



پایان‌نامه‌ی کارشناسی ارشد: نرجس رازقندی، ۱۳۹۶

بهبود نتایج موتور جستجوی مبتنی بر مدل بولی و برداری با استفاده از رویکرد معنایی

نیاز انسان به کسب اطلاعات در مورد یک موضوع خاص را نیازمندی اطلاعاتی می‌نامند. امروزه فضای وب، مخزن گسترده‌ای از داده‌ها را تشکیل داده است که به این ترتیب می‌تواند منبع مناسبی برای یافتن پاسخ به حساب آید. به همین علت، روزانه بسیاری از کاربران تلاش می‌کنند تا صفحات مرتبط با نیازمندی اطلاعاتی خود را در صفحات وب بیابند. برای این منظور، صفحات مختلف را گشوده و محتوای آنها را بررسی می‌کنند. گشودن صفحات یا به صورت تصادفی است و یا با استفاده از دانشی است که از صفحاتی که پیش از این مشاهده کرده اند، حاصل شده است. با توجه به حجم زیاد اسناد وب، بررسی تمام اسناد توسط کاربران ممکن نیست. بنابراین احتمال دارد با وجود صرف زمان بسیار، باز هم به پاسخ مطلوب و قابل قبولی دست نیابند. برای تسهیل جستجو در وب، موتورهای جستجویی طراحی شده اند که کاربر را در رسیدن به پاسخ موردنظر یاری می‌دهند. مدل بولی و مدل برداری دو مدل بهینه در موتور جستجو هستند که در تولید نتایج مرتبط با نیازمندی کاربر تأثیر بسزایی دارند. ما برای بهبود نتایج حاصل شده از موتور جستجو، روشی ارائه می‌دهیم که عملکرد آن به این صورت است که ابتدا کاربر نیازمندی اطلاعاتی خود را به فرم عبارتها و پرسش‌های زبان طبیعی مطرح می‌کند. با اینکه پرسش‌های مطرح شده به زبان طبیعی نیاز اطلاعاتی را به خوبی تشریح می‌کنند ولیکن طبق راهکار پیشنهادی برای کشف دقیق‌تر نیاز کاربر، روابط معنایی بین کلمات پرس و جوی کاربر را هم در نظر گرفته می‌شود. سپس موتور جستجو، اسناد مرتبط با پرس و جوی کاربر را با استفاده از ترکیب مدل بولی و برداری می‌یابد. در مرحله بعد اسناد یافت شده براساس میزان ارتباطشان با پرسش، با تکنیک $tf-idf$ رتبه بندی می‌شوند و بالاخره لیست بلندی از اسناد مرتبط، که بیشترین امتیازها را کسب نموده اند به عنوان پاسخ به کاربر بازگردانده می‌شوند. آزمایش‌ها و نتایج حاصل از این روش پیشنهادی حاکی از آن است که روش پیشنهادی ما روشی مناسب‌تر و دقیق‌تر از تحقیقات قبلی صورت گرفته در این زمینه می‌باشد.

کلیدواژه‌ها: موتور جستجو، مدل بولی، مدل برداری، روابط معنایی

شماره‌ی پایان‌نامه: ۱۲۷۴۱۰۰۶۹۵۲۰۱۰

تاریخ دفاع: ۱۳۹۶/۱۱/۱۳

رشته‌ی تحصیلی: مهندسی کامپیوتر- نرم افزار

دانشکده: فنی و مهندسی

استاد راهنما: دکتر حسن شاکری

استاد مشاور: مهندس علی اکبر نقابی



M.A. Thesis:

improving Search Engine Results Based on Boolean-Vector Model using Semantic Approach

The human need for information about a specific subject is called the information need. Today's web space is a vast reservoir of data, which can be considered as a good source for finding answers. For this reason, many users every day are trying to find pages that are related to their information needs on web pages. To this end, they open various pages and examine their content. Opening pages is either in random order or using the knowledge that has been obtained from the pages previously viewed. Due to the large volume of web documents, it is not possible to check all the documents by users. Therefore, it is possible that they will not receive a satisfactory answer despite the time spent. To facilitate Web search, search engines are designed to help the user reach the desired response. Boolean modeling and modeling are two optimal search engine models that are effective in producing results related to user needs. In order to improve the results of the search engine, we provide a methodology that first performs the user's need for information in the form of natural language queries and questions. Although the questions raised in the natural language well describe the need for information, but according to the proposed solution, for the more detailed discovery of the user's needs, the semantic relations between the user's query words are also taken into account. The search engine then finds the documents associated with the user query using the combination of the boolean and vector model. In the next step, the documents found are ranked by the tf-idf method based on their relevance to the query, and finally a tall list of related documents, which have earned the highest scores, are returned as a response to the user. The experiments and the results of this proposed method suggest that our proposed method is more appropriate and more precise than previous studies in this field.