publications index. A sample of one million medline records was selected randomly and the correlation between the increase in the number of medline records and the increase of head titles in MeSH was assessed to answer the two following questions: 1. How has the quantity of medical subjects developed during time? 2. Currently by the production and publication of how many medical scientific papers a new subject might be added to the headings of MeSH?

Results: Development of MeSH during the time happened as a logarithm in three different phases. Putting these three phases into one, we found a formula to predict development of MeSH and its headings by the increase of publications in the field. Also, we found that nowadays by the publication of more than 250 papers in various subjects of the medical sciences, a new subject is created and added to MeSH linearly.

Conclusion: The results showed that in spite of some ideas, development of scientific subjects is dynamic and is linearly dynamic and will not stop.

Keywords: Medline; Subject Headings; Terminology.

Type of article: Original Research

Received: 22 Apr, 2008 *Accepted: 6 Sep, 2008*

Citation: Tavakolizadehravari M. **The Growth of Medical Sciences Subjects: a Correlation Analysis between Development of MeSH and Medline**. Health Information Management 2007; 4 (2): 192.

Archive of SID

* This article resulted from independent research.

1. Assistant professor, Library and Information Sciences, Yazd University, Yazd, Iran. (Corresponding Author)
E-mail: mravari@yahoo.com

www.SID.ir