

# طراحی یک سیستم توصیه گر ترکیبی معنایی با استفاده از تکنیک‌های پردازش زبان طبیعی فارسی

محسن عباسی، دکتر رضا منصفی و احمد استیری

abasi.mohsen@stu-mail.um.ac.ir, monsefi@um.ac.ir, ahmad.estiri@stu-mail.um.ac.ir

## چکیده

با افزایش سریع اطلاعات، به سیستمی توصیه گر که با دادن پیشنهادات مناسب با علائق یک کاربر بر اساس پیشینه‌ی عملکرد او، وی را از صرف وقت در مرور تمام آیتم‌ها باز دارد، به شدت احساس نیاز می‌شود. سیستم توصیه گر با تحلیل رفتار کاربر خود، اقدام به پیشنهاد مناسب‌ترین اقلام به وی می‌نماید. حوزه‌ی مفاهیم موجود در بسیاری از سیستم‌های توصیه گر فعلی، محدود به دانش موجود در خود سیستم بوده و از منابع عظیم دانش خارج از سیستم، مانند وب معنایی و داده‌های پیوندی استفاده نمی‌شود. در این مقاله سیستم توصیه گر ترکیبی‌ای معرفی شده است که از مفاهیم و روابط موجود بین آنها در پایگاه‌های خارج از خود استفاده می‌کند. در نتیجه یکی از قسمت‌های اصلی سیستم پیشنهادی، استخراج مفاهیم از آیتم‌ها و رفتار کاربران می‌باشد؛ که این کار با تکنیک‌های پردازش زبان طبیعی مانند ریشه‌یابی کلمات کلیدی آیتم‌ها میسر می‌باشد. همچنین در سیستم پیشنهادی از نگاشت معنایی ریشه‌های بدست آمده، به شبکه‌ی واژگان فردوس نت استفاده شده است. با این کار به دلیل گسترش معنایی مجموعه کلمات کلیدی و ویژگی‌های آیتم‌ها، صحت توصیه‌ها افزایش خواهد یافت. در پایان نیز با استفاده از حدود ۸۱ هزار کتاب دسته‌بندی شده استخراجی از سایت Acm و میلیون‌ها نرخ که هزاران کاربر به یک میلیون مقاله/کتاب موجود در مجموعه داده Epinions داده‌اند، نشان داده خواهد شد که مشکل "شروع آهسته" تا حد زیادی قابل حل می‌باشد.

## کلمات کلیدی

سیستم توصیه گر ترکیبی، پردازش زبان طبیعی فارسی، فردوس نت، وب معنایی، ریشه‌یابی

سال ۲۰۱۰ به بعد تعداد ۹۲۵۰ مقاله و تنها در سال ۲۰۱۱

تعداد ۴۱۷۰، در رابطه‌ی مستقیم یا غیرمستقیم با این موضوع، در کنفرانس‌ها و مجلات گوناگون سراسر دنیا پذیرفته شده‌اند.

در سیستم‌های توصیه گر با مقیاس دنیای واقعی به دلیل خرید مقطعی (مانند وسایل الکترونیکی) و یا به طور کلی کثرت تعداد آیتم‌های موجود جهت پیشنهاد، مشتری‌ها نوعاً درصد کمی از آیتم‌ها را نرخ گذاری می‌کنند. این مشکل را پراکندگی می‌نامیم. این مشکل زمانی رخ خواهد داد که تعداد نرخ‌ها در مقایسه با تعداد آیتم‌ها کوچک باشد [3]. مساله "شروع آهسته" نیز مشکل دیگری است که می‌توان آن را نوع خاصی از مشکل پراکندگی به حساب آورد. در ابتدای شروع به کار، سیستم تا مدتی نمی‌تواند پاسخ مناسبی ارائه کند؛ زیرا پایگاه دانش آن از تعداد محدودی نمونه ساخته شده است. این مشکل همچنین در هنگام کار با آیتم جدید و یا کاربر جدید نیز پیش می‌آید. برای بسیاری از مسائل ارائه شده، با گذر زمان و استفاده‌ی مکرر کاربران، پایگاه دانش با استفاده از اطلاعات کاربران تقویت می‌شود ولی در این فاصله‌ی زمانی، کارایی سیستم ممکن است از سطح قابل قبولی برخوردار نباشد.

سیستم‌های توصیه گر بر اساس تخمین نرخ (صریح و یا ضمنی [4]) و چگونگی پیشنهادات به چهار دسته تقسیم می‌شوند:

## ۱- مقدمه

سیستم توصیه گر به عنوان یک سیستم ابتکاری که اطلاعات مفید را پیشنهاد می‌دهد و می‌تواند در دامنه‌های گوناگون بکار رود، در نظر گرفته می‌شود [1]. سیستم توصیه گر یا سامانه‌ی پیشنهادگر با تحلیل رفتار کاربر خود، اقدام به پیشنهاد مناسب‌ترین اقلام (داده، اطلاعات، کالا و ...) به وی می‌نماید. این گونه سیستم‌ها در واقع جهت حل مشکلات ناشی از حجم فراوان و رو به رشد اطلاعات ارائه شده است و به کاربر خود کمک می‌کند تا در میان حجم عظیم اطلاعات، سریع‌تر به هدف خود دست یابد [2]. سیستم‌های توصیه گر به کاربرانی که از بین حجم بالای اطلاعات به دنبال نوعی خاص از اطلاعات مرتبط با اولویت‌هایشان هستند، پیشنهادات شخصی شده‌ای را ارائه می‌دهد. این نوع سیستم‌ها با توانایی‌ای که در جمع‌آوری اطلاعات از رفتار و حرکت کاربران، دسته‌بندی و تفسیر آنها دارند، امکانی فراهم آورده‌اند که کاربران با صرف زمان و انرژی کمتر به اطلاعات مناسب‌تری دسترسی پیدا کنند.

امروزه سعی در ساخت "سیستمی توصیه گر" با درصد خطای کم و سرعت بالا در تمام شرایط به یکی از پرطرفدارترین حوزه‌های تحقیقاتی دانشگاهی تبدیل شده است. به طوری که از