

به نام خدا



وزارت علوم، تحقیقات، و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

خلاصه گزارش طرح پژوهشی سازمانی

عنوان طرح پژوهشی

طراحی روشی هوشمند برای پیشنهاد مستندات علمی به زبان فارسی در پایگاه اطلاعات علمی ایران (گنج) بر پایه سیستم‌های پیشنهاددهنده محتوا محور

تاریخ شروع طرح پژوهشی	۱۴۰۱/۰۳/۱۰	تاریخ پایان طرح پژوهشی	۱۴۰۲/۰۵/۳
-----------------------	------------	------------------------	-----------

نام و نام خانوادگی مجری طرح پژوهشی	مرضیه زرین بال ماسوله
نام و نام خانوادگی همکاران طرح پژوهشی	آزاده محبی

درباره طرح پژوهشی

سیستم‌های پیشنهاددهنده با تشخیص نیاز کاربر پیشنهادهایی متناسب را ارائه کرده و به او در تصمیم‌گیری برای انتخاب موارد مورد نیاز یا مورد علاقه کمک می‌کنند. این سیستم‌ها فهرستی رتبه‌بندی شده از مواردی را که به احتمال زیاد کاربر به آن‌ها نیاز یا علاقه دارد ارائه می‌کند. این فهرست می‌تواند بر اساس پیشنهاد کاربران مشابه، تحلیل رفتار فعلی کاربر، شناسایی موارد مشابه با مورد انتخاب شده کاربر و یا ترکیبی از این موارد شکل گیرد. از چنین سیستم‌هایی در پایگاه‌های جستجوی عمومی و در برخی از پایگاه‌های جستجوی اسناد علمی نظیر اسکوپوس، امرالد و پاب مد نیز از این قابلیت برای ارائه پیشنهاد به کاربران و ارتقاء تجربه آنان استفاده شده است. سامانه گنج در حال حاضر از چنین قابلیت بر خوردار نیست و نیازهای اطلاعاتی کاربر صرفاً از طریق عبارت پرس‌وجو استخراج می‌شود. اما سندی که کاربر از بین اسناد متعدد نتایج جستجو انتخاب می‌کند حاوی اطلاعات ضمنی است که استخراج و استفاده از آن برای ارائه پیشنهادهای مشابه می‌تواند بر تجربه کاربر از استفاده از سامانه گنج تاثیر بسیاری داشته باشد و به او در رسیدن به اهداف جستجو کمک کند. لذا در این پژوهش تمرکز بر روی ساخت چنین سیستمی قرار گرفت.

روش پژوهش

در پژوهش حاضر ابتدا ویژگی‌ها و رویکردهای طراحی سیستم‌های پیشنهاددهنده با تمرکز بر سیستم‌های محتوامحور بررسی شده و سپس با توجه به ویژگی‌های سامانه گنج و داده‌های آن رویکرد مناسب پیشنهاد شد. در ادامه نحوه عملکرد سیستم پیشنهادی ارزیابی گردید. برای ارزیابی چنین سیستم‌هایی به طور معمول از دو مرحله ارزیابی استفاده می‌شود؛ مرحله ۱: ارزیابی براساس داده‌هایی که سیستم براساس آنها طراحی شده و مرحله ۲: ارزیابی براساس کارایی سیستم در پیش‌بینی داده‌های جدید (مشاهده نشده). برای این منظور از دو مجموعه داده استاندارد فارسی محبی و همکاران (۱۴۰۰) با ۳۰۴۷ سند (مجموعه ۱ داده‌ای) و مجموعه داده تصادفی استخراج شده از سامانه گنج با ۳۱۸۳۱ سند (مجموعه ۲ داده‌ای) استفاده شد. سپس اسناد متنی با کمک رویکرد یادگیری عمیق و با بهره‌گیری از الگوریتم SBERT تحلیل شدند. در ادامه با انتخاب یک سند توسط کاربر، اسناد مشابه با کمک معیار فاصله کسینوسی محاسبه و ۵ سند مشابه‌تر به کاربر پیشنهاد شدند.

یافته‌های طرح پژوهشی

ارزیابی سیستم‌های پیشنهاددهنده عموماً به دو شیوه برخط و برون‌خط تقسیم می‌شود؛ ابزارهای برخط شامل سنجش شاخص‌های کلیدی عملکردی (KPI) است که پس از پیاده‌سازی مدل و به صورت واقعی از طریق رفتار کاربران اندازه‌گیری می‌شوند. در شیوه برون‌خط از مجموعه داده متشکل از انتخاب‌ها و یا رتبه‌های کاربران استفاده می‌شود و از آنجا که در این شیوه تعاملی با کاربران واقعی وجود ندارد، با صرف هزینه کم‌تر می‌توان الگوریتم‌های بیشتری را پیاده‌سازی نموده و آزمود. از آنجا که پیاده‌سازی نهایی در محدوده این طرح پژوهشی قرار نداشته و هدف طراحی یک سامانه و یا ارتقاء سامانه گنج برای این منظور نبوده‌است. لذا برای ارزیابی سیستم پیشنهادی از ابزارهای برون‌خط استفاده شد و نتایج گزارش گردید. براساس نتایج بدست آمده در مجموعه ۱ داده‌ای، تقریباً ۶۹ درصد و در مجموعه ۲ داده‌ای، تقریباً ۶۱ درصد نتایج بازیابی شده با سند منتخب مرتبط بوده‌اند. همچنین در مجموعه ۱ داده‌ای، تقریباً ۷۱ درصد و در مجموعه ۲ داده‌ای، تقریباً ۷۵ درصد نتایج مرتبط با سند منتخب بازیابی شده‌اند. و در مجموعه ۱ تقریباً ۷۲ درصد و در مجموعه ۲ تقریباً ۶۱ درصد پیشنهادها به پیشنهادهای واقعی نزدیک بودند.

نتیجه‌گیری و پیشنهادها

سیستم‌های پیشنهاددهنده می‌توانند در بهبود تجربه کاربر و بهتر شدن خدمات ارائه شده نقش بسزایی را داشته باشند اما بکارگیری آنها مستلزم پیش‌نیازهایی است که از جمله آن می‌توان به موارد زیر اشاره کرد:

□ توسعه ابزارهای تحلیل متن خودکار شخصی‌سازی شده براساس مشخصات اسناد متنی موجود در پایگاه گنج. ابزارهای موجود ابزارهایی با کارکرد کلان هستند و برای آنکه دقت افزایش یابد باید براساس نیازهای سامانه‌ها شخصی شوند. البته که امکان بهینه‌سازی برای برخی از این ابزارها وجود دارد اما بهینه‌سازی در شرایطی مقدور است که علاوه بر امکانات پردازشی متناسب، مجموعه داده استاندارد و کافی وجود داشته باشد.

□ وجود مجموعه داده استاندارد و به اندازه یکی از زیربناهای مهم در پیاده‌سازی رویکردها و روش‌های جدید است.

□ وجود امکان پیاده‌سازی آزمایشی و تست سیستم‌های طراحی شده نیز یکی دیگر از زیربناهای مهم در پیاده‌سازی رویکردها و روش‌های جدید و ارتقا کیفیت خدمت‌رسانی سامانه‌هاست.