

به نام خدا



وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

خلاصه گزارش طرح پژوهشی سازمانی

عنوان طرح پژوهشی

طراحی و ساخت یک ماژول نرم افزاری برای استخراج و بازنمایی خودکار پیشنهادهای پژوهش از پایان نامه ها و رساله های پایگاه اطلاعات علمی ایران (گنج)

تاریخ شروع طرح پژوهشی	۱۴۰۰/۱/۲۱	تاریخ پایان طرح پژوهشی	۱۴۰۱/۸/۲۹
نام و نام خانوادگی مجری طرح پژوهشی	نصراله پاک نیت		
نام و نام خانوادگی همکاران طرح پژوهشی	علی اصغر حجت پناه		

درباره طرح پژوهشی

بخش پیشنهادهای پژوهش یکی از بخش های مهم در اسناد و مدارک علمی است. این بخش ارزش زیادی برای پژوهشگران داشته و سرنخ هایی در رابطه با موضوعات پژوهشی و ایده های مناسب برای حل آنها در اختیار پژوهشگران قرار می دهد. پژوهشگران به طور معمول از اطلاعات ارائه شده در این بخش برای انتخاب مساله پژوهش مناسب استفاده می کنند. با توجه به زمان انتشار اسناد علمی، می توان با بررسی بخش پیشنهادهای پژوهش، مساله های پژوهشی به روز یا به عبارتی عامیانه تر، مساله های نیازمند کار بیشتر، را شناسایی کرد و در راستای حل آنها تلاش نمود. با این حال، با توجه به رشد روزافزون مدارک علمی، مطالعه اسناد موجود و استخراج پیشنهادهای پژوهش برای پژوهشگران امری دشوار بوده و روزه روز نیز بر دشواری آن افزوده می شود. با توجه به شناخت کمتر، پژوهشگران تازه کار نیز دشواری بیشتری را در این راستا متحمل می شوند. با توجه به سختی استخراج دستی پیشنهادهای پژوهش از اسناد علمی، در این پژوهش مساله استخراج خودکار این اقلام اطلاعاتی از پایان نامه ها و رساله ها (پارساها)ی موجود در پایگاه اطلاعات علمی ایران (گنج) پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، در نظر گرفته شده و با بررسی روش های ارائه شده بدین منظور در زبان انگلیسی و همچنین ویژگی های زبان فارسی، یک روش خودکار بدین منظور ارائه شده است. روش ارائه شده در قالب یک ماژول نرم افزاری و با

توجه به نیازهای فنی موجود پیاده‌سازی شده است. ماژول نرم‌افزاری طراحی شده تمام‌متن پارساها را به عنوان ورودی دریافت کرده، پیشنهادهای پژوهش را از آن‌ها استخراج نموده و پیشنهادهای استخراج شده را به عنوان فراداده‌ای جدید به مدارک اضافه می‌کند. برای دسترسی کاربران به پیشنهادهای پژوهش استخراج شده، ماژول نرم‌افزاری جدید در قالب یک سامانه جدید به نام «گنجینه پیشنهادهای پژوهش پایان‌نامه‌ها و رساله‌ها» ارائه شده و از طریق آدرس <https://ganjpro.irandoc.ac.ir> قابل دسترس است.

روش پژوهش

در این پژوهش، با مطالعه منابع علمی معتبر روشی پیشنهاد شده است که با استفاده از آن بتوان پیشنهادهای پژوهش ارائه شده در پایان‌نامه‌ها و رساله‌های فارسی‌زبان را شناسایی، استخراج و بازنمایی کرد. برای جمع‌آوری اطلاعات از روش کتابخانه‌ای استفاده شده و نتایج به‌کارگیری روش پیشنهادی روی مجموعه داده‌ای طراحی شده برای این مسئله در این پژوهش، ارزیابی و تحلیل شده است. برای تحلیل نتایج و مقایسه آن‌ها از شاخص‌های مختلفی نظیر دقت، فراخوانی و معیار اف-۱ استفاده شده است. پیاده‌سازی روش ارائه شده با استفاده از زبان برنامه‌نویسی «پایتون» انجام شده است.

یافته‌های طرح پژوهشی

نتایج پژوهش صورت گرفته بیانگر این است که به‌کارگیری روش‌های یادگیری ماشینی برای استخراج پیشنهادهای پژوهش پایان‌نامه‌ها و رساله‌های فارسی‌زبان منجر به نتایج مناسبی می‌شود. همچنین، با توجه به خروجی به دست آمده از آزمایشات صورت گرفته می‌توان نتیجه گرفت که در مقایسه با دسته‌بند بیز ساده، به‌کارگیری دسته‌بندهای رگرسیون لجستیک و ماشین بردار پشتیبان منجر به دستیابی به نتایج بهتری برای مساله استخراج پیشنهادهای پژوهش می‌شود.

نتیجه‌گیری و پیشنهادها

در این پژوهش، به مساله استخراج پیشنهادهای پژوهش از پایان‌نامه‌ها و رساله‌های فارسی‌زبان پرداخته شد و با استفاده از ابزارهای بیز ساده، رگرسیون لجستیک و ماشین بردار پشتیبان یک روش خودکار بدین منظور ارائه شد. روش پیشنهادی تحت عنوان یک ماژول نرم‌افزاری بر روی داده‌های موجود در پایگاه گنج پیاده‌سازی و نتایج حاصل در سامانه «گنجینه پیشنهادهای پژوهش پایان‌نامه‌ها و رساله‌ها» از طریق آدرس اینترنتی <https://ganjpro.irandoc.ac.ir> ارائه شده و به‌روزرسانی می‌شود.

در ادامه این پژوهش، پیشنهاد می‌شود از دیگر روش‌های یادگیری ماشینی مانند ماشین بردار پشتیبان دوقلو، CRF، یادگیری عمیق و روش‌های ترکیبی برای استخراج پیشنهاد‌های پژوهش و مقایسه نتایج با نتایج به دست آمده در این پژوهش استفاده شود. همچنین، پیشنهاد می‌شود از ابزار ارائه شده در این پژوهش برای اهداف علم‌سنجی و برای ارزیابی میزان دنبال شدن پیشنهاد‌های پژوهش ارائه شده در رشته‌ها، دانشگاه‌ها و مقاطع مختلف پرداخت و از نتایج حاصل برای سیاست‌گذاری استفاده کرد.