

بسم الله الرحمن الرحيم



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

طرح پژوهشی

## تحلیل عرضه و تقاضای پایان نامه ها و رساله های حوزه انرژی در پایگاه گنج

مجری

محمد ربیعی

همکار

سمیه فتاحی

مهر ۱۴۰۰

حقوق: پژوهشگاه علوم و فناوری اطلاعات ایران

طرح پژوهشی تحلیل عرضه و تقاضای پایان‌نامه‌ها و رساله‌های حوزه انرژی در پایگاه گنج پیرو قرارداد شماره ۳۳/۳۲۲ مورخ ۹۹/۱/۱۶ میان پژوهشگاه علوم و فناوری اطلاعات ایران (کارفرما) و آقای دکتر محمد ربیعی اجرا شده است. گزارش حاضر یک جلد از مستندات این طرح است.

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

صحت مندرجات گزارش بر عهده مجری طرح پژوهشی است.

تاییدیه:

۳۳/۵۹۵۳

۱۴۰۰/۰۸/۰۱

ندارد

پایان خد/

ایران‌دک در نیم‌قرن دوم کار ثروت همپنان پوران و کوشا همراه هر پژوهش، پژوهشگر، و سیاست‌گذار

۱۱۳۱۲۷ - ۱۱۳۱۲۷

گواهی

انجام طرح پژوهشی:

◇ عنوان: تحلیل عرضه و تقاضای پایان‌نامه‌ها و رساله‌های حوزه انرژی در پایگاه گنج

◇ مجری: دکتر محمد ربیعی

◇ همکار: دکتر سمیه فتاحی

◇ شماره قرارداد: ۳۳/۳۲۲

◇ تاریخ قرارداد: ۱۳۹۹/۱/۱۶

◇ تاریخ شروع: ۱۳۹۹/۱/۱۶

◇ تاریخ پایان: ۱۴۰۰/۷/۲۰

◇ ناظر: دکتر نصراله پاک‌نیت

در پژوهشگاه علوم و فناوری اطلاعات ایران گواهی می‌شود. گزارش پایانی این طرح، بر پایه داوری در نشست ۴۲۹

(تاریخ ۱۴۰۰/۷/۲۷) شورای پژوهشی پژوهشگاه به تصویب رسیده است. لازم می‌دانم مراتب قدردانی خود را به

مجری، همکار، و ناظر این طرح تقدیم دارم.

با آرزوی بهروزی

پرویز شهریاری

معاون پژوهش و آموزش

## تحلیل عرضه و تقاضای پایان‌نامه‌ها و رساله‌های حوزه انرژی در پایگاه گنج

مدیر طرح پژوهشی: دکتر محمد ربیعی، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

نشانی: تهران، خیابان انقلاب، چهارراه فلسطین، شماره ۱۰۹۰، طبقه ۴، اتاق ۴۰۷

صندوق پستی: ۱۳۱۵۷۷۳۳۱۴      تلفن: ۰۲۱-۶۶۴۹۴۹۸۰ (داخلی ۲۴۱)

وبگاه: <https://irandoc.ac.ir/u/m-rabiei>

رایانامه: [m.rabeie@irandoc.ac.ir](mailto:m.rabeie@irandoc.ac.ir)

همکار طرح پژوهشی: دکتر سمیه فتاحی، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

## چکیده

نیاز روزافزون به استفاده از انرژی و گسترش علم و فناوری و تعامل و ارتباط بین پژوهشگران و صنعت موجب شده است تا از یک سو شاهد تعدد و گوناگونی پژوهش‌های صورت پذیرفته در حوزه‌ی انرژی باشیم و از سوی دیگر با تنوع نیازها و تقاضاهای پژوهشی این حوزه مواجه شویم. به منظور مدیریت نظام پژوهش این حوزه، داشتن شناخت کافی نسبت به تقاضاها و نیازهای پژوهشی و توانمندی‌ها و منابع پژوهشی موجود در زمینه انرژی ضروری است. این پژوهش درصدد است راهکار مناسبی را مبتنی بر تحلیل محتوای پژوهش‌های موجود و تقاضاهای جامعه پژوهش در حوزه انرژی و از طریق کاربرست تکنیک‌های متن‌کاوی و پردازش زبان طبیعی و تحلیل رفتار جویشگری کاربران ارائه نماید که از طریق آن کاستی‌های پژوهشی به صورت مستقل از نظر خبرگان و در فرایندی رایانشی شناسایی شود. برای این منظور منابع پژوهشی موجود در پایگاه ایرانداک و نیز جستجوهای صورت پذیرفته توسط کاربران این پایگاه به عنوان داده‌های پژوهش مورداستفاده قرار گرفت. فرایند پژوهش موجب شناسایی ۲۵,۹۵۳ رکورد از منابع اطلاعاتی گنج به عنوان منابع عرضه‌شده و نیز ۴۴۳,۸۹۰ از جستجوهای انجام شده به عنوان منابع تقاضاشده در حوزه انرژی شد و تحلیل کتابشناختی آن نشان می‌دهد که از اواخر دهه ۸۰ عرضه و تقاضای این منابع رشد زیادی داشته‌اند. تحلیل فاصله بین عرضه و تقاضا در موضوعاتی این حوزه، موقعیت هر یک از موضوعات شناسایی شده را در ماتریس عرضه و تقاضا نشان داده است. نتایج حاکی از آن است که موضوعات «محیط‌زیست» و «مدیریت انرژی» از موضوعات داغ حوزه انرژی، موضوعات «نفت» و «گاز» از موضوعات سرد حوزه انرژی، «تکنولوژی»، «اقتصاد انرژی»، «انرژی‌های تجدیدپذیر» و «برق» از موضوعات خاموش حوزه انرژی هستند و موضوع «انرژی خورشیدی» به عنوان موضوع دارای کاستی پژوهش شناسایی شد. پژوهش ارائه‌شده می‌تواند سیاست‌گذاران و مدیران پژوهشی حوزه انرژی را در اتخاذ تصمیمات کلان پژوهشی و اولویت‌بندی پژوهشی یاری رساند. همچنین به دلیل رویکرد میان‌رشته‌ای و روش تلفیقی مورداستفاده، پژوهشگران و علاقه‌مندان حوزه‌های مختلف فناوری اطلاعات، هوش مصنوعی، کتابداری و اطلاع‌رسانی، سیاست‌گذاری علم و فناوری و ... می‌توانند از نتایج آن بهره‌مند شوند.

**کلیدواژه‌ها:** عرضه و تقاضای پژوهش، انرژی، مدل‌سازی موضوعی، متن‌کاوی، پایگاه گنج،

ایرانداک

## قدردانی

مجری این طرح لازم می‌داند مراتب قدردانی و سپاس خود را از تمامی همکارانی که در بخش‌های مختلف این پژوهش همکاری لازم را داشتند به عمل آورد. به صورت ویژه از همکاری جناب آقایان منصور شیدایی و مهندس علی اصغر حجت‌پناه و همکاران ایشان که در تامین منابع مورد نیاز این پژوهش همکاری کردند سپاسگزارم، همچنین از همراهی دوست و همکار گرانقدر جناب آقای دکتر علی نعیمی صدیق نهایت تشکر و قدردانی را دارم.

در پایان از راهنمایی‌ها، زحمات‌ها و دیدگاه‌های ارزنده دکتر نصراله پاک‌نیت که نظارت این طرح پژوهشی را پذیرفتند و با دقت نظر و توجه بسیار آن را به انجام رساندند، سپاسگزارم.

## فهرست مطالب

شماره صفحه

عنوان

|    |                                                    |
|----|----------------------------------------------------|
| ۱  | ۱. مقدمه و پیشینه پژوهش                            |
| ۲  | ۱_۱. مقدمه                                         |
| ۲  | ۱_۲. انرژی                                         |
| ۳  | ۱_۳. علم سنجی                                      |
| ۵  | ۱_۴. حوزه‌های مرتبط با علم سنجی                    |
| ۷  | ۱_۵. مفهوم کاستی پژوهش                             |
| ۷  | ۱_۶. مطالعات علم سنجی                              |
| ۸  | ۱_۷. پایگاه اطلاعاتی گنج                           |
| ۸  | ۱_۸. ضرورت و لزوم انجام پژوهش‌های تقاضا محور       |
| ۱۰ | ۱_۹. اهداف طرح                                     |
| ۱۱ | ۱_۱۰. پیوند با اساس نامه و برنامه راهبردی پژوهشگاه |
| ۱۲ | ۱_۱۱. تعریف واژه‌ها و مفاهیم                       |
| ۱۲ | ۱_۱۲. جمع‌بندی و نتیجه‌گیری                        |
| ۱۴ | ۲. روش‌شناسی پژوهش                                 |
| ۱۵ | ۲_۱. مقدمه                                         |
| ۱۶ | ۲_۲. داده کاوی                                     |
| ۱۷ | ۲_۳. متن کاوی                                      |
| ۲۰ | ۲_۴. کاربرد داده کاوی و متن کاوی در مدیریت پژوهش   |
| ۲۱ | ۲_۵. کاربرد متن کاوی در مرور ادبیات پژوهشی         |

|    |                                                                      |
|----|----------------------------------------------------------------------|
| ۲۳ | ۲_۶. رویکردهای نیازسنجی و اولویت‌یابی پژوهش                          |
| ۲۴ | ۲_۶.۱. رویکرد آسیب‌شناختی                                            |
| ۲۴ | ۲_۶.۲. رویکرد موضوعی و محتوایی                                       |
| ۲۴ | ۲_۶.۳. رویکرد تقاضامحور                                              |
| ۲۴ | ۲_۶.۴. رویکرد عرضه‌محور                                              |
| ۲۵ | ۲_۶.۵. رویکرد نظام‌مند                                               |
| ۲۵ | ۲_۶.۶. رویکرد اداره‌گری / مدیریتی                                    |
| ۲۷ | ۲_۷. روش مدل‌سازی موضوعی                                             |
| ۲۸ | ۲_۷.۱. الگوریتم LDA                                                  |
| ۲۹ | ۲_۸. روش پژوهش                                                       |
| ۳۰ | ۲_۹. فرایند انجام پژوهش                                              |
| ۳۲ | ۲_۱۰. زبان برنامه‌نویسی                                              |
| ۳۴ | ۲_۱۱. سخت‌افزار مورد استفاده                                         |
| ۳۴ | ۲_۱۲. قلمرو پژوهش                                                    |
| ۳۵ | ۲_۱۳. مفروضات پژوهش                                                  |
| ۳۶ | ۲_۱۴. جمع‌بندی و نتیجه‌گیری                                          |
| ۳۷ | <b>۳. شناسایی و تحلیل متون عرضه و تقاضا شده در حوزه انرژی</b>        |
| ۳۸ | ۳_۱. مقدمه                                                           |
| ۳۹ | ۳_۲. فرایند استخراج پارساهای مرتبط با حوزه انرژی                     |
| ۴۰ | ۳_۲.۱. گام ۱ ( دریافت داده‌های نمونه از گنج )                        |
| ۴۱ | ۳_۲.۲. گام ۲ (تولید مجموعه داده اولیه حوزه انرژی)                    |
| ۴۱ | ۳_۲.۳. گام ۳ (شناسایی کلمات کلیدی پرتکرار حوزه انرژی)                |
| ۴۲ | ۳_۲.۴. گام ۴ (جستجوی «انرژی» در اصطلاح‌نامه‌های تخصصی)               |
| ۴۲ | ۳_۲.۵. گام ۵ (شناسایی واژگان اخص مربوط به «انرژی» در اصطلاح‌نامه‌ها) |
| ۴۳ | ۳_۲.۶. گام ۶ (دسته‌بندی و پالایش کلمات مستخرج از اصطلاح‌نامه)        |
| ۴۳ | ۳_۲.۷. گام ۷ (انتخاب کلمات نماینده برای جستجوی حوزه انرژی)           |
| ۴۵ | ۳_۲.۸. گام ۸ ( جستجوی کلمات نماینده و استخراج داده‌ها)               |

|    |                                                                       |
|----|-----------------------------------------------------------------------|
| ۴۶ | ۱_۲_۳. فرایند استخراج تقاضاهای مرتبط با حوزه انرژی                    |
| ۴۷ | ۲_۲_۳. تفکیک جستجوهای کاربران واقعی از ربات‌ها و موتورهای جستجو       |
| ۴۸ | ۳_۲_۳. پردازش متن جستجوها                                             |
| ۵۰ | ۴_۲_۳. شناسایی جستجوهای مرتبط با حوزه انرژی                           |
| ۵۰ | ۳_۳. تحلیل کتابشناختی پارساهای عرضه شده در حوزه انرژی                 |
| ۵۰ | ۱_۳_۳. تحلیل مقطع تحصیلی عرضه‌کنندگان منابع حوزه انرژی                |
| ۵۱ | ۲_۳_۳. تحلیل زمانی منابع عرضه‌شده در حوزه انرژی                       |
| ۵۲ | ۳_۳_۳. تحلیل گروه‌های آموزشی مرتبط با حوزه انرژی                      |
| ۵۳ | ۴_۳_۳. تحلیل دانشگاه‌ها و مراکز پژوهشی عرضه‌کننده پارساهای حوزه انرژی |
| ۵۴ | ۵_۳_۳. تحلیل کلمات کلیدی پرکاربرد در منابع حوزه انرژی                 |
| ۵۶ | ۴_۳. تحلیل کتابشناختی جستجوهای انجام‌شده در حوزه انرژی                |
| ۵۶ | ۱_۴_۳. تحلیل زمانی تقاضاهای حوزه انرژی                                |
| ۵۷ | ۲_۴_۳. بررسی تعداد پاسخ‌های دریافتی کاربر                             |
| ۵۸ | ۳_۴_۳. تحلیل کلمات کلیدی پرکاربرد در جستجوهای حوزه انرژی              |
| ۶۰ | ۵_۳. جمع‌بندی و نتیجه‌گیری                                            |
| ۶۲ | <b>۴. تحلیل موضوعی عرضه و تقاضای پژوهش‌های حوزه انرژی</b>             |
| ۶۳ | ۱_۴. مقدمه                                                            |
| ۶۴ | ۲_۴. مدل‌سازی موضوعی منابع عرضه‌شده حوزه انرژی                        |
| ۶۴ | ۱_۲_۴. تعیین پارامترهای LDA                                           |
| ۶۶ | ۲_۲_۴. تعیین سطح آستانه تشریح                                         |
| ۶۸ | ۳_۴. برچسب‌گذاری موضوعات استخراج شده                                  |
| ۷۰ | ۴_۴. فراوانی منابع عرضه‌شده در موضوعات حوزه انرژی                     |
| ۷۱ | ۵_۴. روند عرضه منابع حوزه انرژی                                       |
| ۷۱ | ۱_۵_۴. روند عرضه منابع موضوع «اقتصاد انرژی»                           |
| ۷۱ | ۲_۵_۴. روند عرضه منابع موضوع «محیط‌زیست»                              |
| ۷۲ | ۳_۵_۴. روند ایند عرضه منابع موضوع «تکنولوژی»                          |
| ۷۲ | ۴_۵_۴. روند عرضه منابع موضوع «انتقال انرژی»                           |

|    |                                                                       |
|----|-----------------------------------------------------------------------|
| ۷۳ | ۴_۵_۵. روند عرضه منابع موضوع «انرژی خورشیدی»                          |
| ۷۳ | ۴_۵_۶. روند عرضه منابع موضوع «انرژی‌های تجدیدپذیر»                    |
| ۷۴ | ۴_۵_۷. روند عرضه منابع موضوع «نفت»                                    |
| ۷۵ | ۴_۵_۸. روند عرضه منابع موضوع «مدیریت انرژی»                           |
| ۷۵ | ۴_۵_۹. روند عرضه منابع موضوع «گاز»                                    |
| ۷۶ | ۴_۵_۱۰. روند عرضه منابع موضوع «برق»                                   |
| ۷۶ | ۴_۱. مدل‌سازی موضوعی منابع تقاضاشده حوزه انرژی                        |
| ۷۷ | ۴_۲. فراوانی منابع تقاضاشده در موضوعات حوزه انرژی                     |
| ۷۸ | ۴_۳. روند تقاضای موضوعات مختلف حوزه انرژی در ماه‌های سال              |
| ۷۹ | ۴_۴. تحلیل فاصله بین عرضه و تقاضای پژوهش                              |
| ۸۲ | ۴_۵. جمع‌بندی و نتیجه‌گیری                                            |
| ۸۴ | <b>۵. نتیجه‌گیری و پیشنهاد پژوهش‌های آتی</b>                          |
| ۸۵ | ۵_۱. مقدمه                                                            |
| ۸۶ | ۵_۲. فرایند و گام‌های طی شده                                          |
| ۸۶ | ۵_۱. پاسخ به پرسش‌های پژوهش                                           |
| ۸۷ | ۵_۱_۱. سوال اول پژوهش                                                 |
| ۸۸ | ۵_۱_۲. سوال دوم پژوهش                                                 |
| ۸۹ | ۵_۱_۳. سوال سوم پژوهش                                                 |
| ۹۰ | ۵_۲. پیش‌بینی عرضه و تقاضای موضوعات حوزه انرژی در سال‌های آتی         |
| ۹۱ | ۵_۳. محدودیت‌های پژوهش                                                |
| ۹۲ | ۵_۴. پیشنهاد پژوهش‌های آتی                                            |
| ۹۲ | ۵_۴_۱. تحلیل توزیع جغرافیایی پژوهش‌ها و تعیین کاستی‌های مکان‌محور     |
| ۹۳ | ۵_۴_۲. شناسایی نیازهای اطلاعاتی پنهان و بیان نشده در تقاضاهای کاربران |
| ۹۴ | ۵_۴_۳. تحلیل چرخه عمر موضوعات و مفاهیم همجوشی اطلاعات                 |
| ۹۵ | ۵_۵. جمع‌بندی و نتیجه‌گیری                                            |
| ۹۶ | <b>۶. فهرست منابع</b>                                                 |
| ۹۶ | ۶_۱. منابع فارسی                                                      |

۹۸

۶\_۲. منابع انگلیسی

۱۰۱

۶\_۳. چکیده انگلیسی

## فهرست شکل‌ها

| عنوان                                                                     | شماره صفحه |
|---------------------------------------------------------------------------|------------|
| شکل ۱-۱- ارتباط بین علم‌سنجی و دیگر شاخه‌های مرتبط با سنجش اطلاعات        | ۶          |
| شکل ۱-۲- فرایند انجام پژوهش                                               | ۳۱         |
| شکل ۱-۳- گام‌های استخراج پارساهای مرتبط با حوزه انرژی                     | ۴۰         |
| شکل ۲-۳- گام‌های استخراج جستجوهای مرتبط با حوزه انرژی                     | ۴۷         |
| شکل ۳-۳- تحلیل مقطع تحصیلی منابع عرضه شده حوزه انرژی                      | ۵۱         |
| شکل ۴-۳- روند زمانی پارساهای عرضه‌شده در حوزه انرژی                       | ۵۱         |
| شکل ۵-۳- سهم هر یک از گروه‌های آموزشی در تامین منابع عرضه شده             | ۵۳         |
| شکل ۶-۳- فهرست ۱۵ دانشگاه با بیشترین تولید پارسا در حوزه انرژی            | ۵۴         |
| شکل ۷-۳- ابر کلمات پرکاربرد با استفاده از کلیدواژه‌های پارساهای عرضه‌شده  | ۵۶         |
| شکل ۸-۳- متوسط تعداد جستجوها در ماه‌های مختلف سال                         | ۵۷         |
| شکل ۹-۳- درصد جستجوهای دارای بیشترین تعداد پاسخ                           | ۵۸         |
| شکل ۱۰-۳- ابر کلمات پرکاربرد با استفاده از کلیدواژه‌های پارساهای عرضه‌شده | ۶۰         |
| شکل ۱-۴- پردازش موازی برای شناسایی برای تعیین میزان سرگشتی مدل            | ۶۵         |
| شکل ۲-۴- میزان سرگشتگی برای تعداد مختلف موضوع                             | ۶۶         |
| شکل ۳-۴- پوشش و کیفیت تشریح در حد آستانه مختلف                            | ۶۸         |

- ۷۰ شکل ۴-۴- فراوانی منابع در موضوعات عرضه شده در حوزه انرژی
- ۷۱ شکل ۴-۵- روند عرضه منابع موضوع «اقتصاد انرژی»
- ۷۲ شکل ۴-۶- روند عرضه منابع موضوع «محیط زیست»
- ۷۲ شکل ۴-۷- روند عرضه منابع موضوع «تکنولوژی»
- ۷۳ شکل ۴-۸- روند عرضه منابع موضوع «انتقال انرژی»
- ۷۳ شکل ۴-۹- روند عرضه منابع موضوع «انرژی خورشیدی»
- ۷۴ شکل ۴-۱۰- روند عرضه منابع موضوع «انرژی های تجدیدپذیر»
- ۷۴ شکل ۴-۱۱- روند عرضه منابع موضوع «نفت»
- ۷۵ شکل ۴-۱۲- روند عرضه منابع موضوع «مدیریت انرژی»
- ۷۵ شکل ۴-۱۳- روند عرضه منابع موضوع «گاز»
- ۷۶ شکل ۴-۱۴- روند عرضه منابع موضوع «برق»
- ۷۸ شکل ۴-۱۵- فراوانی منابع در موضوعات تقاضا شده در حوزه انرژی
- ۷۹ شکل ۴-۱۶- روند تقاضای موضوعات حوزه انرژی به تفکیک ماه های سال
- ۸۰ شکل ۴-۱۷- وضعیت عرضه و تقاضای موضوعات حوزه انرژی
- ۹۱ شکل ۵-۱- پیش بینی عرضه و تقاضای پارساهای حوزه انرژی

## فهرست جدول‌ها

| شماره صفحه | عنوان                                                                     |
|------------|---------------------------------------------------------------------------|
| ۲۶         | جدول ۱-۲- رویکردهای مختلف نیازسنجی و اولویت‌یابی پژوهش                    |
| ۳۴         | جدول ۲-۲- مشخصات سرور مورد استفاده در این پژوهش                           |
| ۴۱         | جدول ۱-۳- کلمات پر کاربرد در داده‌های اولیه حوزه انرژی                    |
| ۴۳         | جدول ۲-۳- کلمات مرتبط و نماینده کلمات مستخرج از اصطلاح‌نامه‌های فنی       |
| ۴۵         | جدول ۳-۳- کلمات پرتکرار و نمایندگان جستجو در مجموعه داده اولیه حوزه انرژی |
| ۴۶         | جدول ۴-۳- نمونه‌ای از پارساهای شناسایی شده مرتبط با حوزه انرژی            |
| ۴۸         | جدول ۵-۳- ویژگی‌های موجود در رکوردهای لاگ                                 |
| ۴۸         | جدول ۶-۳- نمونه‌ای از محتوای ذخیره شده در فایل Log                        |
| ۴۹         | جدول ۷-۳- عبارت‌های کنترلی موجود در متن ذخیره شده در فایل Info لاگ        |
| ۴۹         | جدول ۸-۳- نمونه‌ای از محتوای نهایی (پیش‌پردازش شده) در فایل Log           |
| ۵۵         | جدول ۹-۳- فهرست ۳۰ کلمه کلیدی پرتکرار در منابع عرضه شده حوزه انرژی        |
| ۵۹         | جدول ۱۰-۳- فهرست ۳۰ کلمه کلیدی پرتکرار در منابع جستجو شده حوزه انرژی      |
| ۶۹         | جدول ۱-۴- موضوعات استخراج شده حوزه انرژی و عبارت‌های پرتکرار آن‌ها        |

# فصل اول

## مقدمه و پیشینه پژوهش

## ۱-۱. مقدمه

در سال‌های گذشته، شاهد افزایش چشم‌گیر پژوهش‌هایی هستیم که در حوزه‌های گوناگون علوم انجام شده است، اما رشد کمی پژوهش‌ها، لزوماً با افزایش کیفیت آن‌ها همراه نیست و در بسیاری از موارد انجام این پژوهش‌ها منجر به رفع مشکلات جامعه نشده و کمتر دیده شده است که نتایج و دستاوردهای پژوهشی برای بهبود اوضاع اقتصادی، اجتماعی، فرهنگی و سیاسی جامعه به کار گرفته شود. در این میان شاید برخی از رشته‌های مهندسی و پژوهش‌های کاربردی این بخش وضعیت بهتری را نسبت به دیگر گروه‌های آموزشی داشته باشند (نژاد، حیدری، و پرداز، ۱۳۹۷).

فاصله یا شکاف میان پژوهش‌های دانشگاهی، به‌ویژه پایان‌نامه‌ها و رساله‌ها با نیازهای جامعه یکی از مشکلات اساسی آموزش عالی ایران است. گسترش کمی دانشگاه‌ها و افزایش تعداد دانشجویان تحصیلات تکمیلی موجب آن شده تا از یک سو شاهد انتقادات و واکنش اندیشمندان و متخصصان آموزش و پژوهش باشیم و از سوی دیگر صاحبان صنایع و مدیران اقتصادی از توانمندی اندک دانش‌آموختگان در حل مشکلات و چالش‌های واقعی کشور گل‌ه‌مند باشند (رحمت‌الله، فتاحی، ۱۳۹۰). همچنین در مقوله پژوهش دوباره‌کاری‌های بسیاری را شاهد هستیم که دلیل اصلی آن، ناتوانی استادان و دانشجویان در پیدا کردن مسئله واقعی پژوهش در حوزه‌های گوناگون در جامعه است و در برخی از موارد، جامعه دانشگاهی به جای «مسئله‌یابی» به دنبال «مسئله‌سازی» رفته است. از دیگر سو اولویت‌های پژوهشی که نهادها و سازمان‌ها تعیین نموده و به شکل‌های مختلف به آگاهی پژوهشگران می‌رسانند، جدای از شکل و روش‌های ناکارآمد اطلاع‌رسانی، کمتر به شکل راهبردی، واقع‌بینانه و کاربردی مورد شناسایی قرار گرفته‌اند (رحمت‌الله، فتاحی، ۱۳۹۰).

## ۲-۱. انرژی

انرژی عامل حیاتی صنعت و اقتصاد است که امروزه دیگر ابعاد زندگی مانند فرهنگ، سیاست، آموزش و ... را نیز تحت سیطره خود دارد. ارزشمندی بالای این عامل از یک طرف و مصرف روزافزون و تقاضای بالای بازار برای استفاده از انرژی و نیز کمبود منابع در دسترس انرژی باعث شده است تا به کارگیری آخرین دانش روز در بخش‌های مختلف انرژی گریزناپذیر باشد.

ایران یکی از منابع اصلی انرژی در جهان است. منابع سرشار نفت، گاز و سوخت‌های فسیلی، شرایط مناسب برای تولید انرژی‌های نو مانند انرژی خورشیدی همواره نام ایران را در حوزه انرژی بر سر زبان آورده است. از سوی دیگر ایران یکی از کشورهایی است که با مصرف بسیار زیاد انرژی و صرف هزینه بالای یارانه انرژی مصرفی (۱۰ درصد یارانه سوخت‌های فسیلی دنیا) از این منظر نیز مورد توجه است. از اینرو، سالانه پژوهش‌های زیادی در زمینه‌های مختلف انرژی انجام می‌شود، این پژوهش‌ها در برخی موارد نیازمند استفاده از ابزار، آزمایشگاه‌ها، نیروی انسانی، منابع مالی و منابع طبیعی است و زمان زیادی را نیز صرف خود می‌کنند. به همین دلیل لزوم مدیریت نظام‌مند پژوهش‌های حوزه انرژی بر کسی پوشیده نیست. یکی از مهم‌ترین راهبردهای مدیریت پژوهش، مدیریت عرضه و تقاضای پژوهش است که مانع از شکل‌گیری پژوهش‌های موازی، طولانی‌شدن زمان پاسخ به نیازهای اطلاعاتی و پژوهشی و جلوگیری از هدررفت منابع مختلف پژوهشی می‌شود.

### ۱-۳. علم‌سنجی<sup>۱</sup>

قدمت تولید علم را می‌توان از زمانی متصور بود که بشر برای بقا و بهبود کیفیت زندگی خود در تلاش برای کشف واقعیت‌های جهان برآمد. ثبت و مستندسازی علم نیز پیشینه‌ای بسیار دارد که در عمر خود شاهد تحولات بنیادینی بوده است. اختراع خط، اختراع کاغذ، اختراع ماشین چاپ و ... زمینه لازم را برای تولید انبوه مستندات علمی در سطح جهان فراهم ساخت. اختراع کامپیوتر و ظهور اینترنت انقلابی را در نشر و عمومی‌سازی علم پدید آورد و شیوه‌های نوینی از ارتباطات علمی پایه‌گذاری شد. حجم زیاد از اطلاعات علمی سبب شد تا سیاست‌گذاران و متخصصین اطلاعات، شیوه‌ها و راهکارهای جدیدی را برای سنجش، رصد، مدیریت و برنامه‌ریزی علم و فناوری به کار گیرند. تحلیل استنادی، تکنیک‌های آماری، روش‌های مبتنی بر الگوریتم‌های پیچیده رایانشی، ابزارها و تکنیک‌های تحلیل داده و متن و ... تنها بخشی از این شیوه‌های نوین هستند که گاه به‌تنهایی و گاه به‌صورت ترکیبی موجب شده‌اند تا این حجم از اطلاعات و مستندات قابل رصد و ارزیابی باشند (اسدی، ۱۳۹۳).

---

<sup>۱</sup> Scientometrics

تحلیل علم از طریق روش‌های کمی، هدف اصلی مطالعه علم‌سنجی است. این دانش برای اینکه درک جامعی از فرایند پژوهش علمی به‌عنوان یک فعالیت اجتماعی در اختیار داشته باشد، جنبه‌های کمی تولید، اشاعه و کاربرد اطلاعات علمی را مد نظر قرار داده است (Braun, Glänzel, and Schubert, 1985).

صرف‌نظر از تکنیک‌های مورد استفاده و روش انجام کار، تعاریف گوناگونی از علم‌سنجی ارائه شده است. ویلسون<sup>۱</sup> علم‌سنجی را روشی برای مطالعه جنبه‌های کمی علوم، ارتباطات و سیاست‌های علمی می‌داند (Wilson, 1999). درحالی‌که لیدسدورف<sup>۲</sup> علم‌سنجی را دانش به‌کارگیری ابزارهای کمی برای مطالعه ارتباطات علمی در نظر می‌گیرد (Leydesdorff, 2001). تعاریف عام‌تری که از علم‌سنجی ارائه شده است تمام الگوها و جنبه‌های کمی مرتبط با تولید و اشاعه علم و دانش مبتنی بر فناوری را شامل می‌شود (De Bellis, 2009).

علم‌سنجی از طریق گردآوری و تحلیل داده‌های کمی مرتبط با تولید، توزیع، جستجو و استفاده از مستندات علمی، به توصیف علم و پژوهش می‌پردازد. برای دستیابی به این هدف، علم‌سنجی از متون منتشرشده علمی به‌عنوان عناصر پایه نشر علم بهره می‌گیرد. علم‌سنجی امکان شناسایی و آشکار کردن جریان اطلاعات در حوزه‌های زمانی، مکانی و موضوعی را فراهم می‌کند (اسدی، ۱۳۹۳).

نظریه‌پردازان این حوزه بر این باورند که علم‌سنجی سنتی به بررسی چهار مؤلفه اصلی می‌پردازد که عبارت‌اند از: پدیدآورندگان<sup>۳</sup>، انتشارات علمی، استنادها و ارجاعات. علم‌سنجی از طریق شاخص‌هایی به بررسی این مؤلفه‌ها می‌پردازد، این شاخص‌ها عمدتاً شاخص‌هایی کمی هستند و کمتر به بررسی محتوا و کیفیت آن می‌پردازند، به‌عنوان مثال شاخص‌هایی مانند ضریب تأثیر<sup>۴</sup>، شاخص فوریت<sup>۵</sup>، شاخص کهنگی و نیمه‌عمر<sup>۶</sup> متون از این جمله‌اند. برای رفع این محدودیت، به‌مرور شاخص‌هایی مانند شاخص هیرش یا شاخص اچ<sup>۷</sup> ارائه شده است که علاوه بر تعداد کمی مستندات تولیدشده توسط یک پژوهشگر یا یک مجله به تعداد منابع علمی که به آن پژوهشگر یا مجله استناد داده‌اند نیز دلالت دارد. هرچند که

---

<sup>۱</sup> Wilson, Concepción S

<sup>۲</sup> Leydesdorff, Loet

<sup>۳</sup> Author

<sup>۴</sup> Impact Factor

<sup>۵</sup> Immediacy Index

<sup>۶</sup> Half Life

<sup>۷</sup> H index

معایی مانند اینکه با طول عمر یک مقاله یا یک پژوهشگر هرگز از این شاخص کاسته نخواهد شد و این شاخص همواره رو به رشد است همچنان پابرجاست (تیرگر، یمین‌فیروز، و دهقان، ۱۳۹۳). باوجود قدمت طولانی علم، دانش علم‌سنجی قدمت چندانی ندارد. به‌طور رسمی از نیمه دوم قرن بیستم علم‌سنجی به‌عنوان یک رشته علمی در آمریکا موردتوجه قرار گرفت. پس از پایان جنگ جهانی دوم، روش تحقیق علمی مورد تأکید قرار گرفت و در فرایند انجام پژوهش از روش‌های آماری و تکنیک‌های مبتنی بر اعداد و ارقام بهره گرفته شد. به‌عنوان مثال رشته‌های روان‌سنجی<sup>۱</sup> و زیست‌سنجی<sup>۲</sup> به‌عنوان شاخه‌های کمیت‌پذیر در رشته‌های روانشناسی و زیست‌شناسی بنانهاده شدند؛ علم‌سنجی، کتاب‌سنجی و اطلاع‌سنجی نیز به‌عنوان حوزه‌های سنجش کمی در حوزه کتابداری و اطلاع‌رسانی و مدیریت پژوهش مطرح شدند (امیرحسینی، ۱۳۷۱). تولیدات علمی حوزه علم‌سنجی از نظر موضوعی نسبت به حوزه علم اطلاعات و دانش‌شناسی مستقل شده است و در برخی کشورها ازجمله ایران، علم‌سنجی به‌عنوان یکی از گرایش‌های علم اطلاعات و دانش‌شناسی متولد شده است (خاصه و سهیلی، ۱۳۹۷).

#### ۴-۱. حوزه‌های مرتبط با علم‌سنجی

علم‌سنجی با دیگر حوزه‌های علمی مطرح در حوزه دانش اطلاع‌رسانی هم‌پوشانی دارد. ارتباط علم‌سنجی با علوم مشابه مانند اطلاع‌سنجی<sup>۳</sup>، وب‌سنجی<sup>۴</sup>، مجازسنجی<sup>۵</sup> و کتاب‌سنجی<sup>۶</sup> در شکل ۱-۱ نشان داده شده است.

---

<sup>۱</sup> Psychometrics

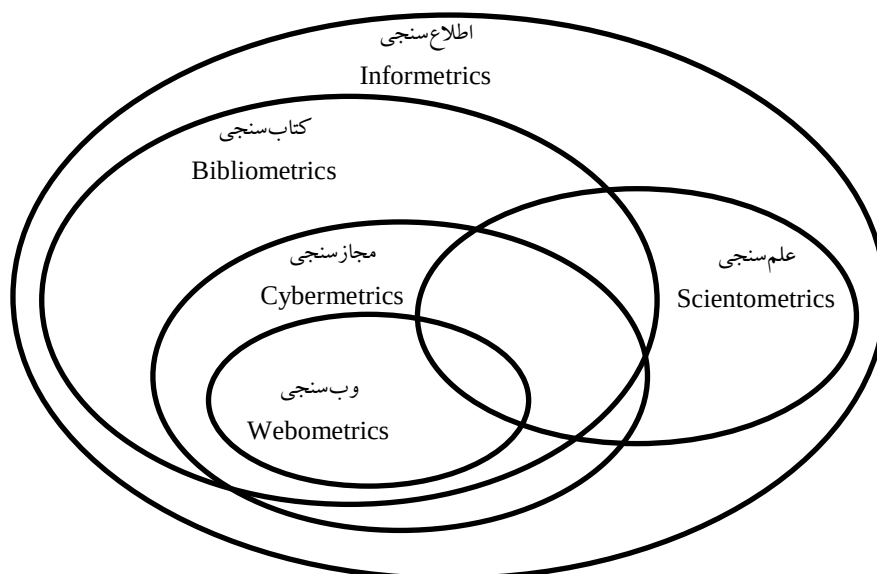
<sup>۲</sup> Biometrics

<sup>۳</sup> Informetrics

<sup>۴</sup> Webometrics

<sup>۵</sup> Cybermetrics

<sup>۶</sup> Bibliometrics



شکل ۱-۱- ارتباط بین علم‌سنجی و دیگر شاخه‌های مرتبط با سنجش اطلاعات (Björneborn and Ingwersen, 2004)

علم‌سنجی شامل تمام مدل‌ها و ابعاد کمی مرتبط با تولید و توزیع دانش علم و فناوری است و اطلاع‌سنجی مطالعه ابعاد کمی اطلاعات در شکل‌های مختلف است و صرفاً به اطلاعات مرتبط با رکوردها یا کتاب‌شناسی محدود نمی‌شود و شامل جوامع اطلاعاتی مختلف است و فقط شامل جامعه علمی نیست درحالی‌که کتاب‌سنجی به زیرمجموعه‌ای از اطلاع‌سنجی که عمدتاً شامل کتاب و مستندات علمی و کتابخانه‌ای است، گفته می‌شود. از سوی دیگر در دنیای دیجیتال، مجازسنجی به سنجش اطلاعات در فضای دیجیتال اشاره دارد و وب‌سنجی آن بخش از فضای مجازی را که در حیطه فضای وب است مدنظر دارد (De Bellis, 2009). شمارش تولیدات علمی و بررسی چگونگی افزایش یا کاهش آن، هدف اصلی بسیاری از مطالعات علم‌سنجی است. این مطالعات گاه افزایش یا کاهش تولیدات علمی و پژوهشی را برای پژوهشگران، سازمان‌ها و کشورها بررسی می‌کند و گاه به بررسی رشد و نزول تولیدات علمی در موضوعات و رشته‌های علمی گوناگون می‌پردازند.

تحلیل استنادی<sup>۱</sup> نیز از شیوه‌های مورد توجه در مطالعه تولیدات علمی و به‌ویژه ارتباطات علمی است که در آن به بررسی میزان استناددهی به منابع علمی پرداخته می‌شود و شامل موارد متعددی مانند تحلیل هم‌نویسی پژوهشگران و نیز تحلیل استناددهی بین متون علمی و ... است. در این روش ارتباطات علمی و همکاری‌های پژوهشی به صورت شبکه‌های پیچیده‌ای از ارتباطات مورد بررسی قرار می‌گیرد که منجر به شناسایی ارتباطات رسمی و غیررسمی بین پژوهشگران می‌شود (اسدی، ۱۳۹۳).

<sup>۱</sup> Citation Analysis

## ۱-۵. مفهوم کاستی پژوهش<sup>۱</sup>

خلا یا کاستی دانش فاصله‌ای است که میان دانش موردنیاز یک سازمان برای انجام وظایف و رسیدن به اهداف تعیین‌شده با دانشی که سازمان از منابع گوناگون دانشی بدست آورده است (Mai, Case, and Given, 2016). این کاستی هم در مورد دانش موجود و آرشو شده در یک نهاد، سازمان یا جامعه می‌تواند مطرح باشد و هم می‌تواند دانش قابل‌استقرار و کارآمد را شامل شود (McBriar et al., 2003). در تعاریف دیگری، کاستی دانش به درکی که از میزان اختلاف بین دانش موجود و کد شده با دانش موردنیاز یک سیستم یا فرد برای رسیدن به اهداف خود است، اطلاق شده است (Junnarkar, 1997; Zhao, Wu, Dai, and Dai, 2015). همین تعریف برای بیان کاستی پژوهش نیز کاربرد دارد و به بیان دیگر کاستی پژوهش، جای خالی پژوهش‌های موردنیاز یک سازمان یا جامعه را در بین منابع پژوهشی در دسترس سازمان یا جامعه مورد نظر قرار می‌دهد. این در حالی است که خلا یا شکاف پژوهش (Research Gap) در مطالعه یک موضوع نشانگر کاستی‌ها و کمبودهای پژوهش‌های پیشین در آن موضوع خاص است و پژوهشگر سعی دارد تا آن کاستی‌ها را برطرف نماید. به عبارتی در مقیاس کل‌نگر عبارت «Research Gap» به معنای بخش‌هایی است که در پژوهش‌های یک جامعه یا سازمان مغفول مانده‌اند یا به قدر کفایت در آن حوزه‌ها پژوهشی انجام نشده است ولی در مقیاس جزءنگر، این عبارت به معنای نواقص و کمبودهایی است که در پژوهش‌های یک موضوع موردنظر وجود دارد و در مستندات پژوهشی مانند مقاله یا پارسا معمولاً پس از مرور ادبیات موضوع به آن اشاره می‌شود. از اینرو در این طرح با رویکرد کل‌نگر به دنبال ارزیابی عرضه و تقاضای پارساها در حوزه انرژی هستیم و بیش از هر چیز به دنبال شناسایی موضوعاتی هستیم که نیازمند پژوهش بیشتر هستند و از واژه «کاستی پژوهش» برای این منظور استفاده خواهد شد.

## ۱-۶. مطالعات علم‌سنجی

مطالعات گسترده‌ای پیرامون علم‌سنجی پژوهش‌های مربوط به موضوعات زیرمجموعه انرژی در دنیا انجام شده است که از آن جمله می‌توان به بررسی مطالعات مربوط به صنعت نفت و گاز (L. Zhu, 2015)، برق (Liu, He, Shi, and Pang, 2015)، انرژی‌های سبز (Ranganathan and Balasubramani, 2014)، سلول‌های خورشیدی (Dutt and Nikam, 2020)، انرژی‌های

---

<sup>۱</sup> Research Gap

(2016)، انرژی‌های تجدیدپذیر<sup>۱</sup> (Romo-Fernández, Guerrero-Bote, and Moya-Anegón, 2013)، زغال سنگ (L. Yang, Wang, Bai, Deng, and Hu, 2018) اشاره نمود. با وجود اهمیت فراوان حوزه انرژی و پژوهش‌های این حوزه برای کشور، مطالعات علم‌سنجی اندکی در حوزه انرژی در ایران انجام شده است. (Davaranpanah, 2012) مطالعات انجام شده در حوزه انرژی اتمی در ایران را در حدها فصل سال‌های ۱۹۹۰ تا ۲۰۱۰ مورد بررسی کتابشاختی قرار داده است و مشخص کرده است که در این مدت پژوهش‌های این حوزه رشد نمایی داشته است و متوسط میزان استناد به مقالات بیش از ۵ مورد بوده است. حدود ۹۳ درصد مقالات از استفاده از همکاری نویسندگان با یکدیگر بدست آمده است و مقالاتی که با همکاری نویسندگان بین‌المللی نوشته شده است، بیش از مقالات داخلی مورد استناد قرار گرفته است. در این میان دانشگاه‌های تهران، تربیت مدرس، آزاد اسلامی و شیراز بیشترین تولیدات علمی را داشته‌اند.

## ۷-۱. پایگاه اطلاعاتی گنج

پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) بیش از نیم قرن است که به گردآوری، سازماندهی و اشاعه منابع پژوهشی کشور می‌پردازد و پژوهشگران نیز از منابع موجود در پایگاه اطلاعاتی ایرانداک<sup>۲</sup> که با نام گنج در جامعه علمی شناخته شده است، بهره می‌برند. بخش قابل توجهی از منابع این پایگاه شامل پایان‌نامه‌های کارشناسی ارشد و رساله‌های دکترا است که در این پژوهش از این دو منبع به نام پارسا<sup>۳</sup> یاد خواهد شد.

پایگاه گنج ایرانداک به‌عنوان مرجع رسمی پایان‌نامه‌ها و رساله‌های دانشجویی ایران، با بیش از یک میلیون رکورد علمی که تقریباً نیمی از آن را پایان‌نامه‌ها و رساله‌ها تشکیل می‌دهند مورد توجه پژوهشگران و دانشجویان حوزه‌های مختلف قرار دارد. در این پایگاه به طور متوسط روزانه بیش از ۱۲۰ هزار جستجو توسط کاربران صورت می‌پذیرد که تحلیل این حجم از تقاضاهای پژوهشگران می‌تواند اطلاعات ارزشمندی را در زمینه عرضه و تقاضای پژوهش‌ها در قالب پایان‌نامه و رساله در اختیار پژوهشگران و مدیران و سیاست‌گذاران علم و فناوری در حوزه‌های مختلف قرار دهد.

## ۸-۱. ضرورت و لزوم انجام پژوهش‌های تقاضا محور

---

<sup>۱</sup> Renewable Energy

<sup>۲</sup> این پایگاه در آدرس <https://ganj.irandoc.ac.ir> در دسترس است.

<sup>۳</sup> پارسا کوته‌نوشته پایان‌نامه و رساله است و به هر دوی این منابع اشاره دارد.

نقشه جامع علمی کشور، دربردارنده برنامه عملیاتی نظام علم و فناوری و نوآوری است و ویژگی‌های این نظام را برشمرده است. در این سند کلان، اولین ویژگی این نظام «ترکیب عرضه‌محوری و تقاضا‌محوری» بیان شده است. بحث عرضه و تقاضای پژوهش و فناوری در این سند به شکل زیر بیان شده است:

«با توجه به اهداف و آرمان‌ها و اولویت‌های بلندمدت نظام و کافی نبودن تقاضاهای بخش‌های اقتصادی و صنعتی از موارد مذکور، برخی حوزه‌های اولویت‌دار باید مورد حمایت ویژه قرار گیرند. این وجه از نظام علم و فناوری معطوف به تولید و عرضه دانش بر مبنای اهداف و آرمان‌های جامعه است. از سوی دیگر افزایش تقاضای نظام‌های فرهنگی، سیاسی، صنعتی و اقتصادی ملی و فراملی و در نتیجه تجاری کردن دانش و فناوری، اهمیت ویژه‌ای در پیشرفت همه‌جانبه و پایدار کشور دارد. بنابراین، الگوی مناسب برای نظام علم و فناوری جامعه ایرانی در این زمینه ترکیبی از الگوهای عرضه‌محوری و تقاضا‌محوری است» (فرهنگی، ۱۳۹۰).

از اینرو لازم است تا توجه ویژه‌ای به پژوهش‌های تقاضا‌محور شده و از انجام پژوهش‌های غیرمولد جلوگیری شود. برای این منظور می‌بایست از یک‌سو شناخت کاملی از تقاضاهای پژوهشی کشور در اختیار داشت و از دیگر سو، مطالعات موجود در کشور را به‌درستی پایش و رصد نمود تا با مقایسه این دو، کاستی‌های پژوهشی کشور مشخص شده و پژوهشگران را برای پر کردن این کاستی‌ها راهبری کرد.

با توجه به موارد مطرح شده در بالا، از مهم‌ترین مواردی که ضرورت انجام این پژوهش را بیان می‌کند می‌توان به موارد زیر اشاره کرد:

- لزوم سنجش وضعیت موجود عرضه و تقاضا در حوزه‌های انرژی برای دستیابی به جایگاه اول منطقه در این حوزه بر اساس سند چشم‌انداز بیست‌ساله کشور؛
- لزوم شناسایی، رصد و تعیین اولویت‌های پژوهشی کشور در حوزه انرژی با اولویت تقاضا‌محوری و تأکید بر انجام پژوهش‌های کاربردی و متناسب با نیازهای جامعه بر مبنای برنامه جامع علمی کشور و با توجه به اهمیت این حوزه در اقتصاد، سیاست، فناوری و دیگر ابعاد کلان مدیریتی کشور؛
- لزوم سنجش و تعیین کاستی‌های پژوهشی در منابع علمی زبان فارسی با توجه به عدم ارائه مقاله به زبان انگلیسی در بسیاری از حوزه‌ها و عدم استخراج مقالات زبان انگلیسی از برخی از پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی.

## ۹-۱. اهداف طرح

شناسایی وضعیت موجود عرضه و تقاضای پژوهش در حوزه‌های مختلف یکی از نیازهای اساسی مدیران و سیاست‌گذاران پژوهشی است. تحلیل وضعیت پژوهش‌های حوزه انرژی به دلیل اهمیت ویژه‌ای که در اقتصاد و سیاست کشور دارد، ضروری است؛ از سوی دیگر، از آنجا که این حوزه از دانش، شامل پژوهش‌های میان‌رشته‌ای زیادی بوده و طیف گسترده‌ای از رشته‌های دانشگاهی و حوزه‌های موضوعی را در بر می‌گیرد، نیازمند ارائه روش‌های هوشمند، جامع و کلی است تا بتوان حجم گسترده پژوهش‌های آن را تحلیل و مدیریت کرد.

پایگاه اطلاعاتی گنج دربردارنده بخش قابل توجهی از پژوهش‌های این حوزه است که در قالب پایان‌نامه و رساله تولید و به جامعه پژوهشی عرضه شده‌اند، همچنین جستجوهای فراوانی به صورت روزانه توسط پژوهشگران در موضوعات مرتبط با حوزه انرژی در این پایگاه انجام می‌شود که حاوی اطلاعات ارزشمندی از موضوعات مورد تقاضای جامعه پژوهشی است. این مطالعه به دنبال آن است تا پس از شناسایی و استخراج منابع پژوهشی عرضه شده و تقاضا شده مرتبط با حوزه انرژی در این پایگاه، شکاف بین موضوعات عرضه شده در پایان‌نامه‌ها و رساله‌های ایران را با موضوعات تقاضا شده توسط جویشگران تحلیل کرده و ارائه نماید.

در این پژوهش مجموعه پایان‌نامه‌ها و رساله‌های موجود در پایگاه گنج و مرتبط با حوزه انرژی به عنوان بعد عرضه در نظر گرفته خواهد شد. برای شناسایی مطالعات مرتبط حوزه انرژی از اصطلاح‌نامه‌های موضوعی، سبدوژگان<sup>۱</sup>، مدل‌سازی موضوعی، تکنیک‌های پردازش زبان طبیعی (NLP)<sup>۲</sup> و متن‌کاوی استفاده خواهد شد و کلمات پرتکرار استخراج می‌شود. سپس موضوعات موجود در این حوزه شناسایی و استخراج می‌شود و تحلیل کتابشناختی این موضوعات از منظرهای مختلف ارائه خواهد شد. سپس جستجوهای که توسط کاربران واقعی در پایگاه گنج صورت پذیرفته، مورد توجه قرار خواهد گرفت و بر اساس موضوعات مستخرج از بخش پیش، جستجوها دسته‌بندی و تحلیل کتابشناختی آنها ارائه خواهد شد. در نهایت ماتریس عرضه و تقاضای موضوعات شناسایی شده در حوزه انرژی ترسیم و هر یک از موضوعات از منظر عرضه و تقاضا بررسی خواهد شد.

از اینرو مهم‌ترین اهداف این پژوهش:

- تحلیل وضعیت عرضه و تقاضای پایان‌نامه‌ها و رساله‌ها در موضوعات مختلف حوزه انرژی؛

---

<sup>۱</sup> Bag of Words

<sup>۲</sup> Natural Language Processing

- شناسایی موضوعات دارای کاستی پژوهش در قالب پایان‌نامه‌ها و رساله‌ها در موضوعات مختلف حوزه انرژی؛  
هستند و برای این منظور تلاش خواهد شد تا به سوالات زیر پاسخ داده شود:  
۱. پایان‌نامه‌ها و رساله‌های عرضه‌شده در حوزه انرژی کدامند و چه موضوعاتی را پوشش می‌دهند؟  
۲. جستجوهای انجام شده در پایگاه گنج در حوزه انرژی کدامند و این تقاضاها در چه موضوعاتی از این حوزه قرار می‌گیرند؟  
۳. شکاف بین موضوعات عرضه شده در پایان‌نامه‌ها و رساله‌ها با موضوعات تقاضا شده توسط جویشران پایگاه اطلاعاتی گنج در حوزه انرژی به چه شکلی است؟

#### ۱-۱۰. پیوند با اساس‌نامه و برنامه راهبردی پژوهشگاه

با توجه به عزم پژوهشگاه بر انجام پژوهش‌های مرتبط با اساس‌نامه، اهداف کلان و برنامه راهبردی پژوهشگاه به منظور ایجاد یکپارچگی و اثربخشی مطالعات، همسویی پژوهش جاری با اساس‌نامه و اسناد بالادستی پژوهشگاه در زیر آمده است:

ماده ۱ اساس‌نامه:

- بند ۶: گسترش و ارائه خدمات مدیریت دانش، مدیریت اطلاعات و اطلاع‌رسانی اطلاعات علم و فناوری؛

از اینرو تحلیل عرضه و تقاضای پژوهش‌های یکی از بخش‌های بسیار مهم کشور مورد نظر ایرانداک قرار گرفته است و این پژوهش نیز در راستای انجام این مهم است. همچنین در راستای مأموریت پژوهشگاه و برطبق برنامه استراتژیک ایرانداک که وظایف ایرانداک را به شکل زیر تعریف می‌کند:

- بند ۲: در مدیریت اطلاعات علم و فناوری کشور مشارکت می‌کند؛

- بند ۵: تصویر و نمای علم و فناوری کشور را بازنمایی می‌کند؛

این پژوهش به دنبال ارائه نمای روشنی از پژوهش‌های حوزه انرژی است. در این برنامه یکی از مهم‌ترین اقدامات که به عنوان کارهای کلیدی مطرح شده است در استراتژی ۲ و کار کلیدی ۱ به شکل زیر تعریف شده است:

- ارتقای آگاهی سیاست‌گذاران (از نتایج طرح)

این کار کلیدی یکی از نتایج اجرای طرح خواهد بود که بر اساس آن سیاست‌گذاران حوزه پژوهش می‌توانند با آگاهی از وضعیت پژوهش‌های حوزه انرژی و بر پایه دانش مستخرج از این طرح تصمیمات کلان خود را بگیرند.

همچنین این پژوهش با طرح‌ها و پژوهش‌های پیشین پیشنهاددهنده:

- شناسایی کاستی‌های پژوهش با استفاده از تکنیک‌های متن‌کاوی و تحلیل رفتار جوی‌شگری کاربران (مورد مطالعه: محیط‌زیست ایران)، رساله دکتر، محمد ربیعی، ۱۳۹۸؛
  - طراحی، ساخت و پیاده‌سازی سامانه عرضه و تقاضای پژوهش (ساعت)، محمد ربیعی، ۱۳۹۵
- ارتباط و پیوند خواهد داشت. ضمن اینکه پژوهش‌های دیگری نیز در پژوهشگاه انجام شده است که از آن جمله می‌توان به:
- تحلیل رفتار جست‌وجوی اطلاعات پژوهشگران در موتور جست‌وجوی سامانه گنج، سمیه فتاحی، ۱۳۹۶
  - مقایسه پژوهش‌ها در حوزه فناوری اطلاعات کشور با روندهای پژوهش فناوری اطلاعات در جهان با استفاده از روش‌های هوشمند تحلیل متون، آرمان ساجدی‌نژاد، ۱۳۹۹
- اشاره کرد و این پژوهش رابطه تنگاتنگی با آنها دارد.

#### ۱-۱. تعریف واژه‌ها و مفاهیم

**پارسا:** منظور پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی است.

**عرضه پارساها:** منظور پایان‌نامه‌های کارشناسی ارشد و رساله‌های تحصیلات تکمیلی است که در پایگاه گنج ایرانداک ثبت شده‌اند.

**تقاضای پارساها:** منظور جستجو‌هایی است که در پایگاه گنج ایرانداک توسط کاربران واقعی صورت می‌پذیرد.

**بازار پژوهش:** ایده بازار پژوهش که مبتنی بر گسترش مفهومی بازار به حوزه پژوهش است بر تعامل عرضه‌کنندگان و تقاضاکنندگان آثار پژوهشی دلالت دارد (توکلی، ۱۳۸۸) که در این مطالعه، بخش پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی به عنوان بخشی اساسی از این بازار مدنظر است. در این بازار بر خلاف بازارهای سنتی که بیشتر متأثر از انگیزه جست و جوی نفع شخصی است، بازیگران به دنبال حداکثر کردن منافع اجتماعی و اخلاقی پژوهش هستند (توکلی، ۱۳۸۸).

#### ۱-۲. جمع‌بندی و نتیجه‌گیری

با توجه به اهمیت ویژه موضوع انرژی در اقتصاد، سیاست، فناوری و دیگر ابعاد راهبردی مدیریت کشور، ضرورت دارد در حوزه پژوهش نیز این حوزه مدنظر قرار گرفته با در نظر گرفتن اهداف و برنامه‌های نقشه جامع علمی کشور به انجام پژوهش‌های تقاضامحور و مولد در این حوزه پرداخته شود.

از اینرو در این پژوهش به دنبال شناسایی و تحلیل موضوعات مرتبط با حوزه انرژی و بسط عرضه و تقاضای پژوهش در هر یک از این موضوعات هستیم. برای این منظور از منابع موجود در پایگاه اطلاعاتی گنج که در بردارنده پارساهای کشور هستند به عنوان پایگاه عرضه منابع پژوهشی این حوزه و از تحلیل جستجوهای پژوهشگران در این حوزه، به شناسایی تقاضاهای پژوهشی حوزه انرژی خواهیم پرداخت.

این پژوهش به دنبال پاسخگویی به سوالات مختلف برای دستیابی به اهدافی چون، تحلیل وضعیت عرضه و تقاضای پارساها در موضوعات مختلف حوزه انرژی و نیز شناسایی موضوعات دارای کاستی پژوهش در این قالب خواهد بود. چگونگی دستیابی به این اهداف و گام‌های مرتبط با آن در بخش بعد که روش‌شناسی پژوهش را تشریح خواهد کرد بیان خواهد شد.

## فصل دوم

### روش‌شناسی پژوهش

## ۲\_۱. مقدمه

در این فصل روش‌شناسی پژوهش شامل بیان روش و رویکرد انتخابی برای پژوهش حاضر بیان خواهد شد و ویژگی‌های نرم‌افزارها و سخت‌افزارهای موردنیاز برای انجام بخش‌های مختلف این مطالعه بیان خواهند شد. همچنین مجموعه داده‌های مورد استفاده و روش انتخاب و استخراج آن‌ها تشریح خواهد شد. در ادامه روش‌ها و مدل‌های مورد استفاده در این پژوهش معرفی شده و دلایل انتخاب و استفاده از آن‌ها بیان خواهد شد. پس از بیان مدل‌ها و الگوریتم‌های یادگیری ماشین و توابع مورد استفاده در این مطالعه، الزامات استفاده از آن‌ها و نیز معرفی تعیین پارامترهای ورودی توابع مختلف مورد استفاده آن‌ها شرح داده شده است.

مفروضات و محدودیت‌هایی که در پژوهش جاری حاکم است در بخش‌های بعدی این فصل آمده است و در نهایت مراحل مختلف انجام پژوهش و خروجی‌های هر مرحله تعیین شده است. در پایان این فصل جمع‌بندی از مطالب بیان شده آمده است.

اولین گام در انجام مطالعات کتاب‌سنجی و تحلیل کلان‌متون اطلاعاتی، شناسایی و تشکیل پیکره<sup>۱</sup> است. به عبارتی لازم است تا منابع اطلاعاتی مرتبط با حوزه مدنظر تحلیل گردآوری شود. در این فصل روش‌های مختلف مربوط به شناسایی متون مربوط به یک حوزه مورد بررسی قرار گرفته و در نهایت متون مرتبط با حوزه انرژی شناسایی و استخراج خواهد شد و پیکره مربوطه تشکیل خواهد شد. این پیکره مبنای انجام تحلیل‌های کتاب‌شناختی و محتوایی در مراحل بعدی کار خواهد بود.

---

<sup>۱</sup> Corpus

تولید و مدیریت پیکره‌های متنی بزرگ با استفاده از تکنیک‌های متن کاوی<sup>۱</sup> و پردازش زبان طبیعی<sup>۲</sup> میسر خواهد بود. برای این منظور از ابزارهای گوناگونی استفاده خواهد شد که در ادامه به بررسی هر یک از روش‌ها و ابزارها پرداخته خواهد شد. پیش از آن لازم است تا در مورد داده کاوی<sup>۳</sup>، متن کاوی و کاربردهای آن در مدیریت اطلاعات علمی و فناوریانه بحث شود.

## ۲-۲. داده کاوی

پیشینه کاربرد عبارت «کشف دانش از پایگاه داده»<sup>۴</sup> در ادبیات متخصصین آمار، تحلیل داده‌ها و مدیریت سیستم‌های اطلاعاتی به سال ۱۹۸۹ برمی‌گردد. با این وجود این مفهوم به صورت جامع در سال ۱۹۹۶ تعریف شد، بر پایه این تعریف کشف دانش از پایگاه داده به فرایند پیدا کردن الگوهای معتبر، مفید و قابل فهم از داده‌ها دلالت دارد (Fayyad, Piatetsky-Shapiro, and Smyth, 1996a). داده کاوی یک دانش میان‌رشته‌ای است که حوزه‌های گوناگونی مانند هوش مصنوعی، مدیریت پایگاه داده‌ها، بصری‌سازی داده‌ها<sup>۵</sup>، یادگیری ماشین<sup>۶</sup>، الگوریتم‌های محاسباتی و آمار را شامل می‌شود (Fayyad et al., 1996a). توسعه این حوزه به سرعت انجام شده است و کاربرد این رشته با شتاب فراوانی رو به گسترش است که دلیل آن نیاز جدی به ایجاد یک رویکرد نوین در تحلیل، درک و بصری‌سازی حجم انبوه اطلاعاتی است که در پایگاه‌های اطلاعاتی تجاری و یا علمی جمع‌آوری می‌شوند (Tsai, 2011). به صورت کلی، کشف دانش از داده‌ها، به فرایند کلی استخراج دانش مفید از داده‌ها گفته می‌شود (غضنفری، علیزاده، و تیمورپور، ۱۳۹۳) درحالی که داده کاوی بخشی از این فرایند است که در آن با استفاده از الگوریتم‌های گوناگون، الگوهای مفید از داده‌ها شناسایی و استخراج می‌شود (Fayyad et al., 1996a). هدف از کشف دانش و داده کاوی یافتن همین الگوها و یا شناسایی رفتارهای داده‌های موجود در پایگاه است که میان حجم عظیمی از داده‌ها پنهان هستند (غضنفری و همکاران، ۱۳۹۳).

---

<sup>۱</sup> Text mining

<sup>۲</sup> Natural Language Processing (NLP)

<sup>۳</sup> Data mining

<sup>۴</sup> Knowledge Discovery in Database (KDD)

<sup>۵</sup> Data Visualization

<sup>۶</sup> Machine Learning

داده‌کاوی قابلیت‌های فراوانی را برای جامعه پژوهشی فراهم کرده است که از آن جمله می‌توان خلاصه‌سازی داده‌ها<sup>۱</sup>، قواعد باهم‌آیی داده‌ها<sup>۲</sup>، رده‌بندی داده‌ها<sup>۳</sup>، پیش‌بینی<sup>۴</sup> و خوشه‌بندی داده‌ها<sup>۵</sup> را نام برد. این کارکردها با استفاده از طیف گسترده‌ای از تکنیک‌ها قابل دستیابی هستند؛ تکنیک‌های مبتنی بر پایگاه داده<sup>۶</sup>، تکنیک‌های مبتنی بر یادگیری ماشین و تکنیک‌های آماری<sup>۷</sup> مهم‌ترین این موارد هستند (Fayyad, Piatetsky-Shapiro, and Smyth, 1996b).

در سالهای اخیر شاهد کاربردهای داده‌کاوی در مدیریت پژوهش هستیم که از آن جمله می‌توان به تحلیل کاستی‌های پژوهشی<sup>۸</sup> (Nazaridoust, Bidgoli, and Rezaeenoor, 2013)، خوشه‌بندی و طبقه‌بندی پژوهش‌ها و تعیین ساختار آن‌ها (Arun, Suresh, Madhavan, and Murthy, 2010; Nieminen, Pölönen, and Sipola, 2013; محمد اسماعیل، ریاحی، و صحبتی‌ها، ۱۳۹۳)، نمایه‌سازی متون (Shamsfard, 2008; Singh and Mohaar, 2016; Vállez, Pedraza-Jiménez, Codina, 2015; Blanco, and Rovira, 2015; نیاکان، ۱۳۸۳) و ... اشاره کرد.

## ۳\_۲. متن‌کاوی

با توجه به اینکه بخش انبوهی از داده‌های ذخیره شده در انبارهای داده را متون غیرساخت‌یافته تشکیل می‌دهند، یکی از شاخه‌های بسیار مهم و کاربردی داده‌کاوی، متن‌کاوی است. هدف متن‌کاوی استخراج دانش از داده‌های متنی غیر ساختاریافته یا نیمه ساختاریافته است. از متن‌کاوی با عناوین دیگری چون داده‌کاوی متنی<sup>۹</sup> (TDM) و کشف دانش از پایگاه‌های متنی<sup>۱۰</sup> نیز نام برده شده است. کاربردهای عمده متن‌کاوی در شناخت متون مشابه، خلاصه‌سازی متون، ترجمه ماشینی متون، انتخاب کلیدواژه یا برچسب‌زنی متون، رده‌بندی اطلاعات متنی و ... است که در فرایند آن از علوم مختلف، به‌ویژه پردازش زبان طبیعی استفاده می‌شود (Choudhary, Oluikpe, Harding, and Carrillo, 2009).

---

<sup>۱</sup> Data Summarization

<sup>۲</sup> Data Association

<sup>۳</sup> Data Classification

<sup>۴</sup> Prediction

<sup>۵</sup> Data Clustering

<sup>۶</sup> Database Oriented Techniques

<sup>۷</sup> Statistical Techniques

<sup>۸</sup> Research Gap Analysis

<sup>۹</sup> Textual Data Mining (TDM)

<sup>۱۰</sup> Knowledge Discovery in Textual Databases

هدف از متن‌کاوی استنتاج دانش ضمنی<sup>۱</sup> است که به صورت پنهان در متن موجود است و پس از این استنتاج، ارائه آن به شکل دانش صریح<sup>۲</sup> اهمیت بالایی دارد. برای دستیابی به این هدف عمدتاً چهار فعالیت مهم در فرایند متن‌کاوی صورت می‌پذیرد. این فعالیت‌ها شامل: بازیابی اطلاعات<sup>۳</sup>، استخراج اطلاعات<sup>۴</sup>، کشف دانش<sup>۵</sup> و فرضیه‌سازی<sup>۶</sup> است. سیستم بازیابی اطلاعات به دنبال آن است تا متن‌های متناسب با حوزه موردنظر را جمع‌آوری کند. سپس در مرحله استخراج اطلاعات، به دنبال یافتن اطلاعاتی با ساختار از پیش تعیین‌شده یا مدل‌شده (مانند استخراج روابط) از این متون هستیم. سیستم کشف دانش، راهکاری برای دستیابی به دانش جدید از اطلاعات استخراج‌شده را فراهم می‌آورد و در نهایت فرضیه‌سازی به تفسیر دانش کشف‌شده در بافت موضوعی مورد بررسی می‌پردازد (F. Zhu et al., 2013).

در مرحله بازیابی اطلاعات با استفاده از تکنیک‌های مختلف از جمله به کارگیری اصطلاح‌نامه<sup>۷</sup> (Mohammad Rabiei, Hosseini-Motlagh, and Haeri, 2017)، بسط جستجو<sup>۸</sup>، به کارگیری شبکه واژگان<sup>۹</sup> تخصصی حوزه مدنظر و ... تلاش می‌شود تا پیکره‌ای از متون مرتبط ایجاد شود و سپس از این پیکره به عنوان منبع اطلاعاتی متن‌کاوی استفاده شود. در تحلیل اطلاعات متنی، پنج رویکرد اصلی وجود دارد که در استخراج اطلاعات از هر یک از این رویکردها یا ترکیبی از آنها استفاده می‌شود. این رویکردها عبارت‌اند از: تحلیل املائی<sup>۱۰</sup>، تحلیل معنایی<sup>۱۱</sup>، تحلیل آماری<sup>۱۲</sup>، تحلیل نحوی<sup>۱۳</sup> و تحلیل کاربری<sup>۱۴</sup> (Losiewicz, Oard, and Kostoff, 2000).

---

<sup>۱</sup> Implicit Knowledge

<sup>۲</sup> Explicit Knowledge

<sup>۳</sup> Information Retrieval

<sup>۴</sup> Information Extraction

<sup>۵</sup> Knowledge Discovery

<sup>۶</sup> Hypothesis Generation

<sup>۷</sup> Thesaurus

<sup>۸</sup> Query Expansion

<sup>۹</sup> WordNet

<sup>۱۰</sup> Orthographic Analysis

<sup>۱۱</sup> Semantic Analysis

<sup>۱۲</sup> Statistical Analysis

<sup>۱۳</sup> Syntactic Analysis

<sup>۱۴</sup> Usage Analysis

در تحلیل املائی به کوچکترین واحدهای معنادار در زبان توجه می شود. مبنای اولیه کار، شناسایی این واحدها در متن است. در بسیاری از زبان ها مانند فارسی و انگلیسی شناسایی مرز کلمات از یکدیگر با استفاده از دستورات پایه ای پردازش زبان طبیعی و استفاده از کاراکتر فاصله، نقطه، علامت تعجب و سؤال و ... ممکن می شود. همچنین در بسیاری از موارد به بررسی n واحد معنایی که به دنبال یکدیگر و به صورت متوالی در متن آمده اند (N\_Gram) پرداخته می شود (Losiewicz et al., 2000).

تحلیل معنایی، به دنبال این است که ارتباط بین ظاهر متن و کلمات تشکیل دهنده آن با معنای پنهان در آن شناسایی شده و این معنا پایه تحلیل های بعدی اطلاعات قرار گیرد. در بسیاری از موارد کلمات متفاوت، معانی همانندی دارند. در این روش از تحلیل ریخت شناسی<sup>۱</sup> نیز استفاده می شود و پسوند های کلمات حذف می شود تا کلمات مختلف ریشه یابی<sup>۲</sup> شده و کلمات دارای معانی مشابه در یک رده شناسایی شوند. تکنیک دیگری که در تحلیل معنایی به کار می رود استفاده از ارتباطات بین کلمات و شناسایی کلمات عام تر<sup>۴</sup> و خاص تر<sup>۵</sup> است که برای این منظور از اصطلاح نامه های موضوعی<sup>۶</sup> و در سطح بالاتر از هستان شناسی<sup>۷</sup> استفاده می شود (Losiewicz et al., 2000).

در تحلیل آماری، به دنبال تحلیل کمی از وجود کلمات گوناگون در متن هستیم. در این روش فراوانی کاربرد کلمات در یک مستند نشان دهنده گرایش یا مفهومی است که آن مستند به دنبال ارائه آن است. در مجموعه های بزرگ، عبارت ها یا کلمات مشابه در حوزه های مختلف حضور دارند اما فراوانی این حضور می تواند موجب رفع ابهام شده و گرایش اصلی حوزه های گوناگون را نشان دهد (Losiewicz et al., 2000).

تحلیل نحوی، به ساختار جملات، جایگاه و نقش کلمات در جمله می پردازد و با استفاده از الگوریتم های مختلف شناسایی عبارت<sup>۸</sup> و یا شناسایی اجزای سخن<sup>۹</sup> نقش کلمات در جملات مشخص

---

<sup>۱</sup> Morphology Analysis

<sup>۲</sup> Suffix

<sup>۳</sup> Stemming

<sup>۴</sup> Broader Term

<sup>۵</sup> Narrower Term

<sup>۶</sup> Domain-Specific Thesaurus

<sup>۷</sup> Ontology

<sup>۸</sup> Parser Algorithm

<sup>۹</sup> Part of Speech (POS)

می‌شود و به این شکل تا حد بسیار زیادی از زبان طبیعی ابهام‌زدایی می‌شود. تحلیل نحوی می‌تواند با دیگر تکنیک‌ها مانند تحلیل معنایی یا تحلیل لغوی<sup>۱</sup> نیز ترکیب شود و به این شکل می‌تواند در زبان‌هایی مانند زبان چینی که مرز مشخصی برای شناسایی کلمات ندارند به شناسایی واحدهای معنادار زبان پردازد (Losiewicz et al., 2000).

تحلیل کاربری به میزان و شکلی که یک متن مورد استفاده کاربران یا سیستم‌های مختلف قرار می‌گیرد توجه دارد، به عنوان مثال زمانی که متنی از طرف صفحات مختلف وبی که عمدتاً موضوع آن‌ها ورزشی است مورد ارجاع<sup>۲</sup> یا پیوند<sup>۳</sup> قرار می‌گیرد این احتمال می‌رود که موضوع آن متن هم ورزشی باشد. همین‌طور رفتار کاربران یا نوع و گرایش آن‌ها در مورد یک متن می‌تواند تا حدی محتوای آن متن را نشان می‌دهد (Losiewicz et al., 2000).

## ۲-۴. کاربرد داده‌کاوی و متن‌کاوی در مدیریت پژوهش

مطالعات زیادی در بحث مدیریت پژوهش با استفاده از ابزارهای متن‌کاوی یا داده‌کاوی انجام شده است. این مطالعات با به کارگیری تحلیل داده‌های پایگاه‌های اطلاعاتی بر آنند تا وضعیت کنونی پژوهش‌ها را رصد کرده و روندهای پژوهش در آینده را ترسیم کنند یا با توجه به این روندهای حوزه‌های مختلف پژوهشی، ارتباطات ایجاد شده بین حوزه‌ها، رشد یا زوال آن‌ها، تأثیر انتشارات علمی در یک حوزه روی شکل‌های دیگر انتشارات همان حوزه (Tu and Seng, 2009)، بخش‌بندی و طبقه‌بندی حوزه‌های بزرگ‌تر به شاخه‌های فرعی و ... را شناسایی و تفسیر کنند. برخی از مهم‌ترین سؤالات گوناگونی که در حوزه مدیریت پژوهش وجود دارد و به کارگیری داده‌کاوی و متن‌کاوی امکان پاسخگویی بالقوه به این سؤالات را فراهم می‌سازد در زیر آمده است (Losiewicz et al., 2000):

- چه علوم، فناوری و پژوهش‌هایی در حال حاضر در حال انجام هستند؟
- چه افرادی این پژوهش‌ها را انجام می‌دهند؟
- زیردامنه‌ها و موضوعاتی که در یک حوزه پژوهشی مطرح هستند کدامند؟
- این پژوهش‌ها در چه مکان‌های جغرافیایی یا سازمان‌هایی در حال انجام هستند؟
- علوم، فناوری و پژوهش‌هایی که مغفول مانده‌اند کدامند؟

---

<sup>۱</sup> Lexical Analysis

<sup>۲</sup> Reference

<sup>۳</sup> Link

- چه حوزه‌هایی بیشترین تمرکز را دارند و چه ارتباطی بین این حوزه‌ها وجود دارد؟
- سمت‌وسوی پژوهش‌هایی که در آینده انجام خواهد شد، چگونه است؟
- و ...

تحلیل و بررسی پژوهش با ابزارها و روش‌های گوناگون اهداف مختلفی را مدنظر دارد که از مهم‌ترین آن‌ها می‌توان به تحلیل روند پژوهش‌ها، شناسایی کاستی‌های پژوهش، نیازسنجی و اولویت‌بندی پژوهش‌ها اشاره کرد. صاحبان صنعت و سیاست‌گذاران پژوهشی<sup>۱</sup> با آگاهی کامل از کاستی پژوهشی می‌توانند در برنامه‌ریزی و تصمیم‌گیری‌های کلان خود حوزه‌هایی را که دچار کمبود منابع علمی و پژوهشی هستند در نظر گرفته و برای جبران این کاستی چاره‌اندیشی کنند.

## ۲\_۵. کاربرد متن‌کاوی در مرور ادبیات پژوهشی

شناسایی پژوهش‌های مشابه و بررسی و تحلیل پیشینه یک پژوهش از اقدامات لازم و ضروری در فرایند انجام یک مطالعه نظام‌مند است، اما با توجه به رشد روزافزون تولید و انتشار پژوهش‌ها، این اقدام بسیار زمان‌بر و پیچیده است. پایگاه‌های متعددی پژوهش‌های علمی را ثبت و سازمان‌دهی نموده و نتایج آن را به شکل‌های گوناگون اشاعه می‌دهند. هر یک از این پایگاه‌ها حجم قابل توجهی از پژوهش‌ها را در حوزه‌های مختلف مدیریت می‌کنند، به همین دلیل ابزارهای گوناگون متن‌کاوی برای شناسایی و تحلیل پژوهش می‌تواند به کار گرفته شود. نتایج یک مطالعه نشان داده است که استفاده از ابزار متن‌کاوی بین ۳۰ تا ۷۰ درصد، زمان انجام مرور ادبیات را کمتر کرده است و در این بین تنها ۵ درصد از پژوهش‌های مرتبط (۹۵٪ بازخوانی) از قلم افتاده‌اند (O'Mara-Eves, Thomas, McNaught, , 2015). (Miwa, and Ananiadou, 2015).

مرور ادبیات، که یکی از بخش‌های اساسی هر مطالعه نظام‌مندی است می‌تواند در مقیاس‌های مختلفی انجام شود. در روش سنتی بخش به بخش مستندات موردنظر توسط پژوهشگر مورد مطالعه قرار می‌گیرد (Nieminen et al., 2013). فرایند کلی مرور ادبیات به شیوه سنتی به شکل زیر است:

پژوهشگر ابتدا فهرستی از کلیدواژه‌هایی را که فکر می‌کند نویسندگان یا خبرگان این حوزه از آن‌ها استفاده می‌کنند را آماده می‌کند، سپس با به کارگیری نرم‌افزارهای کاربردی کتابخانه‌ها یا ابزارهای جستجویی که در پایگاه‌های اطلاعاتی وجود دارد این کلیدواژه‌ها را جستجو می‌کند. در مرحله بعد، نویسنده با استفاده از دانش زمینه‌ای خود تصمیم می‌گیرد که از رکورد موردنظر استفاده کند یا اینکه

---

<sup>۱</sup> Research policy Maker

از آن صرف‌نظر کند. سپس اگر امکان ذخیره متن کامل<sup>۱</sup> آن مستند وجود داشته باشد، آن را ذخیره کرده و به مطالعه و یادداشت مطالب آن می‌پردازد و به سراغ رکورد بعدی می‌رود، این فرایند تا زمانی که پژوهشگر اطمینان حاصل کند که به اندازه کافی در حوزه مدنظر، ادبیات پشتیبان در اختیار دارد ادامه می‌یابد (Delen and Crossland, 2008). پس از گردآوری ادبیات موضوع، پژوهشگر باید با مطالعه آن‌ها ارتباط منطقی بین بخش‌های گوناگون رکوردهای مختلف به دست آورده و در نهایت بتواند از ابعاد مختلف آن‌ها را مورد تحلیل و بررسی قرار دهد. این روش دارای محدودیت‌ها و مشکلاتی است که در ادامه برخی از آن‌ها آمده است (Delen and Crossland, 2008; Nieminen et al., 2013):

- کلیدواژه‌هایی که پژوهشگر برای جستجوی ادبیات مرتبط انتخاب کرده است ممکن است با کلیدواژه‌هایی که نویسندگان و ویراستاران منابع علمی استفاده کرده‌اند تناسب نداشته باشد، از این رو بسیاری از منابع مرتبط با زبانی نخواهند شد و در عوض وجود منابع نامرتبط یا با ارتباط کم در رکوردهای با زبانی شده زمان پژوهشگر را تلف خواهد کرد.
- پژوهشگر به صورت آگاهانه یا ناآگاهانه برخی از منابع مرتبط را مطالعه و بررسی نمی‌کند. از این رو نتایج به دست آمده از مرور ادبیات ممکن است جامعیت و اعتبار کافی را نداشته باشد.
- در این روش پژوهشگر امکان مطالعه و بررسی تعداد محدودی از منابع را دارد و با توجه به گستردگی و تعدد پایگاه‌ها و منابع علمی، بررسی تعداد محدودی از منابع اعتبار کار پژوهشی را زیر سؤال می‌برد.
- تحلیل و بررسی پژوهش‌های مرتبط شناسایی شده از ابعاد مختلف، تحلیل روند پژوهش‌های انجام شده و دسته‌بندی و شناسایی کاستی‌های پژوهشی در روش سنتی کاری دشوار و زمان‌بر است.
- ذهنیت، گرایش و علاقه‌ی پژوهشگر باعث ایجاد سوگیری<sup>۲</sup> در مراحل تحلیل، ارزیابی و بررسی پژوهش‌های پیشین می‌شود و پژوهشگر به صورت خواسته یا ناخواسته تلاش می‌کند تا نتایج بررسی ادبیات موضوع منطبق با خواسته یا ذهنیت قبلی وی باشد.

مشکلات و موارد مطرح شده در بالا باعث شده تا اخیراً پژوهش‌های مختلفی با استفاده از تکنیک‌های متن‌کاوی به دنبال مرور ادبیات پژوهشی باشند. استفاده از متن‌کاوی موجب شده است تا امکان تحلیل

---

<sup>۱</sup> Full Text

<sup>۲</sup> Bias

حجم بسیار بزرگی از مطالعات پیشین صورت پذیرفته در یک موضوع فراهم شود (Nieminen et al., 2013; O'Mara-Eves et al., 2015; Sunikka and Bragge, 2012). در روش‌های مبتنی بر متن کاوی پژوهشگر با تولید پروفایل پژوهشی<sup>۱</sup> می‌تواند مجموعه بزرگی از منابع را به صورت یکجا بررسی و تحلیل کند. در این روش، شناسایی ارتباطات موضوعی و روندهای پژوهش میسر می‌شود، همچنین متن کاوی با کمک موتورهای جستجو امکان استخراج اطلاعات، ارتباطات، فرضیات و روندهای جدیدی را فراهم می‌آورد که به صورت اطلاعات صریح<sup>۲</sup> در متون بیان نشده است و استخراج آن‌ها با استفاده از روش‌های آماری رایج یا مطالعه تک‌تک منابع توسط پژوهشگر امکان‌پذیر نیست (Sunikka and Bragge, 2012). در کنار پژوهش‌های مختلف این حوزه که عمدتاً به بررسی و تحلیل فراداده‌های<sup>۳</sup> رکوردهای پژوهشی می‌پردازند، برخی از پژوهش‌ها مانند (Nieminen et al., 2013) از طریق تحلیل محتوای<sup>۴</sup> متنی مستندات علمی به دنبال یافتن پاسخ پرسش‌های پژوهشی خود هستند.

## ۲\_۶. رویکردهای نیازسنجی و اولویت‌یابی پژوهش

به‌طور کلی رویکردهای گوناگونی در زمینه شناسایی نیازهای پژوهشی و اولویت‌بندی آن‌ها مورد استفاده قرار می‌گیرد که از مهم‌ترین این رویکردها می‌توان به رویکرد آسیب‌شناختی<sup>۵</sup>، رویکرد موضوعی / محتوایی<sup>۶</sup>، رویکرد تقاضامحور<sup>۷</sup>، رویکرد عرضه‌محور<sup>۸</sup>، رویکرد نظام‌مند<sup>۹</sup> و رویکرد اداره‌گری / مدیریتی<sup>۱۰</sup> اشاره کرد (فتحی‌واجارگاه، خسروی، سعادت‌طلب، تورانی، و دیباواجاری، ۱۳۸۹؛ قنادی‌نژاد و حیدری، ۱۳۹۶). این رویکردها لزوماً در تضاد با یکدیگر نیستند و هر یک از آن‌ها از یک بعد به مسئله نیازسنجی و اولویت‌یابی پژوهش نگریسته‌اند. به‌عنوان مثال رویکرد موضوعی / محتوایی از منظر روش انجام و رویکرد تقاضامحور از منظر اهداف به این مقوله می‌پردازند.

---

<sup>۱</sup> Research Profiling

<sup>۲</sup> Explicit Information

<sup>۳</sup> Metadata

<sup>۴</sup> Content Analysis

<sup>۵</sup> Pathological Approach

<sup>۶</sup> Substantive Approach

<sup>۷</sup> Demand Oriented Approach

<sup>۸</sup> Supply Oriented Approach

<sup>۹</sup> Systematic Approach

<sup>۱۰</sup> Administrative / Managerial Approach

## ۲-۶-۱. رویکرد آسیب‌شناختی

در رویکرد آسیب‌شناختی مشکلات و آسیب‌های موجود در یک نظام مورد توجه قرار می‌گیرد. به عبارتی نیازسنجی پژوهش به دنبال شناسایی مسائل عمده و راه‌حل‌های احتمالی است. از این رو این رویکرد نوعی فعالیت حل مسئله است تا با تعریف حوزه‌ها و موضوعات پژوهشی، زمینه حل مشکلات و برطرف کردن کمبودهای موجود را فراهم سازد (Cohen, 1994; فتحی‌واجارگاه و همکاران، ۱۳۸۹).

## ۲-۶-۲. رویکرد موضوعی و محتوایی

این رویکرد بر ماهیت و ساختار دانش موجود در منابع مورد بررسی دلالت دارد. به عبارتی این رویکرد بر مبنای شاخه‌ها و زیرشاخه‌های علمی شناخته‌شده در حوزه مورد نظر بنا شده است. این شاخه‌ها می‌توانند از منابع مورد بررسی استخراج شده باشند و یا از شاخه‌های موضوعی رشته مورد نظر در دیگر منابع (مانند دانشگاه‌ها یا پایگاه‌های علمی) به دست آمده باشند. این رویکرد غنای هر یک از موضوعات و شاخه‌ها را بررسی و تحلیل می‌کند (فتحی‌واجارگاه و همکاران، ۱۳۸۹).

## ۲-۶-۳. رویکرد تقاضامحور

رویکرد تقاضامحور بر اساس توجه به نیازهای مطرح‌شده از طرف متقاضیان یا مصرف‌کنندگان پژوهش، اولویت‌ها را شناسایی و تعیین می‌کند. بر این اساس لازم است تا کاربران و متقاضیان پژوهش به اندازه کافی توانمند بوده و امکان اظهار نظر عملی و مشارکت در تعیین نیازهای پژوهشی را داشته باشند. از آنجایی که در روش‌هایی که از نظر کاربران یا مصرف‌کنندگان پژوهش استفاده می‌کنند، عمدتاً از تعداد محدودی از کاربران نظرخواهی می‌شود، این روش عمدتاً در سطوح پایین نیازسنجی پژوهشی مانند برنامه‌ریزی کوتاه‌مدت پژوهشی یا گزینش طرح‌های پژوهشی استفاده می‌شود (فتحی‌واجارگاه و همکاران، ۱۳۸۹). هرچند که اخیراً پژوهش‌هایی نیز با استفاده از امکانات فناوری اطلاعات در سطح کلان از نظرات کاربران و استفاده‌کنندگان پژوهش بهره‌برده‌اند (Mohammad Rabiei et al., 2017).

## ۲-۶-۴. رویکرد عرضه‌محور

در مقابل رویکرد تقاضامحور، رویکرد عرضه‌محور با تکیه بر منابع علمی و پژوهشی موجود به شناسایی و تحلیل نقاط ضعف و قوت آن‌ها پرداخته و از این روش به عنوان مبنایی برای شناسایی نیازهای پژوهشی و اولویت‌یابی آن‌ها استفاده می‌کند. پژوهش‌های متعددی (Bahadori, Teimourzadeh, Farzaneh, and Nejati, 2011; Gobster, 2014; Gregório et al., 2012; Jurkiewicz, Wendykier, Wojciechowski, Fedoryszak, and Dendek, 2014; Masinter, Small, and Casman, 2014; Nasser and Welch, 2013; O'Mara-Eves et al., 2015; Mohammad Rabiei et al., 2017; Tsai, 2011, 2013) از طریق تحلیل منابع موجود در پایگاه‌های علمی به شناسایی

وضعیت پژوهش در حوزه‌های مختلف پرداخته‌اند. این رویکرد عمدتاً برای تخصیص منابع در سطح کلان برنامه‌های پژوهشی مورد استفاده قرار می‌گیرد (فتحی‌واجارگاه و همکاران، ۱۳۸۹).

## ۲-۶-۵. رویکرد نظام‌مند

این رویکرد بر اساس تعیین یک سری از معیارها و استانداردها صورت می‌پذیرد به این شکل که پس از تعیین و استخراج معیارهای کارایی و اثربخشی پژوهش‌ها، حوزه‌های مختلف پژوهشی و موضوعات شناسایی شده از طریق این معیارها ارزیابی می‌شوند و بر اساس میزان بازده و نتایج واقعی آنها و نیز میزان تأمین اهداف از قبل تعیین شده توسط این پژوهش‌ها، اولویت‌ها تعیین می‌شود (Cohen, 1994; فتحی‌واجارگاه و همکاران، ۱۳۸۹).

## ۲-۶-۶. رویکرد اداره‌گری/مدیریتی

در رویکرد اداره‌گری/مدیریتی، ساختار سازمانی و جایگاه نهادهای حاکمیتی در تعیین نیازها و اولویت‌های پژوهشی حائز اهمیت است. در این روش هر یک از نهادهای بالادستی اولویت‌ها و نیازهای سطح کلان را مشخص می‌کند و سطوح پایین‌تر بر مبنای اولویت‌های کلان ارائه شده، لیستی از نیازها و اقدامات پژوهشی را تهیه می‌کنند.

رویکردهای فوق در برخی موارد بافت و ساختار مورد استفاده در نیازسنجی و اولویت‌یابی پژوهش را مدنظر قرار داده‌اند و در برخی موارد از نظر اهداف نیازسنجی و اولویت‌یابی با یکدیگر متفاوت‌اند. به همین دلیل، در بسیاری از موارد استفاده از یک رویکرد به‌تنهایی نتایج موفقیت‌آمیزی نخواهد داشت و تلفیقی از رویکردهای فوق انتخاب می‌شود و تلاش می‌شود تا با ترکیب نگرش‌های بیان‌شده، نیازهای پژوهشی از ابعاد مختلف شناسایی شده و در اولویت‌بندی پژوهش‌ها، حوزه‌هایی مدنظر قرار گیرد که بیشترین میزان تأثیر را بر جامعه پژوهشی داشته و بهره‌وری حاصل از سرمایه‌گذاری پژوهشی را بیشینه می‌کنند.

جدول ۱-۲ مراحل کلی و نقاط قوت و ضعف رویکردهای مختلف نیازسنجی و اولویت‌یابی پژوهش‌ها را بیان می‌کند. درمجموع، شرایط و امکانات موجود، اهداف مدنظر برای تعیین نیازها و اولویت‌های پژوهشی، اندازه و ابعاد جامعه پژوهشی موردنظر و محدودیت‌ها و ضرورت‌های موجود در انتخاب رویکرد مناسب تعیین‌کننده است.

جدول ۱-۲- رویکردهای مختلف نیازسنجی و اولویت‌یابی پژوهش

| رویکرد<br>نیازسنجی                        | مراحل کلی                                                                                                                                                                                                 | نقاط قوت                                                                                                                                                                   | نقاط ضعف                                                                                                                                                                                                             |
|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ۱-<br>شناسایی<br>نیازها                   | <ul style="list-style-type: none"> <li>• شناسایی مشکلات و آسیب‌های جامعه</li> <li>• تعیین اولویت‌های پژوهشی برای حل مسئله</li> </ul>                                                                      | <ul style="list-style-type: none"> <li>• بهره‌گیری از روش حل مسئله</li> <li>• کمک به رفع مشکلات و نارسایی‌ها</li> </ul>                                                    | <ul style="list-style-type: none"> <li>• نیازمند نظر خبرگان برای شناسایی مشکلات و آسیب‌ها</li> <li>• غیرقابل ارزیابی در کوتاه‌مدت</li> </ul>                                                                         |
| ۲-<br>شناسایی<br>نیازها<br>و<br>اولویت‌ها | <ul style="list-style-type: none"> <li>• شناسایی و تحلیل شاخه‌ها و زیرشاخه‌ها</li> <li>• تعیین زیرشاخه‌های غنی یا فقیر پژوهشی</li> </ul>                                                                  | <ul style="list-style-type: none"> <li>• مبتنی بر محتوای پژوهشی و واقعیت‌ها</li> <li>• امکان تحلیل کلی و جزئی زیرشاخه‌ها</li> </ul>                                        | <ul style="list-style-type: none"> <li>• مبتنی بر پژوهش‌های گذشته</li> <li>• نیازمند پایگاه اطلاعاتی جامع و به‌روز</li> <li>• پیچیدگی‌ها و مشکلات پردازش زبان طبیعی</li> </ul>                                       |
| ۳-<br>تعیین<br>نیازها<br>و<br>اولویت‌ها   | <ul style="list-style-type: none"> <li>• انتخاب گروهی از متقاضیان یا کاربران پژوهش</li> <li>• نظرسنجی متقاضیان و کاربران</li> <li>• تعیین نیازها و اولویت‌ها بر مبنای نتایج نظرسنجی</li> </ul>            | <ul style="list-style-type: none"> <li>• مقبولیت بالا از نظر متقاضیان پژوهش</li> <li>• شناسایی نیازهای متقاضیان</li> <li>• جلوگیری از انجام پژوهش‌های کم‌کاربرد</li> </ul> | <ul style="list-style-type: none"> <li>• نیازمند سطح بالایی از دانش در کاربران</li> <li>• نامناسب برای اولویت‌یابی بلندمدت پژوهش</li> <li>• عدم توجه کافی به پژوهش‌های بنیادی</li> </ul>                             |
| ۴-<br>تعیین<br>نیازها<br>و<br>اولویت‌ها   | <ul style="list-style-type: none"> <li>• محاسبه میزان پژوهش‌های انجام شده در شاخه‌ها و زیرشاخه‌های مختلف</li> <li>• تعیین نیازها و اولویت‌ها بر مبنای تحلیل پژوهش‌های موجود و نظر خبرگان</li> </ul>       | <ul style="list-style-type: none"> <li>• مبتنی بر محتوای پژوهشی و واقعیت‌های موجود</li> <li>• مناسب برای برنامه‌ریزی‌های کلان</li> </ul>                                   | <ul style="list-style-type: none"> <li>• نگاه کاملاً کمی و عدم توجه به نیاز کاربران</li> <li>• مبتنی بر پژوهش‌های گذشته</li> <li>• نیازمند پایگاه اطلاعاتی جامع و به‌روز</li> <li>• وابستگی به نظر خبرگان</li> </ul> |
| ۵-<br>تعیین<br>نیازها<br>و<br>اولویت‌ها   | <ul style="list-style-type: none"> <li>• تعیین معیارها و شاخص‌های اثربخشی پژوهش‌ها</li> <li>• ارزیابی پژوهش‌ها بر اساس این معیارها</li> <li>• تعیین اولویت‌های پژوهشی بر مبنای بازده پژوهش‌ها</li> </ul>  | <ul style="list-style-type: none"> <li>• توجه به کارایی و اثربخشی پژوهش</li> <li>• سنجش با شاخص‌های دقیق کارایی پژوهش</li> </ul>                                           | <ul style="list-style-type: none"> <li>• عدم امکان کمی کردن برخی از بروندهای پژوهشی</li> <li>• پیچیدگی وزن‌دهی به معیارها</li> <li>• عدم توجه کافی به پژوهش‌های بنیادی</li> </ul>                                    |
| ۶-<br>تعیین<br>نیازها<br>و<br>اولویت‌ها   | <ul style="list-style-type: none"> <li>• تعیین ساختار و سلسله‌مراتب مدیریتی</li> <li>• تعیین اولویت‌های کلان توسط مدیران رده‌بالا</li> <li>• تعیین اقدامات و طرح‌های پژوهشی بر مبنای اولویت‌ها</li> </ul> | <ul style="list-style-type: none"> <li>• ساختارمند و توجه به برنامه‌های کلان</li> <li>• امکان کنترل و نظارت بالا</li> </ul>                                                | <ul style="list-style-type: none"> <li>• عدم توجه به نیازهای پژوهشگران</li> <li>• عدم توجه به کیفیت و کارایی پژوهش</li> <li>• انعطاف‌پذیری پایین</li> </ul>                                                          |

## ۲\_۷. روش مدل‌سازی موضوعی

مدل‌سازی موضوعی از روش‌های کارای تحلیل متن است که هدف آن گروه‌بندی مجموعه‌ای از متون مختلف به صورت زیرمجموعه‌های معنادار که موضوع<sup>۱</sup> 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 به جای 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 (John W and Petko, 2013).

مدل‌سازی موضوعی بر اساس این ایده شکل گرفته است که اسناد، ترکیبی از موضوعات هستند و هر موضوع با یک توزیع احتمالی اسناد مختلف را تشریح می‌کند. از سوی دیگر موضوعات ترکیبی از عبارت<sup>۲</sup>‌ها هستند و هر کلمه با یک توزیع احتمالی موضوعات را تشریح می‌کند. در این مدل‌ها موضوعات (و توزیع آن‌ها) به عنوان ساختار نهفته یا متغیرهای نهفته در نظر گرفته می‌شوند. الگوریتم‌های مدل‌سازی موضوعی روش‌های آماری هستند که عبارت‌های اصلی اسناد را تجزیه و تحلیل می‌کنند تا مباحثی را که از طریق آن‌ها ارائه می‌شود تحلیل نموده و چگونه اتصال این موضوعات را به یکدیگر و ارتباط بین عبارت‌ها را تشریح کند. عبارت‌ها می‌توانند کلمات موجود در متن بوده و یا اینکه ترکیبی از کلمات (n-gram) را شامل شوند.

روش‌های استخراج موضوع بر اساس این فرض کار می‌کنند که عبارت‌های موجود در یک متن از نظر معنایی به هم وابسته هستند و معنایی که در متون یک موضوع مسنجم وجود دارد از مجموعه عبارت‌های موجود در مستندات آن به دست آمده است. به عبارتی صرف‌نظر از معانی یا موقعیت قرار گرفتن عبارت در یک متن، هم‌رخدادی عبارت‌ها مورد نظر قرار می‌گیرد و اسناد به صورت مجموعه‌ای از عبارت‌ها در نظر گرفته می‌شود. در مدل‌سازی موضوعی فرض‌های زیر در نظر گرفته می‌شود (Blei, Ng, and Jordan, 2003):

۱. ترتیب حضور مستندات در پیکره بی‌اهمیت است؛
۲. ترتیب قرار گرفتن عبارت‌ها در هر مستند بی‌اهمیت است؛
۳. موضوعات توزیع چندجمله‌ای از عبارت‌های موجود در یک بانک واژگان هستند؛
۴. کلمات یک مستند از تعدادی موضوع پنهان در آن مستند نشأت گرفته است.

---

<sup>۱</sup> Topic

<sup>۲</sup> Term

دو فرض اول که ترتیب مستندات در یک پیکره و ترتیب عبارت‌های موجود در یک مستند را در نظر نمی‌گیرد منجر به خلق مفهومی به نام سبد واژگان<sup>۱</sup> می‌شود که در روش‌های متن‌کاوی به عنوان نماینده‌ای از یک مستند در نظر گرفته می‌شود و روش‌های گوناگونی برای ایجاد و ترسیم این سبد وجود دارد و اصلی‌ترین ورودی الگوریتم‌های مدل‌سازی موضوعی است (Wallach, 2006). فرض سوم و چهارم بیان‌کننده مفهوم موضوع است که توزیعی از عبارت‌های موجود در یک بانک واژگان است. تولید این بانک واژگان نیز از روش‌های گوناگون مانند انتخاب واژگان با بسامد وقوع بالا و یا استفاده از واژه‌نامه‌های استاندارد و یا شکل‌گیری بانک بر اساس نظر خبره است. به عنوان مثال عبارت‌های با احتمال بالای حضور در موضوع ورزش ممکن است «فوتبال»، «والیبال»، «کشتی»، «استقلال»، «علی کریمی» و ... باشند. در صورتی که نیاز به تعیین نام یا برچسب برای یک موضوع وجود داشته باشد، معمولاً از چند کلمه اولی که بیشترین ارتباط را به آن موضوع دارند استفاده می‌شود و یا از خبرگان موضوعی کمک گرفته می‌شود تا موضوع استخراج شده را نام‌گذاری کنند.

## ۲-۱-۲. الگوریتم LDA<sup>۲</sup>

الگوریتم LDA رایج‌ترین و معتبرترین شکل از تکنیک مدل‌سازی موضوعی است که توسط بیلی<sup>۳</sup> و همکارانش در سال ۲۰۰۳ ارائه شد. این مدل بر این فرض بنانهاده شده است که هر مستند شامل مجموعه‌ای از عبارت‌ها (BoW) است که ترکیبی از موضوعات مدنظر نویسنده با به‌کارگیری این عبارت‌ها بیان خواهد شد. هر موضوع نیز توزیعی از عبارت‌های پرتکرار یا عبارت‌های تعیین شده در واژه‌نامه است. از این رو هر مستند را می‌توان با احتمالی (بین ۰ تا ۱) به موضوعات مختلف متعلق دانست (Blei et al., 2003) به عبارتی هر موضوع می‌تواند با این احتمال مستند را تشریح کند. مهم‌ترین نقدی که به این روش می‌شود عدم توجه این روش به ترتیب وقوع عبارت‌ها در یک متن و ترتیب وقوع متون در یک پیکره است. از آنجا که هر سند با وجود برچسب یا موضوعی که برای آن تعیین شده است، می‌تواند موضوعات دیگری را نیز پوشش بدهد (Arun et al., 2010; Chang and Blei, 2009; Lafferty and Blei, 2006; Schubert, Weiler, and Kriegel, 2014) نهایی کردن موضوعات یک سند، مورد توجه است. هر سند نیز از تعدادی عبارت تشکیل شده است و میزان ارتباط یک سند با یک موضوع به میزان اهمیت عبارت‌ها در آن موضوع و میزان اهمیت عبارت‌ها در این سند مرتبط است. به عبارتی اگر میزان اهمیت یا وزن یک عبارت در یک سند را با  $w$  نشان دهیم و یک سند شامل  $m$  عبارت

<sup>۱</sup> Bag of Words

<sup>۲</sup> Latent Dirichlet Allocation

<sup>۳</sup> Blei, David M

باشد که هر یک وزن و اهمیت خاص خود را برای آن سند دارند و با در نظر گرفتن وجود  $n$  مستند در یک پیکره، آنگاه LDA می تواند تعداد  $k$  موضوع را شناسایی کند به شکلی که:

فرمول ۱-۲

$$\begin{matrix} & \text{Doc}_1 & \text{Doc}_2 & \dots & \text{Doc}_n \\ \text{Term}_1 & w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ \text{Term}_2 & w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Term}_m & w_{m,1} & w_{m,2} & \dots & w_{m,n} \end{matrix} \xrightarrow{\text{LDA(BoW(Term,Doc), k)}} \begin{matrix} & \text{Topic}_1 & \text{Topic}_2 & \dots & \text{Topic}_k \\ \text{Term}_1 & \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,k} \\ \text{Term}_2 & \beta_{2,1} & \beta_{2,2} & \dots & \beta_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Term}_m & \beta_{m,1} & \beta_{m,2} & \dots & \beta_{m,k} \end{matrix} * \begin{matrix} & \text{Doc}_1 & \text{Doc}_2 & \dots & \text{Doc}_n \\ \text{Topic}_1 & \theta_{1,1} & \theta_{1,2} & \dots & \theta_{1,n} \\ \text{Topic}_2 & \theta_{2,1} & \theta_{2,2} & \dots & \theta_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Topic}_k & \theta_{k,1} & \theta_{k,2} & \dots & \theta_{k,n} \end{matrix}$$

در فرمول ۱-۲، ماتریس اول وزن ( $w$ ) مربوط به  $m$  عبارت ( $\text{Term}$ ) را در  $n$  مستند ( $\text{Doc}$ ) نشان داده است و در نهایت این ماتریس با حاصل ضرب ماتریسی که اهمیت ( $\beta$ ) عبارت را در  $k$  موضوع ( $\text{Topic}$ ) و اهمیت  $k$  موضوع ( $\theta$ ) را در  $n$  مستند نشان می دهد معادل شده است. از اینرو، نتیجه مدل سازی موضوعی، شناسایی  $k$  موضوع است که در هر موضوع عبارت ها وزن مختص خود را خواهند داشت و از این رو می توان به روش های مختلف مانند در نظر گرفتن چند واژه ای که بیشترین وزن را در یک موضوع دارند و یا کمک گرفتن از خبرگان موضوعی، برای آن موضوع نام یا برجسی را انتخاب کرد. همچنین خروجی حاصل، وزن هر یک از مستندات در یک موضوع را مشخص می کند. هدف اصلی از مدل سازی موضوعی شناسایی موضوعات مستندات (ماتریس  $k*n$ ) است و از آنجا که تعداد موضوعات ( $k$ ) عددی بسیار کوچک تر از تعداد واژه ها ( $m$ ) و تعداد مستندات ( $n$ ) است، استفاده از ماتریس های با اندازه  $k*n$  برای شرح وضعیت و تحلیل مستندات، بسیار قابل فهم تر خواهد بود.

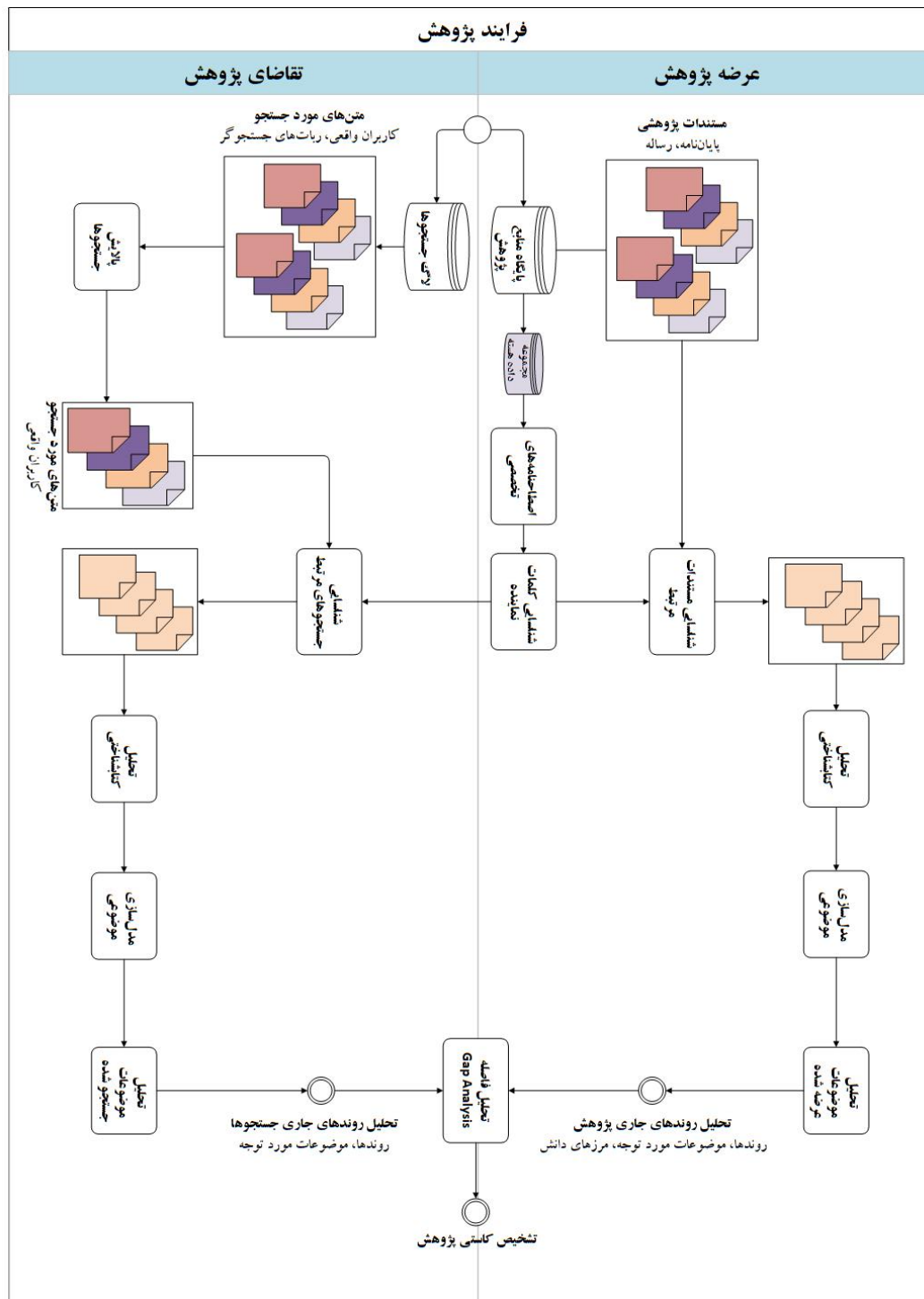
## ۸-۲. روش پژوهش

با توجه به رویکردها و روش های نیازسنجی و اولویت یابی پژوهش که پیشتر بیان شد، این پژوهش از منظر رویکردی تلفیقی از رویکردهای تحلیل موضوعی / محتوایی، رویکرد عرضه محور و رویکرد تقاضا محور را به کار می برد. همچنین مطالعه حاضر از نظر روش اجرا روشی تلفیقی را برگزیده است که با به کارگیری تکنیک های متن کاوی از روش های کمی، کتاب شناختی و تحلیل محتوا بهره می برد. این پژوهش بر مبنای تحلیل رفتار جویشگری کاربران به شناسایی تقاضاهای پژوهشی می پردازد و از

تکنیک‌های پردازش زبان طبیعی و کمک گرفتن از اصطلاح‌نامه‌های موضوعی برای تحلیل اطلاعات پژوهش‌های موجود و نیز رصد تقاضاهای پژوهش استفاده خواهد کرد.

## ۹\_۲. فرایند انجام پژوهش

فرایند انجام پژوهش در شکل ۱-۲ آمده است، در این فرایند، گام‌های پژوهش در دو مسیر مجزا و مرتبط عرضه و تقاضای پژوهش به پیش می‌رود. در بخش عرضه پژوهش، پارساهای موجود در پایگاه ایرانداک گردآوری و تحلیل شده است و در مسیر تقاضای پژوهش، نیازهای پژوهشی کاربران پایگاه اطلاعاتی که با جستجو در این پایگاه بیان کرده‌اند، بررسی شده است. می‌توان ترکیبی از منابع گوناگون را به عنوان منبع تقاضاهای پژوهشی استفاده کرد. این دو منبع به صورت مستقل از یکدیگر در نظر گرفته شده، از این رو اگر منابع اطلاعاتی دیگری وجود داشته باشد که بتوان آن را به عنوان منبع مناسبی برای عرضه و یا تقاضای پژوهش در نظر گرفت، می‌توان از آن منبع در مدل ارائه شده استفاده کرد.



در گام اول لازم است تا رشته‌متنی که کاربران واقعی جستجو کرده‌اند از عبارت‌های جستجو شده توسط ربات‌های جویشگر<sup>۱</sup> یا ربات‌های خزنده<sup>۲</sup> تفکیک شود. ربات‌ها از این عبارت‌ها برای نمایه کردن خودکار اطلاعات وبسایت‌ها استفاده می‌کنند. برای این منظور و نیز برای پاکسازی داده‌ها<sup>۳</sup> مجموعه‌ای از دستورات پیش‌پردازشی<sup>۴</sup> روی داده‌ها اعمال می‌شود. پس از انجام مراحل پیش‌پردازش، با به کارگیری اصطلاح‌نامه‌های تخصصی کلمات نماینده این حوزه شناسایی و با جستجوی این کلمات در پایگاه‌ها، از یک سو منابع پژوهشی مرتبط با حوزه انرژی و از سوی دیگر جستجوهای این حوزه شناسایی شده و بانک داده‌های عرضه و تقاضا تشکیل می‌شود. ربیعی و همکارانش با استفاده از تحلیل لاگ جستجوی کاربران در پایگاه پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی، روند جستجو و تعداد نتایج مربوط به زیرشاخه‌های مختلف حوزه محیط‌زیست را مورد بررسی قرار داده‌اند. در این پژوهش برای تولید و شناسایی مجموعه داده‌های مرتبط با حوزه محیط‌زیست از اصطلاح‌نامه این حوزه استفاده شده است و واژگان مرتبط با محیط‌زیست از این طریق شناسایی و استخراج شده است (Mohammad Rabiei et al., 2017).

سپس با به کارگیری روش‌های تحلیل کتابشناختی<sup>۵</sup>، فراداده‌های عرضه و تقاضای پژوهش‌های انرژی تحلیل خواهد شد. پس از آن با به کارگیری مدل‌سازی موضوعی، موضوعات این حوزه در هر دو بخش شناسایی و جایگاه این موضوعات، موضوعات داغ و روند پژوهش<sup>۶</sup> و ... تعیین خواهد شد. در گام آخر، عرضه و تقاضای پژوهشی تحلیل شده و با بصری‌سازی این نتایج، کاستی‌های موجود در هر یک از موضوعات استخراج و شناسایی شده است. تکنیک‌ها و روش‌های اشاره شده در هر یک از مراحل ارائه شده به صورت مفصل در فصل بعدی تشریح خواهند شد.

## ۲-۱۰. زبان برنامه‌نویسی

برای برنامه‌نویسی بخش‌های مختلف این پژوهش از زبان R استفاده شده است. R یک زبان برنامه‌نویسی شیء‌گرا<sup>۷</sup> و متن‌باز<sup>۸</sup> است که در ابتدا برای محاسبات آمار و ریاضی طرح‌ریزی شده است.

<sup>۱</sup> Search Engine Robots

<sup>۲</sup> Crawling Robots

<sup>۳</sup> Data Cleaning

<sup>۴</sup> Preprocess

<sup>۵</sup> Bibliographic Analysis

<sup>۶</sup> Research Trend

<sup>۷</sup> Object Oriented

<sup>۸</sup> Open Source

پروژه R از سال ۱۹۹۵ در بخش آمار دانشگاه Auckland و توسط Robert Gentleman و Ross Ihaka آغاز شد و به دلیل نام این دو فرد، R نام گرفت. این زبان در حال حاضر توسط یک گروه بین‌المللی نگهداری و به‌روزرسانی می‌شود<sup>۱</sup> و در حوزه‌های مختلف به‌ویژه در زمینه محاسبات آمار، ریاضی، داده‌کاوی و متن‌کاوی، پردازش تصویر و ... مورد استفاده جامعه علمی قرار می‌گیرد.

مهم‌ترین ویژگی‌هایی که موجب انتخاب این زبان برنامه‌نویسی در این پژوهش شد، عبارت‌اند از:

- یک زبان علمی است و قابلیت‌ها و ابزارهای فراوانی در مدیریت داده‌ها، توابع آماری و ریاضیاتی، توابع داده‌کاوی، متن‌کاوی و پردازش زبان طبیعی را داراست و سطح بالایی از همخوانی بین مفاهیم علمی و آکادمیک در حوزه‌های پردازش و تحلیل داده و داده‌کاوی با ابزارها و توابع توسعه داده‌شده برای این محیط وجود دارد، به‌نحوی که بسته‌های نرم‌افزاری، توابع، ابزارها و نحوه کار آن‌ها عمدتاً در مقالات علمی و به زبان آکادمیک تشریح شده و در پایگاه این زبان در دسترس است.
- به دلیل متن‌باز بودن این امکان را فراهم می‌کند تا اجزای داخلی الگوریتم‌ها، مدل‌ها و توابع را مشاهده کرده و امکان تغییر آن‌ها را فراهم می‌کند.
- امکان بسط، توسعه و بومی‌سازی بسته‌ها و توابع در این زبان فراهم است.
- به دلیل ساختار ماتریسی که برای متغیرها در نظر می‌گیرد تطابق زیادی با نگاشت‌های مفهومی در فرایندهای متن‌کاوی و داده‌کاوی دارد و انجام عملیات کشف دانش در آن ملموس است.
- به دلیل متن‌باز بودن گستره زیادی از توابع، ابزارها و راه‌حل‌های موردنیاز در حوزه‌های آماری، پردازش متن و مدل‌سازی را می‌توان در فضای وب جستجو کرده و از آن‌ها استفاده نمود.
- امکان انجام پردازش‌های موازی و توزیع‌شده برای کار با حجم انبوه داده‌ها و پردازش‌های سنگین را دارد که این قابلیت برای داده‌ها و پردازش‌های مربوط به این پژوهش بسیار حیاتی است.
- دارای راهنمای داخلی بسیار خوبی است و در فضای وب گروه‌های علمی زیادی پاسخگوی پرسش‌ها و نیازهای احتمالی خواهند بود.

---

<sup>۱</sup> <http://www.r-project.org>

- رایگان بوده و قابلیت نصب روی سیستم‌عامل‌های گوناگون را دارد و امکان تبادل داده و کد را روی نسخه‌های مختلف داراست.
- به‌منظور سهولت در برنامه‌نویسی R پس از نصب هسته R، امکان نصب پوسته R Studio وجود دارد که از طریق محیط تعاملی آن برنامه‌نویسی ساده‌تر خواهد بود.
- در کنار قابلیت‌ها و مزایای فوق‌زبان R مشکلات و محدودیت‌هایی را نیز داراست که مهم‌ترین آن‌ها به شرح زیر است:
- به نسبت دیگر زبان‌های برنامه‌نویسی تجاری و علمی، منابع، دوره‌های آموزشی، اساتید و خبرگان کمی در داخل کشور از این زبان استفاده می‌کنند و یادگیری و استفاده از این زبان قدری مشکل بوده و برای کار با این زمان در سطح حرفه‌ای لازم است تا تخصص کافی در حوزه برنامه‌نویسی و مبانی مهندسی نرم‌افزار داشته باشد.
- برخی از توابع، ابزارها و بسته‌های نرم‌افزاری توسعه داده شده، با زبان فارسی همخوانی نداشته و لازم است تا برای کار با متون فارسی بازنویسی شوند.
- در این پژوهش از بسته‌های نرم‌افزاری مختلفی در زبان R استفاده شده است که مهم‌ترین آن‌ها بسته پردازش زبان طبیعی (Kurt, 2017)، بسته متن‌کاوی (Feinerer, Hornik, and Feinerer, 2018) و بسته آمار و یادگیری ماشین (Meyer et al., 2018) است.

## ۱۱-۲. سخت‌افزار مورد استفاده

از آنجاکه در بسیاری از بخش‌های این پژوهش نیاز به پردازش‌های سنگین روی حجم بالایی از داده‌ها وجود دارد نیازمند استفاده از یک سرور با قابلیت پردازش و نیز حافظه اصلی بالا هستیم. همچنین به دلیل اینکه انجام بخش‌هایی از پردازش‌ها به صورت سریال امکان‌پذیر نیست و لازم است تا از پردازش موازی در هسته‌های مختلف پردازشگر استفاده شود، لازم است تا سرور از حداقل ۸ هسته پردازش بهره‌مند باشد. ویژگی‌های سرور مورد استفاده در این پژوهش در جدول ۲-۲ فهرست شده است.

جدول ۲-۲ - مشخصات سرور مورد استفاده در این پژوهش

| مشخصات               | اجزا                    |
|----------------------|-------------------------|
| پردازشگر مرکزی (CPU) | Intel Core i7, 4000 GHz |
| تعداد هسته (Core)    | 8 Core                  |
| حافظه اصلی (RAM)     | 32 GB                   |
| سیستم‌عامل           | Windows 8               |

## ۱۲-۲. قلمرو پژوهش

از آنجا که روش ارائه شده بر اساس متن کاوی و تحلیل متون علمی است، هر نوع از منابع علمی شامل کتاب، پایان‌نامه و رساله، مقاله، گزارش علمی و فنی، طرح پژوهشی و ... می‌تواند مورد استفاده قرار گیرد. اما منابع مورد استفاده در این مطالعه شامل پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی مرتبط با موضوع انرژی و نیز داده‌های مربوط به رفتار جوی‌نگری کاربران پایگاه ایرانداک است. از آنجا که بسیاری از مقالات پژوهشی نیز مستخرج از پایان‌نامه‌ها و رساله‌ها هستند، منبع انتخابی می‌تواند نماینده مناسبی برای منابع پژوهشی باشد.

مدل ارائه شده در این مطالعه، مختص زبان خاصی نیست و برخی از تکنیک‌های پایه متن کاوی و پردازش زبان طبیعی در زبان‌هایی که ویژگی‌های ساختاری مشابهی دارند، به صورت مشترک قابل استفاده است. با این حال به دلیل تفاوت‌هایی در ساختار، قواعد و دیگر معیارهای زبانی در بین زبان‌های مختلف، بسیاری از این الگوریتم‌ها باید به صورت اختصاصی و بومی برای هر زبان پیاده‌سازی شوند. این پژوهش از تکنیک‌های متن کاوی برای زبان فارسی استفاده می‌کند و تمامی داده‌های مورد استفاده در آن نیز به زبان فارسی است.

روش ارائه شده در این طرح پژوهشی مخصوص یک رشته و حوزه پژوهشی ویژه نیست، اما به دلیل ویژگی‌های خاص موضوع انرژی و اهمیت آن در اقتصاد، سیاست و دیگر ابعاد راهبردی کشور، پژوهش‌های حوزه «انرژی» در کشور، مبنای کار این مطالعه بوده است.

روش‌ها، ابزار و رویکرد مورد استفاده در این پژوهش مختص جغرافیا و محدوده زمانی خاصی نیست اما به دلیل محدودیت‌هایی که در ارائه و استفاده از منابع اطلاعاتی ایرانداک وجود داشت، جغرافیای اطلاعات مربوط به پژوهش‌هایی است که در ایران انجام شده است و جستجو‌هایی که در پایگاه ایرانداک انجام شده است مبنای کار قرار گرفته است. از آنجا که عمده جستجو‌های انجام شده در این پایگاه از طرف کاربران ایرانی و از داخل کشور انجام می‌شود، تمامی این جستجو‌ها مورد استفاده قرار گرفته است و به دلیل اینکه در بسیاری از موارد کاربران از فیلتر شکن و یا پروکسی برای مخفی سازی و تغییر آدرس خود استفاده می‌کنند امکان محدود سازی کاربران بر اساس آدرس آنها وجود ندارد. از این رو قلمرو این پژوهش ملی بوده و مربوط به کشور ایران است. قلمرو زمانی منابع پژوهشی مورد بررسی مربوط به بازه زمانی سال‌های ۱۳۰۰ تا ۱۳۹۹ است و جستجو‌های کاربران که به عنوان منبع تقاضای پژوهشی منظور شده است مربوط به حدها ۱۳۹۶/۱۲/۱۲ تا ۱۳۹۹/۰۶/۱۷ است.

## ۱۳\_۲. مفروضات پژوهش

شرایط و مفروضات زیر در این پژوهش حاکم است:

۱. پایگاه ایرانداک شامل نزدیک به ۷۰ درصد پارساهای کشور است، با این حال از آنجا که بزرگ‌ترین و جامع‌ترین پایگاه پایان‌نامه‌ها و رساله‌های کشور است، فرض شده است که بازگوکننده پژوهش‌های حوزه‌ی انرژی که یکی از حوزه‌های موجود در این پایگاه است، باشد.
۲. کاربران این پایگاه عمدتاً دارای تحصیلات عالی و یا خبرگان حوزه‌های دانشی گوناگون هستند. این کاربران به دنبال یافتن منابع علمی برای پاسخگویی به نیازهای پژوهشی خود که نمایی از نیازهای پژوهشی جامعه است، هستند.
۳. کاربران انتظار دارند تا پاسخی مناسبی برای نیازهای پژوهشی خود در این پایگاه پیدا کنند، از این رو خواسته‌های خود را در این پایگاه بیان می‌کنند و متون جستجو شده در این پایگاه، بیانگر بخش قابل توجهی از نیازهای اطلاعاتی آنها و نماینده خوبی برای تقاضاهای مطرح شده هستند. از اینرو، از این متون به عنوان شاخص ارزیابی تقاضاهای همه جامعه پژوهشی در حوزه انرژی استفاده شده است.
۴. اطلاعات پایگاه گنج، معتبر است و تمامی رکوردهایی که در بازه موردنظر در قلمرو این پژوهش جستجو شده است، در پایگاه اطلاعاتی به صورت لاگ ذخیره و در این پژوهش استفاده قرار گرفته است.

## ۲-۱۴. جمع‌بندی و نتیجه‌گیری

در این بخش رویکردهای مختلف در نیازسنجی و اولویت‌یابی پژوهش تشریح شد و نقاط قوت و ضعف هر یک از این رویکردها بیان شد. در نهایت با در نظر گرفتن ویژگی‌های منابع اطلاعاتی در دسترس برای تحلیل عرضه و تقاضاهای پارساها در حوزه انرژی، تلفیقی از رویکردهای تحلیل موضوعی/محتوایی، رویکرد عرضه‌محور و رویکرد تقاضا‌محور به عنوان روش پژوهش در نظر گرفته شد. برای این منظور از تکنیک‌های مختلف پردازش زبان طبیعی استفاده خواهد شد. همچنین با توجه به کاربردهای گسترده داده‌کاوی و متن‌کاوی، در پژوهش‌های اخیر علم‌سنجی و لزوم استفاده از روش‌های مبتنی بر یادگیری ماشین در تحلیل حجم انبوه داده‌ها، از این ابزار به منظور شناسایی و استخراج موضوعات گوناگون حوزه انرژی استفاده خواهد شد. روش مدل‌سازی موضوعی به عنوان یک روش مبتنی بر یادگیری ماشین و بر پایه قواعد متن‌کاوی تشریح شد که با استفاده از آن می‌توان موضوعات یک حوزه را شناسایی و استخراج کرد. همچنین فرایند انجام پژوهش به همراه گام‌های مختلف آن تشریح شد. در فصل بعد، بر اساس این فرایند، مراحل مختلف پژوهش انجام خواهد شد و نتایج حاصل در ادامه آن بیان خواهد شد.

## **فصل سوم**

**شناسایی و تحلیل متون  
عرضه و تقاضا شده در حوزه  
انرژی**

### ۱\_۳. مقدمه

در دهه‌های اخیر، شاهد افزایش چشم‌گیر تعداد پژوهش‌هایی هستیم که در زمینه‌های مختلف علوم صورت گرفته است. اما رشد کمی پژوهش‌ها، لزوماً با افزایش کیفیت آن‌ها همراه نبوده است و در بسیاری از موارد دستاوردهای این پژوهش‌ها منجر به رفع مشکلات جامعه نشده و به‌ندرت مشاهده شده است که نتایج و یافته‌های پژوهشی برای بهبود اوضاع سیاسی، اجتماعی، اقتصادی و فرهنگی جامعه به‌کار گرفته شود. در این میان شاید رشته‌های مهندسی و پژوهش‌های کاربردی آن وضعیت بهتری را نسبت به دیگر گروه‌های آموزشی داشته باشند (قنادینژاد، حیدری، و چینی پرداز، ۱۳۹۷).

فاصله یا شکاف میان پژوهش‌های دانشگاهی، به‌ویژه پایان‌نامه‌ها با نیازهای جامعه یکی از چالش‌های اساسی نظام آموزش عالی در ایران است. گسترش کمی دانشگاه‌ها و افزایش پذیرش دانشجویان تحصیلات تکمیلی باعث شده است تا از یک‌طرف شاهد انتقادات 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 (رحمت‌الله فتاحی، ۱۳۹۰). همچنین عرصه پژوهش شاهد دوباره‌کاری‌های بسیاری است که دلیل عمده آن، عدم توانایی استادان و دانشجویان در یافتن مسئله پژوهش در حوزه‌های مختلف موضوعی است و در بسیاری از موارد جامعه دانشگاهی به‌جای «مسئله‌یابی» به دنبال «مسئله‌سازی» شده است.

بر اساس تعریف رایینسون<sup>۱</sup>، 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 (Robinson, Saldanha, and Mckoy, 2011).

---

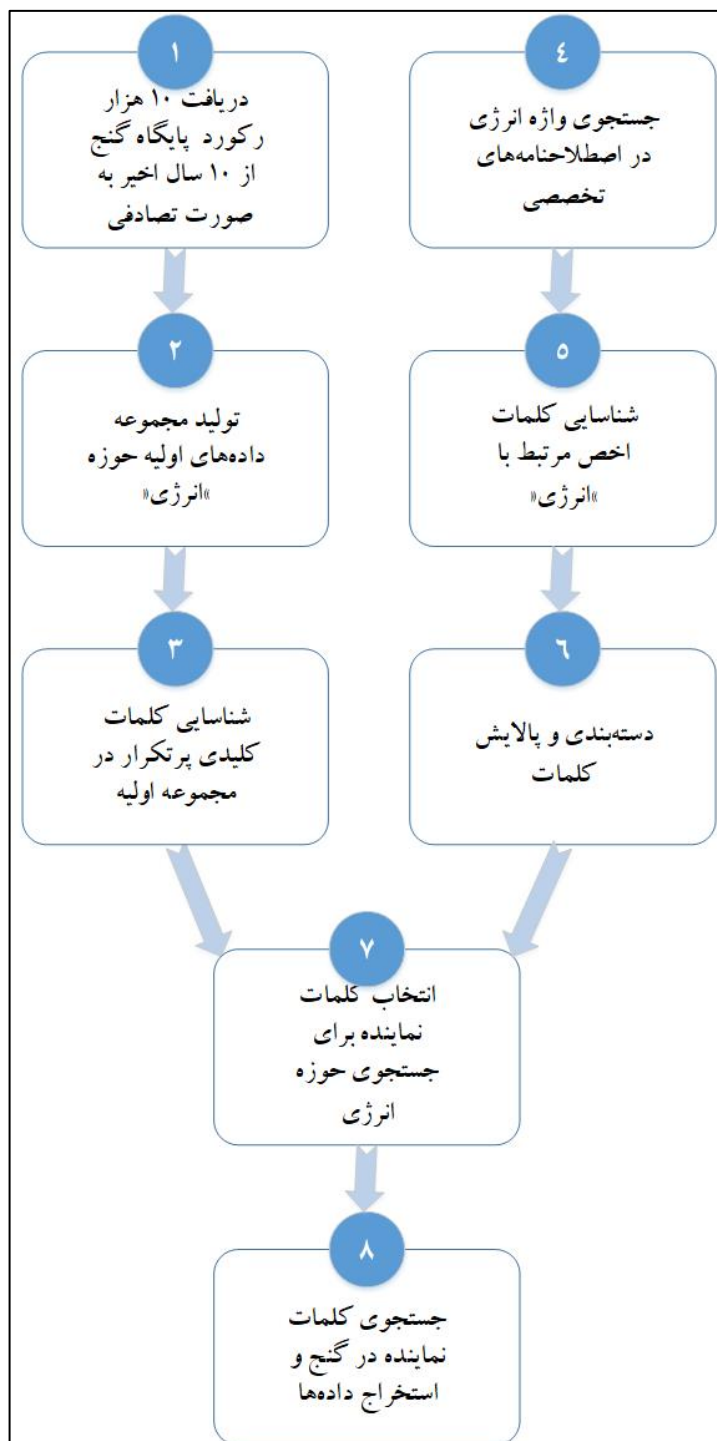
<sup>۱</sup> Robinson, Karen A

به عبارتی بسیاری از پژوهش‌ها 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. اما به نظر می‌رسد که نیازهای پژوهشی فراتر از حلقه پژوهشی است و حوزه‌های دانشی وجود دارد که ممکن است جزو نیازهای پژوهشی باشند اما بخشی از حلقه پژوهشی یا اطلاعاتی نباشند زیرا کمتر کسی به دنبال جستجوی این حوزه‌ها بوده است و در اصل تقاضایی برای این حوزه‌ها وجود نداشته و از این رو در حلقه پژوهشی جای نگرفته‌اند. همچنین بسیاری از نیازهای پژوهشی ممکن است به دلیل پیرو بودن پژوهشی رخ دهد و باعث ایجاد یک فاصله زمانی چند ساله میان پژوهش‌های داغ دنیا و در سطح ملی گردد.

سیاست‌گذاران پژوهش پس از اطلاع از حلقه پژوهشی و نیز پژوهش‌هایی که تقاضای قابل توجهی در جامعه پژوهشی برای آن‌ها وجود ندارد و با بهره‌گیری از دانش و اطلاعات دیگری که در اختیار دارند و نیز در نظر گرفتن سیاست‌های کلان پژوهشی، نیازهای پژوهشی را شناسایی و تعیین می‌کنند. در این مرحله سیاست‌گذار ممکن است سیاست‌ها و برنامه‌هایی را اعمال کند که تقاضای پژوهشی خاصی در جامعه پژوهشی ایجاد شود و این نیاز به پژوهشگران منتقل گردد. در نهایت سیاست‌گذاران و مدیران بر اساس معیارهای مختلف و با توجه به منابع موجود اولویت‌های پژوهشی را در حوزه‌های مختلف تعیین می‌کنند.

### ۲\_۳. فرایند استخراج پارساهای مرتبط با حوزه انرژی

برای استخراج پارساهای مرتبط با موضوع انرژی، گام‌های زیر طی شده است:



شکل ۱-۳- گام‌های استخراج پارساهای مرتبط با حوزه انرژی

در ادامه هر یک از گام‌های موجود در شکل ۱-۳ تشریح خواهد شد.

۱\_۲\_۳. گام ۱ (دریافت داده‌های نمونه از گنج)

در این گام برای آنکه نمونه‌ای از داده‌های گنج در اختیار باشد و بدون استفاده از خبرگان و به روش نیمه خودکار بتوان در نهایت داده‌های مربوط به حوزه انرژی را استخراج کرد، تعداد ۱۰ هزار رکورد که به صورت تصادفی از ۱۰ سال اخیر (هر سال ۱۰۰۰ رکورد) از پایگاه گنج انتخاب و در اختیار مجری طرح قرار گرفت.

### ۲-۲-۳. گام ۲ (تولید مجموعه داده اولیه حوزه انرژی)

در این گام با استفاده از داده‌های نمونه که در گام پیشین در اختیار قرار گرفته بود، با جستجوی عبارت «انرژی» در فیلدهای عنوان و کلیدواژه مجموعه داده‌های اولیه حوزه انرژی شکل گرفت. از این مجموعه به منظور شناسایی عبارت‌های پر کاربرد حوزه انرژی استفاده خواهد شد. این مجموعه شامل ۵۱۸ رکورد بود که نشان می‌دهد تقریباً ۵ درصد از نمونه اولیه قطعاً مرتبط با موضوع انرژی است.

### ۳-۲-۳. گام ۳ (شناسایی کلمات کلیدی پرتکرار حوزه انرژی)

پس از ساخت مجموعه داده اولیه حوزه انرژی، کلمات کلیدی این رکوردها بررسی و فراوانی آنها به دست آمد. به این ترتیب کلمات کلیدی پرتکرار شناسایی شد. ۱۰ واژه پرتکرار این حوزه و فراوانی آنها را در مجموعه داده اولیه برای نمونه در جدول ۳-۱ آمده است اما از ۴۰ واژه پرتکرار این حوزه در گام بعدی استفاده شده است. از آنجا که در ادامه کار به صورت مستقیم از عبارت‌های پرتکرار استفاده نشده است، و در گام هفتم، واژه‌ای که نمایانگر این عبارت باشد مورد استفاده قرار گرفته است، نیازی به ریشه‌یابی این عبارت‌ها وجود ندارد.

جدول ۳-۱- کلمات پر کاربرد در داده‌های اولیه حوزه انرژی

| رتبه | کلمه پرتکرار        | تعداد تکرار |
|------|---------------------|-------------|
| ۱    | انرژی خورشیدی       | ۱۱          |
| ۲    | انرژی               | ۸           |
| ۳    | انرژی تجدیدپذیر     | ۷           |
| ۴    | کاهش مصرف انرژی     | ۷           |
| ۵    | مصرف انرژی          | ۷           |
| ۶    | بهینه سازی          | ۵           |
| ۷    | سلول خورشیدی        | ۵           |
| ۸    | انرژی تاریک         | ۴           |
| ۹    | انرژی‌های تجدیدپذیر | ۴           |
| ۱۰   | خوشه‌بندی           | ۴           |

### ۳\_۲\_۴. گام ۴ (جستجوی «انرژی» در اصطلاح‌نامه‌های تخصصی)

از آنجا که تولید مجموعه داده اولیه حوزه انرژی، صرفاً بر مبنای وجود کلمه «انرژی» در رکوردهای داده‌های نمونه بوده است، قطعاً نمی‌تواند جامعیت کافی را داشته باشد. یکی از روش‌های گردآوری اطلاعات مربوط به یک موضوع علمی استفاده از اصطلاح‌نامه‌های<sup>۱</sup> تخصصی است. در این رویکرد، با به‌کارگیری اصطلاح‌نامه‌های موضوعی یک حوزه، کلمات کلیدی آن حوزه مشخص شده و مرز آن حوزه تعیین می‌شود. بدین ترتیب می‌توان مستندات پژوهشی یک حوزه را شناسایی و از پایگاه‌های مرتبط استخراج کرد.

برای این منظور مدخل «انرژی» در اصطلاح‌نامه‌های تخصصی موجود در پایگاه ایرانداک<sup>۲</sup> جستجو شد. سه نوع رابطه معنایی در اصطلاح‌نامه‌ها وجود دارد. نوع اول این رابطه رابطه هم‌ارزی یا ترجیحی<sup>۳</sup> که واژه‌هایی را به هم ارجاع می‌دهد که مفهوم یکسانی را بیان می‌کنند و می‌توان آنها را به صورت مترادف در نظر گرفت؛ مانند (تکنولوژی ب.ک (به‌کار برید) فناوری). نوع دوم رابطه سلسله‌مراتبی<sup>۴</sup> است که رابطه کلی و جزئی بین مفاهیم را بیان می‌کند. رابطه کلی‌تر یا سطح بالاتر یک عبارت را اصطلاح اعم<sup>۵</sup> (اع) و رابطه جزئی‌تر یا سطح پایین‌تر یک عبارت را اصطلاح اخص<sup>۶</sup> (اخ) می‌نامند؛ مانند (فاجعه هسته‌ای اع فاجعه)، و (فاجعه اخ فاجعه زیست‌محیطی). در نهایت نوع سوم رابطه‌ها در اصطلاح‌نامه رابطه وابسته یا همبسته<sup>۷</sup> است که بین دو واژه مرتبط که نه اعضای یک مجموعه هم‌ارز هستند و نه می‌توان رابطه سلسله‌مراتبی برای آنها متصور بود مطرح می‌شود؛ مانند (مالیات و اقتصاد). جستجوی مدخل «انرژی» در مجموعه اصطلاح‌نامه‌های تخصصی نشان داد که این عبارت در اصطلاح‌نامه‌های کشاورزی، شیمی، علوم زمین، ریاضیات، جامعه‌شناسی و فنی مهندسی حضور دارد که این مساله نشان می‌دهد این حوزه یک حوزه کاملاً بین‌رشته‌ای است.

### ۳\_۲\_۵. گام ۵ (شناسایی واژگان اخص مربوط به «انرژی» در اصطلاح‌نامه‌ها)

از آنجا که هدف از استفاده از اصطلاح‌نامه، شناسایی کلمات زیر مجموعه انرژی است که بتوان در نهایت با استفاده از جستجوی آنها مرز حوزه انرژی را شناسایی نموده و رکوردهای مرتبط را استخراج

<sup>۱</sup> Thesaurus

<sup>۲</sup> این امکان در آدرس <https://esn.irandoc.ac.ir/fa> فراهم است.

<sup>۳</sup> Equivalence or Pererential Relation

<sup>۴</sup> Hierarchical Relation

<sup>۵</sup> Broader Term

<sup>۶</sup> Narrower Term

<sup>۷</sup> Associative Relation

کرد. از بین روابط موجود در اصطلاح‌نامه که در بخش پیش تشریح شد، رابطه اخص مد نظر خواهد بود. از این رو واژگان اخص مربوط به «انرژی» در اصطلاح‌نامه‌های حاوی این عبارت استخراج شد. این واژگان در جدول ۲-۳ آمده است.

### ۶.۲.۳. گام ۶ (دسته‌بندی و پالایش کلمات مستخرج از اصطلاح‌نامه)

همانطور که پیشتر بیان شده روابط سه گانه در اصطلاح‌نامه برای عبارت‌ها وجود دارد که در این طرح صرفاً رابطه اخص مدنظر است. از سوی دیگر بسیاری از عبارت‌های به کار رفته در اصطلاح‌نامه بسیار تخصصی و نادر هستند و بسط مفهوم انرژی به تمامی این عبارت‌ها بسیار زمان‌بر و مشکل خواهد بود. از این رو عبارت‌ها مورد پالایش قرار گرفت و به جای استفاده از تمامی عبارت‌ها، کلمات نماینده‌ای که جستجوی آنها بتواند دربردارنده تمامی کلمات اخص مستخرج از اصطلاح‌نامه باشد مورد نظر قرار گرفت. این کلمات هم برای عبارت‌های مستخرج از اصطلاح‌نامه‌ها و هم برای کلمات پرتکرار شناسایی شده در گام ۳ مورد استفاده قرار گرفت.

### ۷.۲.۳. گام ۷ (انتخاب کلمات نماینده برای جستجوی حوزه انرژی)

همانطور که پیشتر بیان شد، به منظور شناسایی کلمات نماینده جستجو که بتوان با استفاده از آن حوزه انرژی را مشخص کرد، از کلمات پرتکرار موجود در منابع اولیه شناسایی شده حوزه انرژی و نیز اصطلاح‌نامه‌های تخصصی استفاده شد. جدول ۲-۳ کلمات مستخرج از اصطلاح‌نامه‌ها، و نیز کلمات نماینده را نشان می‌دهد.

جدول ۲-۳- کلمات مرتبط و نماینده کلمات مستخرج از اصطلاح‌نامه‌های فنی

| اصطلاح‌نامه         | کلمات مستخرج        | کلمه نماینده | اصطلاح‌نامه           | کلمات مستخرج      | کلمه نماینده |
|---------------------|---------------------|--------------|-----------------------|-------------------|--------------|
| اصطلاح‌نامه ریاضیات | انرژی شبکه          | انرژی        | اصطلاح‌نامه علوم زمین | انرژی آزاد        | انرژی        |
|                     | انرژی نقطه صفر بلور | انرژی        |                       | انرژی بادی        | انرژی        |
|                     | تابع انرژی          | انرژی        |                       | انرژی خورشیدی     | انرژی        |
|                     | تابع انرژی پتانسیل  | انرژی        |                       | انرژی رهایی       | انرژی        |
|                     | تراز انرژی          | انرژی        |                       | انرژی زمین‌گرمایی | انرژی        |
|                     | مشبکه (شیمی)        | -            |                       | انرژی فعال‌سازی   | انرژی        |
| اصطلاح‌نامه شیمی    | انرژی آپیوشی        | انرژی        |                       | انرژی کشندی       | انرژی        |

|                      |               |
|----------------------|---------------|
| انرژی آزاد گیس       | انرژی         |
| انرژی آزاد هلمهولتز  | انرژی         |
| انرژی اتمی شدن       | انرژی         |
| انرژی ارتعاشی        | انرژی         |
| انرژی                | انرژی         |
| الکترون خواهی        |               |
| انرژی الکتریکی       | انرژی         |
| انرژی انتقالی        | انرژی         |
| انرژی اوربیتال       | انرژی         |
| انرژی برهم کنش       | انرژی         |
| انرژی پایداری        | انرژی         |
| انرژی پتانسیل        | انرژی         |
| انرژی پیوند شیمیایی  | انرژی         |
| انرژی تفکیک          | انرژی         |
| انرژی جذب            | انرژی         |
| انرژی چرخشی          | انرژی         |
| انرژی چسبندگی        | انرژی         |
| فلزی                 |               |
| انرژی حلالپوشی       | انرژی         |
| انرژی درونی          | انرژی         |
| انرژی زوج شدن        | انرژی         |
| انرژی شبکه           | انرژی         |
| انرژی شکافت          | انرژی         |
| انرژی شیمیایی        | انرژی         |
| انرژی فعالسازی       | انرژی         |
| انرژی کوانتایی       | انرژی         |
| انرژی گرمایی         | انرژی         |
| انرژی مکانیکی        | انرژی         |
| انرژی مولکولی        | انرژی         |
| انرژی نقطه صفر       | انرژی         |
| انرژی همبستگی        | انرژی         |
| انرژی یونش           | انرژی         |
| اصطلاح‌نامه          | انرژی خورشیدی |
| جامعه‌شناسی          | انرژی هسته‌ای |
| انرژی لرزه‌ای        | انرژی         |
| انرژی موج            | انرژی         |
| انرژی هسته‌ای        | انرژی         |
| انرژی هیدروالکتریکی  | انرژی         |
| -                    |               |
| جنبشی                |               |
| شکافت                | -             |
| منابع انرژی          | انرژی         |
| انرژی خورشیدی        | انرژی         |
| انرژی آزاد گیس       | انرژی         |
| انرژی سطح مشترک      | انرژی         |
| پایداری ترمودینامیکی | -             |
| انرژی بستگی          | انرژی         |
| انرژی بستگی هسته‌ای  | انرژی         |
| انرژی خورشیدی        | انرژی         |
| انرژی فعال‌سازی      | انرژی         |
| انرژی                | انرژی         |
| انرژی زمین گرمایی    |               |
| انرژی                | انرژی         |
| بازیابی گرما         | -             |
| بازیافت انرژی        | انرژی         |
| پایستگی انرژی        | انرژی         |
| پایستگی گرما         | -             |
| پتانسیل الکتریکی     | -             |
| تبدیل انرژی          | انرژی         |
| تبادل انرژی          | انرژی         |
| توان                 | -             |
| تولید گرما           | -             |
| طرح برق آبی          | برق           |
| محرومیت انرژی        | انرژی         |
| مصرف انرژی           | انرژی         |
| مصرف گرما            | -             |
| منبع انرژی           | انرژی         |
| منبع تغذیه الکتریکی  | -             |
| اصطلاح‌نامه کشاورزی  |               |

## تحلیل عرضه و تقاضای پایان‌نامه‌ها و رساله‌های حوزه انرژی در پایگاه گنج □ ۴۵

| انرژی تابش | انرژی | نسبت انرژی        | انرژی |
|------------|-------|-------------------|-------|
|            |       | نگهداری گرما      | -     |
|            |       | هزینه انرژی       | انرژی |
|            |       | هزینه تولید انرژی | انرژی |

همچنین جدول ۳-۳ کلمات پرتکرار شناسایی شده از مجموعه داده اولیه حوزه انرژی و نیز کلمات نماینده جستجو که موجب بازنمایی عبارت‌های پرتکرار خواهد شد را نشان می‌دهد.

جدول ۳-۳- کلمات پرتکرار و نمایندگان جستجو در مجموعه داده اولیه حوزه انرژی

| کلمه پرتکرار        | فراوانی | نماینده      | کلمه پرتکرار            | فراوانی | نماینده |
|---------------------|---------|--------------|-------------------------|---------|---------|
| انرژی خورشیدی       | ۱۱      | انرژی        | طراحی آزمایش            | ۳       | -       |
| انرژی               | ۸       | انرژی        | قابلیت هضم              | ۳       | -       |
| انرژی تجدیدپذیر     | ۷       | انرژی        | کروماتوگرافی گازی       | ۳       | گاز     |
| کاهش مصرف انرژی     | ۷       | انرژی        | ها                      | ۳       | -       |
| مصرف انرژی          | ۷       | انرژی        | بازار برق               | ۳       | برق     |
| بهینه سازی          | ۵       | -            | تمایل به برقراری ارتباط | ۳       | -       |
| سلول خورشیدی        | ۵       | سلول خورشیدی | نگرش                    | ۳       | -       |
| انرژی تاریک         | ۴       | انرژی        | اتلاف انرژی             | ۲       | انرژی   |
| انرژی‌های تجدیدپذیر | ۴       | انرژی        | اقلیم                   | ۲       | -       |
| خوشه‌بندی           | ۴       | -            | امکان سنجی              | ۲       | -       |
| ازدیاد برداشت نفت   | ۴       | نفت          | امنیت انرژی             | ۲       | انرژی   |
| استهلاک انرژی       | ۳       | انرژی        | انتخاب سرخوشه           | ۲       | -       |
| انرژی الکتریکی      | ۳       | انرژی        | انرژی فعالسازی          | ۲       | -       |
| شدت انرژی           | ۳       | انرژی        | آنالیز اقتصادی          | ۲       | -       |
| صرفه جویی انرژی     | ۳       | انرژی        | آنالیز انرژی            | ۲       | انرژی   |
| محیط زیست           | ۳       | -            | برداشت انرژی            | ۲       | انرژی   |
| ایران               | ۳       | -            | بهینه سازی انرژی        | ۲       | انرژی   |
| تزریق گاز           | ۳       | گاز          | بهینه‌سازی چندهدفه      | ۲       | -       |
| خستگی               | ۳       | -            | پیزوالکتریک             | ۲       | -       |
| خلیج فارس           | ۳       | -            | توسعه پایدار            | ۲       | -       |

۸\_۲\_۳. گام ۸ (جستجوی کلمات نماینده و استخراج داده‌ها)

همانطور که در گام های پیشین بیان شد، خروجی ۷ گام پیشین منجر به شناسایی عبارت های «انرژی»، «سلول خورشیدی»، «نفت»، «گاز» و «برق» به عنوان نمایندگان حوزه انرژی شد که انتظار می رود جستجوی این عبارت ها موجب استخراج تمام رکوردهای مرتبط با حوزه انرژی شود. این عبارت ها در سمت عرضه (پارساها) در پایگاه گنج جستجو شد و نتیجه حاصل به عنوان داده های بخش عرضه حوزه انرژی در پژوهش مورد استفاده قرار گرفت.

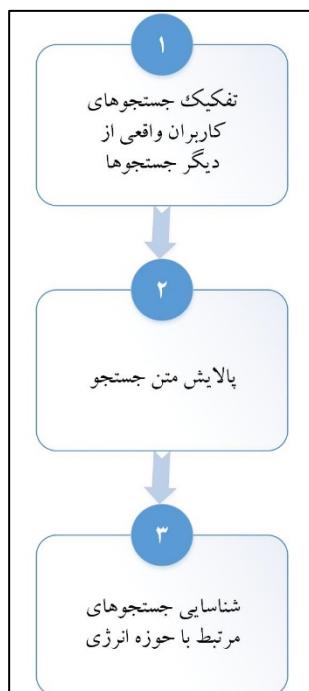
جستجوی بالا در بین پارساها موجب استخراج تعداد ۲۵،۹۵۳ رکورد شد. جدول ۳-۴ نمونه ای از منابع عرضه شده مرتبط با حوزه انرژی را نشان می دهد.

جدول ۳-۴- نمونه ای از پارساهای شناسایی شده مرتبط با حوزه انرژی

| عنوان                                                                                       | گروه         | تاریخ دفاع | کلیدواژه                                                                                                                         | مقطع          | رشته (گرایش) |
|---------------------------------------------------------------------------------------------|--------------|------------|----------------------------------------------------------------------------------------------------------------------------------|---------------|--------------|
| عوامل موثر نوسانات صادرات غیرنفتی ایران                                                     | علوم انسانی  | ۱۳۸۸       | بی ثباتی در آمد، صادرات غیرنفتی                                                                                                  | کارشناسی ارشد | اقتصاد       |
| بررسی عوامل تاثیر گذار بر سیاست خارجی اتحادیه اروپا در قبال جمهوری اسلامی ایران (۲۰۰۹-۲۰۰۱) | علوم انسانی  | ۱۳۸۹       | ایالات متحده آمریکا، حقوق بشر، اتحادیه اروپا، جمهوری اسلامی ایران، خاورمیانه، سیاست خارجی، صلح، جهت گیری، انرژی هسته ای، تروریسم | کارشناسی ارشد | علوم سیاسی   |
| طراحی یک پروتکل مسیریابی کارا از نظر انرژی در شبکه های حسگر بی سیم زیر آب                   | -            | ۱۳۹۸       | مصرف انرژی، شبکه حسگر بی سیم زیر آب، پروتکل های مسیریابی                                                                         | کارشناسی ارشد | مهندسی برق   |
| توسعه مدل ریاضی برای زنجیره تامین حلقه بسته پایدار در چارچوب اصول کارایی و تاب آوری سیستمی  | -            | ۱۳۹۸       | تاب آوری، کارا به انرژی، ریسک، زنجیره تامین حلقه بسته، پایداری                                                                   | دکتر          | مهندسی صنایع |
| طراحی مدرسه ابتدایی پسرانه در بسطام با رویکرد معماری پایدار                                 | فنی و مهندسی | ۱۳۹۴       | انرژی پاک، بام سبز، صفحات فتوولتائیک، دیوار ترومب                                                                                | کارشناسی ارشد | معماری       |

### ۳-۲-۱. فرایند استخراج تقاضاهای مرتبط با حوزه انرژی

برای استخراج تقاضاهای مرتبط با موضوع انرژی، گام های زیر طی شده است:



شکل ۳-۲- گام‌های استخراج جستجوهای مرتبط با حوزه انرژی

در ادامه هر یک از گام‌های موجود در شکل ۳-۲ شرح خواهد شد.

### ۳-۲-۲. تفکیک جستجوهای کاربران واقعی از ربات‌ها و موتورهای جستجو

همان‌طور که در فصل قبلی و در شکل ۲-۱ آمده است، نزدیک به ۲۵ میلیون رکورد در مجموعه داده‌های لاگ جستجو گردآوری شده است. بخشی از این جستجوها مربوط به ربات‌های موتورهای جستجوگر مانند Google، Yahoo و ... است که به‌منظور نمایه کردن<sup>۱</sup> اطلاعات صورت می‌پذیرد. همچنین بخشی از این جستجوها مربوط به جستجوهای است که کاربران در موتورهای جستجویی نظیر گوگل انجام می‌دهند و گوگل با توجه به نمایه‌هایی که از قبل تهیه کرده است اقدام به جستجو در پایگاه گنج می‌کند. از این‌رو لازم است تا جستجوهای که توسط کاربران واقعی صورت پذیرفته است از جستجوهای که توسط موتورهای جستجو و به‌صورت نظام‌مند<sup>۲</sup> انجام شده است، متمایز شوند. برای این منظور لازم است تا روی ویژگی<sup>۳</sup>ها و خصیصه‌هایی که در هر رکورد لاگ ذخیره شده است، تمرکز کنیم. ویژگی‌های موجود در هر یک از رکوردهای لاگ در جدول ۳-۵ آمده است.

<sup>۱</sup> Indexing

<sup>۲</sup> Systematic

<sup>۳</sup> Field

**جدول ۳-۵- ویژگی‌های موجود در رکوردهای لاگ**

| ویژگی      | نام اختصاری | توضیح                                                                          |
|------------|-------------|--------------------------------------------------------------------------------|
| شناسه      | ID          | یک شناسه یکتا برای این رکورد اطلاعاتی است.                                     |
| جستجو      | Info        | متنی است که در فایل لاگ ذخیره شده و شامل محتوای جستجو و تعداد پاسخ است.        |
| IP         | IP          | نشانی دهنده IP دستگاهی است که جستجو از طریق آن دستگاه به پایگاه ارسال شده است. |
| زمان جستجو | Created_at  | تاریخ و ساعت جستجوی کاربر را نشان می‌دهد.                                      |
| شناسه نشست | Session-ID  | شناسه نشستی است که کاربر در آن نشست، جستجوی خود را انجام داده است.             |

در مطالعات پیشینی که در با داده‌های لاگ گنج صورت پذیرفته بود، بخش قابل توجهی از جستجوها (در پایگاه قدیم) مربوط به ربات‌های جستجوگر بود (ربيعی، ۱۳۹۸)، ولی از آنجا که لاگ‌های دریافتی کنونی مربوط به پایگاه جدید است که اجازه جستجو به ربات‌های جستجوگر را نمی‌دهد چنین مشکلی وجود ندارد. برای اطمینان از این موضوع با گروه‌بندی آدرس IP کاربران و مرتب کردن آنها بر اساس فراوانی استفاده از آدرس و جستجوی مالک آن آدرس امکان تفکیک کاربران واقعی از موتورهای جستجو فراهم می‌شود که در مورد داده‌های این پژوهش نیز این کار انجام شد، اما نتیجه حاصل نشان داد که جستجوها عمدتاً از داخل کشور و شرکت‌های ارائه خدمات اینترنتی داخلی بوده است و موتورهای جستجو در بین IP‌های بر تکرار مشاهده نشد.

### ۳-۲-۳. یردازش متن جستجوها

متن جستجویی که در پایگاه اطلاعاتی ذخیره شده است به درستی محتوای جستجوی کاربر را از دیگر اقلام اطلاعاتی که در فایل لاگ ذخیره شده تفکیک نکرده است. به عبارت دیگر فیلدهای مختلفی با یکدیگر ترکیب شده و به صورت یک متن در فیلد Info ذخیره شده‌اند. نمونه‌ای از داده‌های ذخیره شده و شکل ذخیره آنها در جدول ۳-۶ آمده است.

**جدول ۳-۶- نمونه‌ای از محتوای ذخیره شده در فایل Log**

| Created_at          | IP             | Info                                                                                                                                | ID     |
|---------------------|----------------|-------------------------------------------------------------------------------------------------------------------------------------|--------|
| 03/03/2019<br>08:19 | 192.168.12.250 | "\}result_count\"=>14, \"search_title مطالعات یادگیری\"\\\"<=\"\  اجتماعى\\\\\"در عنوان\"}                                          | 38636  |
| 03/03/2019<br>08:20 | 192.168.12.250 | {\"result_count\"=>546, \"search_title\"=>\ \"\\\\\" مجتمع \\\\\"در عنوان\"\\\\\"مسكونى\"}                                          | 153423 |
| 03/03/2019<br>08:20 | 192.168.12.250 | {\"result_count\"=>137, \"search_title\"=>\ \"\\\\\" شبکه اجتماعى\"\\\\\"کارشناسی ارشد \\\\\"محدود به مقطع برابر \\\\\"داده کاوی\"} | 1      |
| 03/03/2019<br>08:20 | 192.168.12.250 | {\"result_count\"=>88, \"search_title\"=>\ \"\\\\\" در \\\\\"پارکینگ\"\\\\\"کلیدواژه\"}                                             | 153029 |

همان‌گونه که در جدول ۳-۶ مشخص شده است، محتوای اصلی جستجو به همراه عبارت‌های کنترلی دیگر در فیلد Info ذخیره شده است. از آنجا که وجه تمایزی بین متن جستجو و عبارت‌های کنترلی در شکل ذخیره‌سازی آنها در این فیلد مشاهده نمی‌شود، امکان جداسازی آنها به دستورات پایه‌ای، ویرایش متن میسر نیست. از این‌رو عبارت‌های کنترلی شناسایی و از متن جستجو حذف شد. این عبارت‌ها در جدول ۳-۷ آمده است.

**جدول ۳-۷- عبارات‌های کنت‌بی موجود در متن ذخیره شده در فیلد Info لاگ**

| عنوان            | محتوای کنترلی              | عنوان          | محتوای کنترلی |
|------------------|----------------------------|----------------|---------------|
| AbstractLabel    | چکیده                      | InTitleLabel   | در عنوان      |
| AllLabel         | همه                        | OrgLabel       | سازمان برابر  |
| AuthorLabel      | در پدیدآور                 | EqualLabel     | برابر         |
| DegreeEqualLabel | مقطع برابر                 | FieldLabel     | رشته          |
| DegreeMSCLabel   | کارشناسی ارشد              | InKeywordLabel | در کلیدواژه   |
| DegreePhDLabel   | دکتری تخصصی PhD            | SubjectLabel   | موضوع         |
| DegreePhD2Label  | دکترای حرفه‌ای عمومی       | TitleLabel     | عنوان         |
| FullTextLabel    | با وضعیت فایل تمام‌متن     | InSubjectLabel | موضوع برابر   |
| FirstPageLabel   | با وضعیت فایل چند برگ نخست | LimitedToLabel | محدود به      |

پس از حذف موارد ذکر شده در جدول ۳-۷ مشخص شد که بیش از ۲۶ درصد از رکوردهای ذخیره شده دارای این متون کنترلی بودند. همچنین تعداد پاسخ‌هایی که توسط موتور جستجو برگردانده شده است، استخراج شد و در نهایت جدول زیر برای فایل لاگ تشکیل شد. همچنین متغیر تعداد پاسخ نیز که در این فیلد ذخیره می‌شد از آن استخراج و به عنوان فیلد مجزایی ذخیره شد. همچنین تاریخ انجام جستجو و ساعت انجام آن که به صورت یکجا در فیلد Created-at ذخیره می‌شد، از یکدیگر مجزا شد. تاریخ ذخیره شده در فایل به صورت تاریخ میلادی بود که برای سادگی تحلیل نتایج، به تاریخ شمسی تبدیل شد. نمونه تغییر یافته جدول ۳-۷ در جدول ۳-۸ آمده است.

جدول ۳-۸- نمونه‌ای از محتوای نهایی (پیش پردازش شده) در فایل Log

| Date       | IP             | AnsCount | Content                 | ID     |
|------------|----------------|----------|-------------------------|--------|
| ۱۳۹۷/۱۲/۱۲ | 192.168.12.250 | ۱۴       | یادگیری مطالعات اجتماعی | 38636  |
| ۱۳۹۷/۱۲/۱۲ | 192.168.12.250 | ۵۴۶      | مجتمع مسکونی            | 153423 |

|            |                |      |                         |        |
|------------|----------------|------|-------------------------|--------|
| ۱۳۹۷/۱۲/۱۲ | 192.168.12.250 | ۱۳۷  | شبکه اجتماعی، داده کاوی | 1      |
| ۱۳۹۷/۱۲/۱۲ | 192.168.12.250 | ۸۸   | پارکینگ                 | 153029 |
| ۱۳۹۷/۱۲/۱۲ | 192.168.12.250 | ۱۰۵۵ | مدیریت آموزشی           | 1      |

### ۴\_۲\_۳. شناسایی جستجوهای مرتبط با حوزه انرژی

برای شناسایی جستجوهای مرتبط با حوزه انرژی از کلمات کلیدی پرتکرار پارساهای مرتبط با انرژی استفاده شده است. در این فرایند ابتدا ۲۰۰ کلمه کلیدی پرتکرار شناسایی شده و سپس کلماتی مانند «بهینه‌سازی»، «شبیه‌سازی»، «الگوریتم ژنتیک»، «ایران»، «مسیریابی»، «شبکه عصبی»، «مدل‌سازی» و ... را که کلمات کلیدی عمومی هستند و مختص حوزه انرژی نیستند را از بین کلمات پرتکرار حذف کردیم که نهایتاً ۱۱۶ کلمه کلیدی خاص حوزه انرژی بدست آمد و نهایتاً جستجوهای که این کلمات در متن جستجوی آنها وجود داشت به عنوان جستجوهای مرتبط با حوزه انرژی در نظر گرفته شد که شامل ۴۴۳۸۹۰ رکورد است که تقریباً معادل ۲ درصد (۰,۰۱۸) از تمامی جستجوهای پایگاه را شامل می‌شود.

### ۳\_۳. تحلیل کتابشناختی پارساهای عرضه شده در حوزه انرژی

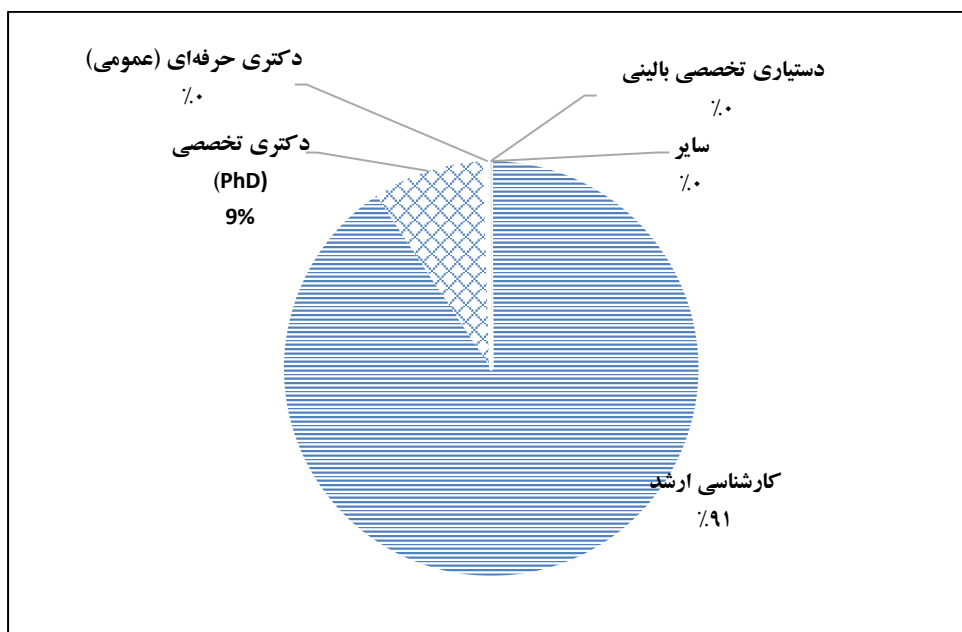
همان‌طور که در (بخش ۳\_۲\_۸) بیان شد، تعداد ۲۵،۹۵۳ رکورد از پارساهای موجود در پایگاه اطلاعاتی به عنوان منابع مرتبط با حوزه انرژی شناخته شدند. در ادامه تحلیل کتابشناختی این منابع ارائه شده است.

#### ۳\_۳\_۱. تحلیل مقطع تحصیلی عرضه‌کنندگان منابع حوزه انرژی

پارساها موجود در پایگاه گنج در مقاطع تحصیلی مختلفی تولید می‌شوند، این مقاطع عبارتند از:

- کارشناسی ارشد
- دکترای تخصصی (Ph.D.)
- دکترای حرفه‌ای (عمومی)
- دستیار تخصصی بالینی
- سایر

منظور از مقطع سایر، شامل مواردی است که مقطع تحصیلی نامشخص است یا اینکه در حالات خاصی منبع مورد نظر مربوط به دانشجویان کارشناسی در مقطع لیسانس و غیره است. شکل ۳-۳ سهم هر یک از این مقاطع را در پارساهای عرضه شده در حوزه انرژی نشان می‌دهد.

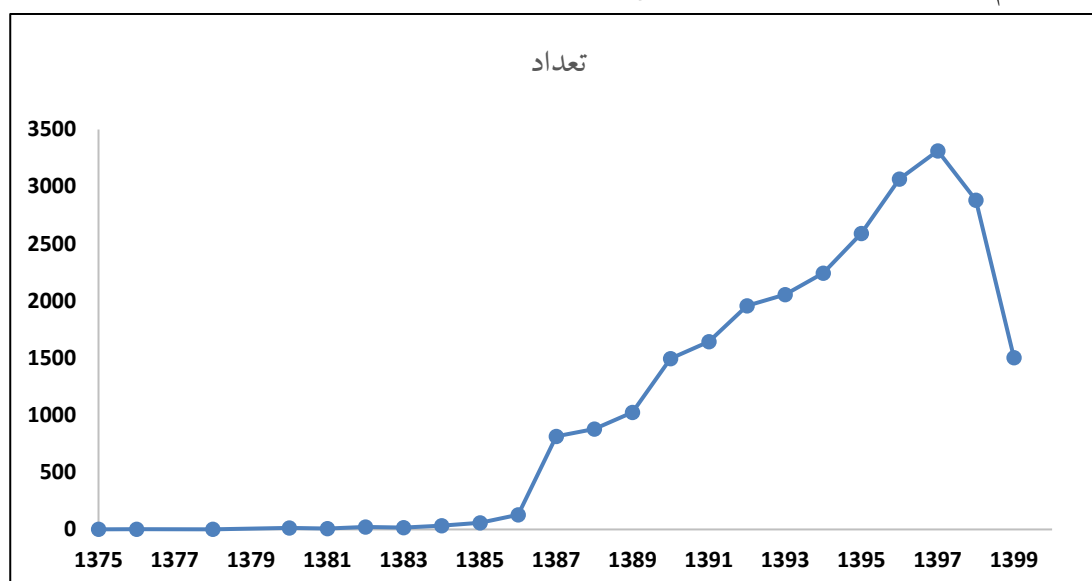


شکل ۳-۳- تحلیل مقطع تحصیلی منابع عرضه شده حوزه انرژی

همانطور که در این شکل مشخص است بیش از ۹۰ درصد از پژوهش‌های عرضه شده در حوزه انرژی مربوط به دانش‌آموختگان مقطع کارشناسی ارشد است و نزدیک به ۹ درصد منابع توسط دانش‌آموختگان مقطع دکترای تخصصی تامین شده است. دیگر مقاطع تحصیلی سهمی کمتر از یک درصد را به خود اختصاص داده‌اند.

### ۳\_۲\_ تحلیل زمانی منابع عرضه‌شده در حوزه انرژی

روند انجام پارساها در حوزه انرژی در شکل ۳-۴ آمده است.



شکل ۳-۴- روند زمانی پارساهای عرضه‌شده در حوزه انرژی

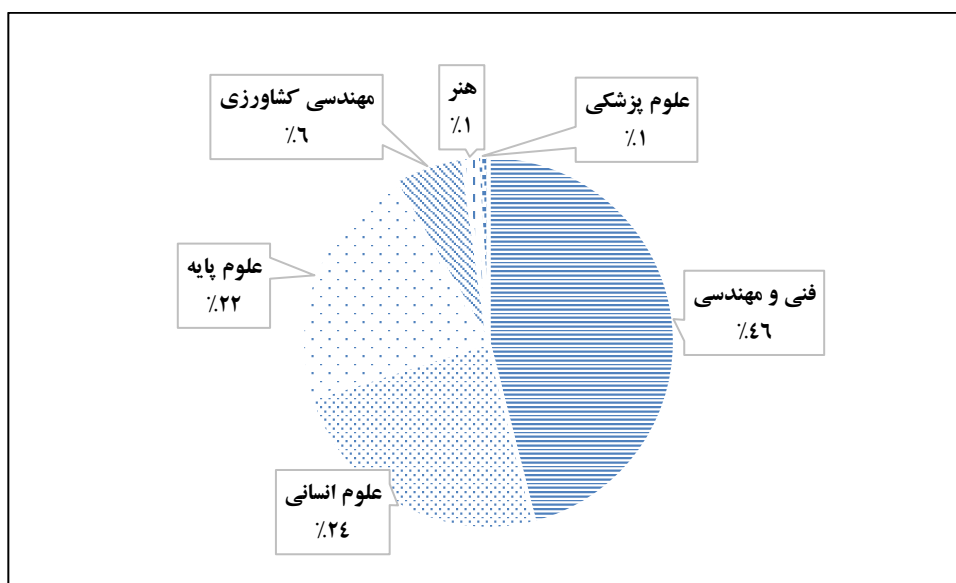
داده‌های موجود در منابع عرضه شده مربوط به پارساهای سال ۱۳۰۰ تا ۱۳۹۹ است، با این حال با توجه به اینکه تنها در حدود ۱ درصد (۲۶ مورد) از این منابع مربوط به سالهای قبل از ۱۳۷۵ است، در شکل ۳-۴، روند عرضه پارساهای این حوزه مربوط به ۲۵ سال گذشته ارائه شده است. همانطور که در این نمودار مشخص است عرضه منابع حوزه انرژی از سال ۱۳۸۷ به بعد رشد فزاینده‌ای داشته است و در سالهای اخیر سالانه بیش از ۳۰۰۰ پایان‌نامه و رساله مرتبط با حوزه انرژی عرضه می‌شود که نشان می‌دهد این حوزه به شدت مورد توجه پژوهشگران است. لازم به ذکر است که از آنجا که منابع دو سال اخیر هنوز به صورت کامل در پایگاه ثبت نشده است، نمودار آن در دو سال اخیر شیب منفی را نشان می‌دهد که انتظار می‌رود با تکمیل اطلاعات سال‌های اخیر این رشد ادامه داشته باشد.

### ۳\_۳. تحلیل گروه‌های آموزشی مرتبط با حوزه انرژی

گروه‌های ۶ گانه آموزشی دانشجوی تحصیلات تکمیلی، یکی از ویژگی‌هایی است که در اطلاعات مربوط به پارساها موجود است. این گروه‌ها عبارت‌اند از:

- گروه فنی و مهندسی
- گروه علوم انسانی
- گروه علوم پایه
- گروه علوم پزشکی
- گروه مهندسی کشاورزی
- گروه هنر

متأسفانه تقریباً در نیمی از رکوردهای مستخرج پایگاه گنج، فیلد مربوط به گروه آموزشی نامشخص است. با این حال، شکل ۳-۵ سهم هر یک از گروه‌های آموزشی را در تامین منابع عرضه شده در حوزه انرژی در بین رکوردهایی که فیلد گروه آموزشی آنها مشخص است را نشان می‌دهد.



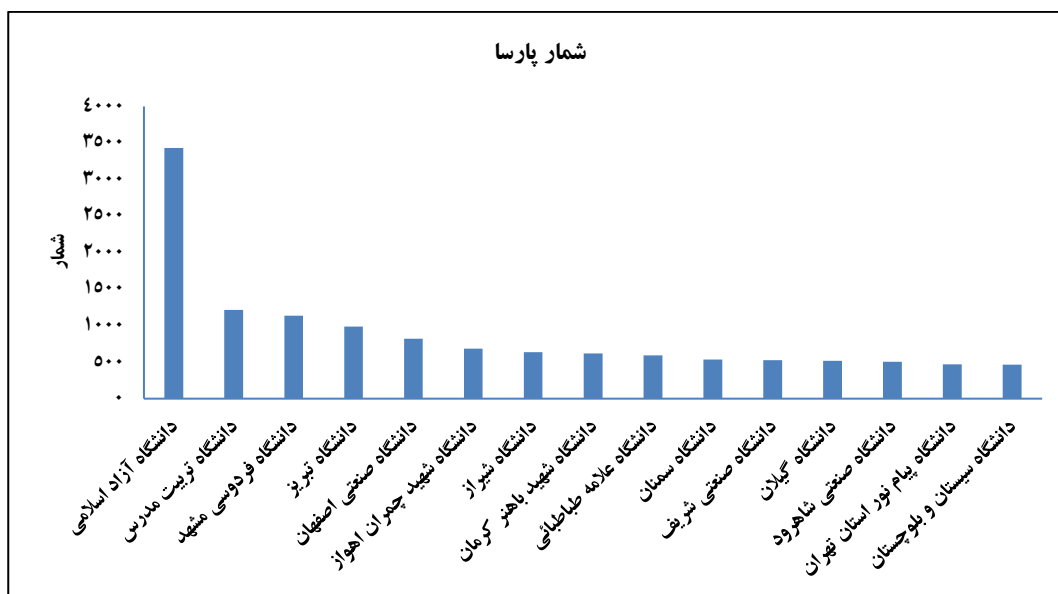
شکل ۳-۵- سهم هر یک از گروه‌های آموزشی در تامین منابع عرضه شده

همانطور که انتظار می‌رفت موضوع انرژی بیش از گروه‌های آموزشی دیگر مربوط به گروه فنی و مهندسی است و نزدیک به نیمی از منابع در این گروه آموزشی تولید شده است. اما گروه علوم انسانی نیز تقریباً یک چهارم (۲۴٪) از پارساهای این حوزه را در اختیار خود دارد و دیگر گروه‌های آموزشی نیز در این حوزه سهم هستند. همین موضوع نشان می‌دهد که حوزه انرژی یک حوزه کاملاً میان‌رشته‌ای است و ابعاد مختلف این حوزه در رشته‌های دانشگاهی گوناگون مورد توجه پژوهشگران است.

### ۳-۳-۴. تحلیل دانشگاه‌ها و مراکز پژوهشی عرضه‌کننده پارساهای حوزه انرژی

دانشگاه‌ها و موسسات پژوهشی مختلفی موظف به ثبت پارساهای دانشجویان خود در پایگاه گنج هستند با این وجود برخی از دانشگاه‌ها تا سالهای اخیر به دلایل گوناگون منابع خود را در این پایگاه به ثبت نرسانده‌اند. بررسی منابع عرضه شده در حوزه انرژی نشان می‌دهد که با تجمیع تمامی واحدهای دانشگاه آزاد اسلامی و در نظر گرفتن این دانشگاه به عنوان یک مجموعه، ۳۷۸ سازمان در حوزه انرژی، پارسا تولید و در پایگاه گنج به ثبت رسانده‌اند.

شکل ۳-۶ فهرست ۱۵ دانشگاه با بیشترین تعداد پارسا در حوزه انرژی را نشان می‌دهد.



شکل ۳-۶- فهرست ۱۵ دانشگاه با بیشترین تولید پارسا در حوزه انرژی

همان‌گونه که در شکل ۳-۶ مشخص است، مجموعه واحدهای دانشگاه آزاد اسلامی با بیش از ۳۴۰۰ پارسا، بیشترین منابع حوزه انرژی را ارائه نموده‌اند؛ این مجموعه در داده‌های گردآوری شده این حوزه شامل ۶۵ واحد دانشگاهی است. دانشگاه تربیت مدرس با بیش از ۱۲۰۰ رکورد و دانشگاه فردوسی مشهد با بیش از ۱۱۳۰ رکورد رتبه‌های بعدی را از نظر تعداد منابع این حوزه به خود اختصاص داده‌اند.

### ۳\_۳\_۵. تحلیل کلمات کلیدی پر کاربرد در منابع حوزه انرژی

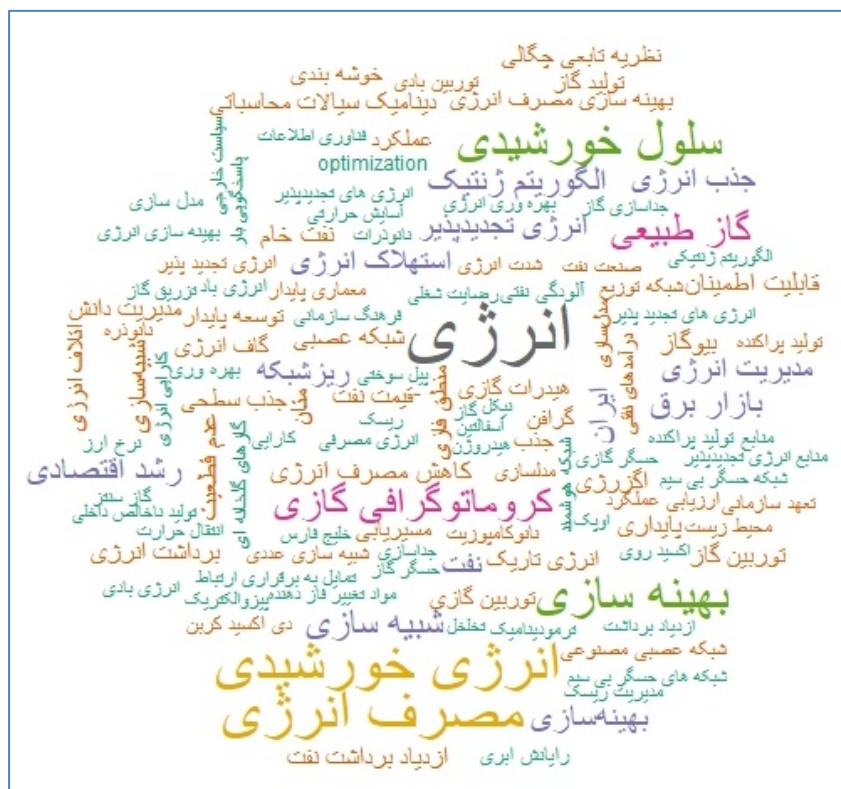
تحلیل کلمات کلیدی پر کاربرد در حوزه انرژی موجب شناسایی این عبارات شده است. جدول ۳-۹ فهرست ۳۰ کلمه کلیدی پر تکرار حوزه انرژی را نشان می‌دهد.

جدول ۹-۳- فهرست ۳۰ کلمه کلیدی پرتکرار در منابع عرضه شده حوزه انرژی

| ردیف | کلمه کلیدی        | تعداد تکرار | ردیف | کلمه کلیدی              | تعداد تکرار |
|------|-------------------|-------------|------|-------------------------|-------------|
| ۱    | انرژی             | ۴۵۸         | ۱۶   | ریز شبکه                | ۱۵۲         |
| ۲    | انرژی خورشیدی     | ۴۰۸         | ۱۷   | جذب انرژی               | ۱۴۸         |
| ۳    | مصرف انرژی        | ۳۹۶         | ۱۸   | انرژی تجدیدپذیر         | ۱۴۳         |
| ۴    | سلول خورشیدی      | ۳۱۶         | ۱۹   | مدیریت انرژی            | ۱۴۲         |
| ۵    | بهینه سازی        | ۳۱۳         | ۲۰   | کاهش مصرف انرژی         | ۱۳۲         |
| ۶    | کروماتوگرافی گازی | ۲۳۰         | ۲۱   | عدم قطعیت               | ۱۲۸         |
| ۷    | گاز طبیعی         | ۲۱۹         | ۲۲   | قابلیت اطمینان          | ۱۲۸         |
| ۸    | شبیه سازی         | ۲۰۱         | ۲۳   | بیو گاز                 | ۱۲۶         |
| ۹    | الگوریتم ژنتیک    | ۱۸۶         | ۲۴   | نفت خام                 | ۱۲۵         |
| ۱۰   | رشد اقتصادی       | ۱۸۴         | ۲۵   | شبیه سازی               | ۱۲۴         |
| ۱۱   | بازار برق         | ۱۸۰         | ۲۶   | شبکه عصبی               | ۱۱۷         |
| ۱۲   | ایران             | ۱۷۳         | ۲۷   | اگزرژی                  | ۱۱۶         |
| ۱۳   | بهینه سازی        | ۱۷۱         | ۲۸   | برداشت انرژی            | ۱۱۶         |
| ۱۴   | نفت               | ۱۵۶         | ۲۹   | دینامیک سیالات محاسباتی | ۱۱۶         |
| ۱۵   | استهلاک انرژی     | ۱۵۳         | ۳۰   | مدیریت دانش             | ۱۱۰         |

شکل ۷-۳ ابر کلمات<sup>۱</sup> پر کاربرد را با استفاده از کلمات کلیدی منابع عرضه شده حوزه انرژی نشان می‌دهد. در این شکل، بزرگی متن نشان‌دهنده فراوانی تکرار کلیدواژه‌هاست و کلیدواژه‌های با بیش از ۵۰ بار تکرار به تصویر کشیده شده است.

<sup>۱</sup> WordCloud



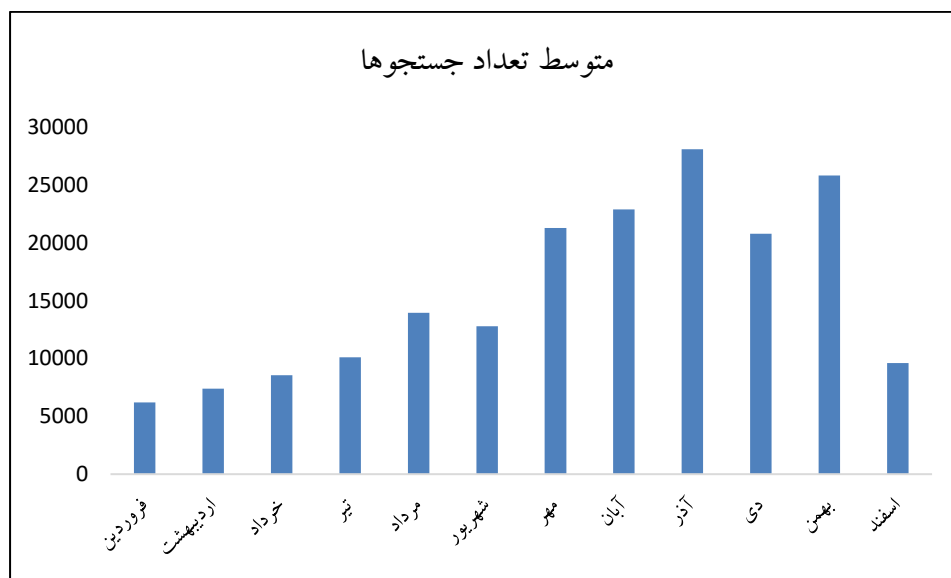
شکل ۳-۷-۱۰ کلمات پر کاربرد با استفاده از کلیدواژه‌های پارساهای عرضه شده

### ۳\_۴. تحلیل کتابشناختی جستجوهای انجام شده در حوزه انرژی

همان‌طور که در بخش (۳\_۲\_۸\_۳\_۴) بیان شد، تعداد ۴۴۳۸۹۰ رکورد از جستجوهای انجام شده در پایگاه به عنوان جستجوهای حوزه انرژی شناسایی شدند. در ادامه تحلیل کتاب‌شناختی این جستجوها ارائه شده است.

### ۳\_۴\_۱. تحلیل زمانی تقاضاهای حوزه انرژی

جستجوهای استخراج شده مرتبط با حوزه انرژی مربوط به بازه زمانی اسفند ۱۳۹۶ تا شهریور ۱۳۹۹ است. از آنجا که تنها مدت زمان ۲ سال و نیم از جستجوها را در اختیار داریم، تحلیل روند سالانه جستجوها نتیجه بخش نخواهد بود، اما میزان جستجوهای انجام شده در ماههای مختلف سال می تواند حاوی اطلاعات مفیدی باشد. برای آنکه روند جستجوها در ماههای مختلف بدست آید، تعداد جستجوهای انجام شده در هر ماه به تعداد دفعاتی که آن ماه در بازه زمانی داده های موجود تکرار شده، تقسیم شده است تا متوسط تعداد جستجوها در هر ماه مشخص شود. شکل ۳-۸ متوسط جستجوهای انجام شده در هر ماه را نشان می دهد.

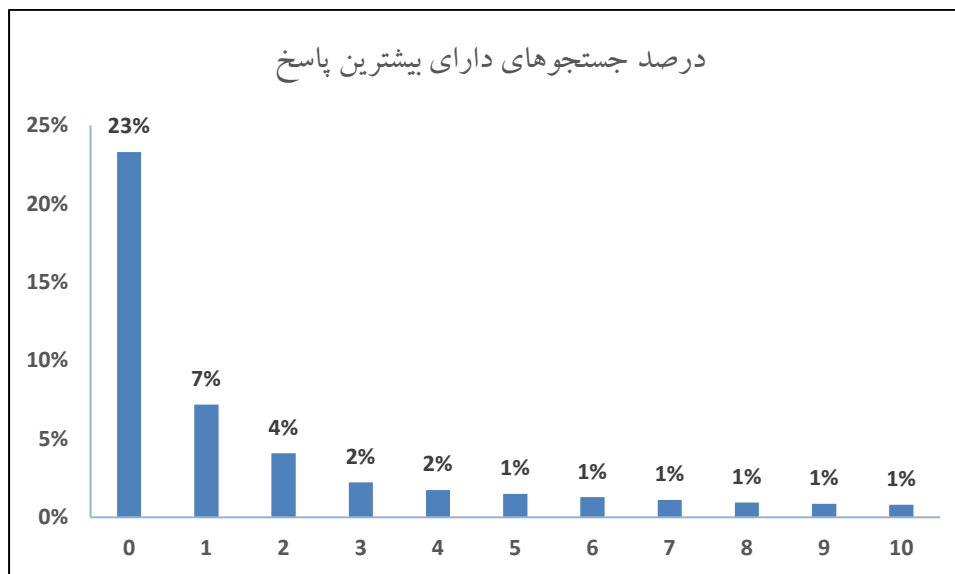


شکل ۳-۸- متوسط تعداد جستجوها در ماه‌های مختلف سال

روند جستجوها در ماه‌های مختلف سال نشان می‌دهد که در نیمه اول سال تحصیلی (ماه‌های فصل پاییز و اوایل زمستان) میزان جستجوها بیش از سایر ماه‌های سال است. این نتیجه در پژوهش ربیعی نیز بدست آمده بود که دلیل آن به دلیل الگوی رفتاری کاربران پایگاه اطلاعاتی است که عمدتاً دانشجویان تحصیلات تکمیلی هستند (ربیعی، ۱۳۹۸).

### ۳-۴-۲. بررسی تعداد پاسخ‌های دریافتی کاربر

همان‌طور که پیشتر در جدول ۳-۸ ذکر شد، یکی از قلم‌های اطلاعاتی که در فیلد Info ذخیره می‌شود، تعداد پاسخ‌هایی است که نشان‌دهنده تعداد رکوردی است که موتور جستجو به عنوان نتیجه به کاربر ارائه کرده است. تعداد پاسخ ارسالی پایگاه به کاربر تا میزان زیادی به تنظیمات موتور جستجوی پایگاه، نوع طراحی پایگاه داده، سیاست‌های ارائه اطلاعات، تنظیمات پروفایل کاربر برای جستجو، شکل و نحوه برنامه‌نویسی و عواملی از این دست مرتبط است، به عنوان مثال محدودیت‌های سخت‌افزاری موجب شده بود تا برای مدتی پایگاه فقط ۶ واژه اول متن جستجو را بازایی کند که این سیاست پس از بهبود منابع سخت‌افزاری تغییر یافته است. همچنین زمانی که طول متن جستجو از میزانی افزایش یابد به نظر موتور جستجو بجای جستجوی دقیق عبارت وارد شده، هر بخش از عبارت را که در پایگاه وجود داشته باشد به عنوان پاسخ برمی‌گرداند و بررسی روی عبارت‌های جستجوی طولانی نشان می‌دهد که در بیشتر مواقع موتور جستجو تمامی رکوردهای موجود در پایگاه را برای این عبارت‌ها به عنوان پاسخ ارائه داده است. از این رو مدنظر قرار دادن این فیلد اطلاعاتی، برای مواردی که تعداد پاسخ زیادی ارائه شده است، غیر قابل اتکا است. شکل ۳-۹ درصد جستجوهای دارای بیشترین تعداد پاسخ را به نسبت دیگر جستجوهای در حوزه انرژی نشان می‌دهد.



شکل ۳-۹- درصد جستجوهای دارای بیشترین تعداد پاسخ

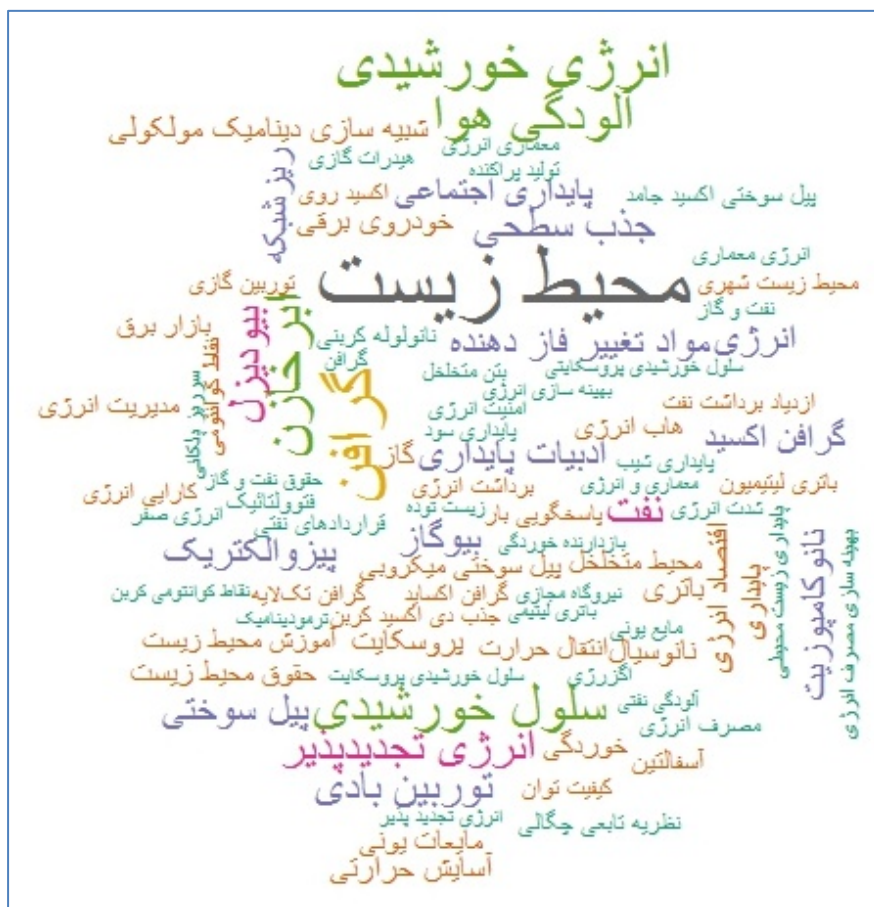
### ۳\_۴\_۳. تحلیل کلمات کلیدی پر کاربرد در جستجوهای حوزه انرژی

تحلیل عبارت‌هایی که بیشترین میزان جستجو را در حوزه انرژی موجب شناسایی این عبارت‌ها شده است. جدول ۳-۱۰ فهرست ۳۰ کلمه کلیدی پر تکرار جستجوها را نشان می‌دهد.

جدول ۱۰-۳- فهرست ۳۰ کلمه کلیدی پر تکرار در منابع جستجو شده حوزه انرژی

| ردیف | کلمه کلیدی      | تعداد تکرار | ردیف | کلمه کلیدی               | تعداد تکرار |
|------|-----------------|-------------|------|--------------------------|-------------|
| ۱    | محیط زیست       | ۲۷۰۸        | ۱۶   | نانو کامپوزیت            | ۸۳۸         |
| ۲    | گرافن           | ۱۹۲۶        | ۱۷   | بیوگاز                   | ۸۲۵         |
| ۳    | ابرخازن         | ۱۵۲۷        | ۱۸   | ریزشبکه                  | ۸۰۹         |
| ۴    | انرژی خورشیدی   | ۱۵۰۰        | ۱۹   | مواد تغییر فاز دهنده     | ۷۷۷         |
| ۵    | آلودگی هوا      | ۱۴۶۶        | ۲۰   | پایداری اجتماعی          | ۷۱۰         |
| ۶    | سلول خورشیدی    | ۱۴۵۱        | ۲۱   | گرافن اکسید              | ۶۸۳         |
| ۷    | نفت             | ۱۱۰۳        | ۲۲   | نانوسیال                 | ۶۷۶         |
| ۸    | انرژی تجدیدپذیر | ۱۰۷۷        | ۲۳   | اقتصاد انرژی             | ۶۴۳         |
| ۹    | بیودیزل         | ۱۰۷۱        | ۲۴   | پروسکایت                 | ۶۳۵         |
| ۱۰   | جذب سطحی        | ۹۹۶         | ۲۵   | آسایش حرارتی             | ۶۲۳         |
| ۱۱   | توربین بادی     | ۹۹۱         | ۲۶   | خودروی برقی              | ۶۲۲         |
| ۱۲   | انرژی           | ۹۳۵         | ۲۷   | باتری                    | ۶۱۰         |
| ۱۳   | پیل سوختی       | ۹۳۴         | ۲۸   | شبه سازی دینامیک مولکولی | ۵۸۹         |
| ۱۴   | ادبیات پایداری  | ۹۱۰         | ۲۹   | گاز                      | ۵۶۷         |
| ۱۵   | پیزوالکتریک     | ۸۶۱         | ۳۰   | پایداری                  | ۵۵۸         |

شکل ۱۰-۳ ابر کلمات پر کاربرد را با استفاده از کلمات کلیدی منابع جستجو شده حوزه انرژی نشان می‌دهد. در این شکل، بزرگی متن نشان‌دهنده فراوانی تکرار کلمات کلیدی است و عبارت‌های با بیش از ۱۰۰ بار تکرار به تصویر کشیده شده است.



شکل ۳-۱۰-۱- ابر کلمات پر کاربرد با استفاده از کلیدواژه‌های پارساهای عرضه‌شده

### ۳-۵. جمع‌بندی و نتیجه‌گیری

در این فصل در مرحله اول به دنبال شناسایی متون مرتبط با حوزه انرژی و تعیین مرز این حوزه از دیگر منابع بودیم تا بتوانیم بدون وابستگی به نظر خبرگان حوزه موضوعی مرز این حوزه را از دیگر منابع موجود شناسایی کرده و منابع مرتبط با آن را استخراج کنیم. برای این منظور ۸ گام اصلی در فرایند استخراج پارساهای مرتبط با حوزه انرژی پیموده شد و در این گام‌ها از کلمات کلیدی شناسایی شده در منابع اولیه حوزه انرژی و نیز اصطلاح‌نامه‌های موضوعی کمک گرفته شد. پس از آن برای استخراج جستجوهای مرتبط با حوزه انرژی و به منظور ایجاد پایگاه منابع تقاضا شده این حوزه، سه گام اصلی طی شد. در نهایت تعداد ۲۵،۹۵۳ رکورد از پارساهای موجود در پایگاه اطلاعاتی به‌عنوان منابع مرتبط با حوزه انرژی در سمت عرضه پژوهش شناخته شد و تعداد ۲۴،۲۰۲،۷۷۴ رکورد از جستجوهای انجام شده در پایگاه گنج، مورد ارزیابی قرار گرفت و در نهایت ۴۴۳،۸۹۰ از آنها سمت تقاضای پژوهش را تشکیل دادند.

تحلیل کتابشناختی منابع عرضه شده در حوزه انرژی موید میان‌رشته‌ای بود این حوزه است، همچنان بررسی‌ها نشان می‌دهد بیش از ۹۰ درصد منابع این حوزه توسط دانش‌آموختگان مقطع کارشناسی ارشد تولید شده است. این پژوهش‌ها از سال ۱۳۸۷ با سرعت فزاینده‌ای تولید و به جامعه پژوهشی کشور عرضه شده است و دانشگاه‌هایی مانند دانشگاه آزاد اسلامی، دانشگاه تربیت مدرس و دانشگاه فردوسی مشهد بیشترین پارساهای این حوزه را عرضه کرده‌اند. روند جستجوهای انجام‌شده در حوزه انرژی نشان می‌دهد که بیشترین میزان جستجوها در نیمه اول سال تحصیلی انجام می‌شود که با توجه به رفتار مخاطبان پایگاه اطلاعاتی دور از انتظار نیست.

در فصل بعد به شناسایی و تحلیل موضوعات پژوهشی حوزه انرژی در هر دو بعد عرضه و تقاضای پژوهش خواهیم پرداخت.

## **فصل چهارم**

**تحلیل موضوعی عرضه و  
تقاضای پژوهش‌های حوزه  
انرژی**

#### ۴\_۱. مقدمه

امروزه حوزه‌های دانشی شامل مفاهیم گوناگونی است که با پیشرفت علم و دانش در ابعاد گوناگون، این مفاهیم با یکدیگر پیوند یافته و مفاهیم و موضوعات جدید را پدید می‌آورند، گاهی به موضوع در طول چرخه عمر خود به موضوعات زیرمجموعه خود تجزیه می‌شود و هر یک از موضوعات زیرمجموعه به عنوان یک موضوع مستقل مسیر خود را می‌پیمایند. در برخی موارد نیز با پدید آمدن رویکردهای میان‌رشته‌ای، موضوعاتی که تا مدتی پیش از یکدیگر جدا بودند با یکدیگر ترکیب شده و موضوع واحدی را پدید می‌آورند که می‌تواند ماهیتی جدا یا وابسته به موضوعات تشکیل‌دهنده خود را داشته باشد.

شناسایی موضوعات موجود در یک حوزه پژوهشی و تحلیل آن‌ها می‌تواند علاوه بر تبیین و تشریح چرخه عمر این موضوعات و تحلیل روندها و چگونگی رشد و نمو و گاه زوال آنها، مسیرهای آتی این موضوعات را نیز پیش‌بینی کند. سیاست‌گذار پژوهشی با آگاهی از تغییراتی که پژوهش‌های یک موضوع در یک حوزه پژوهشی داشته است، می‌تواند مسیر آتی آن را ترسیم کرده و با انتخاب تصمیمات درست در راستای اهداف مورد نظر خود آن‌ها را هدایت کند.

در این فصل موضوعات پژوهشی حوزه انرژی شناسایی و استخراج خواهند شد و عرضه و تقاضای پژوهش در هر یک از این موضوعات تحلیل شده و در نهایت تحلیل فاصله بین عرضه و تقاضا در پژوهش‌های حوزه انرژی با توجه به معیارهای مدنظر در این مطالعه شناسایی و ارائه خواهد شد.

#### ۲-۴. مدل‌سازی موضوعی منابع عرضه‌شده حوزه انرژی

همان‌طور که پیش‌تر در (بخش ۲-۷) بیان شد، از طریق مدل‌سازی موضوعی و استفاده از الگوریتم LDA می‌توان موضوعاتی را که حاوی مستندات مشابه هم هستند شناسایی کرد. شناسایی و دسته‌بندی منابع بر اساس کلیدواژه‌های آنها انجام می‌شود اما از آنجا که برخی از کلیدواژه‌ها ممکن است دارای خطای نگارشی باشند که موجب ایجاد خطا در نتیجه دسته‌بندی خواهد شد، لازم است تا ابتدا بانکی از کلمات کلیدی که به عنوان نماینده در مدل‌سازی موضوعی مورد استفاده قرار می‌گیرند تعیین شود. برای این منظور از کلیدواژه‌هایی که لااقل دوبار در مجموعه داده‌ها تکرار شده‌اند استفاده شده است، با اینکه تعداد کلیدواژه‌های منحصربفرد شناسایی شده در منابع عرضه شده حوزه انرژی نزدیک به ۶۰ هزار عبارت بود، اما تنها ۱۲،۵۰۵ مورد از آنها لااقل دوبار تکرار شده بودند. از این رو این عبارت‌ها در مدل‌سازی موضوعی به عنوان کلمات شاخص در نظر گرفته خواهد شد و براساس آنها مدل‌سازی انجام خواهد شد. پس از تولید ماتریس مستند-واژه با استفاده از واژه‌های دارای بیش از یک‌بار تکرار، ۸۵ درصد از منابع شناسایی شده پوشش داده شد. این به این معناست که در ۸۵ درصد از منابع، دست کم یکی از کلیدواژه‌های شاخص وجود دارد.

پس از آنکه موضوعات منابع عرضه‌شده استخراج شد، در مرحله بعدی منابع جستجو شده حوزه انرژی به موضوعات شناسایی شده نگاشت خواهند شد و سپس موضوعات استخراج شده مورد بررسی و تحلیل قرار خواهد گرفت.

#### ۲-۴-۱. تعیین پارامترهای LDA

اولین پارامتر ورودی LDA، ماتریس سند-واژه است، ستون‌های این ماتریس شامل مستنداتی است که موضوعات مدنظر را در خود جای داده‌اند و سطرها این ستون شامل عبارت‌هایی است که می‌توانند بازنمای مستندات باشند و به عنوان عبارت‌های شاخص در گام قبلی شناسایی شد. ماتریس سند-واژه فراوانی تکرار عبارت‌ها در هر مستند را نشان می‌دهد.

پارامتر ورودی مهم دیگر LDA، مقدار  $k$  است که تعداد موضوعات مدنظر را نشان می‌دهد. روش‌های مختلفی برای تعیین تعداد موضوعات وجود دارد (Blei, 2012; Hall, Jurafsky, and Manning, 2008). مزیت انتخاب تعداد موضوعات زیاد، این است که تمامی حوزه‌های موضوعی پژوهش پوشش داده خواهد شد، از طرفی مزیت تعداد حوزه‌های محدود این است که خبرگان و سیاست‌گذاران درک بهتری از تعداد موضوعات محدود خواهند داشت و تفسیر و تحلیل موضوعات راحت‌تر خواهد بود (De Battisti, Ferrara, and Salini, 2015; M Rabiei, Hosseini-Motlagh, Haeri, and Minaei Bidgoli, 2021). از این رو انتخاب معیاری برای تعیین تعداد موضوعات مناسب

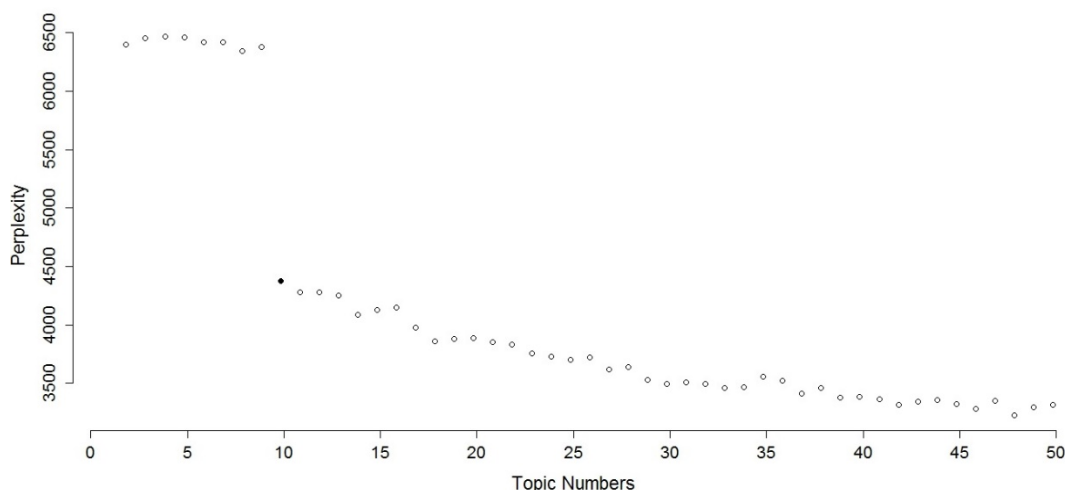
بسیار ضروری است. برای این منظور از معیار سرگشتگی (Perplexity) استفاده می‌شود. سرگشتگی یک معیار اندازه‌گیری در مدل‌های آماری است که برای بررسی میزان مناسب بودن یک توزیع احتمالی یا پیش‌بینی مدل احتمالی یک نمونه مورد استفاده قرار می‌گیرد و می‌توان از آن برای مقایسه مدل‌های احتمالی استفاده کرد. برای ارزیابی مدل‌سازی موضوعی از سرگشتگی اسناد مشاهده نشده استفاده می‌شود که تابعی نزولی از لگاریتم احتمال وقوع اسناد مشاهده نشده است و هر چه مقدار آن کمتر باشد نشان می‌دهد که مدل بهتر است و قابلیت تعمیم‌پذیری بالاتری دارد (Blei et al., 2003) و مقدار بالاتر سرگشتگی نشان‌دهنده این است که عبارت‌ها در زمان یادگیری مدل به موضوعات مناسبی تخصیص نیافته‌اند (De Battisti et al., 2015). برای انتخاب مقدار مناسب  $k$ ، مدل با استفاده از مقادیر مختلف  $k$  (۲ تا ۵۰) اجرا شد و مقدار سرگشتگی برای هر اجرا محاسبه شد. از آنجا که اجرای مدل‌سازی موضوعی روی تعداد زیاد رکوردها (نزدیک ۲۶ هزار رکورد) و با وجود بیش از ۱۲ هزار کلیدواژه که به عنوان ویژگی در نظر گرفته شده است، از نظر پردازشی بسیار پرهزینه خواهد بود. این کار با استفاده از الگوریتم‌های پردازش موازی انجام شد. این الگوریتم در شکل ۴-۱ آمده است.

```
cl = makeCluster(detectCores()-1) # one core will be free
clusterExport(cl=cl,c("Tdtm")) # pass the variables and functions to cores
clusterEvalQ(cl=cl, library(topicmodels))
st<-Sys.time()
perplx2 <- as.matrix(parLapply(cl,seq(2,50,1), function(x)
                                perplexity(LDA(Tdtm,k=x, method = "VEM"))) )
en<-Sys.time()
print(en-st)
stopCluster(cl)

Time difference of 1.081962 hours
```

#### شکل ۴-۱- پردازش موازی برای شناسایی برای تعیین میزان سرگشتی مدل

در کد بالا، ابتدا تعداد هسته‌های سرور تعیین شده است (سرور مورد استفاده ۸ هسته در اختیار داشت) و بجز یک هسته که برای مدیریت سیستم عامل و نیز مدیریت توزیع بار نیاز است، دیگر هسته‌ها در لیست  $cl$  قرار گرفته است. سپس پارامترهای مورد نیاز به تمامی اعضای این لیست ارسال شده است و در نهایت تابع  $parLapply$  برای از طریق توزیع بار بین تمامی هسته‌ها، مقادیر سرگشتگی را برای مدل‌های با تعداد ۲ تا ۵۰ موضوع شناسایی کرده است. این مقادیر در شکل ۴-۲ نمایش داده شده است.



شکل ۴-۲ - میزان سرگشتگی برای تعداد مختلف موضوع

همانطور که پیش‌تر بیان شد، هر چه میزان سرگشتگی کمتر باشد به معنی آن است که مدل به نحو بهتری توانسته است منابع را در موضوعات مختلف توزیع کند. انتخاب تعداد زیاد موضوعات تفسیر نتایج را دشوار خواهد کرد. با توجه به شکل ۴-۲ مشخص است که میزان سرگشتگی در  $k=10$  به شدت کم شده است و پس از آن میزان سرگشتگی با شیب ملایمی کم شده است، از این‌رو نقطه شکست نمودار که همان ۱۰ باشد به عنوان مقدار بهینه برای  $k$  تعیین شد.

#### ۴-۲-۲. تعیین سطح آستانه تشریح

هر موضوعی که با استفاده از LDA استخراج می‌شود، به میزانی به هر مستند مرتبط خواهد بود. به عبارت دیگر مستند  $p$  از طریق موضوع  $t$  با سطح  $\rho(p, t)$  تشریح می‌شود. از این‌رو نیاز به تعیین یک آستانه تشریح شدن برای مقادیر مختلف  $\rho(p, t)$  است تا در صورتی که مستندی با سطح بالاتر از آن آستانه توسط یک موضوع تشریح شد، بتوان آن مستند را به آن موضوع اختصاص داد. در صورتی که برای این آستانه، مقدار پایینی انتخاب شود، هر مستند توسط موضوعات زیادی پوشش داده خواهد شد و نتیجه این خواهد شد که اکثر موضوعات توسط اکثر مستندات تشریح خواهند شد و اگرچه عمده مستندات موجود در پیکره دست‌کم در یک موضوع قرار خواهند گرفت اما موضوع و مستند با ارتباطی بسیار ضعیف به یکدیگر متصل خواهند شد. در عوض در صورتی که این آستانه خیلی سخت‌گیرانه انتخاب شود، برخی از موضوعات با تعداد بسیار کمی از مستندات تشریح خواهند شد و بسیاری از مستندات نیز هیچ موضوعی را تشریح نخواهند کرد. از این‌رو سومین پارامتر مهم در مدل‌سازی

موضوعی تعیین آستانه تشریح<sup>۱</sup> ( $th_e$ ) است. بدین ترتیب زمانی می‌توانیم بگوییم موضوع  $t$  مستند  $d$  را تشریح می‌کند که  $\rho(p, t) \geq th_e$  باشد.

برای تعیین میزان مناسب آستانه تشریح دو معیار «پوشش پیکره<sup>۲</sup>» ( $CC$ ) و «کیفیت تشریح<sup>۳</sup>» ( $EQ$ ) معرفی شده است (De Battisti et al., 2015). پوشش پیکره نسبتی از مستندات موجود در یک پیکره که دست‌کم توسط یک موضوع تشریح شده باشند به تعداد کل مستندات موجود در پیکره است. از این‌رو برای پیکره  $P$  و مجموعه موضوعات استخراج شده  $T$  و آستانه تشریح  $th_e$ ، مجموعه  $E$  شامل تمامی مستندات تشریح شده است و مجموعه  $R$  شامل تمامی زوج‌های  $(p_i, t_j)$  خواهد بود که نشان می‌دهد مستند  $p_i$  توسط موضوع  $t_j$  تشریح شده است.

$$E = \{p_i | i \in \{1, 2, 3, \dots, |P|\}, \exists t_j, j \in \{1, 2, 3, \dots, |T|\}, \rho(p_i, t_j) \geq th_e\} \quad \text{فرمول ۱-۴}$$

$$R = \{(i, j) | i \in \{1, 2, 3, \dots, |P|\}, j \in \{1, 2, 3, \dots, |T|\}, \rho(p_i, t_j) \geq th_e\}$$

بر این اساس  $CC_{P,T}^{th_e}$  که میزان پوشش پیکره  $P$  توسط مجموعه موضوعات  $T$  و با آستانه تشریح  $th_e$  است به شکل زیر تعریف می‌شود:

فرمول ۲-۴

$$C_{P,T}^{th_e} = \frac{|E|}{|P|}$$

همچنین  $EQ_{P,T}^{th_e}$  نشان‌دهنده کیفیت تشریح پیکره  $P$  توسط مجموعه موضوعات  $T$  و با آستانه تشریح  $th_e$  است که این مقدار به شکل زیر به دست می‌آید:

فرمول ۳-۴

$$EQ_{P,T}^{th_e} = \frac{\sum_{(i,j) \in R} \rho(p_i, t_j)}{|R|}$$

از آنجا که هرچه میزان  $th_e$  کمتر باشد، میزان پوشش پیکره بیشتر خواهد بود به عبارتی در آستانه تشریح بسیار پایین، مستندات با سهولت بیشتری توسط موضوعات تشریح خواهند شد اما میزان کیفیت تشریح آن‌ها پایین خواهد بود و هرچه این میزان بیشتر شود کیفیت بهتر شده اما پوشش کمتری را شاهد خواهیم بود، پیدا کردن نقطه‌ای که هر دو مقدار کیفیت و تشریح را بیشینه کند ممکن نخواهد بود،

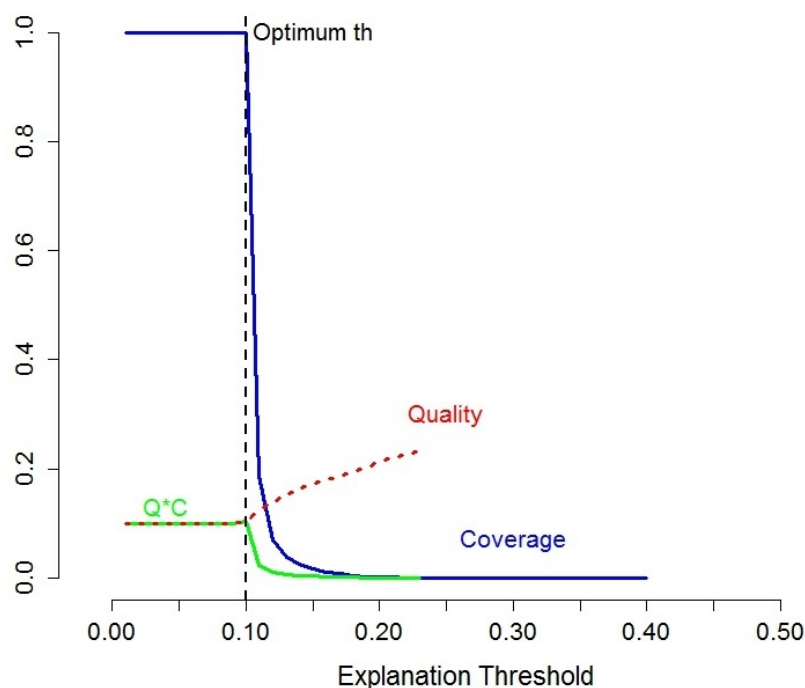
<sup>۱</sup> Explanation threshold

<sup>۲</sup> Corpus coverage

<sup>۳</sup> Explanation quality

از این رو بسته به میزان اهمیت این دو معیار می توان میزان  $th_e$  را تعیین کرد. در این پژوهش فرض شده است که این دو معیار به یک میزان اهمیت دارند؛ به همین دلیل نقطه ای که در آن میانگین این دو مقدار بالاترین میزان را خواهد داشت، مدنظر قرار گرفته است.

شکل ۳-۴ میزان پوشش پیکره و کیفیت تشریح آن و نیز حاصلضرب این دو معیار را در مقادیر مختلف حد آستانه برای مقادیر مختلف  $th_e$  نشان می دهد.



شکل ۳-۴- پوشش و کیفیت تشریح در حد آستانه مختلف

با توجه به شکل ۳-۴ میزان پوشش پیکره تا مقدار ۰/۱۰ برای حد آستانه ۱۰۰ درصد است و پس از آن به شدت میزان پوشش کاهش می یابد و کیفیت تشریح تنها کمی افزایش می یابد. از آنجا که هر دو معیار کیفیت تشریح و پوشش پیکره اهمیت دارد، مقدار حاصلضرب معیارها هم در شکل نمایش داده شده است. و مقدار بهینه برای حد آستانه همان ۰/۱۰ در نظر گرفته شده است. آنجا که  $\sum_{j=1}^{|T|} \rho(p_i, t_j) = 1$ ، مقدار ۰/۱۰ برای حد آستانه تشریح این امکان را یک مستند توسط بیش از یک موضوع تشریح شود.

#### ۳-۴. برچسب گذاری موضوعات استخراج شده

پس از مدل سازی موضوعی، موضوعات ده گانه حوزه انرژی شناسایی شد و کلمات پرتکرار در هر موضوع مشخص شد. به منظور درک بهتر موضوعات بهتر است هر کدام از موضوعات با برچسبی که

مشخص کننده محتوای آن موضوع است شناسایی شود. برای این منظور می‌توان از خبرگان حوزه مورد بررسی خواست تا برچسب مناسب موضوع را با توجه به منابع موجود در آن موضوع و یا منابع شناسایی شده برای موضوع تعیین کنند. در این پژوهش با استفاده از کلمات پرتکرار هر موضوع برچسب آنها پیشنهاد شده است. جدول ۴-۱ موضوعات استخراج شده، برچسب پیشنهادی و عبارت‌های پرتکرار هر موضوع را نشان می‌دهد.

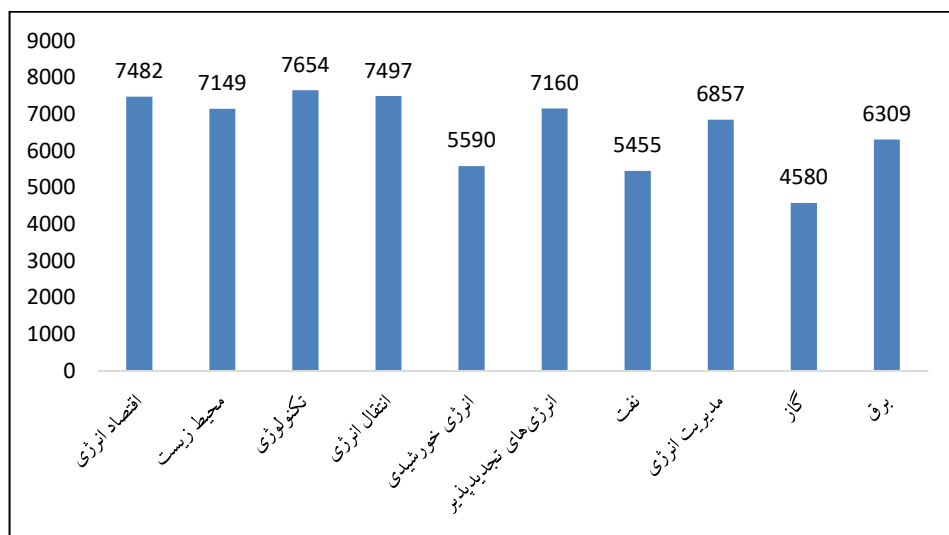
جدول ۴-۱- موضوعات استخراج شده حوزه انرژی و عبارت‌های پرتکرار آنها

| موضوع    | برچسب               | عبارت‌های پرتکرار                                                                                                                                                                                                                                                                                        |
|----------|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Topic 1  | اقتصاد انرژی        | درآمدهای نفتی، مدیریت انرژی، بازار برق، کاهش مصرف انرژی، انرژی تجدیدپذیر، موتور دیزل، سرمایه انسانی، هاب انرژی، خانه هوشمند، پاسخگویی بار، ...                                                                                                                                                           |
| Topic 2  | محیط زیست           | محیط زیست، طول عمر، کاهش مصرف انرژی، نظریه تابعی چگالی، فتولتائیک، انرژی‌های تجدیدپذیر، انرژی‌های نو، بهره‌وری انرژی، چین، پلاسما، هضم بی‌هوازی، بهینه‌سازی انرژی، سیاست، معماری، بهره‌برداری بهینه، اسمز معکوس، ...                                                                                     |
| Topic 3  | تکنولوژی            | هیدرات گازی، مواد تغییر فاز دهنده، شبکه توزیع، توربین گازی، اکسید روی، کلکتور خورشیدی، نانو ذرات، تولید پراکنده، پانل دیتا، ریزساختار، finite element method، منابع تولید پراکنده، پیل سوختی اکسید جامد، سورفکتانت، ...                                                                                  |
| Topic 4  | انتقال انرژی        | برداشت انرژی، ازدیاد برداشت نفت، طول عمر شبکه، کاهش مصرف انرژی، بهینه‌سازی مصرف انرژی، ابرخازن، مدیریت پروژه، پیزوالکتریک، مبدل حرارتی، ریزشبکه، گازهای گلخانه‌ای، energy absorption، میرایی، شبیه‌سازی دینامیک مولکولی، اتلاف انرژی، انرژی پاک، ...                                                     |
| Topic 5  | انرژی خورشیدی       | انرژی خورشیدی، گاف انرژی، سلول خورشیدی، نیتروژن، مواد نانو ساختار، energy gap، جاذب انرژی، روش مونت کارلو، نیروگاه بادی، حسگر گاز، شبکه‌های حسگر بیسیم، اثربخشی، electrochemistry، ...                                                                                                                   |
| Topic 6  | انرژی‌های تجدیدپذیر | انرژی تجدیدپذیر، استهلاك انرژی، انرژی خورشیدی، منابع انرژی تجدیدپذیر، دی‌اکسید کربن، انرژی آزاد گیبس، اتلاف انرژی، توربین گاز، گاز طبیعی، بیوگاز، دیاژنز، مدیریت انرژی، زیست پالایی، جداسازی گاز، هیدروکربن، انرژی‌های تجدیدپذیر، گرانیوی، انرژی بستگی، انتقال حرارت، شدت انرژی، energy dissipation، ... |
| Topic 7  | نفت                 | نفت خام، صنعت نفت، گاز، یارانه، روسیه، foreign energy consumption، petroleum industry، policy، بنزین، گازوئیل، صادرات، nuclear energy، اتحادیه اروپا، گرافن، آمریکا، منابع انرژی، نظام حقوقی، ...                                                                                                        |
| Topic 8  | مدیریت انرژی        | بازار برق، مصرف انرژی، نفت، اتلاف انرژی، مدیریت انرژی، اقتصاد، منابع تولید پراکنده، عملکرد شغلی، سختی، شبکه انتقال نیروی برق، ذخیره‌ساز انرژی، تولید پراکنده، انرژی تجدیدپذیر، نیروگاه مجازی، کیهان‌شناسی، ...                                                                                           |
| Topic 9  | گاز                 | کروماتوگرافی گازی، gas chromatography، خوردگی، nanoparticle، فاز گازی، gaseous phase، اکسایش، نانوفناوری، گاز سنتز، ازدیاد برداشت، performance evaluation، حسگر گاز، ازدیاد برداشت نفت، ...                                                                                                              |
| Topic 10 | برق                 | بازار برق، انرژی مصرفی، دینامیک مولکولی، سلول خورشیدی، تولید انرژی، خواص ترمودینامیکی، انرژی تجدیدپذیر، تهویه مطبوع، صنعت برق‌رسانی، electricity supply industry، توسعه اقتصادی، ...                                                                                                                     |

در تشریح موضوعات استخراج‌شده توجه به چند نکته اهمیت دارد. نخست آنکه همان‌گونه که در جدول ۴-۱ مشخص شده است، برخی از عبارت‌ها در چند موضوع به عنوان عبارت کاربرد حضور دارند و همانطور که در بخش (۴\_۲\_۲) بیان شد، تعیین حد آستانه ۰/۱ باعث خواهد شد تا یک رکورد در چند موضوع حضور داشته باشد، از این رو نامگذاری هر موضوع صرفاً بر اساس مشاهدات عبارت‌های کاربرد انجام شده است و ممکن است از نظر یک فرد خبره نام مناسب دیگری را بتوان به هر موضوع نسبت داد. دوم آنکه موضوعات استخراج شده لزوماً از یکدیگر متمایز نیستند و به عنوان مثال می‌توان موضوع اقتصاد یا محیط‌زیست و مدیریت انرژی را با یکدیگر ترکیب کرد و به عنوان یک موضوع واحد در نظر گرفت. از آنجا که هدف این پژوهش تشریح ساختار مفاهیم حوزه انرژی نیست و به دنبال تحلیل عرضه و تقاضای موضوعات پژوهشی است، تمرکز آن در ادامه روی موضوعات ده‌گانه استخراج شده خواهد بود.

#### ۴\_۴. فراوانی منابع عرضه‌شده در موضوعات حوزه انرژی

با توجه به حدآستانه تعیین شده برای تشریح مستندات توسط موضوعات، فراوانی منابع عرضه شده در موضوعات حوزه انرژی در شکل ۴-۴ آمده است. نکته مهم در این شکل این است که از آنجا که یک سند ممکن است در بیش از یک موضوع قرار گرفته باشد، مجموع منابع موضوعات بیش از تعداد منابعی است که پیش‌تر برای حوزه انرژی شناسایی شد.



شکل ۴-۴- فراوانی منابع عرضه‌شده در موضوعات حوزه انرژی

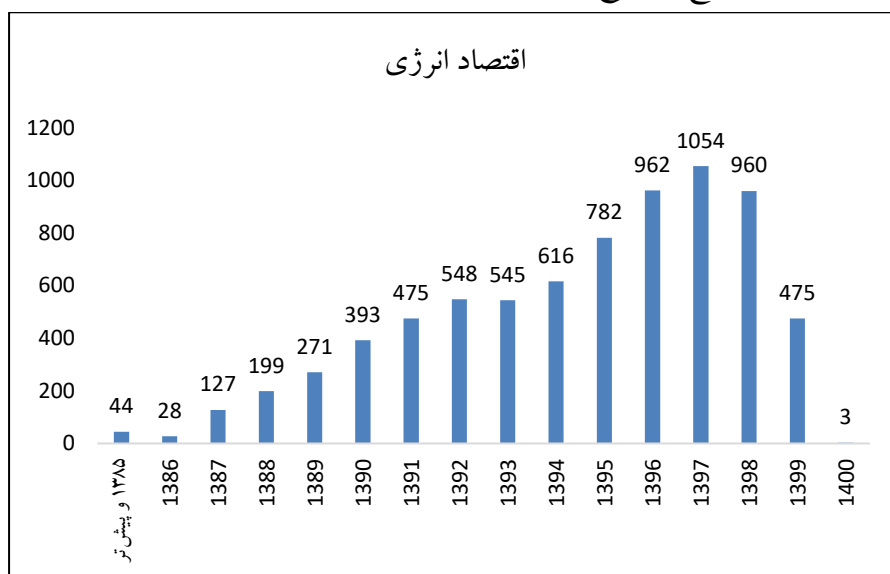
همان‌گونه که در شکل ۴-۴ آمده است بیشترین میزان منابع در موضوعات «تکنولوژی»، «انتقال انرژی»، «انرژی‌های تجدیدپذیر» و «اقتصاد انرژی» عرضه شده است و کمترین میزان منابع ارائه شده در موضوعات «گاز»، «نفت» و «انرژی خورشیدی» است.

#### ۴\_۵. روند عرضه منابع حوزه انرژی

در ادامه روند عرضه منابع حوزه انرژی در موضوعات مختلف تشریح خواهد شد. در این روند منابع پیش از ۱۳۸۵ با یکدیگر جمع شده‌اند، همچنین به ندرت منابعی از سال ۱۴۰۰ نیز در منابع مشاهده شده است که در نمودارها مشاهده می‌شود.

##### ۴\_۵\_۱. روند عرضه منابع موضوع «اقتصاد انرژی»

شکل ۴-۵ روند عرضه منابع موضوع «اقتصاد انرژی» را نشان می‌دهد.

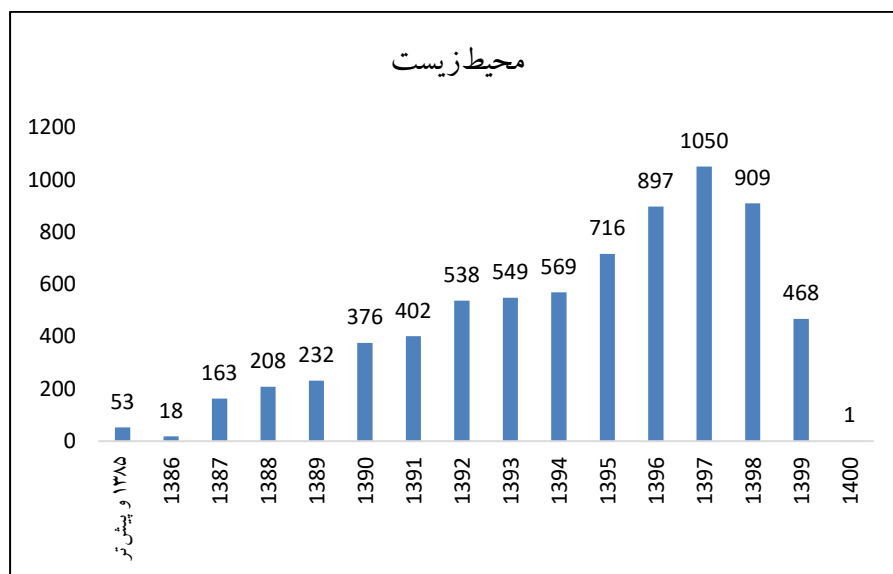


شکل ۴-۵-۱ روند عرضه منابع موضوع «اقتصاد انرژی»

روند فوق نشان می‌دهد که منابع موضوع «اقتصاد انرژی» از اوایل دهه ۹۰ تا سال ۱۳۹۷ رشد داشته‌اند و در سه سال اخیر روند کاهشی داشته است که دلیل آن می‌تواند ناشی از این باشد که منابع تولیدشده سال‌های اخیر هنوز بصورت کامل در پایگاه اطلاعاتی وارد نشده است، چرا که روند فوق با روند موجود در شکل ۴-۳ که مربوط به تمامی منابع عرضه شده حوزه انرژی بود همخوانی دارد.

##### ۴\_۵\_۲. روند عرضه منابع موضوع «محیط‌زیست»

شکل ۴-۶ روند عرضه منابع موضوع «محیط‌زیست» را نشان می‌دهد.

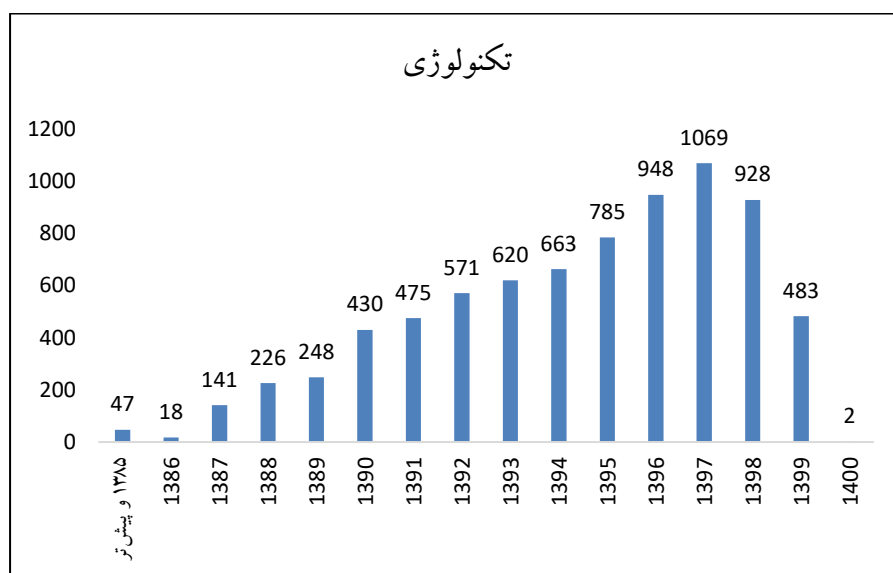


شکل ۶-۶- روند عرضه منابع موضوع «محیط‌زیست»

روند منابع موضوع «محیط‌زیست» نیز همراستا با روند کلی منابع عرضه شده در حوزه انرژی همراستاست و شروع رشد آن از اوایل دهه ۹۰ شروع شده و در سال‌های اخیر نیز شاهد کاهش منابع این موضوع هستیم که می‌تواند به دلیل عدم تکمیل منابع در پایگاه باشد.

#### ۳-۵-۴. روند آینده عرضه منابع موضوع «تکنولوژی»

شکل ۴-۷- روند عرضه منابع موضوع «تکنولوژی» را نشان می‌دهد.

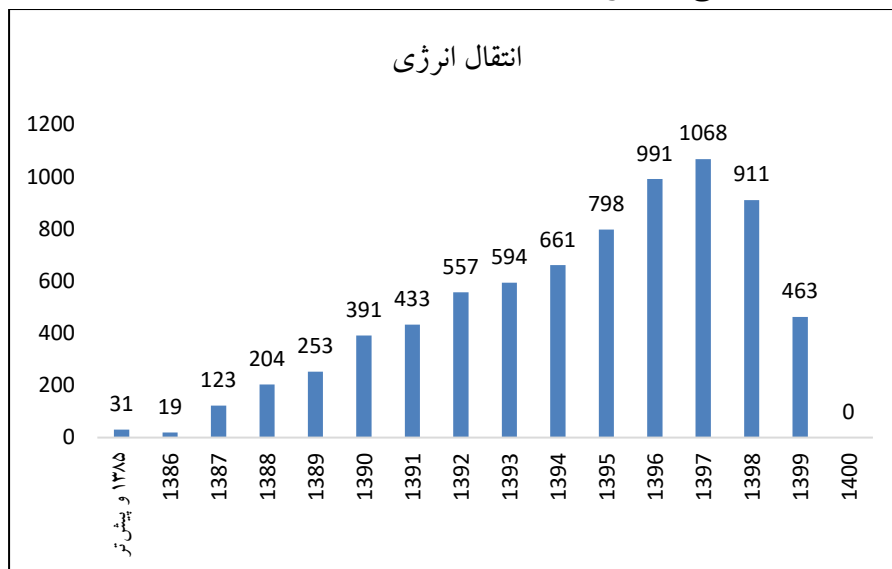


شکل ۷-۷- روند عرضه منابع موضوع «تکنولوژی»

روند منابع موضوع «تکنولوژی» نیز مانند موضوعات قبلی همراستا با منابع کل حوزه انرژی است.

#### ۴-۵-۴. روند عرضه منابع موضوع «انتقال انرژی»

شکل ۴-۸ روند عرضه منابع موضوع «انتقال انرژی» را نشان می‌دهد.

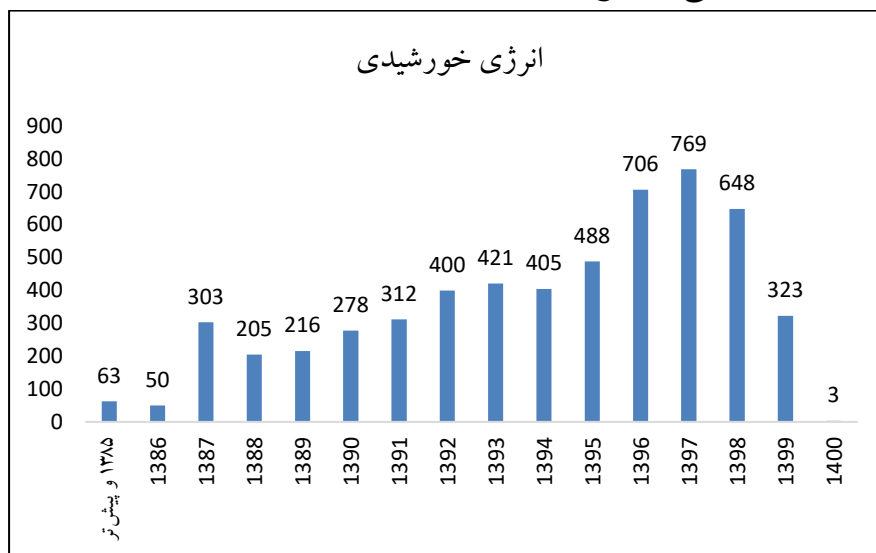


شکل ۴-۸- روند عرضه منابع موضوع «انتقال انرژی»

روند منابع موضوع «انتقال انرژی» نیز از روند کلی موضوعات حوزه انرژی تبعیت می‌کند و در سال‌های اخیر روند افزایشی داشته است.

#### ۴-۵. روند عرضه منابع موضوع «انرژی خورشیدی»

شکل ۴-۹ روند عرضه منابع موضوع «انرژی خورشیدی» را نشان می‌دهد.

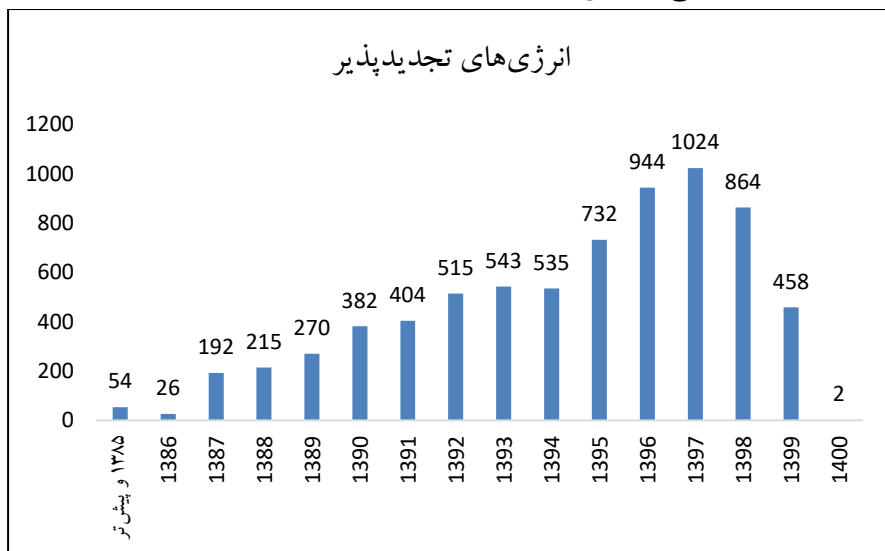


شکل ۴-۹- روند عرضه منابع موضوع «انرژی خورشیدی»

گرچه روند کلی منابع موضوع «انرژی خورشیدی» مانند روند کل حوزه انرژی افزایشی است، اما در سال ۱۳۸۷ به نسبت سه سال پس از آن منابع بیشتری در این موضوع عرضه شده است.

#### ۴-۶. روند عرضه منابع موضوع «انرژی‌های تجدیدپذیر»

شکل ۴-۱۰ روند عرضه منابع موضوع «انرژی‌های تجدیدپذیر» را نشان می‌دهد.

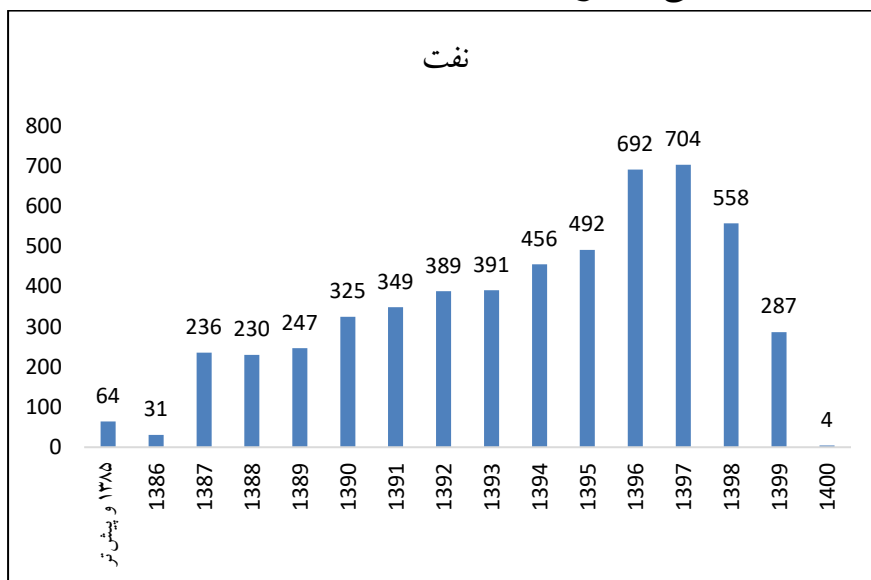


شکل ۴-۱۰- روند عرضه منابع موضوع «انرژی‌های تجدیدپذیر»

روند منابع موضوع «انرژی‌های تجدیدپذیر» نیز از روند کلی موضوعات حوزه انرژی تبعیت می‌کند و در سال‌های اخیر روند افزایشی داشته است.

#### ۴-۵-۷. روند عرضه منابع موضوع «نفت»

شکل ۴-۱۱ روند عرضه منابع موضوع «نفت» را نشان می‌دهد.

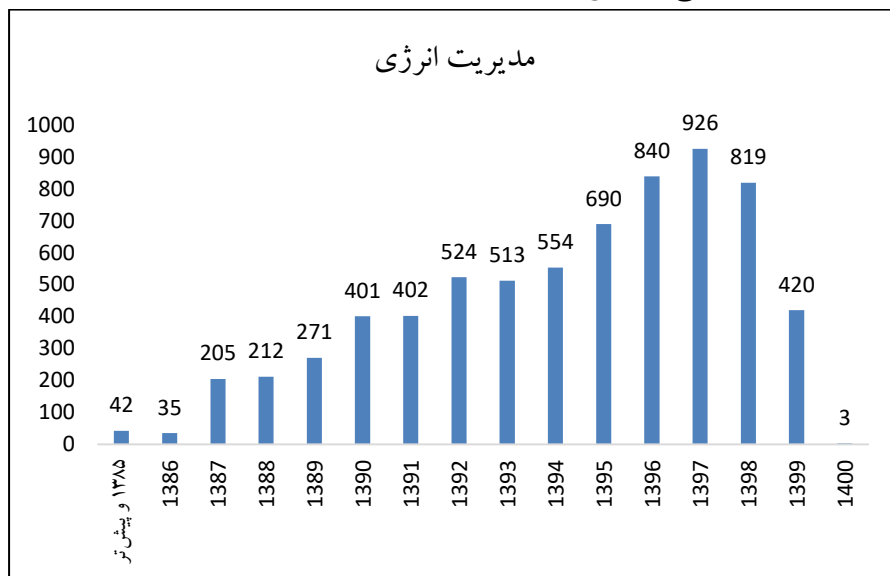


شکل ۴-۱۱- روند عرضه منابع موضوع «نفت»

روند منابع موضوع «نفت» نیز از روند کلی موضوعات حوزه انرژی تبعیت می‌کند و در سال‌های اخیر روند افزایشی داشته است. این روند با شیب نسبتاً کمی از سال ۱۳۸۷ تا ۱۳۹۵ افزایش پیدا کرده است ولی در نیمه دوم دهه ۹۰، شیب رشد آن بیش از گذشته شده است.

#### ۸\_۵\_۴. روند عرضه منابع موضوع «مدیریت انرژی»

شکل ۴-۱۲ روند عرضه منابع موضوع «مدیریت انرژی» را نشان می‌دهد.

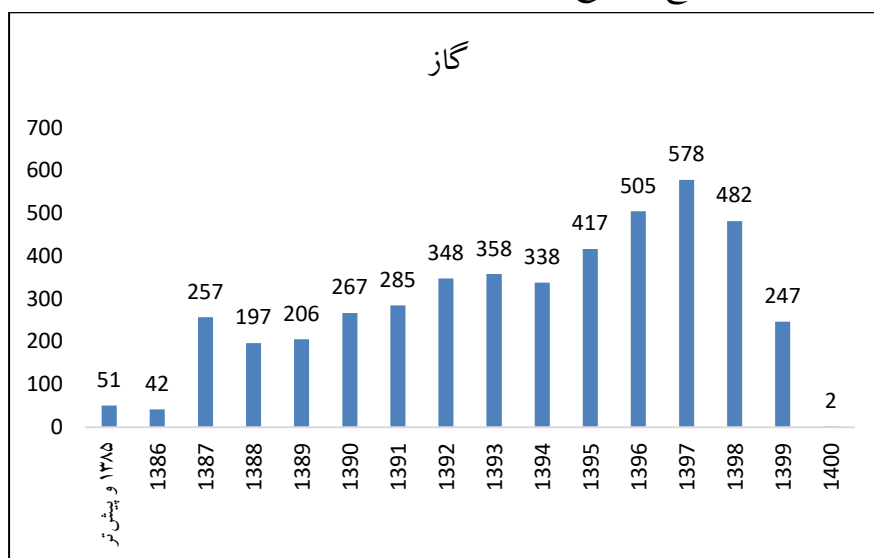


شکل ۴-۱۲- روند عرضه منابع موضوع «مدیریت انرژی»

روند منابع موضوع «مدیریت انرژی» نیز از روند کلی موضوعات حوزه انرژی تبعیت می‌کند و در سال‌های اخیر روند افزایشی داشته است. این روند در سال ۱۳۸۶ به نسبت سال‌های پیش از آن رشد قابل توجه را در منابع عرضه‌شده این موضوع نشان می‌دهد.

#### ۹\_۵\_۴. روند عرضه منابع موضوع «گاز»

شکل ۴-۱۳ روند عرضه منابع موضوع «گاز» را نشان می‌دهد.

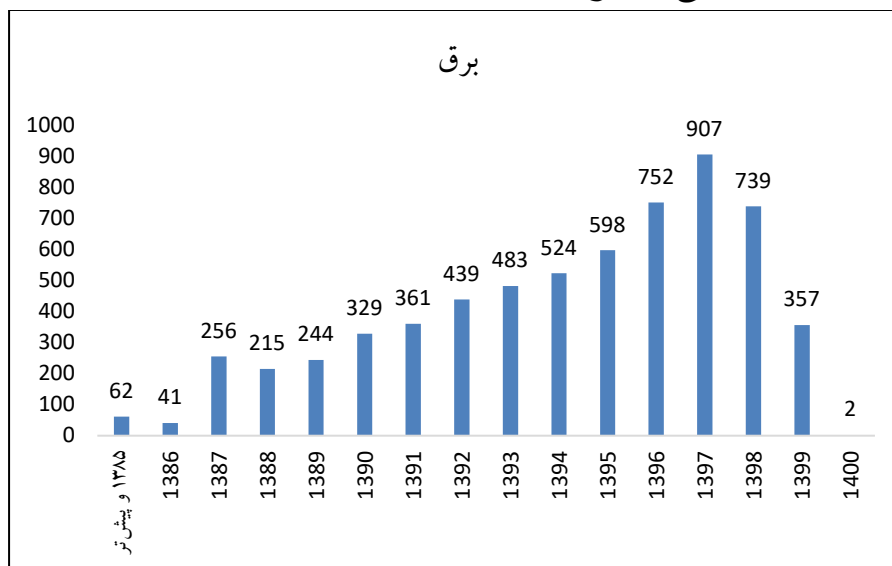


شکل ۴-۱۳- روند عرضه منابع موضوع «گاز»

گرچه روند کلی منابع موضوع «گاز» مانند روند کل حوزه انرژی افزایشی است، اما در سال ۱۳۸۷ به نسبت دو سال پس از آن منابع بیشتری در این موضوع عرضه شده است.

#### ۴-۵-۱۰. روند عرضه منابع موضوع «برق»

شکل ۴-۱۱ روند عرضه منابع موضوع «برق» را نشان می‌دهد.



شکل ۴-۱۴- روند عرضه منابع موضوع «برق»

روند کلی منابع موضوع «برق» هم مانند دیگر موضوعات از روند کل حوزه انرژی تبعیت می‌کند، و تنها تفاوت آن این است که در سال ۱۳۸۷ به نسبت دو سال پس از آن منابع بیشتری در این موضوع عرضه شده است.

تحلیل روند عرضه موضوعات مختلف حوزه انرژی نشان می‌دهد که بجز در برخی از جزئیات، تفاوت چندانی در روند عرضه این موضوعات نسبت به یکدیگر وجود ندارد و تقریباً تمامی آنها از الگوی روند عرضه کل حوزه انرژی تبعیت می‌کنند. در این الگو، میزان تولید پارساها رو به رشد بوده است و این رشد از اوایل دهه ۹۰ شتاب گرفته است. نمودارهای میزان تولیدات سال‌های اخیر را کمتر نشان می‌دهند که دلیل آن به‌روز نبودن پایگاه اطلاعاتی است که نشان می‌دهد منابع اخیراً تولید شده هنوز به صورت کامل در پایگاه ثبت نشده‌اند.

#### ۴-۱. مدل‌سازی موضوعی منابع تقاضاشده حوزه انرژی

پس از آنکه مدل‌سازی موضوعی منابع عرضه شده انجام شد، می‌توان مدل تولید شده را روی داده‌های مربوط به منابع تقاضا شده اجرا کرد و این‌بار منابع تقاضا شده را در همان موضوعات ده‌گانه حوزه انرژی رصد نمود.

برای تعیین بردار ویژگی‌های منابع تقاضا شده، تمای n-gram‌های موجود در عبارت جستجو استخراج شد، سپس آن دسته از n-gram‌هایی که شامل ایست‌واژه<sup>۱</sup> بودند حذف شدند. همانطور که در بخش (۲-۴) عنوان شد، فهرستی از ۱۲،۵۰۵ عبارت به عنوان ستونهای ماتریس مستند-واژه بکار گرفته شد. در اینجا نیز آن دسته از n-gram‌های استخراج شده مورد استفاده قرار گرفت که در این فهرست موجود باشند. به این ترتیب ماتریس مستند واژه تولیدی برای تقاضاهای حوزه پژوهش شامل ۱۲،۵۰۵ ستون و ۴۴۳،۸۹۰ سطر خواهد بود. از آنجا که صرفاً آن بخش از جستجوها که شامل فهرست ۱۲،۵۰۵ عبارت هستند مورد ارزیابی قرار خواهد گرفت پس بخشی از سطرهای این ماتریس برای هیچ یک از ستون‌ها مشاهده نخواهد شد و برای تمامی آنها مقدار صفر را خواهد داشت و امکان تخصیص آن جستجو به هیچ یک از موضوعات شناسایی شده وجود نخواهد داشت، از اینرو لازم است ابتدا سطرهایی که مجموع تمام ستونهای آن صفر است از ماتریس مستند واژه حذف شود.

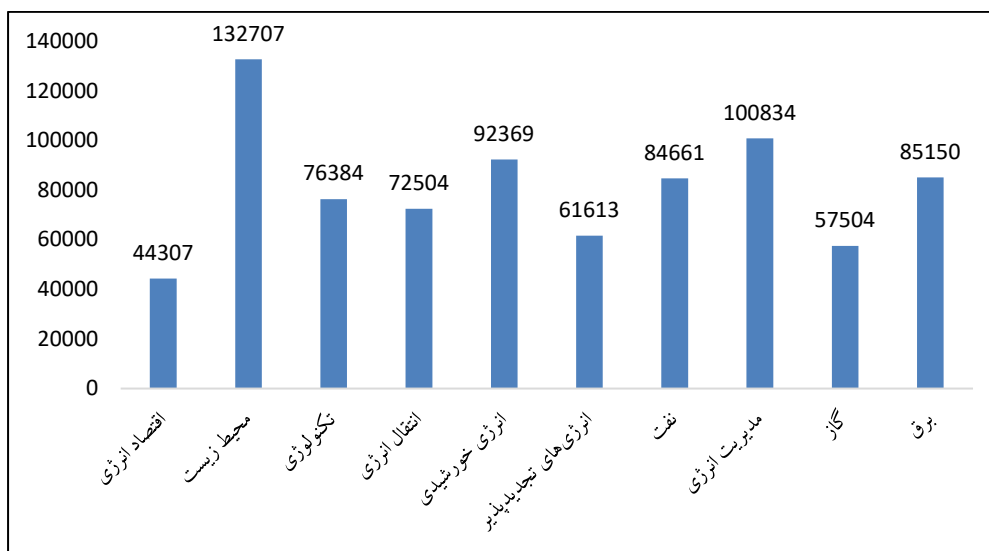
بدلیل حجم بالای این ماتریس، محاسبه مجموع عناصر موجود در هر سطر این ماتریس به سادگی امکان‌پذیر نیست که برای این منظور ابتدا ماتریس به ماتریس‌های کوچکتر تقسیم شد و محاسبات روی ماتریس‌های کوچکتر انجام شد و در نهایت ماتریس‌ها با یکدیگر ادغام شدند. نتیجه نهایی شامل ماتریسی با ۱۲،۵۰۵ ستون و ۲۵۳،۸۷۶ سطر شد.

#### ۲-۴. فراوانی منابع تقاضا شده در موضوعات حوزه انرژی

با توجه به حد آستانه تعیین شده (۰,۱۰) برای تشریح موضوعات، در اینجا نیز توجه به این نکته مهم است که یک جستجو ممکن است در چندین موضوع قرار گرفته باشد. شکل ۴-۱۵ فراوانی موضوعات جستجو شده در حوزه انرژی را در پایگاه گنج نشان می‌دهد.

---

<sup>۱</sup> StopWord



شکل ۴-۱۵- فراوانی منابع در موضوعات تقاضاشده در حوزه انرژی

همان‌گونه که در شکل ۴-۱۵ مشخص است، بیشترین میزان تقاضا در موضوعات «محیط‌زیست» و «مدیریت انرژی» است و کمترین میزان تقاضا به ترتیب مربوط به موضوعات «اقتصاد انرژی»، «گاز» و «انرژی‌های تجدیدپذیر» است.

#### ۴-۳. روند تقاضای موضوعات مختلف حوزه انرژی در ماه‌های سال

در فصل پیش و در بخش (۳-۴-۱) روند تقاضای موضوعات تقاضاشده در حوزه انرژی ارائه شد. در این بخش روند تقاضا به تفکیک هر یک از موضوعات بیان خواهد شد. این روند را در ماه‌های مختلف سال نشان می‌دهد.





شکل ۴-۱۶- روند تقاضای موضوعات حوزه انرژی به تفکیک ماه‌های سال

همان‌گونه که در شکل ۴-۱۶ مشاهده می‌شود موضوعات مختلف تقریباً از روند یکسانی در توزیع جستجوها در ماه‌های مختلف سال تبعیت می‌کنند و طول سال تحصیلی بیش از ایام تعطیلات دانشجویان (مانند نوروز و تابستان) است و نشان‌دهنده این است که عمده کاربران پایگاه اطلاعاتی گنج را دانشجویان تشکیل می‌دهند و تقاضای مورد مطالعه، بیش از آنکه مربوط به صنعت باشد، نیازمندی‌های جامعه پژوهشی و دانشجویان تحصیلات تکمیلی را نشان می‌دهد.

#### ۴-۴. تحلیل فاصله بین عرضه و تقاضای پژوهش

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 هر یک از مقادیر عرضه و تقاضا با توجه به کمیته و بیشینه عرضه و تقاضا نرمال شد. برای این منظور از فرمول زیر استفاده شد.

$$x' = \frac{x - \text{Min}(x)}{\text{Max}(x) - \text{Min}(x)}$$

فرمول ۴-۴

پس از آنکه تمامی مقادیر عرضه و تقاضا با استفاده از فرمول ۴-۴ نرمال شد، 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 زیاد» و «عرضه کم» و نیز با تقسیم محور عمودی به دو بخش «تقاضای زیاد» و «تقاضای کم» 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 ۴-۱۷ این چهار بخش و موقعیت هر یک از موضوعات را نشان می‌دهد.



شکل ۴-۱۷ - وضعیت عرضه و تقاضای موضوعات حوزه انرژی

با توجه به شکل ۴-۱۷، 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 بوده و نیازهای پژوهشی خود را در این موضوعات کاوش می‌کنند. از این رو موضوعات «محیط زیست» و «مدیریت انرژی» که در این ناحیه واقع شده‌اند به عنوان «موضوعات داغ»<sup>۱</sup> در نظر گرفته شده‌اند.

<sup>۱</sup> Hot Topics

موضوعات داغ آن دسته از موضوعات فرض شده‌اند که علاوه بر اینکه به میزان قابل توجهی تولید شده‌اند، به میزان زیادی نیز در جستجوهای کاربران مورد توجه قرار گرفته‌اند. به عبارتی پژوهش‌های این موضوعات همچنان در حال استفاده، بسط و گسترش است. در پژوهشی که پیش‌تر در حوزه عرضه و تقاضای محیط‌زیست انجام شده است، موضوع «انرژی» در موضوعاتی که عرضه زیادی داشته‌اند مشاهده شده است (ربیعی، ۱۳۹۸) و در این پژوهش نیز که حوزه انرژی را مدنظر قرار داده است، آن دسته از پژوهش‌هایی که با رویکردهای مرتبط با محیط‌زیست مساله انرژی را مورد بررسی قرار می‌دهند به عنوان موضوعات با عرضه زیاد شناسایی شده‌اند.

ناحیه سمت راست و پایین شکل ۴-۱۷ دربردارنده موضوعاتی است که اصطلاح «موضوعات خاموش»<sup>۱</sup> برای آن‌ها انتخاب شده است. این بخش شامل موضوعاتی است که دارای تعداد قابل توجهی منبع هستند و به میزان زیادی در آن‌ها پژوهش صورت پذیرفته است، 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 «برق» فاصله بسیار اندکی با ناحیه قبلی دارد و با اندکی سهل‌گیری می‌تواند جزو موضوعات داغ دسته‌بندی شود اما موضوعات «تکنولوژی»، «اقتصاد انرژی»، «انتقال انرژی» و «انرژی‌های تجدیدپذیر» در دسته موضوعات خاموش قرار می‌گیرند.

ناحیه سمت چپ و پایین، ناحیه‌ای است که دربردارنده موضوعاتی 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»<sup>۲</sup> نام‌گذاری شده است. «نفت» و «گاز» در این ناحیه قرار گرفته‌اند. از آنجا که «نفت» تا حدی به ناحیه چهارم نزدیک است وضعیت بهتری نسبت به موضوع «گاز» دارد، اما با توجه به تعداد اندک پژوهش‌های انجام گرفته در موضوعات این ناحیه و نیز عدم توجه کافی جستجوگران پژوهشی حوزه انرژی به موضوعات این ناحیه، می‌توان پیش‌بینی نمود که در صورتی که سیاست‌گذاران پژوهشی برای پژوهش‌های موضوع «گاز» چاره‌اندیشی لازم را نداشته باشند در آینده این موضوع به تدریج حذف خواهد شد و اهمیت خود را به عنوان یک موضوع مستقل از دست خواهد داد.

ناحیه سمت چپ و بالا، آخرین بخش از این ماتریس است. این ناحیه دربردارنده موضوعاتی است که اگرچه مورد توجه جستجوگران قرار گرفته‌اند و تقاضای پژوهشی در آنها بالاست اما به اندازه کافی

<sup>۱</sup> Silent Topics

<sup>۲</sup> Cold Topics

منابع پژوهشی برای آن‌ها وجود ندارد. به عبارتی به ازای تقاضای بالای پژوهشی در این موضوعات، کاربران با منابع پژوهشی کافی مواجه نمی‌شوند. از این رو نام این ناحیه «موضوعات دارای کاستی»<sup>۱</sup> گذاشته شده است که دربردارنده کاستی‌های پژوهش در حوزه انرژی است. «انرژی خورشیدی» و تا حدی «نفت» موضوعاتی است که دارای کاستی پژوهشی هستند.

#### ۵-۴. جمع‌بندی و نتیجه‌گیری

در این فصل موضوعات پژوهشی حوزه انرژی با استفاده از مدل‌سازی موضوعی شناسایی شد که منجر به شناسایی ۱۰ موضوع در این حوزه شد. هر چند که انتخاب برچسب مناسب برای موضوعات یکی از چالش‌های اساسی است و این کار می‌تواند توسط یک خبره موضوعی به شکل دقیق انجام شود، اما صرف‌نظر از عنوانی که به هر موضوع اختصاص داده می‌شود، هدف اصلی این بخش ارزیابی عرضه و تقاضای هر یک از این حوزه‌ها بود. با این حال برای ملموس‌تر شدن مباحث این موضوعات با عناوین «اقتصاد انرژی»، «محیط زیست»، «تکنولوژی»، «انتقال انرژی»، «انرژی خورشیدی»، «انرژی‌های تجدیدپذیر»، «نفت»، «مدیریت انرژی»، «گاز» و «برق» برچسب‌گذاری شد.

رکورد‌های مربوط به بخش عرضه و تقاضا در هر یک از موضوعات ده‌گانه مشخص شد و روند آنها مورد بررسی قرار گرفت. روند عرضه در موضوعات مختلف تفاوت چندانی با یکدیگر نداشت و بجز موارد جزئی تقریباً از الگوی یکسانی پیروی می‌کرد که نشان می‌داد از اوایل دهه ۹۰ عرضه پژوهش در این حوزه رشد بالایی داشته است.

با استفاده از تحلیل فاصله بین عرضه و تقاضا در هر یک از این موضوعات، ماتریسی شامل ۴ منطقه تشکیل شد و موضوعات در این ماتریس نگاشت شد. این ۴ منطقه شامل موضوعات داغ (عرضه و تقاضای بالا)، موضوعات خاموش (عرضه بالا و تقاضای پایین)، موضوعات دارای کاستی (عرضه پایین و تقاضای بالا) و موضوعات سرد (عرضه و تقاضای پایین) بودند. قرارگیری موضوعات در هر یک از این ۴ حوزه وضعیت آنها را در اکوسیستم پژوهش تعیین می‌کند.

سیاست‌گذار پژوهشی با اطلاع از وضعیت عرضه و تقاضای حوزه انرژی و با در اختیار داشتن دیگر منابع اطلاعاتی و آگاهی از سیاست‌های کلان پژوهشی و برنامه‌ها و دستورالعمل‌های بالادستی می‌تواند به نحوی سیاست‌گذاری نماید تا موضوعات مورد نیاز و در اولویت جامعه مورد توجه قرار گرفته و جهت پژوهش‌های این حوزه را به سمت و سوی مطلوب هدایت نماید.

---

<sup>۱</sup> Gap Topics



## **فصل پنجم**

### **نتیجه‌گیری و پیشنهاد پژوهش‌های آتی**

## ۱\_۵. مقدمه

در دهه‌های اخیر، شاهد افزایش چشم‌گیر پژوهش در حوزه‌های مختلف علوم هستیم، نیازمندی صنعت به انجام پژوهش‌های تقاضامحور از یک سو و لزوم مدیریت منابع محدود پژوهشی برای انجام پژوهش‌های مولد و کارآمد و جلوگیری از انجام پژوهش‌های موازی، ناکارآمد یا تکراری موجب شده است تا شناسایی وضعیت موجود عرضه و تقاضای پژوهش در حوزه‌های مختلف یکی از نیازهای اساسی مدیران و سیاست‌گذاران پژوهشی باشد. تحلیل وضعیت پژوهش‌های حوزه انرژی به دلیل اهمیت ویژه‌ای که در اقتصاد و سیاست کشور دارد، امری حیاتی است. بخش قابل توجهی از پژوهش‌های کشور توسط دانشجویان تحصیلات تکمیلی و در قالب پایان‌نامه و رساله انجام می‌شود و پایگاه گنج با در اختیار داشتن این مستندات به منبع بسیار با اهمیتی تبدیل شده است که از یک سو منابع عرضه شده در حوزه‌های دانشی مختلف در آن ثبت و نگهداری می‌شود و از دیگر سو، پژوهشگران برای یافتن پاسخ‌های پژوهشی خود از آن بهره می‌برند.

این پژوهش با تاکید بر حوزه انرژی به عنوان یکی از حوزه‌هایی که اهمیت ویژه‌ای در اقتصاد، سیاست و صنعت کشور دارد با هدف شناسایی و استخراج منابع پژوهشی عرضه شده و تقاضا شده مرتبط با این حوزه در این پایگاه، به دنبال شناسایی و تحلیل شکاف بین موضوعات عرضه شده در پایان‌نامه‌ها و رساله‌های ایران را با موضوعات تقاضا شده توسط جویشگران است.

## ۲-۵. فرایند و گام‌های طی شده

در این پژوهش ابتدا با استفاده از اصلاح‌نامه‌های تخصصی موجود مدخل‌های مرتبط با موضوع انرژی شناسایی و دسته‌بندی شد، با استفاده از این مداخل، عبارت‌هایی که به عنوان نماینده این حوزه هستند و پیش‌بینی می‌شود با جستجوی این عبارت‌ها، منابع مرتبط با حوزه انرژی گردآوری شود تعیین و جستجو در پایگاه اطلاعاتی گنج انجام شد. در نتیجه مجموعه منابع عرضه شده در حوزه انرژی شناسایی و استخراج شد که نتیجه آن استخراج تعداد ۲۵،۹۵۳ رکورد از پارساهای موجود بود. سپس کلمات کلیدی این مجموعه مورد بررسی قرار گرفت و عبارت‌های پرتکرار شناسایی شد. برای این منظور از تکنیک‌های پردازش زبان طبیعی (NLP) و متن کاوی استفاده شد. همچنین تمامی جستجوهای انجام شده در این پایگاه در حدها ۱۳۹۶/۱۲/۱۲ تا ۱۳۹۹/۰۶/۱۷ که شامل تعداد ۲۴،۲۰۲،۷۷۴ رکورد بود، گردآوری شد. پس از محدود کردن جستجوهای انجام شده به کلمات کلیدی و خاص حوزه انرژی که از منابع عرضه شده به دست آمده بود، آن دسته از جستجوهای که شامل این عبارت‌ها بود و شامل ۴۴۳،۸۹۰ رکورد می‌شد شناسایی و به عنوان منابع تقاضا شده حوزه انرژی مدنظر قرار گرفت.

در مرحله بعدی اطلاعات کتاب‌شناختی پارساهای عرضه شده حوزه انرژی مورد بررسی و تحلیل قرار گرفت، این تحلیل شامل بررسی مقطع تحصیلی منابع، تحلیل زمانی منابع، بررسی گروه‌های آموزشی دانشجویان تولید کننده پارساها، ارائه آمار دانشگاه‌ها و مراکز تولید کننده پارساها و در نهایت تحلیل کلمات کلیدی موجود در آن‌ها بود. همچنین کتاب‌شناختی جستجوهای حوزه انرژی در پایگاه گنج نیز ارائه شد. این تحلیل شامل تحلیل زمانی، بررسی تعداد پاسخ‌هایی که موتور جستجو برای هر یک از تقاضاها ارائه نموده است و تحلیل کلمات کلیدی موجود در متن جستجوها بود.

در گام بعدی با استفاده از مدل‌سازی موضوعی، منابع شناسایی شده در بخش عرضه و تقاضای پژوهش به موضوعات تشکیل دهنده آنها افزای شدند. تعداد ۱۰ موضوع برای حوزه انرژی شناسایی شد و منابع تشکیل دهنده آنها مورد ارزیابی و تحلیل قرار گرفت. این عمل در هر دو بعد عرضه و تقاضای انرژی انجام شد. در نهایت با استفاده از نرمال‌سازی مقادیر مربوط به منابع عرضه و تقاضا شده، ماتریس عرضه و تقاضای پارساهای حوزه انرژی به منظور تحلیل فاصله میان عرضه و تقاضا تشکیل شد. این ماتریس موضوعات حوزه انرژی را بر اساس میزان زیاد یا کم بودن تقاضا و عرضه پژوهش در ۴ منطقه قرار داد و موقعیت هر یک را تشریح کرد.

## ۱-۵. پاسخ به پرسش‌های پژوهش

هدف از انجام این پژوهش

- تحلیل وضعیت عرضه و تقاضای پایان‌نامه‌ها و رساله‌ها در موضوعات مختلف حوزه انرژی؛
  - شناسایی موضوعات دارای کاستی پژوهش در قالب پایان‌نامه‌ها و رساله‌ها در موضوعات.
- بود، برای رسیدن به این هدف لازم بود تا به سوالات زیر پاسخ داده شود:
۱. پایان‌نامه‌ها و رساله‌های عرضه‌شده در حوزه انرژی کدامند و چه موضوعاتی را پوشش می‌دهند؟
  ۲. جستجوهای انجام شده در پایگاه گنج در حوزه انرژی کدامند و این تقاضاها در چه موضوعاتی از این حوزه قرار می‌گیرند؟
  ۳. شکاف بین موضوعات عرضه شده در پایان‌نامه‌ها و رساله‌ها با موضوعات تقاضا شده توسط جویشگران پایگاه اطلاعاتی گنج در حوزه انرژی به چه شکلی است؟
- در ادامه به بررسی هر یک از سوالات و پاسخی که پژوهش انجام شده برای هر یک از سوالات بالا دارد خواهیم پرداخت.

#### ۵-۱-۱. سوال اول پژوهش

سوال اول این پژوهش به دنبال شناسایی پارساهای حوزه انرژی است. برای پاسخ به این سوال گام‌های هشت‌گانه که در شکل ۳-۱ آمده است طی شد. در این گام‌ها ابتدا نمونه‌ی اولیه از داده‌های پایگاه گنج دریافت شد که شامل ۱۰ هزار رکورد از منابع حوزه‌های مختلف و به صورت تصادفی از ده سال اخیر بود. با استفاده از این داده‌ها مجموعه داده اولیه حوزه انرژی که شامل منابعی می‌شد که کلمه «انرژی» در عنوان یا کلیدواژه آنها وجود داشت. این مجموعه شامل ۵۱۸ رکورد بود که تقریباً ۵ درصد منابع استخراج شده را شامل می‌شد.

بر اساس این منابع کلمات کلیدی پرتکرار حوزه انرژی استخراج شد، از سوی دیگر مدخل‌هایی که با عبارت «انرژی» از اصطلاح‌نامه‌های تخصصی مختلف وجود داشت بررسی شد و کلمات مرتبط با آنها استخراج شد. از ترکیب کلمات کلیدی استخراج شده از مجموعه داده اولیه و کلمات استخراج شده از اصطلاح‌نامه‌ها، فهرستی از کلمات نماینده این حوزه مشخص شد، در ادامه پس از پالایش عبارت‌های بدست آمده، عبارتی که جستجوی آن می‌توانست کلمات نماینده را استخراج کند تعیین شد و در نهایت مجموع این عبارت‌ها برای جستجو در پایگاه گنج گردآوری شد. در گام‌نهایی این عبارت‌ها در پایگاه گنج جستجو شد و نتیجه حاصل به عنوان داده‌های بخش عرضه حوزه انرژی در پژوهش مورد استفاده قرار گرفت. این جستجوی در بین پارساها موجب استخراج تعداد ۲۵،۹۵۳ رکورد شد که در فصل چهارم این منابع به صورت کامل مورد تحلیل قرار گرفت و تحلیل کتابشناختی آنها ارائه شد.

پس از شناسایی منابع عرضه شده، نوبت به شناسایی موضوعات آنها شد که برای این منظور از مدل‌سازی موضوعی بهره گرفته شد. در مدل‌سازی موضوعی هر موضوع شامل مستندات است و هر مستند با تابع توزیع احتمالی توسط یک موضوع تشریح می‌شود. همچنین هر عبارت دارای تابع توزیع احتمالی است و هر عبارت نیز با یک احتمال به یک موضوع تعلق خواهد داشت. در انتخاب تعداد موضوعات، فاکتورهای مختلفی تعیین کننده است، اگر تعداد موضوعات بسیار زیاد باشد تفسیر آنها دشوار خواهد بود، از طرفی تعداد اندک موضوعات باعث خواهد شد تا موضوعات بیش از حد در یکدیگر ترکیب شده باشند و اشتراک آنها زیاد باشد. برای این منظور از معیار سرگشتگی استفاده می‌شود و تعداد بهینه موضوعات از این طریق شناسایی خواهد شد که با توجه به این معیار عدد ۱۰ برای تعداد موضوعات بدست آمد. در نهایت منابع در این ده موضوع توزیع شدند و بر اساس عبارت‌هایی که بیشترین درجه تعلق را به هر موضوع داشتند این موضوعات نامگذاری شد. این ده موضوع عبارتند از:

«اقتصاد انرژی»، «محیط زیست»، «تکنولوژی»، «انتقال انرژی»، «انرژی خورشیدی»، «انرژی‌های تجدیدپذیر»، «نفت»، «مدیریت انرژی»، «گاز» و «برق».

از آنجا که تمام منابع توسط تمام موضوعات توسط یک ضریب احتمالی بین ۰ تا ۱ تشریح می‌شوند، انتخاب یک حد آستانه برای تشریح شدن یک منبع توسط یک موضوع اهمیت دارد. برای این منظور می‌توان از دو فاکتور پوشش پیکره و کیفیت تشریح برای تعیین حد آستانه استفاده کرد. پوشش پیکره نشان می‌دهد که چه بخشی از پیکره توسط دست کم یک موضوع تشریح شده است و هر چه این مقدار بیشتر باشد مطلوب‌تر خواهد بود و کیفیت تشریح مشخص می‌کند که به طور متوسط منابع با چه ضریب احتمالی توسط موضوعات تشریح شده‌اند و هر چه این میزان نیز بالاتر باشد مطلوب‌تر خواهد بود. اما هر چه حد آستانه بالاتری برای تشریح انتخاب شود، کیفیت تشریح بیشتر خواهد شد اما پوشش پیکره کمتر خواهد بود و برعکس. از اینرو امکان انتخاب حد آستانه مطلوب با استفاده از این دو پارامتر مهیا خواهد بود. نتایج بدست آمده حد آستانه ۰/۱ را بدست آورد.

نکته قابل توجه این است که با توجه به حد آستانه انتخاب شده، بسیاری از منابع در چند موضوع قرار می‌گیرند و توسط موضوعات مختلف تشریح خواهند شد.

## ۵-۲. سوال دوم پژوهش

دومین سوال این پژوهش، معطوف به شناسایی جستجوهای انجام شده در پایگاه گنج است که مرتبط با حوزه انرژی باشد و همچنین به دنبال دسته‌بندی این حوزه به موضوعات مرتبط با آن است.

برای پاسخ به این سوال نیز گام‌های شکل ۳-۲ طی شد. در مرحله اول تمامی جستجوهای انجام شده در پایگاه گنج در بازه زمانی ۱۳۹۶/۱۲/۱۲ تا ۱۳۹۹/۰۶/۱۷ که شامل تعداد ۲۴،۲۰۲،۷۷۴ رکورد بود، گردآوری شد. متن جستجوی ذخیره شده در پایگاه گنج، علاوه بر آنکه محتوای جستجو را شامل می‌شد شامل عبارت‌های کنترلی مختلفی نیز بود که نشان می‌داد جستجو به چه شکلی، در کدام یک از فیلدهای پایگاه (مثلا در چکیده) و تحت چه شرایطی (مثلا با فایل تمام‌متن) انجام شده است، همچنان تعداد نتایجی که موتور جستجو برای این جستجو برگردانده بود نیز در متن جستجو ذخیره شده بود. از اینرو لازم بود تا این متن پالایش شده و محتوای جستجو از عبارت‌های کنترلی تفکیک شود. این پالایش به دلیل اینکه باید بود روی بیش از ۲۴ میلیون رکورد انجام شود، پیچیدگی‌های خاص خود را داشت که با استفاده از پردازش موازی انجام شد و محتوای جستجو استخراج شد.

در ادامه با توجه به آنکه کلمات پرتکرار متون مرتبط با حوزه انرژی از منابع عرضه‌شده استخراج شده بود، آن‌دسته از جستجوها که شامل کلمات پرتکرار حوزه انرژی بود، شناسایی شده و به عنوان منابع تقاضا شده در حوزه انرژی مدنظر قرار گرفت. این جستجوها شامل ۴۴۳،۸۹۰ رکورد شد. برای آنکه بتوان این جستجوها را نیز مشابه با منابع عرضه‌شده در موضوعات شناسایی شده نگاشت کرد، یادگیری مدل با استفاده از منابع عرضه‌شده انجام شد و سپس همان موضوعات از منابع جستجو شده استخراج شد و موضوعات ده‌گانه انرژی در بعد تقاضا تعیین شد. نکته مهم اینکه با توجه به حد آستانه‌ای که برای تعلق یک عبارت به یک موضوع در نظر گرفته شد، بسیاری از جستجوها در چند موضوع قرار می‌گیرند.

### ۵-۱-۳. سوال سوم پژوهش

برای پاسخ به سوال سوم این پژوهش، لازم است تا ابتدا میزان عرضه و تقاضای هر یک از موضوعات نرمال‌سازی شود. این فرایند با نگاشت هر یک از متغیرها بین ۰ تا ۱ انجام شد که به این ترتیب بیشترین فراوانی مستندات در عرضه و تقاضا مقدار ۱ و کمترین آن مقدار ۰ گرفته است و دیگر عناصر بر همین اساس توزیع شده‌اند. در ادامه ماتریس عرضه و تقاضا با ۴ بخش تهیه شد که این بخش‌ها شامل عرضه زیاد و تقاضای زیاد (موضوعات داغ)، عرضه زیاد و تقاضای کم (موضوعات خاموش)، عرضه کم و تقاضای زیاد (موضوعات دارای کاستی) و عرضه کم و تقاضای کم (موضوعات سرد) است.

توزیع موضوعات در این ماتریس نشان می‌دهد که موضوعات «محیط زیست» و «مدیریت انرژی» در مجموعه «موضوعات داغ» جای دارند، این موضوعات علاوه بر اینکه به میزان قابل توجهی تولید شده‌اند، به میزان زیادی نیز در جستجوهای کاربران مورد توجه قرار گرفته‌اند. به عبارتی پژوهش‌های این موضوعات همچنان در حال استفاده، بسط و گسترش هستند. موضوعات «برق»، «تکنولوژی»، «اقتصاد

انرژی»، «انتقال انرژی» و «انرژی‌های تجدیدپذیر» در دسته موضوعات خاموش قرار می‌گیرند. موضوعات «نفت» و «گاز» در مجموعه «موضوعات سرد» قرار گرفته‌اند که با توجه به تعداد اندک پژوهش‌های انجام گرفته در موضوعات این ناحیه و نیز عدم توجه کافی جستجوگران پژوهشی حوزه انرژی به موضوعات این ناحیه، می‌توان پیش‌بینی نمود که ادامه این روند موجب خواهد شد که در آینده این موضوع به تدریج حذف شده و اهمیت خود را به عنوان یک موضوع مستقل از دست خواهد داد. در نهایت موضوع «انرژی خورشیدی» و با کمی اغماض موضوع «نفت» موضوعاتی است که دارای کاستی پژوهشی هستند.

## ۲-۵. پیش‌بینی عرضه و تقاضای موضوعات حوزه انرژی در سال‌های آتی

پیش‌بینی عرضه و تقاضای پژوهش‌های حوزه انرژی در قلمرو این پژوهش نیست و رویکردهای مختلفی و جامعی برای پیش‌بینی آن وجود دارد که امکان بررسی و ارائه آنها در این مطالعه نمی‌گنجد. متغیرهای گوناگونی در عرضه و تقاضای پژوهش در سال‌های آتی تاثیرگذار است و ممکن است به دلایل مختلف سیاست‌گذاری پژوهش مسیر آن را دچار تغییر و تحول کند، اما صرف نظر از این متغیرها توجه به یک نکته مهم می‌تواند تا حدی روند عرضه و تقاضا را در سال‌های آتی مشخص کند. نکته اساسی توجه به نوع مخاطبان پایگاه اطلاعاتی گنج است که در این پژوهش به عنوان منبع مورد استفاده قرار گرفته است. از آنجا که عمده کاربران پایگاه اطلاعاتی گنج شامل دانشجویان تحصیلات تکمیلی است که در حال انجام رساله یا پایان‌نامه خود هستند می‌توان چنین فرض کرد که کاربری که موضوع خاصی را در پایگاه اطلاعاتی گنج جستجو می‌کند، به احتمال زیاد روی همان موضوع یا مفاهیم نزدیک به آن در حال انجام رساله یا پایان‌نامه خود است و یا اینکه به این موضوع علاقه‌مند است و ممکن است همین موضوع را به عنوان پیشنهادیه ارائه نموده و در آینده روی این موضوع پژوهش انجام دهد. از آنجا که مدت زمان انجام رساله دکتری در حدود ۲ تا ۳ سال پس از پیشنهادیه است و مدت زمان انجام پایان‌نامه در حدود یک سال پس از پیشنهادیه است، از این رو می‌توان اینگونه فرض کرد که در فاصله زمانی ۱ تا ۳ سال آینده، شاهد تولید و عرضه موضوعاتی خواهیم بود که در حال حاضر در پایگاه مورد جستجو قرار می‌گیرند. شکل ۵-۱ این مطلب را نشان می‌دهد.



شکل ۵-۱- پیش‌بینی عرضه و تقاضای پارساهای حوزه انرژی

همان‌طور که در شکل ۵-۱ مشخص شده است، موضوعات دارای کاستی به سبب آنکه تقاضای زیادی پایگاه و پیرامون آنها شکل می‌گیرند این امید وجود دارد که پژوهشگرانی که آن را دنبال می‌کنند در سال‌های آینده در این موضوع تولیداتی را به جامعه پژوهشی عرضه کنند، از این‌رو عرضه این موضوعات در سال‌های آینده بیشتر خواهد شد و انتظار می‌رود که در سال‌های بعد این موضوعات به موضوعات داغ تبدیل شوند.

با همین منوال، موضوعات داغ که دارای عرضه و تقاضای زیاد هستند، انتظار می‌رود که در سال‌های آتی نیز جزو موضوعات داغ باقی بمانند. این در حالی است که موضوعات خاموش که شامل ۵ مورد از موضوعات حوزه انرژی است، به سبب تقاضای پژوهشی کم، احتمالاً علاقه‌مندان کمتری داشته و در سال‌های آتی نیز شاهد تولید اندک آنها خواهیم بود. از این‌رو این احتمال وجود دارد که این موضوعات در سال‌های آتی به موضوعات سرد تبدیل شوند. همچنین موضوعات سرد هم این احتمال را دارند که سال‌های آینده همچنان در انزوا باقی بمانند.

### ۳\_۵. محدودیت‌های پژوهش

وجود محدودیت‌هایی در این پژوهش مانع از آن شد که بتوان تحلیل‌های عمیق‌تری از عرضه و تقاضای پژوهش در این حوزه را ارائه نمود که در صورت برطرف شدن امکان بسط پژوهش و تحلیل‌های بیشتر وجود داشت. از مهم‌ترین محدودیت‌های این پژوهش می‌توان به موارد زیر اشاره کرد:

- محدودیت در بازه زمانی داده‌های جستجوی کاربران  
داده‌های جستجو صرفاً مربوط به حدود ۲ سال و نیم پیش است و به همین دلیل امکان تحلیل روند زمانی تغییرات الگو یا موضوع جستجوها فراهم نیست. در صورتی که این بازه مربوط به دوره طولانی‌تری باشد، می‌توان تغییرات آن را در گذر زمان براساس موضوعات مختلف رصد نمود.
- محدودیت در فیلدهای اطلاعاتی پارساها  
امکان ارائه برخی از فیلدهای اطلاعاتی وجود نداشت و تحلیل پدیدآوران پارساها می‌توانست اطلاعات ارزشمندی را در مورد افراد تاثیرگذار هر یک از موضوعات حوزه انرژی در اختیار قرار دهد. همچنین با استفاده از تحلیل شبکه ارتباطات پدیدآوران و مقایسه آن با ارتباطات بین موضوعات امکان تحلیل روند تغییرات موضوعات و رشد یا زوال آنها در چرخه عمر موضوعات پژوهشی آشکار می‌شود.

#### ۴-۵. پیشنهاد پژوهش‌های آتی

انجام این پژوهش سرآغازی برای انجام مطالعات دیگر در حوزه تحلیل منابع اطلاعاتی با رویکرد عرضه و تقاضای آن در اکوسیستم پژوهش است. از جمله پیشنهادهایی که می‌تواند توسط علاقه‌مندان این حوزه در پژوهش‌های آتی مورد توجه قرار گیرد، می‌توان به موارد زیر اشاره کرد:

##### ۴-۵-۱. تحلیل توزیع جغرافیایی پژوهش‌ها و تعیین کاستی‌های مکان‌محور

بسیاری از پژوهش‌های علم‌سنجی ناگزیر هستند که به بررسی جغرافیای مرتبط با تولیدات علمی پردازند و مقایسه بین کشورها، استان‌ها، شهرها و مناطق جغرافیایی را برای تعیین چگونگی توزیع جغرافیایی پژوهش‌ها مدنظر قرار دهند (حمیدی، اصنافی، و عصاره، ۱۳۸۷). اطلاعات جغرافیایی رابطه بسیار نزدیکی با مصورسازی اطلاعات<sup>۱</sup> دارد که هدف اصلی آن نمایش ارتباطات موجود بین مناطق جغرافیایی پیرامون اطلاعات موجود در یک حوزه خاص است. به عنوان مثال نمایش ارتباطات بین شهرهای مختلف جهان که در حوزه‌ی خاصی به تولید پژوهش پرداخته‌اند از طریق پیوندها و گره‌ها در یک شبکه مبتنی بر گراف نمونه‌ای از مصورسازی اطلاعات است که با در نظر گرفتن ابعاد مختلف علم

و با به‌کارگیری سیستم‌های اطلاعات جغرافیایی<sup>۱</sup> منجر به تولید نقشه دانش<sup>۲</sup> خواهد شد (اسدی و جلالی‌منش، ۱۳۹۲). نقشه دانش کاربردهای فراوانی در برنامه‌ریزی‌های کلان و سیاست‌گذاری علم و فناوری دارد و بوسیله آن توزیع علم و فناوری قابل تبیین و نمایش خواهد بود. تجزیه تحلیل محدودیت‌ها و تعیین اولویت‌ها می‌تواند با توجه به منطقه جغرافیایی انجام شود و مؤلفه بسیار مهم در این حوزه آگاهی یافتن از نیازهای کاربران مناطق مختلف است.

از این‌رو یکی از موضوعاتی که می‌تواند در پژوهش‌های آتی این حوزه مدنظر قرار گیرد، تحلیل توزیع جغرافیایی پژوهش‌ها و شناسایی کاستی‌های پژوهش بر مبنای جغرافیا است. این مسئله می‌تواند سیاست‌گذاران و مدیران پژوهشی را تقسیم جغرافیایی منابع پژوهشی کمک کرده و نیازهای پژوهشی مناطق مختلف کشور را در مطالعات آمایش سرزمینی به‌صورت دقیق تشریح و ترسیم نماید.

#### ۵-۴-۲. شناسایی نیازهای اطلاعاتی پنهان و بیان نشده در تقاضاهای کاربران

همان‌گونه که در بخش مفروضات پژوهش بیان شد، در مطالعه تقاضاهای پژوهشی فرض اولیه بر این است که خواسته‌های کاربران با نیازهای اطلاعاتی آن‌ها منطبق است. هر چند که تردیدی وجود ندارد که تقاضاهای کاربران بخشی از نیازهای اطلاعاتی کاربران است (بابایی، ۱۳۸۶) اما باید تأکید کرد که بخشی از نیازهای اطلاعاتی نیز به دلایل مختلف از جمله ناآگاهی آنان از موضوع پژوهشی، ناامیدی آن‌ها از یافتن پاسخ مناسب از منبع اطلاعاتی، الزام و اجبار بیرونی به ارائه تقاضاهایی که نیاز واقعی کاربر و جامعه نیست و ... هرگز به عنوان تقاضا یا خواسته بیان نشود.

از سوی دیگر کاربران خواسته‌های فردی خود را در محدوده منابع موجود مطرح می‌کنند. این بدان معناست که کاربران نیازهای دیگری نیز دارند که به دلیل در دسترس نبودن منابع اطلاعاتی و یا درک و نگرشی که آن‌ها نسبت به منبع اطلاعاتی در دسترس خود دارند، از بیان آن‌ها خودداری می‌کنند (بابایی، ۱۳۸۶). همچنین یکی از موانع اساسی در عدم بیان درست نیازهای اطلاعاتی در تقاضاهای کاربر، فقدان یا ضعف سواد اطلاعاتی<sup>۳</sup> کاربران است. سواد اطلاعاتی مجموعه مهارت‌هایی است که فرد را قادر می‌سازد نیاز اطلاعاتی خود را تشخیص داده و توانایی جستجوی کارآمد در منابع اطلاعاتی را داشته و پس از انجام جستجو اطلاعات به دست آمده را ارزیابی کرده و به‌منظور تولید اطلاعات جدید پیوند لازم بین اطلاعات جدید را با دانش قبلی خود برقرار سازد (Baji, Bigdeli, Parsa, and Haeusler, 2018).

---

<sup>۱</sup>Geographical Information System (GIS)

<sup>۲</sup>Knowledge Map

<sup>۳</sup>Information Literacy

از این رو بخشی از نیازهای پژوهشی به صورت نیاز اطلاعاتی نهفته<sup>۱</sup> است و شامل نیازهای اطلاعاتی می‌شود که کاربر از وجود آن‌ها آگاه نیست و صرفاً زمانی می‌تواند در صورت وجود خلأ اطلاعاتی از آن‌ها بهره‌برداری کند که آن اطلاعات در معرض دید او قرار بگیرند. این نوع از نیاز نشان‌دهنده این نکته مهم است که مفهوم نیازهای پژوهشی تنها معطوف به نیازهای کنونی افراد نیست و آنچه فرد به صورت بالقوه به آن نیاز دارد، مورد توجه قرار می‌گیرد. همچنین بخشی از نیازهای پژوهشی به صورت نیاز اظهارنشده<sup>۲</sup> است و به دلایل مختلف این نیاز بیان نشده است و در نتیجه امکان تحلیل و ارزیابی آن نیز فراهم نخواهد بود.

یکی از موضوعاتی که می‌تواند در پژوهش‌های آتی مورد توجه پژوهشگران حوزه علم‌سنجی و تحلیل اطلاعات قرار گیرد، کشف نیازهای پنهان و اظهارنشده پژوهشی است که می‌تواند در کنار هدف پژوهش حاضر که شناسایی کاستی‌های پژوهش از منظر تقاضاهای بیان شده کاربران پایگاه اطلاعات علمی است، اطلاعات جامع‌تری از نیازهای پژوهشی کشور را استخراج کرده و برای سیاست‌گذاری پژوهش کشور در اختیار سیاست‌گذاران این حوزه قرار داد.

#### ۵-۴-۳. تحلیل چرخه عمر موضوعات و مفاهیم همجوشی اطلاعات<sup>۳</sup>

بررسی قدرت تفکیک موضوعات نزدیک به هم و همچنین ارزیابی روند همگرایی یا واگرایی موضوعات پژوهشی مرتبط باهم و نیز چگونگی تشکیل، رشد، ترکیب و تجزیه موضوعات تشکیل دهنده یک حوزه پژوهشی با استفاده از روش ارائه شده در این پژوهش، می‌تواند در مطالعات آتی مورد بررسی قرار بگیرد. این موضوع که در علم اطلاعات در مباحث همجوشی اطلاعات یا ادغام اطلاعات مورد استفاده قرار می‌گیرد بر تولید اطلاعات با ارزش از طریق ادغام و ترکیب داده‌ها یا اطلاعات تمرکز دارد و اهمیت بالایی در شناخت علم و تحلیل چستی و چگونگی تشکیل مفاهیم جدید علمی دارد. همچنین این موضوع در حل مسائل پیچیده از طریق ادغام اطلاعات منابع مختلف و دستیابی به دانش با سطح بالاتر که از طریق هم‌افزایی اطلاعات مختلف به دست آمده است دلالت دارد (Leung, Ji, and Ma, 2013).

بررسی چگونگی ظهور کلیدواژه‌ها یا مفاهیم جدید در یک حوزه پژوهشی، مدت‌زمانی که طول می‌کشد تا یک عبارت جدید به عنوان یک موضوع پژوهشی مورد استقبال قرار گیرد، زمان و چگونگی بلوغ و تکامل یک موضوع پژوهشی، زایش یا کاهش ابعاد یک موضوع و ... مفاهیمی است که می‌توان

---

Dormant Need<sup>۱</sup>

Unexpressed Need<sup>۲</sup>

Information fusion<sup>۳</sup>

در این منظر مورد بررسی قرار داد. از این رو یکی از موضوعات مورد توجه در پژوهش‌های آتی می‌تواند تحلیل چرخه عمر موضوعات و همچنین بررسی مفاهیم همجوشی اطلاعات در حوزه‌های مختلف پژوهشی با استفاده از تحلیل رفتار جویشرانه کاربران پایگاه اطلاعاتی باشد.

#### ۵\_۵. جمع‌بندی و نتیجه‌گیری

در این فصل جمع‌بندی مباحث مطرح شده در فصول گذشته ارائه شد، همچنین فرایند و گام‌های طی شده در پژوهش به اختصار شرح داده شد و اهداف و پرسش‌های پژوهش بار دیگر مرور شد. در ادامه پاسخ هر یک از پرسش‌ها با توجه به نتایج بدست آمده در مراحل پیشین ارائه شد. با توجه به نوع مخاطبان پایگاه اطلاعاتی ایرانداک، عرضه و تقاضای پژوهش در حوزه انرژی در سال‌های آتی پیش‌بینی شد و در ادامه محدودیت‌هایی که در فرایند انجام این پژوهش وجود داشت تشریح شد. همچنین پیشنهادهایی برای انجام مطالعات بعدی در این زمینه ارائه شد که می‌تواند مورد استفاده پژوهشگران و علاقه‌مندان آن قرار گیرد.

# فهرست منابع

## ۱-۶. منابع فارسی

- اسدی، سعید. ۱۳۹۳. سنجش و مصور سازی پراکندگی جغرافیایی ثروت علمی: مفاهیم و کاربردها، ویرایش اول، تهران: پژوهشگاه علوم و فناوری اطلاعات ایران.
- اسدی، سعید و عمار جلالی منش. ۱۳۹۲. نگاشت و مصور سازی پراکندگی جغرافیایی ثروت علمی در ایران. پژوهشنامه پردازش و مدیریت اطلاعات، ۲۸(۴)، ۹۱۷-۹۴۳.
- امیرحسینی، مازیار. ۱۳۷۱. کتابسنجی و اطلاع سنجی. فصلنامه کتاب، ۳، ۱۸۳-۲۰۹.
- بابایی، محمود. ۱۳۸۶. نیازسنجی اطلاعات (شماره ۳). تهران: پژوهشگاه اطلاعات و مدارک علمی ایران.
- توکلی، محمد جواد. ۱۳۸۸. اقتصاد پژوهش؛ آسیب شناسی معادلات اقتصادی حاکم بر بازار پژوهش. عیار پژوهش در علوم انسانی (پژوهش سابق)، ۱۳۸۸(۲)، ۱۴۹-۱۹۱.
- تیرگر، آرام، موسی یمین فیروز، و زهرا دهقان. ۱۳۹۳. ارزیابی کیفی بروندهای علمی پژوهشگران با استفاده از واژه‌های کلیدی مشابه. مجله علم سنجی کاسپین، ۱، ۱۳-۱۸.
- حمیدی، علی، امیررضا اصنافی و فریده عصاره. ۱۳۸۷. بررسی تحلیلی و ترسیم ساختار انتشارات علمی تولید شده در حوزه‌های کتاب سنجی، علم سنجی، اطلاع سنجی و وب سنجی در پایگاه Web of Science طی سالهای ۱۹۹۰-۲۰۰۵. کتابداری و اطلاع رسانی، ۱۱(۲)، ۱۶۱-۱۸۲.
- خاصه، علی اکبر و فرامرز سهیلی. ۱۳۹۷. ترسیم چشم انداز پژوهش در علم سنجی و حوزه‌های سنجشی وابسته. فصلنامه علمی پژوهشی پردازش و مدیریت اطلاعات، ۳۳(۳)، ۹۴۱-۹۶۶.
- ربیعی، محمد. ۱۳۹۸. شناسایی کاستی‌های پژوهش با به کارگیری تکنیک‌های متن کاوی و تحلیل رفتار جوی‌شگری کاربران مورد مطالعه: محیط‌زیست ایران. (ر ساله دکترا)، دانشگاه علم و صنعت ایران، تهران.
- شورای عالی انقلاب فرهنگی. ۱۳۹۰. سند نقشه جامع علمی کشور. تهران: شورای عالی انقلاب

غضنفری، مهدی، سمیه عزیزاده و بابک تیمورپور. ۱۳۹۳. داده کاوی و کشف دانش. تهران: انتشارات دانشگاه علم و صنعت ایران.

فتاحی، رحمت‌الله. ۱۳۹۰. حکایت استفاده کاربردی از پژوهش‌ها در ایران، ۴۶. فصلنامه علوم و فناوری اطلاعات، ۷۷۷-۷۷۹.

فتحی‌واجارگاه، کورش، علی‌اکبر خسروی، آیت‌الله سعادت‌طلب، حیدر تورانی و طلعت دیباواجاری. ۱۳۸۹. طراحی مکانیزم اولویت‌یابی پژوهشی برای نظام آموزش و پرورش ایران (مورد آموزش و پرورش تهران). مطالعات تربیتی و روان‌شناسی، ۱۱(۲)، ۱۸۵-۲۱۰.

قنادی‌نژاد، فرزانه و غلامرضا حیدری. ۱۳۹۶. اولویت‌های پژوهشی و انتخاب موضوع در علم اطلاعات و دانش‌شناسی: چار.

قنادی‌نژاد، فرزانه، غلامرضا حیدری و رحیم و چینی پرداز، ۱۳۹۷. تحلیل محتوای متون مربوط به اولویت‌های پژوهشی در علم اطلاعات و دانش‌شناسی. پژوهشنامه کتابداری و اطلاع‌رسانی، ۸، ۷۴-۵۵.

محمداسماعیل، صدیقه، عارف ریاحی و فریبا صحبتی‌ها. ۱۳۹۳. ارزیابی کمی و کیفی مجلات ایران در پایگاه استنادی اسکوپوس طی سال‌های ۲۰۱۲-۲۰۲۰. مجله علم‌سنجی کاسپین، ۱، ۳۳-۳۹.

نیاکان، شهرزاد. ۱۳۸۳. نمایه‌سازی ماشینی. تهران: مرکز اطلاعات و مدارک علمی ایران.

## ۲-۶. منابع انگلیسی

- Arun, R., Suresh, V., Madhavan, C. V., and Murthy, M. N. 2010. *On finding the natural number of topics with latent dirichlet allocation: Some observations*. Paper presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining.
- Bahadori, M., Teimourzadeh, E., Farzaneh, A., and Nejati, M. 2011. Prioritizing research needs: insights from a healthcare organization in Iran. *Archives of Pharmacy Practice*, 2(3), 135.
- Baji, F., Bigdeli, Z., Parsa, A., and Haeusler, C. 2018. Developing information literacy skills of the 6th grade students using the Big 6 model. *Malaysian Journal of Library & Information Science*, 23(1), 1-15.
- Björneborn, L., and Ingwersen, P. 2004. Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.
- Blei, D. M. 2012. Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Braun, T., Glänzel, W., and Schubert, A. 1985. *Scientometric indicators: A 32 country comparative evaluation of publishing performance and citation impact*: World Scientific.
- Chang, J., and Blei, D. M. 2009. *Relational topic models for document networks*. Paper presented at the International conference on artificial intelligence and statistics.
- Choudhary, A. K., Oluikpe, P., Harding, J. A., and Carrillo, P. M. 2009. The needs and benefits of Text Mining applications on Post-Project Reviews. *Computers in Industry*, 60(9), 728-740.
- Cohen, J. I. 1994. *Biotechnology priorities, planning, and policies: a framework for decision making*: International Service for National Agricultural Research.
- Davarpanah, M. R. 2012. Scientometric analysis of nuclear science and technology research output in Iran. *journal of Scholarly Publishing*, 43(4), 421-439.
- De Battisti, F., Ferrara, A., and Salini, S. 2015. A decade of research in statistics: a topic model approach. *Scientometrics*, 103(2), 413-433.
- De Bellis, N. 2009. *Bibliometrics and citation analysis: from the science citation index to cybermetrics*: Scarecrow Press.
- Delen, D., and Crossland, M. D. 2008. Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, 34(3), 1707-1720.
- Dutt, B., and Nikam, K. 2016. Scientometric analysis of global solar cell research.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996a. From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.

- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996b. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34.
- Feinerer, I., Hornik, K., and Feinerer, M. I. 2018. Package 'tm'. *Corpus*, 10, 1.
- Gobster, P. H. 2014. (Text) Mining the LANDscape: Themes and trends over 40 years of Landscape and Urban Planning. *Landscape and Urban Planning*, 126, 21-30.
- Gregório, G., Tomlinson, M., Gerolin, J., Kieling, C., Moreira, H. C., Razzouk, D., and Mari, J. d. J. 2012. Setting priorities for mental health research in Brazil. *Revista Brasileira de Psiquiatria*, 34(4), 434-439.
- Hall, D., Jurafsky, D., and Manning, C. D. 2008. *Studying the history of ideas using topic models*. Paper presented at the Proceedings of the conference on empirical methods in natural language processing.
- John W, M., and Petko, B. 2013. Introduction-Topic models: What they are and why they matter. *Poetics*, 41(6), 545-569. doi: <https://doi.org/10.1016/j.poetic.2013.10.001>.
- Junnarkar, B. 1997. Leveraging collective intellect by building organizational capabilities. *Expert Systems with Applications*, 13(1), 29-40.
- Jurkiewicz, J., Wendykier, P., Wojciechowski, K., Fedoryszak, M., and Dendek, P. J. 2014. Common Map of Academia: augmenting bibliography research information data. *Procedia Computer Science*, 33, 74-79.
- Kurt, H. 2017. Natural Language Processing Infrastructure (Version 0.1-11): CRAN. Retrieved from <https://cran.r-project.org/web/packages/NLP/index.html>
- Lafferty, J. D., and Blei, D. M. 2006. *Correlated topic models*. Paper presented at the Advances in neural information processing systems.
- Leung, Y., Ji, N.-N., and Ma, J.-H. 2013. An integrated information fusion approach based on the theory of evidence and group decision-making. *Information Fusion*, 14(4), 410-422.
- Leydesdorff, L. 2001. *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*: Universal-Publishers.
- Losiewicz, P., Oard, D. W., and Kostoff, R. N. 2000. Textual data mining to support science and technology management. *Journal of Intelligent Information Systems*, 15(2), 99-119.
- Mai, J.-E., Case, D. O., and Given, L. M. 2016. *Looking for information: A survey of research on information seeking, needs, and behavior*: Emerald Group Publishing.
- Masinter, A., Small, M., and Casman, E. 2014. Research prioritization using hypothesis maps. *Environment Systems and Decisions*, 34(1), 49-59.
- McBriar, I., Smith, C., Bain, G., Unsworth, P., Magraw, S., and Gordon, J. L. 2003. Risk, gap and strength: key concepts in knowledge management. *Knowledge-Based Systems*, 16(1), 29-36.

- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., . . . Meyer, M. D. 2018. Package 'e1071'. from <https://cran.r-project.org/web/packages/e1071/index.html>
- Nasser, M., and Welch, V. 2013. Prioritization of systematic reviews leads prioritization of research gaps and needs. *Journal of clinical epidemiology*, 66(5), 522.
- Nazaridoust, M., Bidgoli, B. M., and Rezaeenoor, J. 2013. Providing a Triangular Model for Gap Analysis Case Study: Iran Khodro Company.
- Nieminen, P., Pölönen, I., and Sipola, T. 2013. Research literature clustering using diffusion maps. *Journal of Informetrics*, 7(4), 874-886.
- O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., and Ananiadou, S. 2015. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews*, 4(1), 1.
- Rabiei, M., Hosseini-Motlagh, S.-M., and Haeri, A. 2017. Using text mining techniques for identifying research gaps and priorities: a case study of the environmental science in Iran. *Scientometrics*, 110(2), 815- 842.
- Rabiei, M., Hosseini-Motlagh, S.-M., Haeri, A., and Minaei Bidgoli, B. 2021. Evolution of IT, management and industrial engineering research: A topic model approach. *Scientia Iranica*, 28(3), 1830-1852.
- Ranganathan, C., and Balasubramani, R. 2014. Scientometric profile of research activities on green energy: An Indian perspective. *International Journal of Research in Applied, Natural and Social Sciences*, 2(1), 23-30.
- Robinson, K. A., Saldanha, I. J., and Mckoy, N. A. 2011. Development of a framework to identify research gaps from systematic reviews. *Journal of clinical epidemiology*, 64, 1325-1330.
- Romo-Fernández, L. M., Guerrero-Bote, V. P., and Moya-Anegón, F. 2013. Co-word based thematic analysis of renewable energy (1990–2010). *Scientometrics*, 97(3), 743-765.
- Schubert, E., Weiler, M., and Kriegel, H.-P. 2014. *Signitrend: scalable detection of emerging topics in textual streams by hashed significance thresholds*. Paper presented at the Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.
- Shamsfard, M. 2008. *Towards Semi Automatic Construction of a Lexical Ontology for Persian*. Paper presented at the LREC.
- Singh, R., and Mohaar, G. S. 2016. *Fast Document Indexing Using Aho-Corasick State Machine*. Paper presented at the Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on.
- Sunikka, A., and Bragge, J. 2012. Applying text-mining to personalization and customization research literature—Who, what and where? *Expert Systems with Applications*, 39(11), 10049-10058.

- Tsai, H.-H. 2011. Research trends analysis by comparing data mining and customer relationship management through bibliometric methodology. *Scientometrics*, 87(3), 425-450.
- Tsai, H.-H. 2013. Knowledge management vs. data mining: Research trend, forecast and citation approach. *Expert Systems with Applications*, 40(8), 3160-3173.
- Tu, Y.-N., and Seng, J.-L. 2009. Research intelligence involving information retrieval—An example of conferences and journals. *Expert Systems with Applications*, 36(10), 12151-12166.
- Vállez, M., Pedraza-Jiménez, R., Codina, L., Blanco, S., and Rovira, C. 2015. A semi-automatic indexing system based on embedded information in HTML documents. *Library Hi Tech*, 33(2), 195-210.
- Wallach, H. M. 2006. *Topic modeling: beyond bag-of-words*. Paper presented at the Proceedings of the 23rd international conference on Machine learning, Pittsburgh, Pennsylvania, USA.
- Wilson, C. S. 1999. Informetrics. *Annual Review of Information Science and Technology (ARIST)*, 34, 107-247.
- Yang, D., Guo, J.-e., Li, J., Wang, S., and Sun, S. 2020. Knowledge Mapping in Electricity Demand Forecasting: A Scientometric Insight. *arXiv preprint arXiv:2003.10055*.
- Yang, L., Wang, Q., Bai, X., Deng, J., and Hu, Y. 2018. Mapping of trace elements in coal and ash research based on a bibliometric analysis method spanning 1971–2017. *Minerals*, 8(3), 89.
- Zhao, W., Wu, R., Dai, W., and Dai, Y. 2015. *Research paper recommendation based on the knowledge gap*. Paper presented at the 2015 IEEE International Conference on Data Mining Workshop (ICDMW).
- Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A., . . . Shen, B. 2013. Biomedical text mining and its applications in cancer research. *Journal of biomedical informatics*, 46(2), 200-211.
- Zhu, L., Liu, X., He, S., Shi, J., and Pang, M. 2015. Keywords co-occurrence mapping knowledge domain research base on the theory of Big Data in oil and gas industry. *Scientometrics*, 105(1), 249-260.

## **ABSTRACT**

The growing need to use energy and the spread of science and technology and the interaction between researchers and industry has led to the multiplicity and complexity of research and the diversity of research needs and demands in this field. In order to manage the research system in this field, it is necessary to have sufficient knowledge about the demands and needs of research and capabilities and existing research resources in the field of energy.

This research tries to provide a suitable solution based on the of the content analysis of existing research and the demands of the research society in the field of energy and through the application of text mining and natural language processing techniques and analysis of users' behavior in an expert\_independent approach. For this purpose, the resources available in Irandoc database as well as the search queries performed by the users of this database were used as research data.

The research process identified 25,953 records of Ganj repository as supplied resources and 443,890 of queries performed as demanded. The bibliographic analysis shows that since 2010 the supply and demand in this domain have grown a lot. The gap analysis of the supply and demand amount ten identified topics of energy has shown their position in the supply and demand matrix. The results show that the topics of "Environment" and "Energy Management" are hot topics in the field of energy, "Oil" and "Gas" are cold topics, "Technology", "Energy Economy", "Renewable Energy" and "Electricity" are the silent topics in the field of energy and the topic of "Solar Energy" was identified as research gap.

The presented research can help policy makers and research managers in the field of energy in making major research decisions to determine research needs and priorities. Also, due to the interdisciplinary approach of this study, it could be applied by researchers and those interested in various fields of information technology, artificial intelligence, librarianship and science and technology policy makers.

### **Keywords:**

Research Supply and Demand, Energy, Topic Modeling, Text Mining, Ganj Repository, IranDoc



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

Research project

# **Supply and Demand Analysis of Energy Theses and Dissertations in the Ganj Repository**

**By:  
Mohammad Rabiei**

**and  
Somayeh Fatahi**

**September 2021**