سومین کنفرانس ملی مهندسی و مدیریت محیط زیست
The 3rd National Conference on Environmental Engineering and Management (3CEEM) 31 May 2021
۱۰ خرداد ۱۴۰۰
موسسه آموزش عالی تجن
وزارت علوم، تحقیقات و فناوری
مرکز تحقیقات مهندسی محیط زیست مازندآب
ISC

تحلیل عرضه و تقاضای پایان‌نامه‌های حوزه محیط‌زیست در ایران[۱]

محمد ربیعی*

استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، m.rabiei@irandoc.ac.ir

چکیده

به‌منظور مدیریت نظام پژوهش حوزه محیط‌زیست، داشتن شناخت کافی نسبت به تقاضاها و نیازهای پژوهشی و توانمندی‌ها و منابع پژوهشی این حوزه ضروری است. این پژوهش درصدد است راهکار مناسبی را مبتنی بر تحلیل محتوای پژوهش‌های موجود و تقاضاهای جامعه پژوهشی از طریق کاربست تکنیک‌های متن‌کاوی[۲] و پردازش زبان طبیعی[۳] و تحلیل رفتار جویشگری کاربران این حوزه ارائه نماید که از طریق آن کاستی‌های پژوهشی به‌صورت مستقل از نظر خبرگان حوزه محیط‌زیست و در فرایندی رایانشی شناسایی شود. برای این منظور بیش از ۵۰۰ هزار رکورد پایان‌نامه‌ها و رساله‌های موجود در پایگاه ایرانداک و نیز بیش از ۲۵۰ میلیون رکوردجستجوهای صورت پذیرفته توسط کاربران این پایگاه تحلیل و در نهایت رکوردهای مرتبط حوزه محیط‌زیست شناسایی شد. سپس تحلیل کتابشناختی این منابع ارائه شد و نیز با استفاده از مدل‌سازی موضوعی ۲۰ موضوع در این حوزه شناسایی و استخراج شد که پس از ارزیابی عرضه و تقاضای این موضوعات، مشخص شد که موضوعات «صنعت»، «رودخانه»، «حقوق بین‌الملل» و «زباله» در حوزه محیط‌زیست دارای کاستی پژوهش هستند و نیازمند توجه بیشتر مدیران و سیاست‌گذاران پژوهشی حوزه محیط‌زیست هستند.

واژه‌های کلیدی: محیط‌زیست، شناسایی کاستی پژوهش، پردازش زبان طبیعی، علم‌سنجی، متن‌کاوی.

---

[۱] این پژوهش مستخرج از رساله دکترای نویسنده با عنوان « شناسایی کاستی‌های پژوهش با به‌کارگیری تکنیک‌های متن‌کاوی و تحلیل رفتار جویشگری کاربران (موردمطالعه: محیط‌زیست ایران) » از دانشگاه علم و صنعت ایران است.

[۲] Text Mining

[۳] Natural Language Processing (NLP)

۱- مقدمه

فاصله یا شکاف میان پژوهش‌های دانشگاهی[۱]، به‌ویژه پایان‌نامه‌ها و رساله‌ها با نیازهای جامعه یکی از چالش‌های اساسی نظام آموزش عالی در ایران است. گسترش کمّی دانشگاه‌ها و افزایش پذیرش دانشجویان تحصیلات تکمیلی باعث شده است تا از یک طرف شاهد انتقادها و واکنش اندیشمندان و خبرگان آموزش و پژوهش باشیم و از طرف دیگر کارفرمایان صنایع و اقتصاد از توانمندی اندک دانش‌آموختگان در حل مسائل واقعی کشور گله‌مند باشند [۱]. همچنین عرصه پژوهش شاهد دوباره‌کاری‌های بسیاری است که دلیل عمده آن، عدم توانایی استادان و دانشجویان در یافتن مسئله پژوهش در حوزه‌های مختلف موضوعی موردنیاز جامعه است و در بسیاری از موارد جامعه دانشگاهی بجای «مسئله‌یابی» به دنبال «مسئله‌سازی» شده است. از سوی دیگر اولویت‌های پژوهشی که بیشتر وزارتخانه‌ها و سازمان‌ها مشخص کرده و به نحوی به اطلاع پژوهشگران می‌رسانند، صرف‌نظر از شکل و روش‌های ناکارآمد، کمتر به شکل واقع‌بینانه و کاربردی مورد شناسایی قرارگرفته‌اند [۱].

خلأ یا کاستی دانش[۲] فاصله موجود میان دانش موردنیاز یک سازمان برای انجام وظایف و رسیدن به اهداف تعیین‌شده با دانشی که سازمان از طریق منابع مختلف دانشی خود از آن بهره‌مند است را بیان می‌کند [۲]. کاستی دانش هم می‌تواند در مورد دانش موجود و آرشیوشده در یک سازمان، نهاد یا جامعه صدق کند و هم می‌تواند دانش قابل‌استقرار و کارآمد را مدنظر داشته باشد [۳]. در تعریفی دیگر، کاستی دانش درکی از میزان اختلاف بین دانش موجود و کد شده با دانش موردنیاز یک سیستم یا فرد برای رسیدن به اهداف خود است [۴, ۵]. همین تعریف برای بیان کاستی پژوهش نیز کاربرد دارد و به عبارتی جای خالی پژوهش‌های موردنیاز یک سازمان یا جامعه را در بین منابع پژوهشی در دسترس سازمان یا جامعه را شامل می‌شود. این در حالی است که خلأ یا شکاف پژوهش (Research Gap) در مطالعه یک موضوع مبین کاستی‌ها و نواقص پژوهش‌های پیشین در آن موضوع خاص است و پژوهشگر سعی دارد تا آن کاستی‌ها را برطرف نماید. به عبارتی در مقیاس کل‌نگر واژه «Research Gap» به معنای حوزه‌هایی است که در پژوهش‌های یک جامعه یا سازمان مغفول مانده‌اند یا به اندازه کافی در آن حوزه‌ها پژوهش صورت نپذیرفته است. اصطلاح «خلأ پژوهشی» و «نیاز پژوهشی»[۳] در بسیاری از پژوهش‌ها بجای یکدیگر استفاده می‌شود و این مسئله در بیشتر اوقات خللی بر مفهوم مدنظر پژوهشگر وارد نمی‌کند اما در مدیریت پژوهش این دو مقوله تفاوت محسوسی با یکدیگر دارند که لازم است در این قسمت این تفاوت بیان شود. بر اساس تعریف رابینسون[۴]، کاستی پژوهشی، محدودیت اطلاعاتی در حوزه یا موضوعی است که مانع از نتیجه‌گیری در مورد یک سؤال پژوهشی می‌شود و نیاز پژوهشی محدودیت اطلاعاتی است که تصمیم‌گیری سیاست‌گذار را محدود می‌کند [۶]. بر همین اساس پژوهش‌های متعددی نیز بحث خلأ را در دانش یا اطلاعات مطرح کرده و بر اساس این خلأ، نیاز پژوهشی را تعیین کرده‌اند [۷-۱۲]. به عبارتی بسیاری از پژوهش‌هایی که با عنوان تعیین یا شناسایی نیازهای پژوهشی مطرح شده‌اند و در فرایند شناسایی از تحلیل الگوهای جستجوی اطلاعات کاربران پایگاه‌های اطلاعاتی یا کاربران کتابخانه‌ها استفاده کرده‌اند در بهترین حالت به تعیین یا شناسایی تقاضاهای پژوهشی دست یافته‌اند[۲].

به دلیل جغرافیای خاص ایران، محیط‌زیست به‌عنوان یکی از چالش‌های اساسی کشور همیشه مطرح بوده است و بر همین اساس از سال‌ها پیش، رشته‌ها و گرایش‌های مرتبط با آن در نظام آموزش عالی ایران ایجاد شده است. به همین خاطر سالانه پژوهش‌های متعددی در این حوزه در کشور انجام می‌شود و به خاطر لزوم ارائه پیوست‌های زیست‌محیطی در بسیاری از پروژه‌های کلان، همچنان نیازهای پژوهشی متعددی نیز در این حوزه مطرح می‌شود [۱۳]. ذات میان‌رشته‌ای بودن این حوزه از علم موجب شده است تا رده‌بندی پژوهش‌ها که لازمه تحلیل اطلاعات و شناخت از دارایی‌های پژوهشی و نیازهای پژوهشی آن است، پیچیده‌تر باشد. اهمیت بررسی و تحلیل پژوهش‌های حوزه محیط‌زیست و نیز میان‌رشته‌ای بودن این حوزه موجب شده است تا محیط‌زیست به‌عنوان نمونه موردمطالعه در این پژوهش مدنظر قرار گیرد.

---

[۱] Research gap

[۲] Knowledge Gap

[۳] Research Needs

[۴] Robinson, Karen A

پایگاه ایرانداک که به نام «گنج» در اختیار کاربران قرار دارد و مهم‌ترین کاربران آن را پژوهشگران و دانشجویان تحصیلات تکمیلی تشکیل می‌دهند منابع گوناگون علمی از رشته‌های علمی مختلف را گردآوری، سازمان‌دهی و اشاعه می‌کند. تمرکز اصلی این پایگاه بر سازمان‌دهی پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی است و بزرگ‌ترین و تنها مرجع رسمی است که وظیفه ثبت، سازمان‌دهی و اشاعه پایان‌نامه‌ها و رساله‌ها را به عهده دارد. این پایگاه در حال حاضر بیش از ۱ میلیون رکورد علمی را شامل می‌شود که در حال حاضر بیش از ۵۰۰ هزار رکورد آن مربوط به پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی داخل کشور و بیش از ۱۵ هزار رکورد آن مربوط به پایان‌نامه‌ها و رساله‌های ایرانیان خارج از کشور است. به‌طور متوسط روزانه بیش از ۱۳۰ هزار جستجو در این پایگاه انجام می‌شود [۱۴]، منبع مورد استفاده در این پژوهش برای استخراج منابع مرتبط با حوزه محیط‌زیست این پایگاه است.

هدف اصلی این پژوهش ارائه روشی رایانشی و مستقل از خبرگان و مبتنی بر متن‌کاوی در پایگاه اطلاعاتی منابع پژوهشی است که با استفاده از تحلیل حجم انبوهی از جستجوهایی که کاربران در موضوع محیط‌زیست انجام می‌دهند و مقایسه آن با منابع پژوهشی موجود در پایگاه، بتوان کاستی‌های پژوهشی این حوزه را شناسایی نمود، به‌نحوی‌که سیاست‌گذاران و مدیران پژوهشی با آگاهی از آن و با در اختیار داشتن دیگر منابع اطلاعاتی بتوانند نیازهای پژوهشی را تعیین نموده و اولویت‌دهی پژوهش‌های حوزه محیط‌زیست را به نحوی کارا و اثربخش انجام دهند.

۲- پیشینه پژوهش

گستره پژوهش‌های مربوط به علم‌سنجی در حوزه‌های مختلف بسیار وسیع است. بااین‌حال با هدف شناسایی رویکردها و روش‌های مختلف شناسایی کاستی‌ها، نیازسنجی و اولویت‌بندی پژوهش در ادامه به برخی از این پژوهش‌ها خواهیم پرداخت.

گوبستر[۱] در پژوهش خود با بررسی مقالات مربوط به ۴۰ سال از مجلات مرتبط با معماری و طراحی شهری توانسته است مقالات را در ۳ زیرشاخه‌ی اصلی تقسیم‌بندی کند و با شناسایی کلیدواژه‌های مشترک بین مقالات در این ۳ زیرشاخه، موضوعات داغ آن‌ها را به‌صورت نقشه واژه[۲] نشان داده است؛ او در پژوهش خود از ترکیبی از متن‌کاوی و بصری‌سازی برای نمایش بهتر ساختار محتوایی پژوهش‌ها و چگونگی توزیع و چگالی موضوعات داغ[۳] در هر یک از حوزه‌ها استفاده کرده است [۱۵]. همچنین برای تحلیل روند رشد مقوله مدیریت، تجارت و اقتصاد پایدار، رکوردهای مرتبط با این حوزه، از پایگاه اطلاعاتی کتابخانه بریتانیا[۴] استخراج شده است. این رکوردها از طریق جستجوی عبارت «*Sustain» در زیرمجموعه‌های مدیریت و کسب‌وکار انتخاب شده‌اند و نتایج نشان داده است که هم در تعداد منابع و هم در تعداد ارجاعات، مقوله پایداری رشد سریعی داشته است. همچنین روندها نشان می‌دهد که مفهوم «توسعه پایدار» از اواسط دهه اول قرن ۲۱ به سمت «پایداری به‌منزله یک تمرین مدیریتی» تغییر رویه داده است [۱۶]. مطالعه کتاب‌شناختی اطلاعات مربوط به مقالات ۲۰ ساله حوزه «ارزیابی زیست‌محیطی» موجب شناسایی روندها، موضوعات و کشورهای فعال در این حوزه شده است [۱۷]. همچنین پژوهش دیگری با استفاده از خوشه‌بندی کلمات کلیدی مقالات حوزه مهندسی برق، مهندسی مکانیک، مهندسی عمران، علوم تربیتی، الهیات و مدیریت حوزه‌های در حال رشد و حوزه‌هایی که به بلوغ رسیده‌اند را شناسایی نموده است [۱۸].

پژوهش‌های متعددی در کشور با عنوان «تحلیل محتوا» و با استفاده از منابع پایگاه‌های علمی فارسی صورت پذیرفته‌اند، اما عمده این پژوهش‌ها صرفاً به تحلیل آماری فراداده‌های متون علمی پرداخته‌اند و بررسی محتوای متون علمی که نیازمند استفاده از تکنیک‌های پیچیده پردازش زبان طبیعی است مدنظر نبوده است [۱۹, ۲۰]. به‌عنوان نمونه، خانم عصاره و همکاران [۲۰] تعداد ۹۴۲ مقاله فارسی مجلات علمی، پژوهشی و علمی ترویجی در حوزه علم اطلاعات و دانش‌شناسی را در بازه زمانی ۱۳۸۴ تا ۱۳۸۸ موردبررسی قرار داده‌اند. در بسیاری از موارد، این پژوهش‌ها صرفاً به توصیف کتاب‌شناختی منابع یک حوزه پژوهشی پرداخته است، مانند آنچه که در توصیف پایان‌نامه‌های رشته کتابداری و اطلاع‌رسانی [۲۱]، بررسی پایان‌نامه‌های رشته جهانگردی [۲۲] و یا ارائه آمار پایان‌نامه‌های

[۱] Gobster Paul H.
[۲] Term Map
[۳] Hot Topics
[۴] British Library

مربوط به موضوع انقلاب اسلامی [۲۳] و مانند انجام شده است. در این پژوهش‌ها صرفاً اطلاعات فراداده‌های پایان‌نامه‌ها مصورسازی شده است و صحبتی از نیازسنجی پژوهش و یا شناسایی کاستی به میان نیامده است.

در پژوهشی که ربیعی و همکاران با به‌کارگیری لاگ جستجوی کاربران در پایگاه اطلاعاتی ایرانداک در حوزه محیط‌زیست و بر مبنای روش CRISP-DM[۱] انجام داده‌اند، تعداد ۱۰۶,۳۱۴ رکورد مربوط به جستجوهای انجام شده در حوزه محیط‌زیست را مورد تحلیل قرار داده‌اند و میزان جستجوی هر یک زیرشاخه‌های محیط‌زیست و تعداد پاسخ موتور جستجو به هر یک از آن‌ها را مبنای تحلیل قرار داده‌اند و درنهایت زیرشاخه‌ها را در چهار گروه «جستجوی زیاد و پاسخ زیاد»، «جستجوی زیاد و پاسخ کم»، «جستجوی کم و پاسخ زیاد» و «جستجوی کم و پاسخ کم» دسته‌بندی کرده است [۱۳]. هرچند که این پژوهش از تقاضاهای کاربران برای شناسایی کاستی‌های پژوهش استفاده کرده است اما مبنای کار صرفاً بعد تقاضای پژوهش بوده است و فرض بر این بوده است که پاسخ موتور جستجو برای تعداد منبع قابل‌عرضه قابل‌اتکا است. در این پژوهش شناسایی جستجوهای مربوط به حوزه محیط‌زیست از طریق تعیین تعدادی کلیدواژه مرتبط انجام شده است.

بدین ترتیب شکاف‌ها و خلأهای پژوهش‌های پیشینی که در زمینه شناسایی کاستی‌های پژوهشی ارائه شده‌اند عبارت‌اند از:

- توجه کمتر به روش‌های مبتنی بر اکتشاف دانش از محتوای پژوهشی و تأکید بر تحلیل‌های آماری از اطلاعات کتابشناختی پژوهش‌ها،
- اکتفا به نظر تعداد محدود خبرگان در شناسایی کاستی‌ها و نیازهای پژوهشی با پذیرش محدودیت‌ها و نواقص روش‌های مبتنی بر نظرسنجی از خبرگان،
- توجه بیش‌ازحد به بعد عرضه پژوهش و شناسایی تقاضاهای پژوهش از طریق تحلیل و طبقه‌بندی پژوهش‌های عرضه‌شده که مانع از شناخت موضوعات موردنیاز جدید و ناب مطرح نشده در پژوهش‌های عرضه‌شده می‌شود،
- پو شش تعداد محدودی از منابع پژوه شی و تحلیل نمونه‌های کوچک از منابع پژوه شی به دلیل محدودیت‌های موجود در رویکردها و روش‌های مورد استفاده و کاربرد اندک روش‌های رایانشی و هوشمند استخراج دانش در روش‌های پیشین،
- کمبود شدید پژوهش‌هایی که عرضه و تقاضای منابع علمی حوزه محیط زیست را رصد نمایند.

۳- روش‌شناسی پژوهش

این پژوهش از منظر رویکردی تلفیقی از رویکردهای تحلیل موضوعی/ محتوایی، رویکرد عرضه‌محور و رویکرد تقاضامحور را به کار می‌برد. همچنین مطالعه حاضر از نظر روش اجرا روشی تلفیقی را برگزیده است که با به‌کارگیری تکنیک‌های متن‌کاوی از روش‌های کمّی، کتاب‌شناختی و تحلیل محتوا بهره می‌برد. این پژوهش بر مبنای تحلیل رفتار جویشگری کاربران به شناسایی تقاضاهای پژوهشی حوزه محیط‌زیست می‌پردازد و از تکنیک‌های پردازش زبان طبیعی برای تحلیل اطلاعات پژوهش‌های موجود و نیز رصد تقاضاهای پژوهش این حوزه استفاده خواهد کرد که در ادامه تکنیک‌های مورد استفاده معرفی خواهند شد. این تکنیک‌ها متناسب با نیازها و بافت پژوهش، بهبودیافته و شامل نوآوری‌هایی خواهند بود که درنهایت در فصل بعد روی داده‌ها پیاده‌سازی خواهند شد.

رده‌بندی متن[۲]، تخصیص یک سند متنی به یک رده از پیش تعریف‌شده را گویند [۲۴]. در این میان رده‌بندی متون پژوهشی به دلیل کاربرد فراوانی که در نمایه‌سازی متون داشته و نیز امکان تحلیل و رصد دارایی پژوهشی[۳] یک جامعه پژوهشی را برای پژوهشگران و سیاست‌گذاران فراهم می‌کند، دارای اهمیت ویژه‌ای است. در این زمینه تاکنون روش‌های ماشینی یا نیمه‌ماشینی مختلفی مبتنی بر یادگیری ماشین ارائه شده است.

فرایند شناسایی متون مرتبط با حوزه محیط‌زیست، با استفاده از متن‌کاوی در پایان‌نامه‌ها و رساله‌های این حوزه و با به‌کارگیری رویکرد رده‌بندی با استفاده از ماشین بردار پشتیبان تک‌رده‌ای انجام شده است. برای این منظور پس از استخراج مجموعه اولیه داده‌های

---

[۱] Cross Industry Standard Process for Data Mining

[۲] Text Classification

[۳] Research Asset

حوزه محیط‌زیست، مجموعه داده‌های هسته این حوزه به‌منظور استفاده در فرایند یادگیری ماشین ایجاد شده است. سپس از تکنیک‌های پردازش زبان طبیعی در مراحل پیش‌پردازش و تولید ماتریس بردار ویژگی‌ها استفاده شده است و درنهایت مجموعه داده‌های آزمون ایجاد و بر اساس پارامترهای مختلف، مدل مورد ارزیابی و بهبود قرار گرفته است. درنهایت با استفاده از این مدل، مجموعه داده‌های مربوط به محیط‌زیست از بین جستجوهای انجام شده و نیز منابع پژوهشی موجود در پایگاه استخراج شده است.

برای هر مستند لازم است تا سبدی از واژه‌های مرتبط تهیه شود که این سبد نماینده مناسبی از محتوای آن سند باشد. برای این منظور فرایندی ارائه شده است که با استفاده از N-gramهای مستخرج از کلیدواژه و عنوان، سبد واژه‌ها تشکیل خواهد شد. N-gram مجموعه‌ای شامل N عنصر پشت سر هم در یک نوشتار یا صحبت را گویند که می‌توانند حرف، آوا، واژه و... باشند [۲۵].

$$n_gram_i(St,n)=word_i word_{i+1} ... word_{i+n-1} \quad (۱)$$

با توجه به فرمول (۱) در این پژوهش منظور از N-gram، یک عبارت شامل N واژه پشت سرهم است. عناوین Uni-gram، Bi-gram، Three-gram و Quad-gram به ترتیب برای n=1,2,3,4 مصطلح است. در ادامه برای آنکه علاوه بر تعداد دفعات تکرار عبارت، اهمیت تعداد اجزای تشکیل‌دهنده (تعداد کلمات) نیز در این پارامتر مدنظر قرار گیرد، ارزش عبارت از حاصل‌ضرب این دو عامل به‌دست‌آمده است.

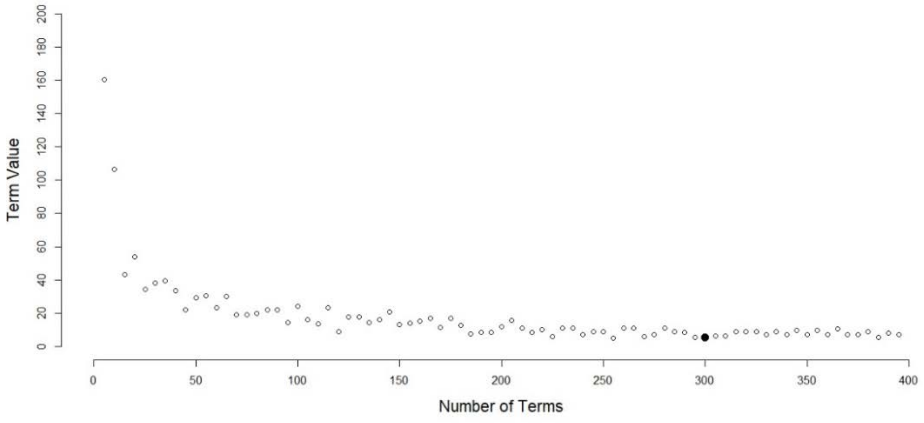
$$A(Term)=\{S|\forall i\in S, Term\in BoW(Doc_i)\}$$

$$Index(Term) = \max_{S\in A(Term)} |S| \quad (۲)$$

$$Words(Term)= \max n | Term\in n_gram_1(Term,n)$$

$$TermValue(Term) = Words(Term) * Index(Term)$$

شکل ۳-۱ مقادیر ارزش عبارت را بر اساس تعداد عبارت‌ها در ماتریس بردار ویژگی‌ها نشان می‌دهد.



شکل ۱-مقدار ارزش عبارت برای عبارت‌های پرتکرار ماتریس ویژگی‌ها

همان‌طور که در شکل ۱مشخص شده است، ارزش عبارت در ماتریس بردار ویژگی برای ۱۰۰ عبارت اول دارای نوسان‌های قابل‌ملاحظه است و برای عبارت‌های پس از ۳۰۰ عبارت پرارزش این نوسان بسیار اندک است و نمودار آن تقریباً نزدیک به صفر است، ازاین‌رو ۳۰۰ عبارت پرارزش در این ماتریس به‌عنوان عبارت‌های بردار ویژگی‌های متون انتخاب شده است. سپس ماتریس بردار ویژگی

به عنوان ورودی در روش ماشین بردار پشتیبان[1] تک (SVM) ارائه شده است و با این ترتیب منابع مرتبط با حوزه محیط‌زیست شناسایی شده است. در این شناسایی هر یک از رکوردها تا میزانی مرتبط با حوزه محیط‌زیست شناسایی شده‌اند به همین دلیل تعداد منابع شناسایی شده زیاد است اما از آنجا که وزن این ارتباط در ادامه فرایند مدنظر قرار گرفته است مشکلی بابت شناسایی منابع کمتر مرتبط بوجود نخواهد آمد. در نهایت از میان ۵۱،۳۹۱۵ پایان‌نامه و رساله موجود در پایگاه، تعداد ۱۰۶،۷۰۹ به نحوی با حوزه محیط‌زیست مرتبط شناخته شدند و از میان بیش از ۲۵۰ میلیون رکورد جستجو، پس از پالایش جستجوها بیش از ۷۶ میلیون رکورد به عنوان جستجوهای کاربران واقعی (بجز رباتها و موتورهای جستجوگر) شناسایی شدند که از این میان ۱،۱۹۲،۱۵۳ رکورد (حدود ۱/۵ درصد) از جستجوهای انجام‌شده مربوط به حوزه محیط‌زیست بوده است.

جدول ۱: نمونه‌ای از جستجوهای شناسایی شده مرتبط با حوزه محیط‌زیست

شناسه	متن جستجو	تعداد پاسخ	زمان جستجو
۱۲	("بیابان‌زایی", "خراسان", "فرسایش بادی")	۲	۱۳۹۱/۱۱/۲۸
۴۱۱	پدیده گرد و غبار	۱۴	۱۳۹۱/۱۱/۲۸
۸۲۳	فاضلاب صنعتی	۶۸۵	۱۳۹۱/۱۱/۲۸
۲۷۳۵	فرسایش و رسوب در سرند چای	۰	۱۳۹۱/۱۱/۲۸
۵۵۱۶	بررسی بهره‌وری و تابع تقاضای آب در تولید گندم در منطقه رامشیر	۰	۱۳۹۱/۱۱/۲۸
۵۷۷۰	("رسوب معلق","شبکه عصبی")	۲۷	۱۳۹۱/۱۱/۲۸

پس از شناسایی منابع لازم است تا آنها بصورت دسته‌بندی شده در قالب موضوع[2]های مختلف مورد تحلیل قرار گیرند. مدل‌سازی موضوعی[3] از روش‌های کارای تحلیل متن است که هدف آن گروه‌بندی مجموعه‌ای از متون مختلف به‌صورت زیرمجموعه‌های معنادار که موضوع نامیده می‌شوند است. روش‌های مدل‌سازی موضوعی با حداقل دخالت انسان این فرایند را انجام می‌دهند و همین نکته باعث جذابیت و استفاده روزافزون آن‌ها در تحلیل حجم عظیمی از متون شده است. در این روش بجای استفاده از عناوین از قبل تعیین شده، تنها با تعیین تعداد موضوعات، زیرمجموعه‌های موضوعی استخراج شده و میزان ارتباط هر سند با هر زیرمجموعه نیز تعیین می‌شود [۲۶]. برای مدل‌سازی موضوعی از الگوریتم LDA که رایج‌ترین شکل از مدل‌سازی موضوعی است و توسط بیلی[4] و همکارانش در سال ۲۰۰۳ ارائه شد، استفاده شده است. این مدل بر این فرضیه بنانهاده شده است که هر مستند شامل مجموعه‌ای از کلمات (BoW) است که ترکیبی از موضوعات مدنظر نویسنده از طریق این کلمات بیان شده است. هر موضوع نیز توزیعی از کلمات پرتکرار است. ازاین‌رو هر مستند را می‌توان با میزانی از ارتباط به موضوعات مختلف متعلق دانست [۲۷].

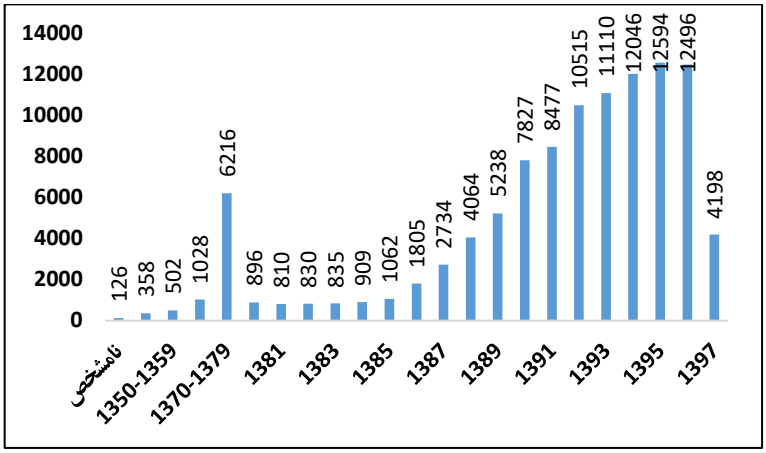
۴- یافته‌های پژوهش

۴-۱- تحلیل زمانی پژوهش‌های حوزه محیط‌زیست

تحلیل زمانی روند انجام پایان‌نامه‌ها و رساله‌های مرتبط با حوزه محیط‌زیست در شکل ۲ آمده است. این بررسی در بیش از نیم‌قرن گذشته در ایران، نشان می‌دهد که تا سال ۱۳۷۰ به‌طور متوسط کمتر از سالانه ۱۰۰ مورد پایان‌نامه مرتبط با مسائل زیست‌محیطی تولیدشده است که عمدتاً مربوط به رشته‌های داروسازی، بهداشت، دامپزشکی و عمران است. این مقدار برای سال‌های بین ۱۳۷۰ تا

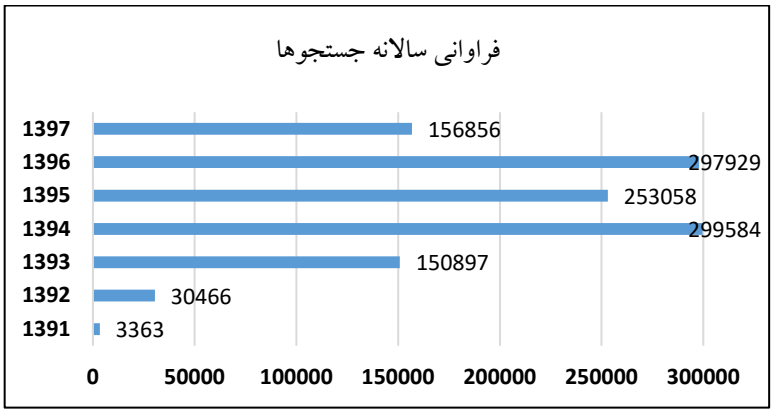
[1] Support Vector Machine (SVM)
[2] Topic
[3] Topic Modeling
[4] Blei, David M

۱۳۸۰ رشد ۶ برابری داشته و به‌طور متوسط سالانه بیش از ۶۰۰ پایان‌نامه در این حوزه تولیدشده است. این رشد سریع تا سال ۱۳۹۲ ادامه داشته است و در پنج سال اخیر تقریباً در محدوده متوسط ۱۲۰۰۰ پایان‌نامه در هر سال ثابت مانده است.



شکل ۲-روند زمانی پایان‌نامه‌های مرتبط با محیط‌زیست

در شکل ۳، تحلیل زمانی جستجوهای مرتبط با محیط زیست آمده است. جستجوهای استخراج شده مرتبط با حوزه محیط‌زیست مربوط به بازه زمانی بهمن‌ماه ۱۳۹۱ تا مهرماه ۱۳۹۷ است.

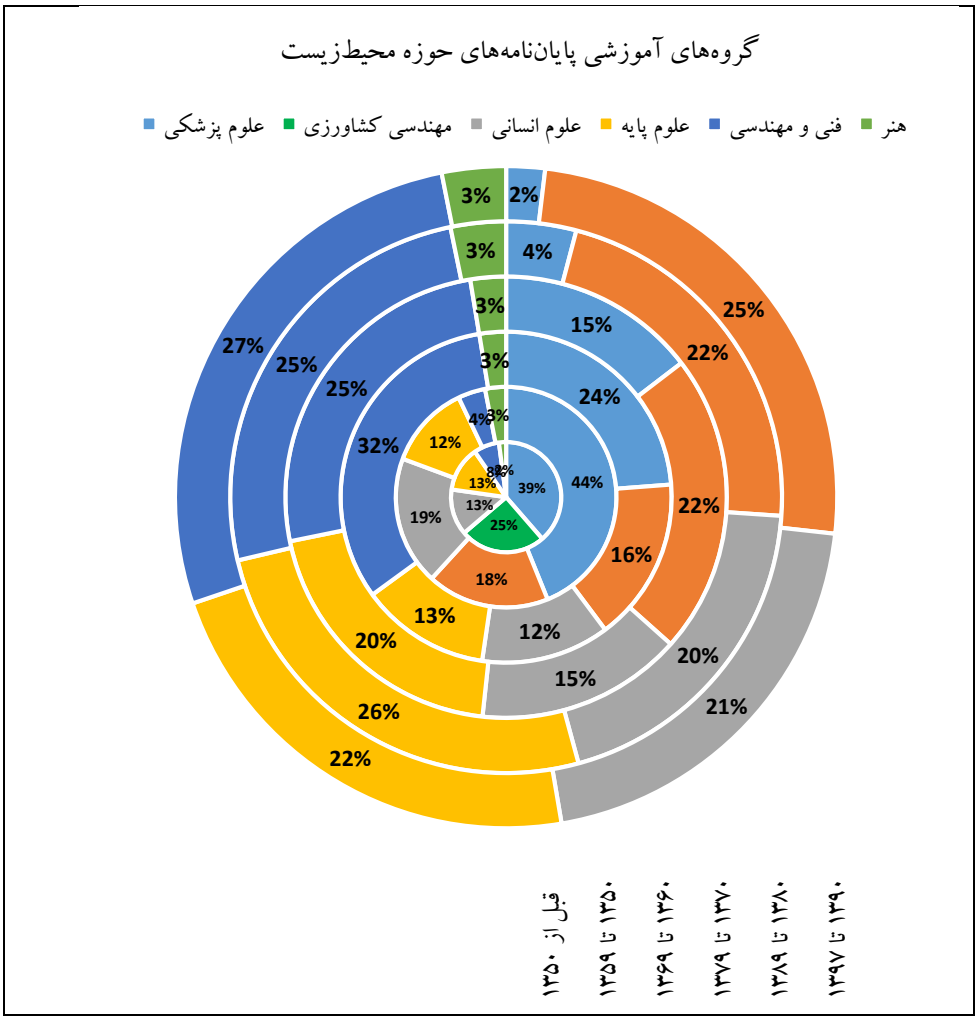


شکل ۳- نمودار فراوانی سالانه جستجوهای حوزه محیط‌زیست

توجه به این نکته ضروری است که به دلیل آنکه داده‌های سال اول (۱۳۹۱) فقط مربوط به بازه کمتر از ۲ ماه بوده و سال آخر (۱۳۹۷) داده‌های ۷ ماه اول از سال را شامل می‌شود، رکوردهای مربوطه کمتر از سال‌های دیگر استاطلاعات شکل ۳ تائید کننده این موضوع است که محیط‌زیست یک دانش کاملاً میان‌رشته‌ای بوده و تمامی گروه‌های آموزشی به میزان مختلف در دانش این حوزه سهیم هستند. میزان مشارکت گروه‌های آموزشی در دوره‌های زمانی مختلف متفاوت بوده است.

۴-۲- گروه‌های آموزشی عرضه‌کننده منابع حوزه محیط‌زیست

بررسی شکل ۴ نشان می‌دهد که گروه هنر ۳ درصد از منابع این حوزه را تولید کرده است و یک‌چهارم مطالعات این حوزه در گروه آموزشی کشاورزی صورت پذیرفته است و این دو گروه در طول زمان، تفاوت قابل‌توجهی در میزان مشارکت خود در تولیدات پژوهشی این حوزه نداشته‌اند. مشارکت گروه آموزشی علوم انسانی و علوم پایه در طول زمان افزایش یافته است و سهم ۱۳ درصدی هر یک از آن‌ها در مطالعات قبل از ۱۳۵۰ به بالای ۲۰ درصد در مطالعات دهه اخیر حوزه محیط‌زیست رسیده است.



شکل ۴- سهم گروه‌های آموزشی در پژوهش‌های محیط‌زیست در هر دوره زمانی

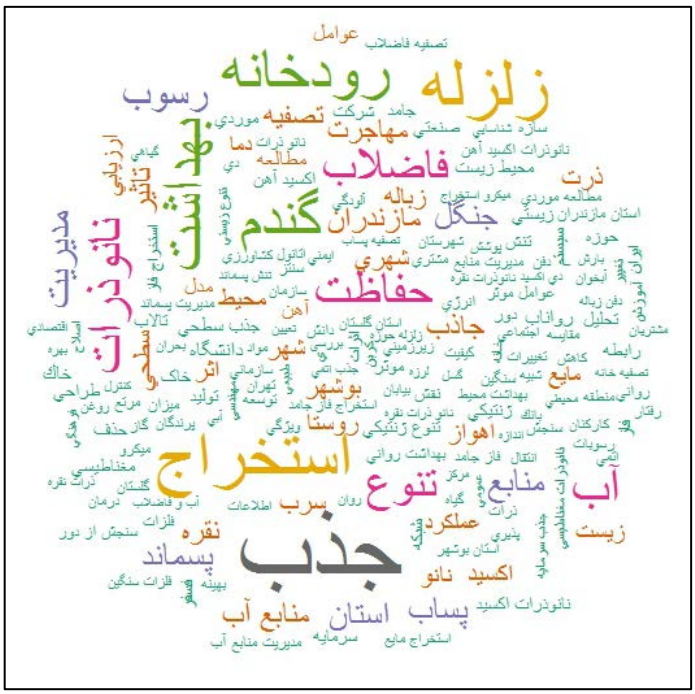
نکته قابل‌توجه در مشارکت گروه‌های آموزشی در منابع پژوهشی حوزه محیط‌زیست، تغییر ماهیت مطالعات این حوزه از علوم پزشکی به سمت علوم فنی و مهندسی است. سهم گروه فنی و مهندسی از مطالعات این حوزه تا سال ۱۳۶۰ بین ۴ تا ۷ درصد بوده است، اما در دهه‌های اخیر رویکرد مهندسی در مطالعات مربوط به حوزه محیط‌زیست توسعه پیدا کرده است و در حال حاضر این گروه بیش از تمامی گروه‌های آموزشی (۲۷ درصد) در تأمین منابع پژوهشی این حوزه نقش دارد. این در حالی است که گروه آموزشی علوم پزشکی که تا سال ۱۳۶۰ نزدیک به ۴۰ درصد از مطالعات این حوزه را پوشش داده، در سال‌های اخیر کمتر به این موضوع پرداخته است و در حال حاضر کمتر از ۵ درصد مطالعات این حوزه را پوشش می‌دهد. یکی از دلایل این تغییر، تشکیل رشته‌های مهندسی محیط‌زیست در دانشکده‌های فنی و مهندسی در دهه‌های اخیر بوده است. همچنین رشته‌های مهندسی مانند مهندسی عمران، مهندسی شیمی، مهندسی مکانیک، مهندسی صنایع و ... بخش قابل‌توجهی از مطالعات این حوزه را در دهه‌های اخیر انجام داده‌اند.

۴-۳- تحلیل عبارت‌های پرکاربرد حوزه محیط‌زیست

نتیجه بررسی عبارت‌های پرکاربرد در منابع مربوط به حوزه محیط‌زیست در شکل۵ بصورت ابر واژگان کلمات پرکاربرد منابع حوزه محیط‌زیست آمده است. در این شکل اندازه هر عبارت نشان‌دهنده اهمیت و فراوانی تکرار آن عبارت در متون مربوط به محیط‌زیست است.

شکل ۵- ابر واژگان کلمات پرکاربرد منابع عرضه شده حوزه محیط‌زیست

همچنین شکل ۶ ابر واژگان [۱] کلمات پرکاربرد جستجوهای حوزه محیط‌زیست را نشان می‌دهد.



شکل ۶- ابر واژگان جستجوهای مرتبط با محیط‌زیست

از این عبارت‌ها می‌توان در بررسی اینکه جستجوهای استخراج شده تا چه حدی با مباحث موجود در حوزه محیط‌زیست تطابق دارند، استفاده کرد. بررسی شهودی این عبارت‌ها تأییدکننده این است که جستجوهای انتخاب‌شده کاملاً مرتبط با موضوعات مختلف قابل‌بحث در حوزه محیط‌زیست هستند.

[۱] WordCloud

۴-۴- تحلیل موضوعات شناسایی شده در حوزه محیط‌زیست

پس از مدل‌سازی موضوعی، ۲۰ موضوع در این حوزه شناسایی شد. سپس برای هر موضوع با شمارش عبارت‌های پرتکرار موجود در سبد واژگان مستنداتی که با میزانی بیش از th_e در آن موضوع تشریح شده‌اند عبارت‌های نماینده هر موضوع به دست آمد که در جدول ۲ فهرست شده‌اند. با توجه به این عبارت‌ها، به هر یک از موضوعات برچسبی تخصیص داده شده است. در این پژوهش، برچسب‌گذاری صرفاً به دلیل ملموس‌تر و قابل‌درک‌تر شدن تفسیر موضوعات انجام شده است و تأثیری در نتایج نخواهد داشت.

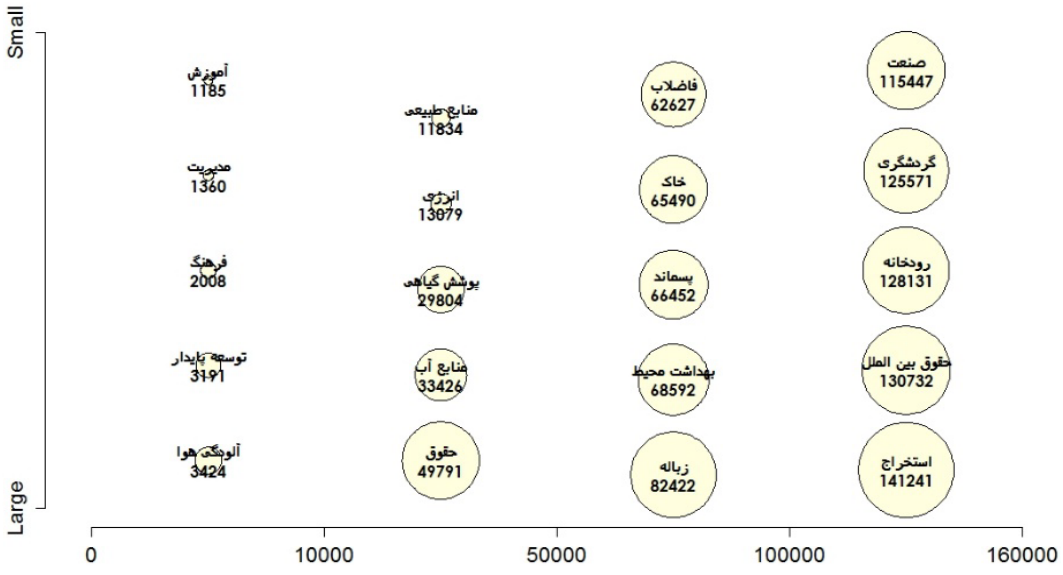
جدول ۲- عبارت‌های پرتکرار موضوعات استخراج شده

موضوع	برچسب	عبارت‌های پرتکرار
T1	خاک	خاک، آلودگی، سرب، فلزات سنگین، آرسنیک، جذب، کروم، آلودگی خاک، کادمیم، عناصر سنگین، آهن، نیکل، کادمیوم
T2	استخراج	استخراج، میکرواستخراج، استخراج فاز جامد، پیش‌تغلیظ، جداسازی، کروماتوگرافی گازی، نانوذرات، طراحی آزمایش، سرب
T3	بهداشت محیط	بهداشت محیط، ارزیابی اثرات زیست‌محیطی، توسعه، بهداشت، بازیافت، ارزیابی زیست‌محیطی، ایمنی، آلودگی، زباله، کمپوست، توسعه پایدار، جمعیت، صنعت
T4	انرژی	انرژی، مصرف انرژی، نیروگاه، توسعه، انرژی خورشیدی، خلیج فارس، رشد اقتصادی، انرژی تجدیدپذیر، زنجیره تأمین سبز، اقلیم، پایداری، معماری سبز
T5	حقوق	مسئولیت مدنی، توسعه، تخریب، حقوق، حقوق محیط‌زیست، نظام حقوقی، مشارکت، توسعه پایدار، رفتار زیست‌محیطی، منابع طبیعی، اخلاق زیست‌محیطی
T6	فاضلاب	تصفیه، فاضلاب، پساب، تصفیه فاضلاب، فضای سبز، نیترات، تصفیه پساب، محیط‌زیست شهری، شیمی سبز، آلاینده، کروم، آمونیاک
T7	آلودگی هوا	دی اکسید کربن، رشد اقتصادی، آلودگی، آلودگی هوا، منحنی زیست‌محیطی کوزنتس، انتشار دی اکسید کربن، آلاینده، پلی آکریل آمید، آلدهیدها، سازگار با محیط‌زیست
T8	گردشگری	گردشگری، توسعه، اکوتوریسم، سنجش از دور، اکولوژیکی، صنعت، گردشگری روستایی، نیتروژن، توسعه شهری، کاربری زمین، بوم‌سامانه
T9	توسعه پایدار	پایداری، پایداری زیست‌محیطی، توسعه، اماکن ورزشی، عوامل زیست‌محیطی، توسعه پایدار، اثرات زیست‌محیطی، وانادیم، جمعیت، مازندران
T10	مدیریت	آموزش محیط‌زیست، توسعه پایدار، زنجیره تامین سبز، مدیریت زیست‌محیطی، مدیریت زنجیره تامین سبز، سیستم مدیریت زیست‌محیطی، حفاظت، مشارکت، پایداری، برنامه‌ریزی در سی، ارزشیابی، مصرف انرژی
T11	زباله	زباله، آلودگی محیط‌زیست، شیرابه، environmental pollution، لندفیل، بیوگاز، آلودگی، Pollution، دفع زباله، خاک، GEOCHEMISTRY، pollution، زمین‌شیمی
T12	پسماند	پسماند، آلاینده، نانوذرات، مسئولیت اجتماعی، سیمان، مدیریت پسماند، نانوکامپوزیت، امواج فراصوت، تخریب، کارخانه سیمان، مقاومت فشاری، بازیافت
T13	حقوق بین‌الملل	environmental protection، نظام حقوقی، حقوق بین‌الملل، دریای خزر، international law، system legal، توسعه روستایی، sustainable development، تنوع زیستی، biodiversity، پارس جنوبی
T14	رودخانه	رودخانه، جریان زیست‌محیطی، ماهی، کیفیت آب، منابع آب، دریاچه ارومیه، اکولوژیکی، رسوب، شوری، مهاجرت، رودخانه کارون
T15	منابع طبیعی	طبیعت، تخریب، منابع طبیعی، تخریب محیط‌زیست، حفاظت، مدل تخریب، معماری سبز، محیط طبیعی، اسکان عشایر، پایداری، ارزیابی اثرات توسعه

موضوع	برچسب	عبارت‌های پرتکرار
T16	پوشش گیاهی	گرد و غبار، زمین شناسی، پوشش گیاهی، رسوبات، آلاینده، زمین شناسی زیست‌محیطی، رسوب، فرسایش خاک، سیلاب، بیابان‌زایی، گیاه، اهواز
T17	صنعت	توسعه، صنعت، عملکرد محیط‌زیست، آهن، رشد اقتصادی، توسعه مالی، فلزات سنگین، انرژی، ایزوترم، آزادسازی تجاری، نیترات، آلاینده
T18	فرهنگ	رفتار زیست‌محیطی، توسعه، آموزش زیست‌محیطی، سبک زندگی، توسعه پایدار، فرهنگ زیست‌محیطی، حفاظت، نگرش زیست‌محیطی، ارزش افزوده سرمایه فکری، آموزش محیط‌زیست
T19	منابع آب	منابع آب، آب زیرزمینی، آبخوان، کیفیت آب، آب آشامیدنی، تالاب، مدیریت منابع آب، آب سطحی، آب زیر زمینی، رودخانه
T20	آموزش	سواد زیست‌محیطی، ارزیابی ریسک، آگاهی زیست‌محیطی، توسعه، آموزش محیط‌زیست، ایمنی، گردشگری، اکولوژیکی، حفاظت، کشاورزی، ابزارهای بازاریابی سبز، آزمون انتخاب

همان‌طور که پیش‌تر بیان شد، تقاضاهای پژوهش بر مبنای جستجوی کاربران پایگاه اطلاعاتی شکل گرفته است. برای شناسایی موضوعات مرتبط با جستجوهای انجام شده، مدل تولید شده در مدل‌سازی موضوعی با پارامترهای تعیین شده، روی سبد واژگان استخراج شده برای هر جستجو اجرا شده است. حد آستانه تشریح (th_e) مقدار ۰/۳ در نظر گرفته شد به همین دلیل یک جستجو می‌تواند توسط بیش از یک موضوع (بیشینه توسط سه موضوع) تشریح شود.

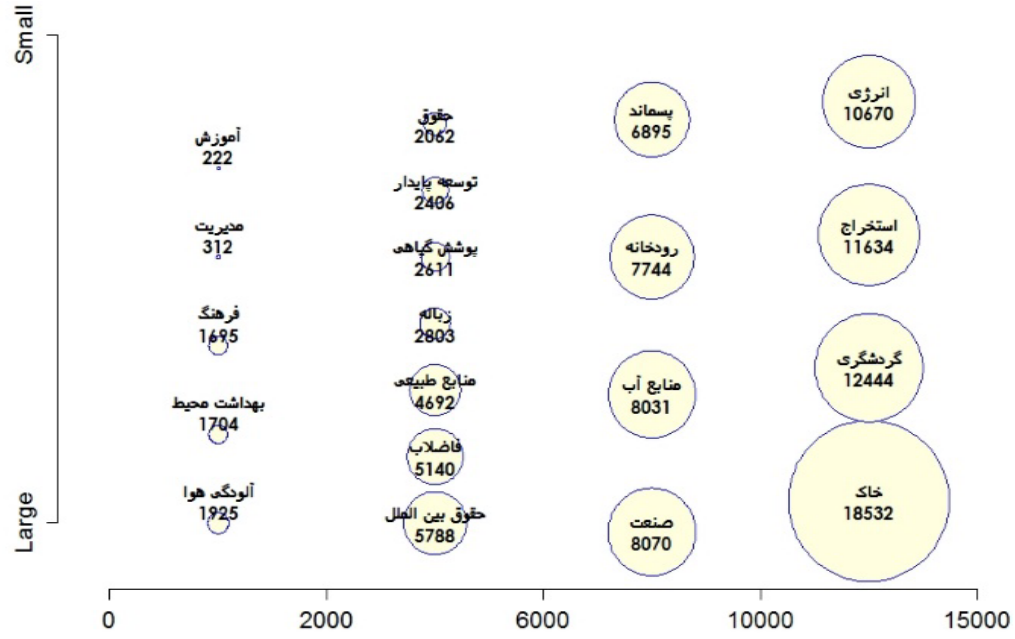
پس از اجرای مدل، موضوعات بیست‌گانه محیط‌زیست در بین جستجوهای کاربران شناسایی شد. در این بین موضوع «آموزش» با کمی بیش از ۱۰۰۰ جستجوی مرتبط، کوچک‌ترین موضوع مربوط به محیط‌زیست و موضوع «استخراج» با بیش از ۱۴۵۰۰۰ جستجوی مرتبط، بیشترین میزان تقاضا را به خود تخصیص داده است. شکل ۷ سهم هر یک از موضوعات مورد جستجو را نشان می‌دهد.



شکل ۷- سهم هر یک از موضوعات از جستجوهای انجام شده در حوزه محیط‌زیست

همان‌طور که در شکل ۷ آمده است، بیشترین میزان جستجوهای کاربران در حوزه محیط‌زیست مرتبط با موضوعات «استخراج»، «حقوق بین‌الملل»، «رودخانه»، «گردشگری» و «صنعت» است و کمترین میزان جستجو نیز مربوط به موضوعات «آموزش»، «مدیریت»، «فرهنگ»، «توسعه پایدار» و «آلودگی هوا» است. توجه به این نکته اهمیت دارد که بسیاری از جستجوها به‌طور همزمان توسط چند موضوع پوشش داده شده‌اند. سهم هر یک از موضوعات از تمامی مستنداتی که مرتبط با حوزه محیط‌زیست شناخته شده‌اند

در شکل ۸ مشخص شده است. در این شکل اندازه دایره‌ها متناسب با عدد نوشته شده در آن‌ها و نشان‌دهنده تعداد منابع هر موضوع است.



شکل ۸- سهم هر یک از موضوعات از منابع شناسایی شده حوزه محیط‌زیست

توجه به این نکته ضروری است که شکل ۸ مربوط به تمامی منابع شناخته‌شده مرتبط با این حوزه است و لزوماً توسط پژوهشگران رشته‌های محیط‌زیست انجام نشده است. به عبارتی در این شکل توزیع تمامی پژوهش‌هایی که مرتبط با این حوزه شناخته شده‌اند به تصویر کشیده شده است.

۴-۵- تحلیل فاصله بین عرضه و تقاضای پژوهش‌های حوزه محیط زیست

به‌منظور تحلیل فاصله بین عرضه و تقاضای موضوعات استخراج شده، لازم است تا میزان عرضه و تقاضای هر یک از موضوعات در مرحله اول نرمال‌سازی شود. برای این منظور از تقسیم تعداد منابع عرضه یا تقاضا شده هر موضوع به بزرگ‌ترین مقدار استفاده شد. درنتیجه تمامی مقادیر در مقیاس بین صفر تا یک قرار گرفت. سپس موضوعات بر مبنای موقعیت خود در دو محور عرضه و تقاضای پژوهش نگاشت شدند. با تقسیم محور افقی عرضه به بخش «عرضه بالا» و «عرضه پایین» و نیز با تقسیم محور عمودی به دو بخش «تقاضای بالا» و تقاضای «پایین» ماتریسی شامل چهار بخش به دست خواهد آمد. مقدار ۰/۵ به عنوان مرز بین عرضه و تقاضای بالا و پایین تعیین شده است. شکل ۹ این چهار بخش و موقعیت هر یک از موضوعات را نشان می‌دهد.

شکل ۹- وضعیت کنونی عرضه و تقاضای موضوعات حوزه محیط‌زیست

با توجه به شکل ۹، ناحیه سمت راست و بالای این نمودار دربردارنده موضوعاتی است که دارای عرضه بالای پژوهشی هستند و همچنین پژوهشگران زیادی نیز در این زمینه در حال جستجو هستند. از این‌رو موضوعات «استخراج» و «گردشگری» که در این ناحیه واقع شده‌اند به عنوان «موضوعات داغ» در نظر گرفته شده‌اند. موضوعات داغ آن دسته از موضوعات فرض شده‌اند که علاوه بر اینکه به میزان قابل‌توجهی تولید شده‌اند، به میزان زیادی نیز در جستجوهای کاربران موردتوجه قرار گرفته‌اند. به عبارتی پژوهش‌های این موضوعات همچنان در حال استفاده، بسط و گسترش است، در پژوهش پیشین در این حوزه، موضوع «مواد» که تشابه بسیار زیادی با موضوعی که در این پژوهش با نام «استخراج» از آن یاد شده است در گروه پژوهش‌های با میزان بالای جستجو و نرخ بالای پاسخ شناسایی شده است [۱۳]. ناحیه سمت راست و پایین شکل ۹ دربردارنده موضوعاتی است که اصطلاح «موضوعات خاموش[۱]» برای آن‌ها انتخاب شده است. این بخش شامل موضوعاتی است که دارای تعداد قابل‌توجهی منبع هستند اما در حال حاضر جستجوگران علاقه زیادی به استفاده از این منابع ندارند. به نظر می‌رسد برخی از پژوهش‌های قدیمی و نیز حوزه‌های پژوهشی که تا حدی اشباع شده باشند در این بخش قرار می‌گیرند. موضوع «خاک» فاصله بسیار اندکی با ناحیه قبلی دارد و با اندکی سهل‌گیری می‌تواند جزو موضوعات داغ دسته‌بندی شود اما موضوع «انرژی» با وجود اینکه فاصله زیادی با مرز ناحیه بعدی ندارد، می‌تواند در دسته موضوعات خاموش قرار گیرد، پژوهش قبلی موضوعات «خاک» و «جنگل» را در موضوعات دارای جستجوی پایین و نرخ پاسخ بالا شناسایی کرده است [۱۳].

ناحیه سمت چپ و پایین، ناحیه‌ای است که دارای تعداد اندک پژوهش انجام شده بوده و همچنین این موضوعات مورد استقبال جستجوگران نیز قرار نگرفته‌اند. این ناحیه که دقیقاً نقطه مقابل پژوهش‌های داغ قرار گرفته است با عنوان «موضوعات سرد[۲]» نام‌گذاری شده است. «مدیریت»، «آموزش»، «آلودگی هوا»، «فرهنگ»، «توسعه پایدار»، «منابع طبیعی»، «پوشش گیاهی»، «حقوق» و «منابع آب» در این ناحیه قرار گرفته‌اند. همچنین موضوعات «پسماند»، «فاضلاب» و «بهداشت محیط» نیز اگرچه در فاصله بسیار کمی از ناحیه چهارم قرار گرفته‌اند اما از نظر تقسیمات موردبحث در این بخش، جزو موضوعات سرد قابل‌طبقه‌بندی هستند. با توجه به تعداد اندک پژوهش‌های انجام‌گرفته در موضوعات این ناحیه و نیز عدم توجه کافی جستجوگران پژوهشی حوزه محیط به موضوعات این ناحیه، می‌توان پیش‌بینی نمود که در صورتی که سیاست‌گذاران پژوهشی برای این موضوعات چاره‌اندیشی لازم را نداشته باشند در آینده این موضوعات به‌تدریج حذف خواهند شد و اهمیت خود را به عنوان یک موضوع مستقل از دست خواهند داد. پژوهش پیشینی که در زمینه اولویت‌یابی موضوعات پژوهشی حوزه محیط‌زیست از طریق تحلیل لاگ جستجوی کاربران در ایران انجام شده است،

---

[۱] Silent Topics

[۲] Cold Topics

حوزه‌های «حقوق»، «مدیریت» و «اقتصاد» را به عنوان حوزه‌هایی که در جستجوهای کاربران دارای نرخ پاسخ پایین و نرخ جستجوی پایین است، شناخته شده است [۱۳]. ناحیه سمت چپ و بالا، آخرین بخش از این ماتریس است. این ناحیه دربردارنده موضوعاتی است که اگرچه موردتوجه جستجوگران قرار گرفته‌اند و تقاضای پژوهشی در آنها بالاست اما به اندازه کافی منابع پژوهشی برای آن‌ها وجود ندارد. به عبارتی به ازای تقاضای بالای پژوهشی در این موضوعات، کاربران با منابع پژوهشی کافی مواجه نمی‌شوند. ازاین‌رو نام این ناحیه «موضوعات دارای کاستی[1]» گذاشته شده است که دربردارنده کاستی‌های پژوهش در حوزه محیط‌زیست است. «رودخانه»، «صنعت»، «حقوق بین‌الملل» و «زباله» مهم‌ترین موضوعاتی است که دارای کاستی پژوهشی هستند و همان‌طور که پیش‌تر گفته شد، «پسماند»، «فاضلاب» و «بهداشت محیط» نیز به این ناحیه بسیار نزدیک بوده و می‌توان آن‌ها را نیز به عنوان کاستی‌های پژوهش در حوزه محیط‌زیست قلمداد کرد. پژوهش پیشین این حوزه، «آموزش» و «آلودگی» را در دسته پژوهش‌های با نرخ جستجوی بالا و نرخ پاسخ پایین شناسایی کرده است [۱۳] که البته این پژوهش مربوط به چند سال اخیر است و به نظر نرخ جستجو در این حوزه‌ها تغییر کرده است اما نرخ پاسخ تقریباً به همان شکل باقی‌مانده است.

۵- نتیجه‌گیری

در این پژوهش منابع پژوهشی حوزه محیط‌زیست موجود در پایگاه اطلاعاتی ایرانداک (گنج) و نیز جستجوهای انجام‌شده توسط کاربران، در یک فرایند رایانشی مورد تحلیل قرار گرفت، ازآنجاکه دسترسی به نمونه داده‌هایی که نمایانگر حوزه‌های مختلف پژوهشی باشند و روش‌های یادگیری ماشین بتواند از طریق کشف الگو از آن‌ها، به تشخیص و رده‌بندی متون پژوهشی بپردازد، بسیار مشکل و پرهزینه بوده و در برخی موارد غیرممکن است، رده‌بندی تک‌رده‌ای به‌عنوان یک راه‌حل ارائه شد و برای این منظور از ماشین بردار پشتیبان تک‌رده‌ای استفاده شد. پس از یادگیری ماشین، مجموعه داده‌های لاگ جستجو توسط الگوریتم‌های پردازش موازی مورد پیش‌پردازش قرار گرفت و در مرحله بعد با استفاده از رده‌بندی تک‌رده‌ای جستجوهای مرتبط با محیط‌زیست شناسایی شد. به روش مشابه سبد واژگان مربوط به پایان‌نامه‌ها تشکیل شد. سپس تحلیل کتاب‌شناختی جستجوهای انجام‌شده و تحلیل کتابشناختی پایان‌نامه‌های تولیدشده ارائه شد. در ادامه با استفاده از مدل‌سازی موضوعی و پس از تعیین پارامترهای بهینه آن، ۲۰ موضوع مختلف در حوزه محیط‌زیست شناسایی شد و منابع عرضه‌شده و تقاضا شده این حوزه در این موضوعات دسته‌بندی شد. دسته‌بندی به نحوی انجام شد که امکان قرارگیری یک رکورد در بیش از یک دسته وجود داشت. درنهایت با نگاشت موضوعات در دو بعد عرضه و تقاضا، موضوعات به چهار دسته کلی «موضوعات داغ» دارای عرضه و تقاضای بالا، «موضوعات سرد» دارای عرضه و تقاضای پایین، «موضوعات خاموش» دارای عرضه بالا و تقاضای پایین و «موضوعات دارای کاستی» که دارای عرضه پایین و تقاضای بالا بودند تقسیم شده و پیرامون هر یک از آن‌ها بحث شد. روند تغییرات تقاضاها در سه سال اخیر نسبت به سه سال قبل از آن به عنوان نرخ تغییرات تقاضا در نظر گرفته.

سیاست‌گذاران پژوهشی با مطالعه و رصد این تغییرات و نیز بررسی عرضه و تقاضای پژوهش در حوزه محیط‌زیست علاوه بر شناخت وضع موجود پژوهش‌های این حوزه و با استفاده از دیگر منابع اطلاعاتی خود می‌توانند در زمینه تقویت یا بهبود وضعیت پژوهش‌های موضوعات مختلف این حوزه سیاست‌گذاری نموده و با شناخت کافی از نیازهای جامعه پژوهشی، اولویت‌های پژوهشی این حوزه را تعیین نمایند. همچنین پیش‌بینی مسیر پژوهش‌های آتی این حوزه را برای آن‌ها ممکن می‌سازد تا با اتخاذ سیاست‌های مناسب امکان مدیریت آن فراهم گردد

مراجع

[1] فتاحی،ر. حکایت استفاده کاربردی از پژوهش‌ها در ایران، فصلنامه علوم و فناری اطلاعات، شماره ۴۶، ص ۷۷۷-۷۷۹، ۱۳۹۰.

[2] J.-E. Mai, D. O. Case, and L. M. Given, *Looking for information: A survey of research on information seeking, needs, and behavior*: Emerald Group Publishing, 2016.

[1] Gap Topics

[3] I. McBriar, C. Smith, G. Bain, P. Unsworth, S. Magraw, and J. L. Gordon, "Risk, gap and strength: key concepts in knowledge management," *Knowledge-Based Systems*, vol. 16, pp. 29-36, 2003.
[4] B. Junnarkar, "Leveraging collective intellect by building organizational capabilities," *Expert Systems with Applications*, vol. 13, pp. 29-40, 1997.
[5] W. Zhao, R. Wu, W. Dai, and Y. Dai, "Research paper recommendation based on the knowledge gap," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, 2015, pp. 373-380.
[6] K. A. Robinson, I. J. Saldanha, and N. A. Mckoy, "Development of a framework to identify research gaps from systematic reviews," *Journal of clinical epidemiology*, vol. 64, pp. 1325-1330, 2011.
[7] W. W. Hay Jr, T. N. Raju, R. D. Higgins, S. C. Kalhan, and S. U. Devaskar, "Knowledge gaps and research needs for understanding and treating neonatal hypoglycemia: workshop report from Eunice Kennedy Shriver National Institute of Child Health and Human Development," *The Journal of pediatrics*, vol. 155, 2009.
[8] N. Kräuchi, P. Brang, and W. Schönenberger, "Forests of mountainous regions: gaps in knowledge and research needs," *Forest Ecology and Management*, vol. 132, pp. 73-82, 2000.
[9] J. J. Manyà, "Pyrolysis for biochar purposes: a review to establish current knowledge gaps and research needs," *Environmental science & technology*, vol. 46, pp. 7939-7954, 2012.
[10] H. A. Tilson, "Developmental neurotoxicology of endocrine disruptors and pesticides: identification of information gaps and research needs," *Environmental health perspectives*, vol. 106, p. 807, 1998.
[11] P. S. Young, J. J. Cech, and L. C. Thompson, "Hydropower-related pulsed-flow impacts on stream fishes: a brief review, conceptual model, knowledge gaps, and research needs," *Reviews in Fish Biology and Fisheries*, vol. 21, pp. 713-731, 2011.
[12] D. Eerkes-Medrano, R. C. Thompson, and D. C. Aldridge, "Microplastics in freshwater systems: a review of the emerging threats, identification of knowledge gaps and prioritisation of research needs," *Water research*, vol. 75, pp. 63-82, 2015.
[13] M. Rabiei, S.-M. Hosseini-Motlagh, and A. Haeri, "Using text mining techniques for identifying research gaps and priorities: a case study of the environmental science in Iran," *Scientometrics*, vol. 110, pp. 815- 842, 2017.
[14] ایرانداک، پایگاه اطلاعاتی گنج به آدرس http://ganj.irandoc.ac.ir ، تاریخ دسترسی (۱۳۴۰۰/۲/۱۰)
[15] P. H. Gobster, "(Text) Mining the LANDscape: Themes and trends over 40 years of Landscape and Urban Planning," *Landscape and Urban Planning*, vol. 126, pp. 21-30, 2014.
[16] J. G. Cullen, "Educating Business Students About Sustainability: A Bibliometric Review of Current Trends and Research Needs," *Journal of Business Ethics*, pp. 1-11, 2015.
[17] L. Wei and Z. Yang, "Bibliometric analysis of global environmental assessment research in a 20-year period," *Environmental Impact Assessment Review*, vol. 50, pp. 158-166, 2015.
[18] خطیر،ا، محبی، آ و گنجه‌فر، س. ارائه راهکاری خودکار بر اساس متن‌کاوی برای شناخت و تحلیل روند تحقیقات حوزه‌های علمی، فصلنامه علمی پژوهشی مدیریت اطلاعات، شماره ۴، ۱۳۹۷.
[19] شرفی، ع و نورمحمدی، ح. تعیین پوشش موضوعی اولویت‌های پژوهشی علم اطلاعات و دانش‌شناسی، پژوهشنامه کتابداری و اطلاع رسانی، شماره ۴، ص ۱۶۹-۱۸۲، ۱۳۹۳.
[20] عصاره، ف، فرج‌پهلو، ع و سیامکی، ص. پژوهش های علم اطلاعات و دانش شناسی در ایران: تحلیل محتوا، مجله علم‌سنجی کاسپین، شماره ۲، ص ۷-۱۳، ۱۳۹۳.
[21] J. Azaranfar, "Content analysis of Library and Information Science theses and dissertations held with IRANDOC 1994-2004," *Iranian Journal of Information Processing and Management*, vol. 21, pp. 19-39, 2006.
[22] M. A. Zaki, "Content Analysis of Iranian Thesis in Tourism," *Iranian Journal of Information Processing and Management*, vol. 24, pp. 51-78, 2008.
[23] M. A. Zaki, "An Investigation on Research on Islamic Revolution in Iran All of the Post-Graduate Dissertations of State Universities Between 1979-2004," *Iranian Journal of Information Processing and Management*, vol. 26, pp. 121-143, 2010.
[24] F. Sebastiani, "Machine learning in automated text categorization," *ACM computing surveys (CSUR)*, vol. 34, pp. 1-47, 2002.
[25] G. Erkan and D. R. Radev, "Lexrank: Graph-based lexical centrality as salience in text summarization," *Journal of artificial intelligence research*, vol. 22, pp. 479-457, 2004.
[26] M. John W and B. Petko, "Introduction-Topic models: What they are and why they matter," *Poetics*, vol. 41, pp. 545-569, 2013.
[27] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, pp. 993-1022, 2003.