



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

تعیین اولویت‌های پژوهش در حوزه محیط زیست با تحلیل رفتار پژوهشگران

محمد ربیعی

عضو هیئت علمی پژوهشگاه علوم و فناوری اطلاعات ایران

۱۳۹۵

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

داده‌های مربوط به رفتار کاربران در وبسایت‌ها، یکی از منابع مهم اطلاعاتی است که تحلیل آنها می‌تواند با شناخت واقعی از نیازمندی‌های کاربران، دانش ارزشمندی را در حوزه‌های مدیریت و برنامه‌ریزی، طراحی کسب و کار، بهبود کانال‌های ارتباطی، ارتقای نظام‌های خدمت‌رسانی و ... فراهم آورد. پایگاه گنج^۱ به عنوان گنجینه اطلاعات علمی کشور دربردارنده بیش از ۸۰۰ هزار رکورد علمی است که بخش عمده‌ای از آن مربوط به پایان-نامه‌ها و رساله‌های تحصیلات تکمیلی است و بطور متوسط روزانه نزدیک به ۵۰ هزار جست‌وجو در این پایگاه صورت می‌پذیرد. از آنجا که کاربران این پایگاه را عمدتاً پژوهشگران و دانشجویان تحصیلات تکمیلی در رشته‌های مختلف تشکیل می‌دهند، تحلیل رفتار این کاربران در پایگاه اطلاعاتی ایرانداک ما را در شناسایی نیازهای جامعه پژوهشی کشور توانمند می‌سازد. با تحلیل این رفتار می‌توان به سوالات زیر پاسخ داد:

- چه دانش و فناوری در حال انجام است؟
- چه کسی یا کسانی آن را انجام می‌دهند؟
- مرز دانش و فناوری کجاست؟
- مکان‌های انجام آن کجاست؟
- چه دانش و فناوری خیلی مدنظر نیست؟
- مهم‌ترین حوزه‌های مورد تاکید کدامند؟
- چه ارتباطی بین این حوزه‌ها وجود دارد؟
- چه حوزه‌هایی به عنوان حوزه‌های پشتیبان عمل می‌کنند؟

¹ ganj.irandoc.ac.ir

- جهت‌گیری پژوهش‌های جدید به کدام سو است؟
 - نوآوری‌ها و ابداعات چه هستند؟
 - خلاء پژوهشی در چه حوزه‌هایی مشهود است؟
 - روند تغییر تولیدات و نیازهای پژوهشی در طول زمان به چه شکلی است؟
- حوزه محیط زیست به عنوان یکی از چالش‌های اساسی کشور نیازمند برنامه‌ریزی جامع و دقیق در ابعاد مختلف است که پژوهش به عنوان یکی از مهم‌ترین این ابعاد مطرح است. با تحلیل رفتار کاربران حوزه محیط زیست در پایگاه اطلاعاتی گنج و رصد جست‌وجوهای کاربران این حوزه و بکارگیری تکنیک‌های پردازش زبان طبیعی (NLP)^۲ و متن کاوی^۳، روندهای جست‌وجوی کاربران، محتوای موجود در پایگاه اطلاعاتی و ... مورد بررسی قرار گرفت و اولویت‌های پژوهشی در حوزه محیط زیست شناسایی شد.
- در این پژوهش رفتار جست‌وجوگران در ۳۰ ماه گذشته در پایگاه اطلاعاتی مورد کنکاش قرار گرفت که در این میان نزدیک به ۲۰۰ هزار جست‌وجو در حوزه محیط زیست انجام شده بود با تحلیل این جست‌وجوها و روند تغییرات آنها و نیز مقایسه موضوعات مورد جست‌وجو با محتوای اطلاعاتی موجود، خلاء پژوهشی^۴ موجود در حوزه محیط زیست شناسایی شد. در ادامه بخش‌هایی از نتایج اصلی این پژوهش به صورت خلاصه بیان شده است.

جدول‌های زیر فراوانی عبارت‌های جست‌وجو شده در حوزه محیط زیست را نشان می‌دهد. همان‌طور که در این جدول‌ها مشخص است پس از واژه‌های سازنده عبارت «محیط زیست» بیشترین واژه‌های مورد جست‌وجو واژه‌های «هوا»، «ارزیابی»، «آب» و «حقوق» هستند.

^۲ Natural Language Processing

^۳ Text mining

^۴ Research Gap

فراوانی عبارت‌های جست‌وجو شده در حوزه محیط زیست

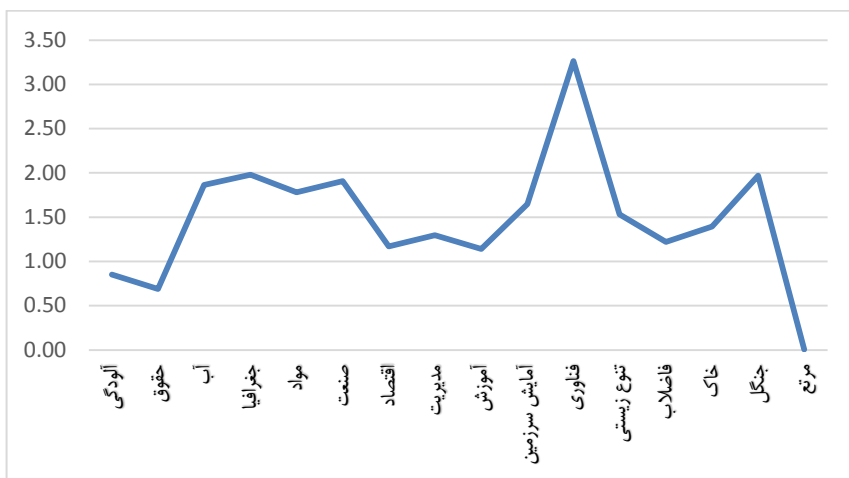
عبارت‌های تک‌واژه‌ای (Uni_Gram)					
ردیف	عبارت	فراوانی	ردیف	عبارت	فراوانی
۱	زیست	۶۱۱۶۲	۱۱	مدیریت	۲۶۰۳
۲	محیط	۳۳۰۵۵	۱۲	آموزش	۲۲۴۵
۳	آلودگی	۲۶۲۱۳	۱۳	خاک	۱۹۵۴
۴	محیطی	۱۸۶۰۲	۱۴	حفاظت	۱۹۵۰
۵	زیستی	۱۳۲۲۹	۱۵	نفتی	۱۹۱۶
۶	هوا	۵۸۰۹	۱۶	انجمن	۱۴۳۴
۷	ارزیابی	۵۰۳۱	۱۷	مسئولیت	۱۴۱۳
۸	آب	۳۹۵۲	۱۸	سنگین	۱۳۲۹
۹	حقوق	۳۵۵۵	۱۹	زیستگاه	۱۲۶۵
۱۰	شناسی	۳۴۴۶	۲۰	تهران	۱۲۲۷

عبارت‌های دو واژه‌ای (Bi_Gram)					
ردیف	عبارت	فراوانی	ردیف	عبارت	فراوانی
۱	محیط زیست	۳۱۶۳۰	۱۱	آموزش محیط	۱۵۲۸
۲	زیست محیطی	۱۸۲۷۶	۱۲	آلودگی محیط	۱۳۵۹
۳	آلودگی هوا	۵۶۰۰	۱۳	آلودگی خاک	۱۱۶۲
۴	زیست شناسی	۲۴۵۴	۱۴	ارزیابان محیط	۱۰۹۰
۵	آلودگی آب	۲۱۹۰	۱۵	فلزات سنگین	۱۰۱۳
۶	تنوع زیستی	۱۶۹۰	۱۶	کود زیستی	۹۵۸
۷	حقوق محیط	۱۶۲۰	۱۷	ارزیابی زیست	۹۲۹
۸	آلودگی صوتی	۱۶۱۶	۱۸	مدیریت زیست	۸۶۴
۹	حفاظت محیط	۱۵۷۳	۱۹	زیست تخریب	۷۵۳
۱۰	آلودگی نفتی	۱۵۵۴	۲۰	مسئولیت مدنی	۷۰۳

عبارت‌های سه واژه‌ای (Tri_Gram)					
ردیف	عبارت	فراوانی	ردیف	عبارت	فراوانی
۱	اثرات زیست محیطی	۲۹۸۶	۱۱	تخریب محیط زیست	۶۳۶
۲	حقوق محیط زیست	۱۶۱۴	۱۲	حسابداری محیط زیست	۶۰۳
۳	حفاظت محیط زیست	۱۵۷۱	۱۳	حفظ محیط زیست	۵۹۲
۴	آموزش محیط زیست	۱۵۱۰	۱۴	زیست تخریب پذیر	۵۷۴
۵	ارزیابی اثرات زیست	۱۳۹۲	۱۵	آلودگی آب زیرزمینی	۵۵۹
۶	آلودگی محیط زیست	۱۲۶۷	۱۶	کیفیت محیط زیست	۵۳۰
۷	ارزیابان محیط زیست	۱۰۹۰	۱۷	رفتار زیست محیطی	۵۱۶
۸	مدیریت زیست محیطی	۸۵۲	۱۸	محیط زیست شهری	۴۶۰
۹	ارزیابی زیست محیطی	۸۴۹	۱۹	آلودگی زیست محیطی	۴۴۹
۱۰	مدیریت محیط زیست	۶۴۶	۲۰	آلودگی فلزات سنگین	۴۱۳

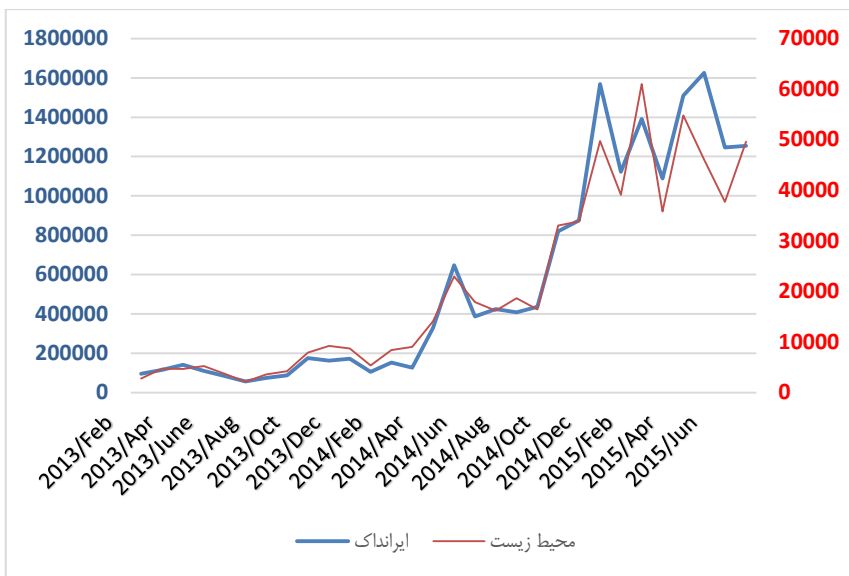
عبارت‌های چهار واژه‌ای (Four_Gram)					
ردیف	عبارت	فراوانی	ردیف	عبارت	فراوانی
۱	ارزیابی اثرات زیست محیطی	۱۳۸۶	۱۱	زمین شناسی زیست محیطی	۳۱۱
۲	ارزیابان محیط زیست هگمتانه	۱۰۹۰	۱۲	بررسی اثرات زیست محیطی	۲۸۱
۳	انجمن ارزیابان محیط زیست	۱۰۹۰	۱۳	ایمنی بهداشت محیط زیست	۲۶۰
۴	حقوق بین الملل محیط	۵۴۶	۱۴	پژوهشگاه ملی مهندسی ژنتیک	۲۴۶
۵	دانشکده عمران محیط زیست	۴۷۱	۱۵	ملی مهندسی ژنتیک وزیست	۲۴۱
۶	دانشکده مرتع آبخیزداری شیلات	۳۵۲	۱۶	مهندسی ژنتیک وزیست فناوری	۲۴۱
۷	آبخیزداری شیلات محیط زیست	۳۵۲	۱۷	جاذب بیولوژیکی تالاب مصنوعی	۱۹۹
۸	مرتع آبخیزداری شیلات محیط	۳۵۲	۱۸	دفن پسماند استفاده بستر	۱۹۲
۹	انجمن محیط زیست کومش	۳۲۶	۱۹	بستر متخلخل جاذب بیولوژیکی	۱۹۲
۱۰	ارزیابی ریسک زیست محیطی	۳۱۷	۲۰	مراکز دفن پسماند استفاده	۱۹۲

شکل یک نسبت تعداد پاسخ به تعداد جست‌وجو را در پایگاه اطلاعاتی ایرانداک نشان می‌دهد، همانگونه که از این شکل مشخص است حوزه‌ی «فناوری» بیشترین میزان پاسخ به نسبت تعداد جست‌وجو را دربر دارد، در حالیکه «مرتع»، «حقوق»، «آلودگی» و «آموزش» کمترین میزان پاسخ را به نسبت تعداد جست‌وجوهای کاربران داشته‌اند.

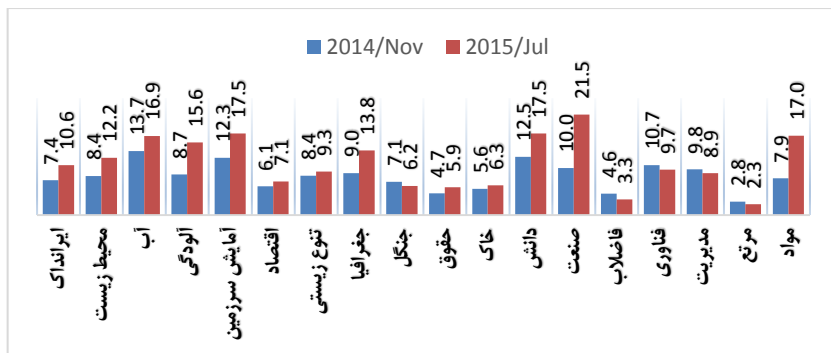


شکل ۱. نسبت تعداد پاسخ به تعداد جست‌وجو در حوزه‌های مختلف محیط زیست

شکل دو رشد جست‌وجوهای انجام شده در مدت زمان ۳۰ ماه را نشان می‌دهد. رشد میزان جست‌وجوهای کل پایگاه اطلاعاتی ایرانداک با رنگ آبی و محیط زیست با رنگ قرمز مشخص شده است. میزان این جست‌وجوها به دلیل تغییر در سیاست‌های اطلاع‌رسانی، تغییرات نرم‌افزار، رشد تعداد کاربران و ... تقریباً ۱۰ برابر شده است و همانطور که از نمودار بالا مشخص است رشد میزان جست‌وجوهای محیط زیست نیز تقریباً پیرو رشد جست‌وجوهای کل پایگاه بوده است. برای تشریح بیشتر این رشد به شکل سه نگاه کنید.

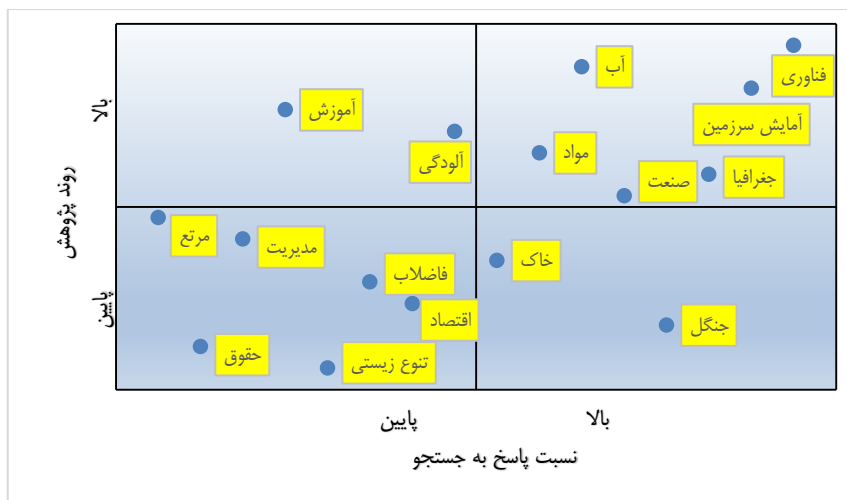


شکل ۲. مقایسه رشد جست و جوهای کل پایگاه ایراندک با رشد جست و جوهای محیط زیست



شکل ۳. مقایسه رشد جست و جوها در کل پایگاه ایراندک با محیط زیست و هر یک از حوزه های آن

پس از بکارگیری تکنیک‌های داده کاوی، می‌توان حوزه‌های زیر مجموعه محیط زیست را بر مبنای نسبت پاسخ موجود و روند پژوهش (نشان‌دهنده تمایل پژوهشگران به پژوهش در حوزه‌ای خاص) به چهار منطقه تقسیم کرد. منطقه دارای روند پژوهشی بالا و نسبت پاسخ پایین که شامل حوزه‌های «آموزش» و «آلودگی» است، بیشترین اولویت پژوهش را دارد. در مقابل آن منطقه دارای روند پژوهشی پایین و نسبت پاسخ بالا که شامل «خاک» و «جنگل» است دارای افزونگی پژوهش است و از اولویت کمتری برخوردار است. منطقه دارای نسبت پاسخ پایین و روند پژوهش پایین در بردارنده حوزه‌هایی است که از گذشته پژوهش در این موضوعات کم‌رنگ بوده و پژوهشگران هم تمایل چندانی به پژوهش در این حوزه‌ها ندارند، به عبارتی جامعه پژوهشی نسبت به این حوزه‌ها غافل بوده است. منطقه دارای نسبت پاسخ بالا و روند بالا شامل حوزه‌هایی است که هرچند میزان قابل توجهی از پژوهش‌ها را در خود جاداده است اما همچنان در این حوزه‌ها به میزان قابل توجهی پژوهش انجام می‌شود.



شکل ۴. اولویت پژوهش در حوزه محیط زیست بر مبنای خلأ پژوهشی و روند پژوهش