

## چکیده طرح پژوهشی

ارائه روشی هوشمند برای استخراج کلیدواژه از مستندات علمی زبان فارسی براساس سیستم‌های پیشنهاددهنده

### آزاده محبی

با توجه به افزایش حجم تولید و ثبت مستندات علمی، به خصوص در پایگاه گنج، نیاز است که فرایند نمایه‌سازی با سرعت بیشتری انجام شود. هم اکنون علاوه بر متوسط نرخ ثبت روزانه ۲۰۰ عنوان پایان نامه در روز در گنج، حدود ۱۴۰ هزار سند موجود است که هنوز نمایه نشده‌اند. در حال حاضر فرایند نمایه‌سازی در ایرانداک براساس دانش و تخصص نمایه‌سازان انجام میشود و روش‌های ماشینی برای پردازش خودکار متون و نمایه‌سازی آنها استفاده نمی‌شود. در بسیاری از پایگاه‌های علمی در دنیا از روش‌های ماشینی و خودکار در کلیه فعالیت‌های فرایند نمایه‌سازی یا بخشی از آنها استفاده می‌شود. تعدادی از این روش‌ها بر مبنای تحلیل آماری متون و استفاده از روش‌های یادگیری ماشین هستند، تعدادی بر مبنای تحلیل معنایی متون بواسطه اصطلاحنامه‌های تخصصی و هستان‌شناسی، و در تعدادی دیگر از این روش‌ها از تلفیق هر دو استفاده می‌شود.

در همین راستا، هدف از انجام این طرح پژوهشی تسریع فرایند نمایه‌سازی با استفاده از روش‌های یادگیری ماشین و تحلیل آماری متون است. برای این منظور روشی ارائه می‌گردد که از طریق آن برای نمایه‌سازی یک سند جدید، مجموعه‌ای از کلیدواژه‌های مرتبط، به نمایه‌ساز پیشنهاد شود تا نمایه‌ساز سریعتر بتواند از بین آنها کلیدواژه‌های مناسب را در نهایت انتخاب کند. پیشنهاد کلیدواژه به نمایه‌ساز، براساس رویکرد سیستم‌های پیشنهاددهنده و با اتکا به دانش ضمنی موجود در مجموعه مستندات نمایه شده قبلی صورت می‌گیرد.

کلیدواژه‌ها: سیستم‌های پیشنهاددهنده؛ استخراج کلیدواژه؛ نمایه‌سازی خودکار؛ استدلال موردبیان

Key words: Recommender systems, Keyword extraction, Automatic Indexing, case-based reasoning