

به نام خدا



وزارت علوم، تحقیقات، و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران

(ایرانداک)

خلاصه گزارش برای پژوهشگران

عنوان طرح: تحلیل رفتار جستجوی اطلاعات پژوهشگران در موتور جستجوی سامانه‌ی گنج

سال: ۱۳۹۶-۱۳۹۵

اسامی مجری و همکاران: سمیه فتاحی

معرفی پژوهش

یکی از رسالت‌های مهم هر سامانه‌ی اطلاعاتی ارائه بهترین خدمات متناسب با نیازهای کاربران می‌باشد. این امر باعث افزایش رضایت کاربران و علاقه‌مندی آنان به استفاده از سامانه‌های اطلاعاتی می‌گردد. امروزه موتورهای جستجو یکی از مهمترین ابزارهای بازیابی اطلاعات و بستری برای عرضه‌ی انواع اطلاعات برای بسیاری از کاربران هستند. دانشجویان، اساتید دانشگاه و پژوهشگران به طور روزانه به جستجو در موتورهای جستجو پرداخته و نیازهای اطلاعاتی خود را برآورده می‌سازند. پایگاه اطلاعاتی گنج (گنجینه علمی ایرانیان) پژوهشگاه علوم و فناوری اطلاعات ایران، با برخورداری از ۸۷۰ هزار رکورد علمی، امکان جستجو برای پژوهشگران را در پایان‌نامه‌ها، نشریات علمی داخلی، مقالات، همایش‌ها، طرح‌های پژوهشی و گزارش‌های دولتی فراهم می‌کند. واضح است که کاربران تحت تأثیر عوامل مختلف از شیوه‌ها و روش‌های گوناگون برای کسب اطلاعات استفاده می‌نمایند و رفتارهای جستجوی اطلاعات متفاوتی را از خود بروز می‌دهند. هدف اصلی پژوهش حاضر، تحلیل رفتار جستجوی اطلاعات کاربران در موتور جستجوی سامانه‌ی گنج می‌باشد. تحلیل رفتار جستجوی کاربران می‌تواند به درک بهتر نیازهای آنان کمک کند.

روش پژوهش

سامانه‌ی ملی اطلاعات پایان‌نامه‌ها/رساله‌های دانش‌آموختگان داخل کشور تحت عنوان سامانه‌ی گنج در پژوهشگاه علوم و فناوری اطلاعات ایران، پیاده‌سازی شده است. بخش پایگاه‌داده‌ی آن با استفاده از نرم‌افزار MySQL پیاده‌سازی شده است. کاربران از طریق این سامانه قادر به جستجوی اطلاعات مورد نیاز خود هستند. در زمانی که کاربران از موتور جستجوی این سامانه استفاده می‌کنند، اطلاعات جستجوی آنان نیز توسط نرم‌افزار MySQL در فایل لاگ، ثبت و ذخیره می‌شود. فایل لاگ موتور جستجو سامانه گنج شامل اطلاعاتی از قبیل شناسه کاربر، عبارت جستجو، تعداد دفعات جستجو، مکان دسترسی کاربر، زمان و تاریخ جستجو است. در این پژوهش بررسی‌ها بر روی لاگ‌های سامانه‌ی گنج در سه ماه سوم سال ۱۳۹۵ انجام گرفت. حجم رکوردهای این لاگ فایل ۱۱۲۵۲۹۷۴ تعداد است. با استفاده از زبان SQL، جستارهایی بر روی این لاگ فایل اجرا شد و اطلاعات مورد نیاز استخراج شد. در واقع، یک فایل خام لاگ در اختیار این پژوهش قرار گرفت، که با استفاده از زبان SQL، جستارهایی جهت استخراج نگارش شد. لاگ فایل مورد بررسی در این پژوهش، در قالب سه ماه مهر، آبان و آذر تفکیک شد. جهت تحلیل بهتر و منطقی‌تر، لاگ فایل جستجوی کاربران را در سه حالت بررسی کرده‌ایم. حالت اول کل جستجوهای که بر روی سامانه‌ی گنج انجام گرفته است و حالت دوم تنها جستجوهای مورد بررسی قرار گرفت که از طریق تالار جستجوی پژوهشگاه علوم و فناوری اطلاعات ایران، یعنی جستجوی کاربرانی که به صورت حضوری و از داخل پژوهشگاه اقدام به جستجو کرده‌اند. حالت سوم بررسی جستجوهای است که از داخل کشور صورت گرفته‌اند.

## یافته‌ها

یافته‌های این پژوهش نشان می‌دهد که ۹۰ درصد جستجوها با ip خارج از کشور و حدود ۱۰ درصد از جستجوها به داخل کشور مربوط است که از این میزان رقمی کمتر از یک درصد به ip های داخل پژوهشگاه مربوط است و نشان می‌دهد که جستجوها به صورت حضوری انجام گرفته است. همچنین نتایج نشان می‌دهد که بیشترین حجم جستجوها در بین ساعات ۷ الی ۹ صبح و ۵ الی ۷ بعدازظهر است. از طرفی جستجو بین رشته‌های مختلف علمی مورد بررسی قرار گرفت و نتایج نشان می‌دهد به عنوان مثال افراد علاقه‌مند به موضوعات فنی مهندسی و فناوری اطلاعات بیشتر در صبح و ساعات ۸ الی ۱۲ جستجوهای خود را انجام می‌دهند. این در حالی است که موضوعات مربوط به علوم انسانی بیشتر در ساعات ۱۲ الی ۱۶ مورد جستجو است.

از لحاظ عبارت پرتکرار در جستجوها، لاگ فایل بررسی شد. اکثر واژه‌های جستجو شده شامل عبارت‌هایی هستند که به مکان‌ها مربوط می‌شوند. همچنین واژه‌های پرتکرار در هر زمینه علمی استخراج شد و یافته نشان می‌دهد که افرادی که به صورت حضوری به پژوهشگاه مراجعه کرده‌اند، عبارات دقیق تری را برای جستجو انتخاب کرده‌اند و جستجوی هدفمندتری را داشتند.

همچنین ویژگی‌هایی مانند طول عبارت جستجو شده، زمان جستجو و تعداد دفعات جستجو در نظر گرفته شد و الگوریتم‌های خوشه‌بندی بر روی داده‌ها اعمال شد. نتایج نشان داد که راساس میانگین و انحراف از معیار، یک خوشه افرادی هستند که روزهای ابتدای هفته و ساعات نخستین روز را برای جستجو انتخاب می‌کنند و دفعات جستجوی آنها زیاد است. اما دسته دیگر افرادی هستند که روزهای آخر هفته و ساعات نزدیک به ظهر را برای جستجو انتخاب می‌کنند و تعداد جستجوهای آنها کمتر است.

## نتیجه گیری و پیشنهادها

در این پژوهش به تحلیل لاگ‌های کاربران سامانه‌ی گنج پرداخته شد و نتایج این تحلیل نشان می‌دهد که بیش از ۹۹٪ کاربران تمایل دارند تا از بیرون پژوهشگاه به جستجو بپردازند و کمتر از ۱٪ از آنها به دلیل دسترسی به متن کل پایان‌نامه و رساله در تالار جستجوی پژوهشگاه مراجعه حضوری دارند. این امر خود بیانگر آنست که موتور جستجوی پایگاه گنج باید به نحو مطلوب تری عمل نماید تا با توجه به اینکه درصد بسیاری از پژوهشگران در تهران حضور دارند، بتواند تالار جستجو تعداد کاربر بیشتری را جذب نماید. همچنین با توجه به اینکه هر کاربر در مراجعه حضوری به‌طور متوسط پس از جستجوی پنج کلید واژه فرایند جستجوی خود را متوقف می‌نماید نیاز است که پایگاه گنج به نحو جذاب تری طراحی گردد تا پژوهشگر بتواند در حوزه پژوهش خود دانش بیشتری را کسب نماید.

ویژگی‌های ذخیره شده در لاگ‌ها، تنها شامل عبارت مورد جستجو، زمان جستجو، تعداد نتایج جستجو و محل جستجو است. برای پژوهش‌های آتی می‌توان به ارتباط کلیدواژه‌ها، خوشه‌بندی آنها و تشکیل گراف برای تحلیل بیشتر و دقیق تر رفتار کاربران اشاره نمود. همچنین می‌توان ویژگی‌های بیشتری را در لاگ‌ها ذخیره نمود از جمله اینکه کلیک کاربر بر روی کدام نتایج بوده است، زمان کلیک کاربر بر روی هر نتیجه، مدت زمان حضور کاربر بر روی هر یک از نتایج، جنسیت کاربران، مقطع تحصیلی، سن آنها غیره. واضح است که با ذخیره کردن ویژگی‌های بیشتر از کاربران و رفتار آنها در زمان تعامل با سامانه گنج، تحلیل‌های بیشتری می‌توان ارائه کرد.