

شناسایی و تحلیل واژگان عمومی در منابع وب: رویکردی نو به بسط عبارت جستجو با استفاده از زبان طبیعی در موتورهای کاوش

چکیده

بسط جستجو با استفاده از واژه‌های عمومی که همراه کلیدواژه‌های موضوعی در منابع و متون موجود در وب ظاهر می‌شوند می‌تواند موجب افزایش میزان دقت و ربط در نتایج بازیابی در موتورهای کاوش شود. به این منظور، پژوهشی در دو مرحله انجام گرفت. یافته‌های مرحله نخست نشان داد که ۱۴/۶٪ از واژه‌های عمومی میان دو حوزه مشترک هستند، یعنی عمومی مطلق به شمار می‌آیند، ۸۵/۴٪ واژه‌های عمومی وابسته به حوزه موضوعی می‌باشند، ۶۵٪ درصد واژه‌های عمومی پیش و ۳۵٪ از آنها پس از کلیدواژه‌های موضوعی در متون ظاهر می‌شوند. همچنین، از نظر نوع واژه‌ها، ۶۲/۵٪ غیرموضوعی و ۳۷/۵٪ نیمه موضوعی به شمار می‌روند. در مرحله دوم پژوهش، با اضافه کردن واژه‌های عمومی به کلیدواژه‌های موضوعی اولیه و انجام دوباره جستجو (یعنی بسط جستجو) در ۴ حالت جستجوی کلیدواژه‌ای عمومی، عبارت دقیق، عنوان دقیق، و نشانی اینترنتی دقیق در گوگل، مشخص شد که این کار موجب بازیابی نتایج بسیار دقیقتر و مرتبط‌تر می‌شود. نتایج نشان داد که تفاوت معناداری در نتایج جستجوی کلیدواژه‌ای و عبارتی دقیق میان دو حوزه وجود دارد. همچنین، میان دو حوزه از نظر نتایج جستجوی دقیق عنوانی و نشانی اینترنتی، تفاوت معنادار مشاهده شد. همچنین، آزمون نسبت بسامد نتایج بازیابی در حالت‌های مختلف نسبت به جستجوی کلیدواژه‌ای نشان داد که تفاوت معناداری میان دو حوزه وجود دارد. در مجموع، نتایج پژوهش نشان داد که بسط جستجو در حالت جستجوی عنوانی و نشانی اینترنتی در هر حوزه موضوعی می‌تواند بسیار موفقیت‌آمیز باشد. به همین جهت، موتورهای جستجو، می‌توانند پیش‌فرض جستجو را به دو حالت عنوانی و نشانی اینترنتی محدود کنند. همچنین، می‌توان یک سیاهه از واژگان عمومی ایجاد کرده و به منزله یک ابزار هوشمند در موتورهای کاوش تعبیه کرد تا در بسط جستجو مورد استفاده کاربران قرار گیرد.

واژه‌های کلیدی: بسط جستجو، بازیابی اطلاعات در وب، زبان جستجو، واژگان عمومی، موتورهای کاوش