

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 توسعه یک سامانه واژه‌نامه جدید با به کارگیری کلیدواژه‌های پایان‌نامه‌ها و رساله‌های فارسی زبان

تاریخ شروع طرح پژوهشی	۱۴۰۲/۳/۲۲	تاریخ پایان طرح پژوهشی	۱۴۰۳/۱۱/۳
نام و نام خانوادگی مجری طرح پژوهشی	نصراله پاک‌نیت		
نام و نام خانوادگی همکاران طرح پژوهشی	ساناز قربانلو، ستاره مدرسی		

درباره طرح پژوهشی

بسیاری از پژوهشگران در زمان تدوین و نگارش پیشنهادها، پایان‌نامه‌ها، و رساله‌های تحصیلات تکمیلی و همچنین برای انتشار مقالات علمی خود نیاز به انتخاب منبعی مورد اعتماد دارند تا ترجمه‌های مناسبی را برای واژه‌های تخصصی خود پیدا کنند. یکی از این منابع، واژه‌نامه‌های تخصصی هستند که توسط انجمن‌ها و مراکز علمی تهیه و منتشر می‌شوند. سامانه‌های بسیاری، از جمله سامانه واژه‌نامه‌های ایرانداک، نیز وجود دارند که امکان استفاده از این واژه‌نامه‌های تخصصی را به صورت برخط در اختیار کاربران قرار می‌دهند. با این حال، پژوهشگران همچنان با چالش‌های زیادی روبه‌رو هستند. وجود ترجمه‌های فارسی عامه‌پسندتر و بهتر برای برخی واژگان و اصطلاحات تخصصی، عدم وجود ترجمه برای تمام واژگان و اصطلاحات تخصصی به‌ویژه در موضوعات به‌روز، و عدم وجود واژه‌نامه‌های تخصصی برای تمام زمینه‌ها و موضوعات تخصصی برخی از مهم‌ترین چالش‌ها در این زمینه هستند. در راستای حل مساله بیان شده، در این پژوهش یک سامانه واژه‌نامه جدید (قابل دسترس از طریق آدرس <https://vocab.irandoc.ac.ir>) طراحی شده که هدف اصلی آن استفاده از کلیدواژه‌های گزارش شده توسط دانش‌آموختگان تحصیلات تکمیلی در پایان‌نامه‌ها و رساله‌های خود برای توسعه خودکار واژه‌نامه است. در مقایسه با دیگر سامانه‌های موجود، توسعه واژه‌نامه در سامانه جدید خودکار بوده و ثبت هر پایان‌نامه یا رساله در پایگاه‌های اطلاعاتی پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) منجر به توسعه واژه‌نامه خواهد شد. از طریق سامانه جدید، پژوهشگران می‌توانند با بررسی و مقایسه تعداد پایان‌نامه‌ها/رساله‌های مختلف استفاده‌کننده از ترجمه‌ها و معادل‌های متفاوت، موارد بهتر و متداول‌تر را شناسایی و استفاده کنند. علاوه بر دریافت ورودی از پایان‌نامه‌ها و رساله‌ها، در سامانه جدید امکان دریافت واژه‌ها و اصطلاحات، و ترجمه‌های آن‌ها از طریق واژه‌نامه‌های تخصصی موجود و همچنین پیشنهاد کاربران تایید صلاحیت شده نیز با هدف هر چه کامل‌تر بودن پایگاه داده سامانه فراهم شده است. کاربران همچنین می‌توانند با محدودسازی نتایج بر اساس اقلام اطلاعاتی متفاوتی مانند نوع منبع ورودی (شامل پایان‌نامه/رساله، واژه‌نامه، و پیشنهاد کاربران)، موضوع، رشته، سال، نام سازمان، نام واژه‌نامه و

غیره، ترجمه‌های مرسوم واژگان و اصطلاحات تخصصی در منابع، موضوعات، رشته‌ها، بازه‌های زمانی، سازمان‌ها و واژه‌نامه‌های متفاوت را مشاهده نمایند.

روش پژوهش

در این پژوهش، در ابتدا با رویکردی کتابخانه‌ای به بررسی مطالعات صورت گرفته در زمینه موضوع پژوهش پرداخته شد. در ادامه روشی برای استخراج و جمع‌آوری کلیدواژه‌های فارسی و انگلیسی پایان‌نامه‌ها و رساله‌های فارسی‌زبان به همراه برخی اقلام اطلاعاتی دیگر متناظر با مدارک دربردارنده آن‌ها از پایگاه‌های اطلاعاتی ایرانداک توسعه یافت. روش پیشنهادی با اعمال پیش‌پردازش‌های موردنیاز مانند یکسان‌سازی نویسه‌ها بر روی داده‌های جمع‌آوری شده به پاک‌سازی داده‌ها می‌پردازد. سپس، با استفاده از اطلاعات آماری و فراداده‌های واکنشی شده همراه با کلیدواژه‌ها در مجموعه داده‌های منتج از پیاده‌سازی روش، پایگاه داده سامانه واژه‌نامه با استفاده از SQL server 2019 ایجاد گردید. لازم به توضیح است که پیاده‌سازی روش موردنظر برای استخراج اطلاعات با استفاده از زبان برنامه‌نویسی پایتون صورت گرفته است. در پایان پیاده‌سازی بخش‌های سمت سرور و سمت کاربر سامانه با استفاده از زبان برنامه‌نویسی C# و فناوری‌های سمت کاربری مانند bootstrap, JQuery, Kendo انجام شد.

یافته‌های طرح پژوهشی

در این طرح پژوهشی، با توجه به مشکلات سامانه واژه‌نامه فعلی ایرانداک، به طراحی و توسعه یک سامانه واژه‌نامه جدید پرداخته شد. در این راستا، پیشینه پژوهشی و اجرایی موضوع مورد بررسی قرار گرفت و در ادامه، جزئیات سامانه جدید ارائه شد. سامانه جدید از طریق سه منبع واژگان و اصطلاحات تخصصی فارسی و معادل انگلیسی آن‌ها را دریافت کرده و به کاربران عرضه می‌کند. استفاده از مجموعه داده‌های غنی موجود در پایگاه‌های اطلاعاتی ایرانداک مهم‌ترین منبع برای جمع‌آوری واژگان است. در این راستا، پس از بررسی انطباق کلیدواژه‌های گزارش شده در زمان ثبت مدارک با موارد بیان شده در تمام متن مدارک، موارد منطبق وارد سامانه می‌شوند. همچنین، برنامه‌ای هم‌گام‌ساز بدین منظور طراحی شده که به‌روزرسانی خودکار سامانه با توجه به پایان‌نامه‌ها/رساله‌هایی که در آینده در پایگاه‌های اطلاعاتی ایرانداک ثبت می‌شوند را میسر می‌نماید. منبع ورودی دوم در نظر گرفته شده واژه‌نامه‌های تخصصی موجود در ایرانداک هستند که از طریق سامانه جدید نیز در دسترس خواهند بود. منبع ورودی سوم، کاربران تایید صلاحیت شده سامانه خواهند بود که می‌توانند واژگان و معادل تخصصی آن‌ها را پیشنهاد داده و به غنی‌سازی منابع اطلاعاتی سامانه کمک کنند. با ارائه امکان محدودسازی نتایج با استفاده از اقلام اطلاعاتی متفاوتی با توجه به نوع منبع ورودی اطلاعات (شامل پایان‌نامه/رساله، واژه‌نامه، و پیشنهاد کاربران)، مانند موضوع، رشته، سال، نام سازمان، نام واژه‌نامه و غیره؛ سامانه واژه‌نامه جدید امکان مشاهده معادل‌های مرسوم واژگان و اصطلاحات تخصصی در منابع، رشته‌ها، بازه‌های زمانی، واژه‌نامه‌ها و موضوعات متفاوت را در اختیار کاربران قرار می‌دهد.

نتیجه‌گیری و پیشنهادها

به عنوان پیشنهادات اجرایی و پژوهشی، موارد زیر برای توسعه این پژوهش پیشنهاد می‌شوند:

۱- با جمع‌آوری مقالات و دیگر انواع مدارک علمی در پایگاه‌های اطلاعاتی ایرانداک، می‌توان از کلیدواژه‌های موجود در این نوع

مدارک علمی نیز استفاده کرد و پایگاه داده سامانه جدید را غنی تر نمود.

- ۲- در این پژوهش از انطباق کلیدواژه‌های گزارش شده در زمان ثبت مدارک با موارد بیان شده در تمام متن مدارک برای اطمینان از تناظر کلیدواژه‌های فارسی و انگلیسی یک پایان‌نامه/رساله استفاده شد. با این حال، با توجه به مشکلات بیان شده در مورد ورود اطلاعات توسط دانش‌آموختگان در سامانه ثبت، این رویکرد منجر به ریزش زیاد رکوردهای اطلاعاتی (زوج کلیدواژه فارسی و انگلیسی) می‌شود. استفاده از مدل‌های زبانی برای اطمینان از تناظر بین کلیدواژه‌های فارسی و انگلیسی می‌تواند به عنوان راه‌کاری جایگزین در نظر گرفته شده و با استفاده از آن از تناظر کلیدواژه‌های فارسی و انگلیسی اطمینان حاصل یافت.
- ۳- با توجه به مثال‌های مختلف ارائه شده از استفاده از معادل‌های فارسی متفاوت برای یک اصطلاح تخصصی در جدول ۱-۱ و بررسی پایگاه اطلاعات علمی ایران (گنج) می‌توان متوجه شد که بازایی اطلاعات در حال حاضر در این سامانه با مشکل روبرو است و به‌کارگیری سامانه واژه‌نامه جدید می‌تواند منجر به بهبود نتایج بازایی در پایگاه گنج شود. این مهم را می‌توان بدین صورت انجام داد که در زمان جستجو در این پایگاه، از معادل انگلیسی واژگان و دیگر معادل‌های فارسی آن‌ها نیز برای بازایی کامل تر اطلاعات استفاده نمود.
- ۴- با توجه به تعداد کم پایان‌نامه‌ها/رساله‌ها در زبانی غیر از فارسی و انگلیسی، سامانه خروجی این طرح پژوهشی تنها شامل واژه‌نامه تخصصی برخط فارسی-انگلیسی است. با این حال و علی‌رغم محدود بودن پایان‌نامه‌ها به زبان‌های دیگر، می‌توان با به‌کارگیری این مدارک در کنار واژه‌نامه‌های دو زبانه موردنظر به توسعه سامانه جدید پرداخت.
- ۵- یکی از مواردی که می‌تواند بر ارزش سامانه واژه‌نامه جدید بیافزاید، تجهیز آن به جملات فارسی و انگلیسی است؛ یعنی به ازای هر واژه، جمله/جملاتی فارسی یا انگلیسی در بردارنده آن واژگان نیز در اختیار کاربران قرار داده شود. با توجه به وجود کلیدواژه‌ها در چکیده فارسی و انگلیسی پایان‌نامه‌ها/رساله‌ها می‌توان این مهم را نیز در آینده به سامانه افزود.
- ۶- شناسایی و نمایش واژگان و عباراتی که از نظر مفهومی با یکدیگر مرتبط هستند به کاربران از دیگر مواردی است که با توجه به ارتباط کلیدواژه‌های پایان‌نامه‌ها/رساله‌ها با یکدیگر می‌توان از آن برای ارائه خدمات بیشتر به کاربران استفاده نمود.