

## معرفی یک نظام پیشنهاد دهنده کلیدواژه‌های مرتبط در تدوین پرسوجو

محمد رضا فلاحتی قدیمی فومنی

دانشیار گروه پژوهشی زبانشناسی رایانه‌ای، مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری، شیراز، ایران.

نام نویسنده مسئول:

محمد رضا فلاحتی قدیمی فومنی

تاریخ دریافت: ۱۳۹۸/۱۱/۲۳

تاریخ پذیرش: ۱۳۹۹/۲/۹

چکیده

هدف از انجام پژوهش حاضر آن بود تا نرم‌فزاری تهیه و در نظام رایسست تلفیق شود که بتواند: (۱) جستجوی موضوعی را در پایگاه مقالات تمام‌متن فارسی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری محقق سازد و (۲) پس از درونداد حداقل دو کاراکتر در جعبه پرسوجو، پنج کلیدواژه پیشنهادی پرسوجو را برای استفاده احتمالی کاربر ارائه نماید. برای انجام تحقیق از مخزن کلیدواژه‌های رایسست (البته، صرفاً کلیدواژه‌های فارسی) به عنوان منبع اصلی داده تحقیق استفاده شد. این کلیدواژه‌ها بررسی و برای استفاده در پژوهش حاضر اصلاح و آماده‌سازی شدند. مدخل‌های مخزن از ترکیبات پیچیده به ترکیبات ساده تبدیل شد. سپس الگوریتم ساخت نرم‌افزار پروژه حاضر تهیه و به برنامه تبدیل گردید. این نرم‌افزار حاوی دو بخش اساسی است. بخش اول مربوط به کاربر سیستم است که به عنوان پشتیبان سیستم کار می‌کند و لازم است روزآمدسازی اطلاعات نرم‌افزار را به انجام برساند. بخش دوم که تولید نسخه جدید از صفحه مربوط به پایگاه تمام‌متن مقالات فارسی است و نرم‌افزار اصلی در آن تلفیق گردیده است. کاربران عادی در حقیقت نسخه اصلاح شده پایگاه تمام‌متن مقالات فارسی را روئیت خواهند کرد. این نرم‌افزار کاربران را قادر می‌سازد: (۱) جستجوی خود را بر اساس حوزه‌های موضوعی مورد علاقه به انجام برسانند و (۲) با درونداد حداقل دو کاراکتر، کلیدواژه‌ها و پرسوجوهای پیشنهادی (پنج مورد) را از سیستم دریافت نمایند. پس از ساخت نرم‌افزار، نتیجه کار توسط سه متخصص علم اطلاعات و دانش‌شناسی با مدرک کارشناسی ارشد و پنج کاربر نهائی از مقطع تحصیلی کارشناسی ارشد مورد ارزیابی قرار گرفت. افراد مورد آزمایش اذعان داشتند که نسخه جدید در مقایسه با نسخه قدیمی موفق‌تر است. دوم آنکه به دلیل ارائه تعداد کمتری از رکوردها، سبب صرفه‌جویی در وقت کاربر یا اطلاع‌رسان می‌شود. سوم آنکه به دلیل انجام جستجو در حوزه‌های موضوعی خاص، باعث افزایش دقت بازیابی می‌شود و چهارم آنکه با ارائه پیشنهادها سبب کاهش مشکلات تایپی و نیز جهت‌دهی به انتخاب کلیدواژه‌ها توسط کاربر می‌گردد.

**واژگان کلیدی:** زبانشناسی رایانه‌ای، پایگاه اطلاعاتی مقالات تمام‌متن فارسی، فراخوان خودکار کلیدواژه‌ها، نظام پیشنهاددهنده کلیدواژه.

## مقدمه

امروزه در عصر انفجار اطلاعات زندگی می‌کنیم، عصری که در آن مجموعه متنوعی از منابع اطلاعاتی چاپی و غیرچاپی بر روی پایگاه‌های اطلاعاتی قرار گرفته است. تعداد این پایگاه‌ها نیز آنچنان رو به فزونی است که گاه مخاطبان برای دستیابی به پایگاه مورد نظر و یا پایگاهی کیفی در حوزه‌ای خاص، ناگزیرند زمان زیادی را صرف کنند. همان‌گونه که برای مطالعه منابع چاپی اسلوب خاصی وجود دارد، برای استخراج و مطالعه محتوای علمی پایگاه‌های اطلاعاتی نیز اصول و ضوابط خاصی وجود دارد. پایگاه‌های اطلاعاتی با ساختاری خاص تهیه می‌شوند و هدف عمده آنها ارائه اطلاعات یا تأمین نیازهای اطلاعاتی کاربران به آسان‌ترین شکل ممکن می‌باشد. از این رو، پایگاه‌های اطلاعاتی به صورت مستمر در حال روزآمدسازی و ارتقاء می‌باشند و پیوسته می‌کوشند با لحاظ نمودن متغیرهای لازم به گونه‌ای عمل نمایند تا کاربران اطلاعات خود را سهل‌تر، در زمانی کوتاه‌تر و با دقتی بیشتر بازیابی نمایند.

در پایگاه‌های اطلاعاتی، ورود کلیدواژه‌ها و عبارات جستجو، نخستین گام در فرایند بازیابی اطلاعات محسوب می‌گردد. به بیانی دیگر، تا زمانی که کاربر کلیدواژه یا عبارتی را به عنوان واژه(های) پرس‌وجو در جعبه پرس‌وجو وارد نکند، جستجوی انجام نمی‌پذیرد و بنابراین نیاز اطلاعاتی کاربر مرتفع نمی‌شود. بنابراین، کیفیت واژه‌های درون‌داد شده در جعبه پرس‌وجو از اهمیتی خاص برخوردار است. از طرفی افرادی که در تعیین کلیدواژه‌ها و درون‌داد آن‌ها در جعبه جستجو دخالت دارند نیز متفاوت و متعدّدند. از این لحاظ می‌توان دو گروه عمده را شناسایی نمود:

(۱) کاربران عمومی. محققان، اساتید، دانشجویان و تمامی افرادی که به نوعی اطلاعاتی را جستجو می‌نمایند، کاربر عمومی محسوب می‌گردند. کاربران عمدتاً موضوعاتی در محدوده تخصص خود را جستجو می‌کنند، با این حال، بیشتر اوقات در خصوص روال‌های اطلاع‌یابی و ترفندهای تعیین کلیدواژه، اطلاعات محدودی دارند.

(۲) محققان و متخصصان علم اطلاعات و دانش‌شناسی. این گروه به طور عمده در مراکز اطلاع‌رسانی، کتابخانه‌ها، کتابخانه‌های مجازی و خودکار و سازمان‌های مشابه حضور دارند و با اشرافی که به روال اطلاع‌یابی دارند، به کاربر در تعیین کلیدواژه یاری رسانده و گاه خود عملیات جستجو را به نیابت از او به انجام می‌رسانند. این دسته گاه آموزش‌هایی را نیز برای کاربران تدارک می‌بینند که به ارتقاء مهارت‌های اطلاع‌یابی کاربران منجر می‌گردد. وجه اشتراک هر دو نوع کاربر (عمومی و متخصص اطلاع‌رسانی) آن است که هر دو باید کلیدواژه‌هایی را در پایگاه اطلاعاتی وارد نمایند و وجه افتراق آنها نیز در کیفیت کلیدواژه‌هایی است که به سیستم درون‌داد می‌شود.

البته در مراکز اطلاع‌رسانی از ابزار کمکی نظیر تجارب پیشین متخصصان اطلاع‌رسانی، دانش زمینه‌ای کاربران، واژه‌نامه‌های تخصصی، اصطلاحنامه‌ها، سرعنوانهای موضوعی و ... به عنوان ابزار کمکی برای استخراج کلیدواژه‌ها استفاده می‌شود. با این همه، استفاده از ابزار جانبی چنانچه فرایند اطلاع‌یابی را طولانی نماید، یک مزیت محسوب نمی‌شود و از این رو است که پایگاه‌های اطلاعاتی و طراحان آنها همواره می‌کوشند روال اطلاع‌یابی را سهل‌تر، دقیق‌تر و کاربر پسندتر نمایند.

مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری (رایسست) پایگاه‌های اطلاعاتی متعددی را در خود جای داده است. برخی از این پایگاه‌ها (مانند پایگاه مقالات تمام‌متن فارسی و پایگاه نشریات الکترونیکی فارسی) از پایگاه‌های با مخاطبین بالا محسوب می‌گردند (مهراد، ۱۳۹۲). بنابراین، ارتقاء زیرساخت‌های پایگاه‌هایی از این دست و افزایش سهولت اطلاع‌یابی می‌تواند بر رضایت و میزان استفاده کاربران تأثیر مستقیم داشته باشد. نگاهی به موتور جستجوی گوگل نشان می‌دهد که این نظام در طی سال‌های گذشته توانسته خدمات متعددی (بیش از ۱۵۰ مورد) را به جامعه علمی ارائه نماید (موهان، ۲۰۱۴). یکی از این خدمات، سرویس گوگل ساجست (Google Autocomplete/Google Suggest) می‌باشد. بر این اساس و با توجه به تأثیر این خدمت در تسهیل و تسریع فرایند جستجو، هدف پژوهش حاضر به‌گونه‌ای ترسیم شد تا چنین خدمتی در نظام رایسست و برای پایگاه اطلاعاتی مقالات تمام‌متن فارسی، پیاده‌سازی گردد و بنابراین هدف آن بود تا نرم‌افزاری تولید و در نظام رایسست تلفیق گردد تا کاربران پایگاه مقالات تمام‌متن فارسی بتوانند: (۱) مقالات مورد نظر خود را در یک حوزه موضوعی خاص جستجو نمایند و (۲) با وارد کردن حداقل دو کاراکتر، کلیدواژه‌های برگرفته از مخزن کلیدواژه‌های رایسست (پنج مورد) جهت انتخاب به کاربران پیشنهاد شود.

البته مانند هر پژوهش دیگری در تحقیق حاضر نیز محقق با محدودیت‌هایی مواجه بوده است که در ادامه برخی از مهم‌ترین آن‌ها به اختصار ذکر می‌شود:

با توجه به تنوع پایگاه‌های اطلاعاتی موجود در رایسست از نظر محتوا، زبان و زیرساخت‌ها، در تحقیق حاضر، صرفاً بر روی پایگاه اطلاعاتی مقالات تمام‌متن فارسی تمرکز شد چرا که مخزن کلیدواژه‌های رایسست به این پایگاه مربوط بود. ضمن آنکه پرداختن به پایگاه‌های انگلیسی خود به ملزومات دیگری نیاز داشت که ترجیح داده شد در گام نخست صرفاً در حوزه پایگاه مربوطه فعالیت شود.

در این پژوهش برای وزن‌دهی به کلیدواژه‌ها از چهار متغیر با وزن یکسان استفاده شد که عبارت بودند از: فراوانی پرس‌وجو، جدیدترین تاریخی که پرس‌وجو در آن تاریخ زده شده است، تاریخ جدیدترین مقاله‌ای که کلیدواژه در آن ظاهر شده است و فراوانی کلیدواژه در پایگاه مقالات تمام‌متن فارسی مرکز منطقه‌ای. علت استفاده از وزن یکسان آن بود که پیرو بازخورد اخذ شده از سیستم، با همین روش نتیجه مناسب حاصل گردید. هر چند این چهار متغیر از برجسته‌ترین، معروف‌ترین و پرکاربردترین متغیرها در این حوزه از تحقیق به حساب می‌آیند، متغیرهای دیگری نیز وجود دارد که به دلیل پیچیدگی فرایند تحلیل، از آن‌ها در پژوهش حاضر استفاده نشد. رضایت کاربران در مرحله ارزیابی، از عملکرد سیستم نیز عامل دیگری بود که باعث شد متغیرهای دیگر، مورد جست‌وجو و استفاده قرار نگیرد.

با توجه به اینکه برای تحقق جستجوی موضوعی، تفکیک حوزه‌های موضوعی مقالات امری ضروری بود، همزمان با آغاز طرح مکاتبات لازم در این خصوص به انجام رسید تا از سوی گروه پژوهشی ارزیابی و توسعه منابع مرکز منطقه‌ای و در قالب یک پژوهش دیگر، تفکیک موضوعی نشریات نیز محقق گردد. بنابراین، کیفیت تفکیک موضوعی در پژوهش حاضر با کم و کیف تفکیک موضوعی نشریات در آن پژوهش ارتباط مستقیم خواهد داشت.

هر چه اطلاعات مربوط به رکوردهای اطلاعاتی مقالات فارسی کامل‌تر باشد عملکرد نرم‌افزار نیز به همان نسبت بهتر خواهد شد و به همین نسبت به هر میزان که این رکوردهای اطلاعاتی ناقص باشند، به همان میزان نیز از عملکرد سیستم کاسته خواهد شد. بررسی‌هایی که محقق حاضر به انجام رساند نشان داد که ۱۳ درصد مقالات فارسی تمام‌متن فاقد کلیدواژه بوده‌اند. با توجه به اینکه کلیدواژه، پایگاه داده اصلی در پژوهش حاضر بوده است، بدیهی است نبود کلیدواژه برای ۱۳ درصد از مقالات یکی از محدودیت‌های پژوهش حاضر محسوب می‌گردد. بدین معنی که ۱۳ درصد مقالات ناگزیر از جریان تحلیل کنار گذاشته شد.

## پیشینه پژوهش

در خصوص نظام‌های پیشنهاد دهنده کلیدواژه و پرس‌وجو در سطح دنیا کارهای متعددی انجام شده است (هر چند فعالیت‌های انجام شده در این خصوص برای زبان فارسی و پایگاه‌های اطلاعاتی فارسی محدود می‌باشد). البته، اطلاعات و زیرساخت‌های نظام‌های موجود بر عموم محققان آشکار نیست. در حقیقت، در موارد بسیار محققان باید به طور غیرمستقیم در خصوص نحوه عملکرد اینگونه نظام‌ها اطلاعات اخذ کنند. وایتینگ و جوز (۲۰۱۴، ص. ۳) بیان می‌دارند: «با وجود اهمیت بسیار زیاد نظام‌های پیشنهاد دهنده پرس‌وجو، برای ارتقاء این گونه نظام‌ها آنطور که باید مطالعه صورت نگرفته است که دلیل آن هم عمدتاً در دسترس نبودن اطلاعات پرس‌وجوها برای عامه محققان و انحصاری بودن این گونه داده‌ها برای شرکت‌های خاص می‌باشد». هر چند اطلاعات ما در خصوص نرم‌افزارهای عمده و رایجی چون گوگل محدود می‌باشد، در سال‌های گذشته پایان‌نامه‌ها و مقالات متعددی در این زمینه نگاشته شده که می‌تواند به عنوان منبع مفیدی از اطلاعات مورد استفاده قرار گیرد. از این رو، در ادامه برخی از کارهای عمده انجام شده که با محتوای پژوهش حاضر ارتباط بیش‌تری نیز دارند مرور خواهند شد.

سولیوان (۲۰۱۱) در مقاله مفصل خود با عنوان «How google instant's autocomplete suggestions work»

به تفصیل به بررسی مختصات گوگل ساجست پرداخت. مهم‌ترین نکاتی که وی به آن‌ها اشاره کرد عبارت بودند از: استفاده از جستجوهای واقعی به عنوان مبنائی برای ارائه پیشنهادات؛ حساسیت به مکان جغرافیایی و زبان؛ سابقه جستجوها؛ نحوه رتبه‌بندی پیشنهادات؛ اصلاحات؛ تازگی اطلاعات؛ حذف پیشنهادات؛ خطای برنامه و نقش کاربران در جهت‌دهی به پیشنهادات. همچنین سینگ استاک و گرتز (۲۰۱۱) نرم‌افزاری را پیشنهاد دادند که به زمان درونداد پرس‌وجو در طول شبانه روز

حساس بود. به عبارت دیگر، نرم‌افزار می‌توانست بر حسب اینکه در چه زمانی از شبانه روز پرس‌وجو وارد شده، رتبه‌بندی نظام‌های پیشنهاد دهنده پرس‌وجو را ارائه نماید. به‌تایا، مجومدار و میترا (۲۰۱۱) نیز برای استخراج و سپس رتبه‌بندی پرس‌وجوها شیوه‌هایی را ارائه نمودند و در انجام این مهم بر روی عبارات عامی تمرکز کردند که از اسناد نمایه شده استخراج شده بود. گلبندی، کاتزیر، کورن و لمپل (۲۰۱۳) یک مدل رگرسیون را پیشنهاد دادند که می‌توانست ضمن شناسایی پرس‌وجوهای مطرح، فرایند شناسایی و عرضه پرس‌وجوهای پیشنهادی را ارتقاء بخشد. استریژوسکایا، بایتین، گالینسکایا و سردیوکوف (۲۰۱۲) نیز دقت پیش‌بینی پرس‌وجو را در مدل‌های معروف مختلف برای جعبه پرس‌وجوی موتور جست‌وجوی یاندکس (Yandex) اندازه‌گیری کردند. در اثری دیگر، بار-یوسف و کراس (۲۰۱۱) در مقاله خود روشی بافت-محور را برای پیشنهاد خودکار پرس‌وجو ارائه دادند. طبق نظر این دو محقق، پیشنهاد خودکار پرس‌وجو وقتی که بر مبنای درون‌داد یک زنجیره کوتاه (یک یا دو کاراکتر) انجام شود، نتیجه چندان رضایت بخش نخواهد بود. از این رو، آنها نشان دادند که با استفاده از متغیر «پرس‌وجوهای اخیر کاربر» می‌توان پیشنهاد‌های بهتری را حتی با درون‌داد یک یا دو کاراکتر، از سیستم دریافت کرد. آنها یک متغیر حساس به بافت را با نام «نزدیک‌ترین گزینه» (Nearest Completion) پیشنهاد دادند. کاری که این متغیر انجام می‌دهد این است که پرس‌وجوهای را که کاربر پیشتر وارد کرده با پرس‌وجوی حاضر مقایسه می‌کند و شبیه‌ترین پرس‌وجوی پیشین با پرس‌وجوی حاضر را ارائه می‌دهد. این محققان برای محاسبه میزان شباهت نیز پرس‌وجوها و بافت‌ها را به عنوان بردارهایی در نظر گرفتند که بر روی هر کدام وزن پرس‌وجوها مشخص شده بود. سپس با استفاده از روش Cosine میزان شباهت پرس‌وجوی قبلی به کاراکترهای درون‌داد شده فعلی تعیین شد. نگاشت پرس‌وجوها به بردارها از طریق یک روش پیشنهادی جدید گسترش پرس‌وجو انجام شد. در این روش، پرس‌وجو از طریق بررسی درخت پیشنهاد پرس‌وجو که ریشه آن نیز در پرس‌وجو است، گسترش می‌یابد. وقتی که پرس‌وجوهای کاربر به پرس‌وجوی اخیر که کاربر در حال تایپ آن است مربوط باشد، پس از تایپ یک کاراکتر، متغیر «نزدیک‌ترین گزینه» ۴۸ درصد بیش‌تر از مقدار متوسط این متغیر در حالت عادی خواهد بود. ضمناً، وقتی که بافت نامرتب باشد مقدار متغیر «نزدیک‌ترین گزینه» برابر با صفر خواهد بود. برای رفع این مشکل این دو محقق از متغیر «نزدیک‌ترین گزینه» در ترکیب با متغیر دیگری با نام «مشهورترین گزینه» (Most Popular Completion) استفاده کردند. بررسی‌ها نشان داد که استفاده از این مدل ترکیبی در مقایسه با استفاده از هر یک از این دو مدل به صورت مجزا به میزان ۳۱/۵ درصد کیفیت پرس‌وجو را بهبود می‌دهد.

وایتینگ و جوز (۲۰۱۴) در مقاله خود مدلی انعطاف‌پذیر را برای پیشنهاد خودکار پرس‌وجو ارائه دادند. طبق نظر آنها، پیشنهاد دهنده خودکار پرس‌وجو باید دارای قابلیت‌های زیر باشد: (۱) با وارد کردن کمترین طول زنجیره حروف فعال شود و (۲) پرس‌وجوهای پیشنهادی را اولویت و رتبه‌بندی کند. غالباً نظام‌های پیشنهاد دهنده خودکار جستجو، برای رتبه‌بندی پرس‌وجو از معیار مقبولیت و شهرت پیشین استفاده می‌کنند. از این رو، این گونه سیستم‌ها غالباً برای پرس‌وجوهای که پیشتر نیز جستجو شده‌اند موفق عمل می‌کنند. اما کاربران ممکن است پرس‌وجوها و موضوعاتی را هم وارد نمایند که پیشتر آنها را جستجو نکرده باشند و بنابراین هیچ تاریخچه‌ای در سیستم برای آن پرس‌وجوها ذخیره نشده باشد. در نتیجه، سیستم باید هم انعطاف‌پذیر و هم حساس به زمان باشد. وایتینگ و جوز سه رویکرد عملی را برای رتبه‌بندی پیشنهادها ارائه دادند. رویکرد اول میزان معروفیت پرس‌وجو را در طی ۲ تا ۲۸ روز گذشته بررسی و تعیین می‌کند. رویکرد دوم توزیع معروفیت پرس‌وجو را در  $N$  تعداد پرس‌وجو با یک زنجیره حروف خاص تعیین می‌کند. سرانجام، رویکرد سوم معروفیت پرس‌وجو را در کوتاه مدت و بر اساس روندهایی که به تازگی مشاهده شده‌اند، پیش‌بینی می‌کند. این محققان متغیرهای ضروری برای ارتقاء کرائی سیستم را با سه بانک رایگان پرس‌وجو به نام‌های MSN، AOL، و Sogou و برای ۲ تا ۵ کاراکتر بررسی کردند. مقایسه مدل پیشنهادی این محققان با مدل‌های قبلی نشان داد که با استفاده از متغیر جدید وقتی که سیستم بر مبنای درون‌داد ۲ تا ۳ کاراکتر کار می‌کند، به میزان ۹/۲ درصد در کیفیت پرس‌وجوها بهبود حاصل می‌شود. این بهبود عملکرد از آنجا ناشی می‌شود که مدل پیشنهادی این محققان به دلیل انعطاف‌پذیری قادر بوده است پرس‌وجوهای را که هیچ تاریخچه‌ای از آنها در سیستم وجود ندارد با کمک سه متغیر معرفی شده به گونه‌ای بهتر پردازش نماید.

کاستریناکیس و زیتزیکاس (۲۰۱۰) در مقاله خود روش‌هایی را برای بهبود نظام‌های پیشنهاد دهنده پرس‌وجو ارائه نمودند و در ارائه این روش‌ها بر روی افزایش تعداد و کیفیت پرس‌وجوهای پیشنهادی متمرکز شدند. آنها روشی جدید را برای تفکیک (Partitioning) ساختار داده درونی پیشنهاد کردند که ضمن حفظ تعداد پرس‌وجوهای پیشنهادی قادر است سرعت نظام و تعداد پرس‌وجوهای پیشنهادی را افزایش داده و سنجش عملکرد آن را نیز تسهیل نماید. در مدل رتبه‌بندی که این محققان ارائه دادند هرچه پیشنهادی بتواند به پیشنهادهای مرتبط بیشتری ختم شود، آن پیشنهاد از اهمیت و اولویت بیشتری برخوردار خواهد بود و لذا به عنوان انتخاب‌های ارجح در رأس پیشنهادها ظاهر می‌شود. شهرت پرس‌وجو و نیز تعداد پیشنهادهایی که پیشنهاد حاضر را به عنوان پیشنهاد در ساختار خود دارند نیز دو عامل مهم دیگری است که محققان حاضر در فرایند رتبه‌بندی پرس‌وجوها از آن‌ها بهره بردند.

ژیائو، کین و وانگ (۲۰۱۳) در مقاله خود اذعان داشتند که بکارگیری نظام‌های پیشنهاد پرس‌وجو می‌تواند در وقت کاربر صرفه‌جویی قابل توجهی ایجاد نماید. این محققان به موضوع تحمل خطاهای تاپی کاربران در حین تکمیل پرس‌وجو در نظام‌های پیشنهاد دهنده پرس‌وجو پرداختند. این محققان ضمن اشاره به روش‌های موجود برای حل مشکل خطاها، عنوان داشتند این روش‌ها با مشکلات و محدودیت‌هایی نیز رو به رو هستند. به عنوان مثال، آن‌ها عنوان داشتند که تعداد مدخل‌هایی که با حروف یکسان شروع می‌شوند بسیار زیاد است و این امر فرایند بررسی مدخل‌ها را کند می‌کند. بر این اساس، محققان در مقاله حاضر خود یک الگوریتم تولید همسایگی جدید (Novel Neighborhood Generation Based Algorithm) را با نام IncNGTrie پیشنهاد دادند که ضمن پرداختن به خطاهای نگارشی در جعبه پرس‌وجو، از سرعتی معادل دو برابر سرعت نظام‌های مشابه دیگر برخوردار است. این نظام از تعداد محدودی گره (Node) فعال استفاده می‌کند و بدین ترتیب فرایند پردازش پرس‌وجو در زمان و فضای کم‌تر به انجام می‌رسد.

کیراسلا (۲۰۰۷) به تفصیل کاربردهای مختلف نظام پیشنهاد دهنده پرس‌وجوی گوگل ساجست را در صفحات ۶ تا ۸ مقاله خود تشریح کرد. وی بیان داشت که این نظام به کاربران کمک می‌کند تا با وارد کردن بخش کوچکی از پرس‌وجو، پرس‌وجوهای پیشنهادی را از نظام دریافت دارند و بدین ترتیب صرفه‌جویی زیادی در زمان کاربر حاصل می‌گردد. ضمناً، او اعلام کرد که می‌توان از این نظام به طور غیرمستقیم هم استفاده کرد. برای مثال، چنانچه کاربر املائی درست کلمه یا عبارتی را نداند و فقط چند حرف آغازین را بداند می‌تواند با وارد کردن آن حروف بازخورد نظام را دریافت دارد. البته، منظور اصلی کیراسلا از استفاده غیرمستقیم از این نظام در موتورهای کاوش آن است که می‌تواند خطاهای املائی کاربران را تصحیح نماید و با وجود خطای املائی در زنجیره حروف وارد شده، کلمات درست را به روئیت کاربر رسانده و در صورت تأیید کاربر، متناسب با آن نتایج جستجو را ارائه نماید. وی در مورد دقت این نظام نیز به ذکر توضیحی می‌پردازد بدین ترتیب که نظام از پرس‌وجوهای مشهور و رایج به عنوان یک متغیر اساسی برای پیشنهاد پرس‌وجوها استفاده می‌کند که بر دقت اطلاعات پیشنهادی می‌افزاید.

وایتینگ، مک مین و جوز (۲۰۱۳) بر این نکته تأکید کردند که با توجه به اینکه در نظام‌های پیشنهاد دهنده پرس‌وجو، از سوابق پرس‌وجوی کاربر به عنوان یک عامل کلیدی استفاده می‌شود، وقایع و حوادث اخیر که بیشتر مورد جستجو قرار نگرفته یا کم جستجو شده شانس زیادی را برای ظهور در فهرست پیشنهادها نخواهند داشت. بر این اساس، این محققان روشی را برای محدود و کوتاه کردن تاریخچه جستجوها ارائه دادند تا بتوانند اتفاقات جدید را به عنوان پیشنهاد احتمالی بهتر منعکس نمایند. بر این اساس، آن‌ها زمان بررسی تاریخچه کلیدواژه‌ها را ۲ تا ۱۴ روز در نظر گرفتند و طول زنجیره متنی را نیز ۲ تا ۵ کاراکتر تعیین کردند. تحلیل‌ها نشان داد که بین طول پرس‌وجوی درون‌داد شده توسط کاربر و عملکرد نظام پیشنهاد دهنده کلیدواژه وقتی که دوره‌های زمانی مختلف مورد بررسی قرار بگیرد، رابطه خطی و مستقیمی وجود دارد. محققان به ویژه گزارش کردند که وقتی طول کاراکترهای وارد شده کمتر باشد، با کاهش طول تاریخچه مورد بررسی، عملکرد نظام پیشنهاد دهنده پرس‌وجو بهتر می‌شود و بر عکس هر چه طول کاراکترهای درون‌داد شده در جعبه پرس‌وجو بیشتر باشد، با افزایش طول تاریخچه مورد بررسی، عملکرد نظام ارتقاء خواهد یافت.

بیالزاک، مازورزیک و زیپیورسکی (۲۰۱۴) در خصوص ارسال داده‌های پنهان از طریق گوگل ساجست به مطالعه پرداختند. این محققان در مقاله خود ضمن تشریح ویژگی‌ها و مزایای این امکان در موتور کاوش گوگل، روشی را با عنوان استگ ساجست

(StegSuggest) پیشنهاد کردند. هدف شبکه‌های استیگنوگرافی (Steganography) آن است که در خلال تعاملات کاربران با شبکه، اطلاعات و داده‌های محرمانه را مخفی می‌کند. در این روش از پیشنهاد‌های ارائه شده توسط گوگل ساجست به عنوان یک ناقل اطلاعات پنهان استفاده شد. محققان برای توجیه روش خود، شبکه‌ی ترافیکی تولید شده توسط گوگل ساجست را مورد تجزیه و تحلیل قرار دادند و از این طریق کرائی مدل خود را توجیه کردند. آن‌ها از تحلیل ترافیک شبکه همچنین برای کشف دو مختصه Window Scale و Timestamp استفاده کردند که گوگل ساجست بر اساس آن‌ها کار می‌کند. ارزیابی مدل ارائه شده نشان داد که می‌توان ۱۰۰ بیت استیگنوگرام را در هر فهرست پیشنهادی که توسط گوگل ساجست ارسال می‌شود، وارد نمود.

کایز، لیانگ و دو ریک (۲۰۱۴) نیز در مقاله خود ضمن تشریح مفهوم و مختصات نظام‌های پیشنهاد دهنده پرس‌وجو، از شهرت پرس‌وجوها در گذشته به عنوان یکی از مختصه‌های رایج در این گونه نظام‌ها یاد کردند. با این همه، آن‌ها تأکید کردند که موضوع شهرت یک پرس‌وجو می‌تواند به عوامل متعددی بستگی داشته باشد. آن‌ها به دو عامل عمده در این زمینه اشاره کردند: (۱) امکان تغییر مختصه شهرت با گذشت زمان و (۲) امکان تغییر مختصه شهرت در میان کاربران مختلف. با این مقدمه آن‌ها بیان داشتند که در رتبه‌بندی پیشنهادها لازم است این عوامل نیز در نظر گرفته شوند. آن‌ها ضمن استفاده از مدل‌های حساس به زمان و حساس به کاربر موجود، روشی ترکیبی و برگرفته از این دو مدل را ارائه نمودند. نتایج به دست آمده از این مدل ترکیبی بر روی دو مجموعه داده متفاوت حاکی از آن بود که عملکرد این مدل ترکیبی از عملکرد هر یک از دو مدل پیشین (حساس به زمان و حساس به کاربر) به ترتیب به میزان ۳ و ۷ درصد بهتر بوده است.

وارد، هان و فیست (۲۰۱۲) نحوه استفاده از نرم‌افزار پیشنهاد دهنده پرس‌وجو را در واسط‌های جستجو بررسی کردند. آن‌ها در صدد بودند اولویت‌های کاربران را برای استفاده از این نظام در بافت جستجوهای دانشگاهی تعیین کنند. کاربردهایی که پیرو آزمایش‌های به عمل آمده استخراج گردید به شرح زیر بود: همه کاربران بر این باور بودند که یکی از کاربردهای عام این نظام، اصلاح املائی پرس‌وجو است؛ کاربران برخلاف روال مرسوم غالباً با روش آزمایش و خطا کلیدواژه‌ها یا عبارات پرس‌وجوی خود را انتخاب می‌کنند و از این منظر نظام پیشنهاد دهنده پرس‌وجو می‌تواند بسیار مفید باشد؛ استفاده از این نظام می‌تواند سبب سرعت بخشیدن به فرایند جستجو و اطلاع‌یابی شود و سرانجام استفاده از این گونه نظام‌ها می‌تواند تمرکز بیشتری را در فرایند جستجو سبب گردد. البته، دانشجویان پیشنهاد‌هایی را نیز ارائه نمودند. برای مثال، برخی اعلام کردند چنانچه تعداد پرس‌وجوهای پیشنهادی بیشتر شود نظام از عملکرد بهتری برخوردار خواهد بود.

خسروی، فتاحی، پربرخ و دیتانی (۱۳۹۲) در مقاله بسیار مؤثر و کاربردی خود، کارآمدی کلیدواژه‌ها و عبارت‌های پیشنهادی موتور کاوش گوگل را در بسط جستجو و افزایش ربط از دیدگاه دانشجویان تحصیلات تکمیلی بررسی نمودند. طبق نظر این محققان، موتورهای کاوش قابلیت جدیدی به نام کلیدواژه‌ها و عبارت‌های پیشنهادی را پیاده نموده‌اند. این که این کلیدواژه‌ها تا چه میزان با نیاز کاربران همخوان است، مسئله‌ای اساسی است. آن‌ها پژوهش کاربردی خود را به روش پیمایشی به انجام رساندند. داده‌های تحقیق با استفاده از پرسشنامه و به هنگام جستجوی کاربران در گوگل به دست آمد. در مجموع از یک نمونه ۶۰ نفری از دانشجویان مربوط به دو حوزه علوم انسانی/اجتماعی و علوم پایه/فنی-مهندسی برای گردآوری داده‌ها استفاده به عمل آمد. در ضمن، با هر یک از آزمودنی‌ها نشست‌هایی به صورت جداگانه ترتیب داده شد. یافته‌های این پژوهش نشان داد که بین میزان ربط نتایج بازیافتی حاصل از عبارت جستجوی اولیه و نتایج بازیافتی حاصل از بسط جستجوی مبتنی بر کلیدواژه‌ها/عبارت‌های پیشنهادی، تفاوت معناداری وجود دارد. یعنی، کلیدواژه‌ها/عبارت‌های پیشنهادی گوگل در بهبود ربط بازیافت‌ها مؤثر بوده است. همچنین بین نتایج بازیافتی حاصل از استفاده از عبارت جستجوی اولیه و نتایج بازیافتی حاصل از بسط جستجو بین دو حوزه مورد بررسی تفاوت معناداری مشاهده نشد. این محققان در پایان به منظور بهبود نتایج بازیافتی توصیه‌هایی را ارائه دادند، به ویژه آنکه توصیه کردند «با توجه به کارآمدی کلیدواژه‌ها و عبارت‌های پیشنهادی در تدوین درست عبارت جستجو و کمک به کاربران در بازیابی اطلاعات مرتبط تر ... طراحان نظام‌های بازیابی اطلاعات از جمله نرم‌افزارهای کتابخانه‌ای و پایگاه‌های اطلاعاتی (مقاله‌ها و پایان‌نامه‌ها) که قابلیت جستجوی کلیدواژه‌ای دارند، در به کارگیری این قابلیت در رابط کاربر نرم‌افزارهای خود اقدام نمایند» (همان، ص. ۱۴۸).



شکوهی (۲۰۱۳) در مقاله خود مختصاً پیشنهاد پرس‌وجو را به عنوان یکی از برجسته‌ترین ویژگی‌های موتورهای کاوش جدید معرفی کرد. او بیان داشت که نظام‌های پیشنهاد دهنده پرس‌وجو پیرو درونداد یک یا چند کاراکتر در جعبه پرس‌وجو فعال می‌شوند و اینکه این نظام‌ها غالباً از متغیر جستجوهای پیشین به عنوان یک متغیر اساسی استفاده می‌کنند. نکته جالبی که شکوهی به آن اشاره کرد این است که شهرت پرس و جویها برای کاربران مختلف و نیز محیط‌های مختلف، متفاوت است. برای مثال، او به محیط‌های اینستاگرام و ایمدیب (Imdb) اشاره می‌کند که محیط اول بیشتر مورد علاقه خانم‌های جوان و محیط دوم بیشتر مورد توجه آقایان است و بنابراین شخصی‌سازی پرس‌وجوها می‌تواند متغیری بسیار مهم قلمداد گردد. او در مقاله خود چارچوبی نظارت شده (Supervised) را برای شخصی‌سازی رتبه‌بندی پرس‌وجوها ارائه داد. او یک رویکرد برچسب‌دهی جدیدی را نیز ارائه داد که می‌تواند رتبه‌بندی‌های شخصی‌سازی شده را یاد گرفته و تولید نماید. او ضمن بررسی متغیرهای مختلف به این نتیجه رسید که موقعیت مکانی و نیز سابقه طولانی مدت یک فرد در جستجو، دو متغیری هستند که می‌توانند در فرایند شخصی‌سازی رتبه‌بندی‌ها بیشترین نقش را ایفا نمایند. او آزمایش خود را بر روی پرس‌وجوهای AOL و نیز Bing انجام داد. نتایج حاکی از آن بود که رتبه‌بندی نظارت شده که توسط مشخصه‌های شخصی‌سازی شده ارتقاء یافته‌اند در مقایسه با آن دسته از نظام‌های متداول فعلی که فاقد مشخصه‌های شخصی‌سازی شده‌اند، از عملکرد بهتری (تا ۹ درصد) برخوردارند. شکوهی و رادینسکی (۲۰۱۲) و نیز شکوهی (۲۰۱۱) برای افزایش کارایی نظام‌های پیشنهاد دهنده پرس‌وجو، الگوهای بدست آمده پرس‌وجو را در گذشته مدل‌بندی کردند.

طاهریان و حیاتی (۱۳۸۸) نیز در پژوهشی دیگر الگوی جستجوی کلیدواژه‌ای با دریافت کمک را با الگوی جستجوی بدون دریافت کمک در موتور کاوش گوگل از دیدگاه کاربران بررسی کردند. نتایج به دست آمده توسط محققان حاکی از آن بود که حدود ۷۰ درصد از پیشنهاد‌های گوگل ساجست بتا به عنوان یک الگوی کلیدواژه‌ای با دریافت کمک، از ربط متوسط به بالا برخوردار بود. ضمناً، استفاده از این امکان افزایش رضایت کاربران را از کاربرپسند بودن سیستم و عملکرد کلی نظام سبب گردید.

فتاحی (۱۳۸۵) در پژوهش خود با استفاده از رویکرد بسط جستجو و با بهره‌گیری از زبان طبیعی، پژوهشی را پیرامون موتور کاوش گوگل انجام داد. نتایج به دست آمده از این پژوهش نشان داد که استفاده از واژگان عمومی برای بسط جستجو، به تولید نتایج مرتبط‌تر و دقیق‌تر می‌انجامد و از این رو استفاده از این نوع واژگان در فرایند جستجو توصیه می‌شود. کلانتر و منصفی (۱۳۸۱) مبحث فیلترینگ داده‌ها و اطلاعات را با توجه به حجم گسترده اطلاعات در وب مورد توجه قرار دادند. آن‌ها ضمن اشاره به رشد انفجاری اینترنت به ویژه وب جهان گستر، یافتن اطلاعات مورد نیاز کاربران را دغدغه‌ای اساسی بر شمرند. آن‌ها تأکید کردند با وجود اینکه ابزار کلاسیک یافتن اطلاعات مانند موتورهای کاوش تا اندازه‌ای در رفع این مشکل کمک نموده، ولی هنوز جایگاه ابزارهایی که دارای ویژگی‌های شخصی‌سازی، تطبیق‌پذیری با تغییر نیازها و بازخورد متناسب باشند، خالی است. در این مقاله، محققان با تلفیق دو فناوری فیلترینگ اطلاعات و عامل‌ها، نظامی را با نام (WIFS) Web Information Filtering System پیشنهاد دادند. در این نظام محققان از الگوریتم جدیدی استفاده کردند که برای فیلترینگ اطلاعات از بازخورد مثبت و منفی کاربر استفاده می‌کند. همچنین از الگوریتم‌های ژنتیک برای به دست آوردن پرس‌وجوی بهینه که نمایش دهنده نیازهای اطلاعاتی کاربر باشد، استفاده شده است. با توجه به منبع برشمرده و اندک‌شمار بودن پژوهش‌های کاربردی در خصوص نرم‌افزارهای پیشنهاد دهنده کلیدواژه و پرس‌وجو در ایران و برای زبان فارسی، لازم است در این زمینه پژوهش‌های گسترده‌ای به انجام برسد. انجام پژوهش حاضر نیز دقیقاً با چنین هدفی صورت پذیرفت.

## روش پژوهش

برای انجام تحقیق حاضر از مخزن کلیدواژه‌های مقالات رابست به عنوان منبع اصلی داده تحقیق استفاده شد. کلیدواژه‌های مربوط به مقالات فارسی و انگلیسی در این مخزن وجود دارد که البته برای انجام پژوهش حاضر صرفاً از کلیدواژه‌های فارسی استفاده به عمل آمد. با توجه به اینکه هدف اصلی از انجام پژوهش حاضر ساخت نرم‌افزار پیشنهاددهنده

کلیدواژه و تلفیق آن در نظام رایسست با هدف ارتقاء این نرم‌افزار بومی بوده است، از زبان برنامه نویسی #C به عنوان اولین ابزار مورد استفاده در این پژوهش استفاده به عمل آمد. ضمناً با توجه به اینکه مخزن کلیدواژه‌های رایسست عمدتاً از طریق اطلاعات مقالات علمی تغذیه می‌شود، از پایگاه تمام‌متن مقالات فارسی نیز به عنوان دومین ابزار تحقیق استفاده به عمل آمد. به عبارت دیگر، سعی شد پایگاه تمام‌متن مقالات فارسی مبنا قرار گیرد و امکانی جدید (امکان جست‌وجوی موضوعی و نیز پیشنهاد پرس‌وجو) به این پایگاه اضافه شود.

برای انجام پژوهش حاضر به ترتیب کارهای زیر به انجام رسید: نخست، مخزن کلیدواژه‌های رایسست به عنوان داده پشتیبان مورد بررسی و واری‌گری قرار گرفت و راهکارهای استفاده از اطلاعات آن برای پژوهش حاضر بررسی شد. ضمناً، در این مخزن برای استفاده بهینه از اطلاعات آن در پروژه حاضر اصلاحات لازم انجام شد ضمن آنکه مدخل‌های مخزن از ترکیبات پیچیده به ترکیبات ساده تبدیل شد. همچنین الگوریتم ساخت نرم‌افزار رایسست به عنوان داده پشتیبان مورد بررسی قرار گرفت. سپس الگوریتم ساخت نرم‌افزار پژوهش حاضر تهیه شد. این نرم‌افزار دارای بخش‌های متعددی است که این بخش‌ها را می‌توان از نظر نوع کاربر به دو بخش عمده تقسیم کرد: بخش مربوط به کاربر سیستمی و بخش مربوط به کاربر نهایی. در حقیقت، تمامی بخش‌هایی که به پردازش داده‌ها و روزآمدسازی بانک داده نرم‌افزار مربوط می‌شوند، صرفاً از سوی کاربر سیستمی مورد استفاده قرار می‌گیرند که به عنوان بخش پشتیبان عمل می‌کند. شرط موفقیت نرم‌افزار در رفع نیاز اطلاعاتی کاربر نهایی آن است که کاربر سیستم، نخست پردازش‌های لازم را به انجام رسانده و نرم‌افزار را به صورت ماهیانه روزآمد کرده باشد. این بخش هیچگاه به رویت کاربر عادی که همان کاربر نهایی است، نمی‌رسد. بخش دوم همان تولید نسخه جدید از صفحه مربوط به پایگاه تمام‌متن مقالات فارسی است. به عبارت دیگر، نرم‌افزار تولید شده در این پژوهش در نسخه اولیه پایگاه تمام‌متن مقالات فارسی تلفیق گردید (منظور از نسخه اولیه، پایگاه تمام‌متن مقالات فارسی پایگاهی است که پیش از انجام این پژوهش در وبگاه مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری به آدرس [www.ticest.ac.ir](http://www.ticest.ac.ir) وجود داشته است یعنی همان نسخه‌ای که کاربران برای دسترسی به متن کامل مقالات تمام‌متن فارسی وارد آن می‌شده‌اند. به همین ترتیب مراد از نسخه جدید، نسخه ارتقاء یافته همین پایگاه است. به این ترتیب که با تولید نرم‌افزار ساخته شده و تلفیق آن در پایگاه تمام‌متن مقالات فارسی، قابلیت‌های جدیدی در معرض استفاده کاربران قرار می‌گیرد. بدین معنی که کاربران نهایی می‌توانند برخلاف نسخه قبلی پایگاه، جستجوی خود را در یک حوزه موضوعی خاص مانند علوم انسانی، علوم فنی و مهندسی، علوم پزشکی و ... به انجام برسانند، ضمن آنکه برخلاف نسخه قبلی کاربران امکان می‌یابند با درونداد حداقل ۲ کاراکتر پرس‌وجوهائی (۵ مورد) را به صورت پیشنهاد دریافت دارند. بنابراین، پایگاه مربوطه در وبگاه مرکز وجود داشته اما دو امکان ارائه شده و تلفیق شده در آن پایگاه کاملاً جدید محسوب می‌شود.

## تحلیل داده‌ها

### مراحل ساخت نرم‌افزار

#### مرحله پیش پردازش: توسعه نرم‌افزار رایسست جهت ارتباط بین پرس‌وجو و نوع پایگاه داده مربوطه

هر مقاله علمی دارای ارکان مختلفی است که از آن جمله می‌توان به عنوان، کلیدواژه‌ها، چکیده، سال انتشار، نام مجله، شماره، جلد و ... اشاره نمود. با توجه به جایگاه اساسی کلیدواژه‌ها در نظام‌های ذخیره و بازیابی اطلاعات و با توجه به اینکه این مختصه از بدو راه‌اندازی رایسست، در نرم‌افزار این مرکز موجود بوده است، از کلیدواژه‌ها به عنوان مبنائی برای ساخت نرم‌افزار پیشنهاددهنده پرس‌وجو استفاده شد.

برای ایجاد این امکان نیاز بود تا مختصاتی تعریف شوند تا بتوان در مراحل بعدی از آن‌ها استفاده کرد. سپس نیاز بود این مختصات استانداردسازی و پردازش شوند تا در نهایت بتوانیم بر اساس آن‌ها تصمیم بگیریم که چه کلیدواژه‌هایی باید به کاربران پیشنهاد شود که با موضوعات مورد نظر آن‌ها بیش‌ترین قرابت موضوعی را داشته باشد. یکی از این مختصات «پرس‌وجوها» بودند که کاربران در طی سال‌ها در نرم‌افزار رایسست وارد نموده‌اند. نکته اینجاست که پیش از اجرای پروژه حاضر، پرس‌وجوها در رایسست ذخیره می‌شدند ولی معلوم نبود که در چه پایگاهی (در موتور جستجوی جامع یا در پایگاهی دیگر) توسط کاربر



درونداد شده‌اند. به همین خاطر، مرحله‌ی پیش پردازش در پژوهش حاضر گنجانده شد. کاری که در مرحله‌ی پیش پردازش به انجام رسید این بود که کدبندی رایسست توسعه یافت. نحوه‌ی انجام این توسعه نیز به این ترتیب بود که در پایگاه داده‌ی رایسست بین هر پرس‌وجو و پایگاهی (پایگاه تمام‌متن مقالات فارسی و یا موتور جستجوی جامع) که آن پرس‌وجو در آن درونداد گردیده، لینک مستقیم برقرار شد.

در کنار پرس‌وجوهائی که در پایگاه مقالات تمام‌متن فارسی زده می‌شود، کاربران گاه ترجیح می‌دهند برای جستجوی اطلاعات خود، از موتور جستجوی جامع که درهمکرد اطلاعات پایگاه‌های مختلف در آن وجود دارد نیز استفاده نمایند. در موتور جستجوی جامع همه نوع فرآورده‌ی اطلاعاتی شامل مقاله، کتاب، طرح و ... قابل جستجو است. با توجه به اینکه کاربران غالباً از مقالات موجود در پایگاه استفاده می‌نمایند و ضمناً گاه برای جستجو، همه‌ی موارد (کتاب، مقاله، طرح و ...) را انتخاب می‌کنند، بسیاری از موارد نمی‌توان تشخیص داد که کاربر برای دسترسی به چه نوع فرآورده‌ی اطلاعاتی، در موتور جستجوی جامع، پرس‌وجویی خاص را درونداد کرده است. به همین خاطر و با توجه به اینکه تعداد کلیدواژه‌های درونداد شده در موتور جستجوی جامع قابل ملاحظه بود، محقق تصمیم گرفت برای استفاده از ظرفیت کلیدواژه‌های موجود در این پایگاه، پرس‌وجوهایی را که در موتور جستجوی جامع زده شده بود نیز به اسم کلیدواژه‌های مرتبط با پایگاه اطلاعاتی مقالات تمام‌متن فارسی، ثبت نماید. به بیان ساده، کلیدواژه‌های درونداد شده در موتور جستجوی جامع به عنوان کلیدواژه‌های پایگاه تمام‌متن مقالات فارسی شناخته شدند و بنابراین از کلیدواژه‌های درونداد شده (پرس‌وجوها) در هر دو پایگاه (شامل موتور جستجوی جامع و پایگاه مقالات تمام‌متن فارسی) به عنوان مخزن داده‌های تحقیق استفاده شد.

### مرحله ۱: طراحی و پیاده‌سازی فاز اول نرم‌افزار فراخوان خودکار کلیدواژه (پردازش پرس‌وجوها)

پرس‌وجوها در موتورهای جستجوی منطقی، ترکیباتی منطقی‌اند یعنی در ساختار آن‌ها عملگرهایی نظیر AND و OR

به کار رفته است. تصویرهای ۱ و ۲ به ترتیب عملگرهای AND «و» و OR «یا» را در ساختار پرس‌وجوها نشان می‌دهند.

SearchWord	date
علم و ندی	30:31.4
علم و ندی	31:18.1
علم و ندی	31:51.9
علم و ندی	32:23.5
سرفصل و کنلاری	33:38.0
سرفصل و کنلاری	34:06.4
بکزی و کافنی	37:26.8
بکزی و کافنی	39:03.1
توری و انگر	47:16.5
بکزی و کافنی	52:54.2
بکزی و کافنی	53:16.9
هیرات و گاز	17:25.0
هیرات و گاز	18:05.3
هیرات و گاز	18:30.0
خرافه و یوسنی	01:52.2
عزت و نقس	08:31.0
عزت و نقس	09:10.1
عزت و نقس	09:43.9
عزت و نقس	10:58.6
عزت و نقس	12:56.7
عزت و نقس	14:36.9
عزت و نقس	15:15.7
عزت و نقس	19:22.8
ناور و های و نخر و منطقی	20:20.2
ناور های و نخر و منطقی	20:37.7
سیمس و نکرید و جانی	55:35.3
اعاید و البکرتک و نرغ	34:52.0
اعاید و البکرتک و نرغ	35:06.7
شبر و و نرغ	39:08.6
studies in conservation	47:20.4

تصویر ۱. پرس‌وجوی حاوی عملگر AND «و» (۴۳۰۴۴۲ مورد)

1	SearchWord date
2	(فینیسیم یا فینیزیم)
3	(فینیسیم یا فینیزیم)
4	زن یا زنان) و اسلام
5	زن یا زنان) و اسلام و اجتماعی
6	زن یا زنان) و اسلام و اجتماعی
7	زن یا زنان) و اسلام
8	مشارکت و (زن یا زنان)
9	مشارکت و (زن یا زنان)
10	مشارکت و (زن یا زنان)
11	مشارکت و (زن یا زنان)
12	مشارکت و (زن یا زنان)
13	(فینیسیم یا فینیزیم)
14	مشارکت و (زن یا زنان)
15	مشارکت و (زن یا زنان)
16	مشارکت و (زن یا زنان)
17	مشارکت و (زن یا زنان)
18	(رفاه یا افراط یا افراطی) و (قرآن یا اسلام یا اسلامی)
19	(رفاه یا افراط یا افراطی) و (قرآن یا اسلام یا اسلامی)
20	(طعام یا غذا) و قرآن
21	رفاه و (دین یا دینی یا مذهب یا مذهبی یا ادیان)
22	رفاه و (دین یا دینی یا مذهب یا مذهبی یا ادیان)

## تصویر ۲. پرس‌وجوی حاوی عملگر OR «یا» (۱۰۳۱۷ مورد)

بنابراین، برای دریافت پرس‌وجوهای ترکیبی کاربران لازم بود که هر پرس‌وجوی پیچیده منطقی به پرس‌وجوهای ساده شکسته شود. انجام این مهم از طریق طراحی و پیاده‌سازی فاز اول نرم‌افزار فراخوان خودکار کلیدواژه محقق گردید. در ادامه، مختصات فاز اول تشریح می‌شود.

برای شکستن پرس‌وجوهای پیچیده منطقی به پرس‌وجوهای ساده، نرم‌افزاری مبتنی بر ویندوز طراحی و پیاده‌سازی شد. از آنجائی که موتور جستجوی رایسست، یک موتور جستجوی بولی است، پرس‌وجوها از یک ترکیب منطقی برخوردار می‌باشند و به همین دلیل نیاز بود که این ترکیبات بر اساس نوع عملگرها به ترکیبات ساده‌تر شکسته شوند. بدین صورت که بر اساس ترکیبات AND و OR منطقی، نوع درخواست کاربر کاملاً متفاوت خواهد بود. این نرم‌افزار، پرس‌وجوها را بر اساس این ترکیبات تفکیک کرده و پرس‌وجوهای جدید را تولید می‌کند و در پایگاه داده مخصوص به خود که برای آن طراحی شده است، به همراه میزان فراوانی پرس‌وجو و آخرین تاریخ درخواست، به روزرسانی می‌کند. همچنین این نرم‌افزار اطلاعات آخرین پرس‌وجوی پردازش شده را برای پردازش‌های آتی ذخیره می‌نماید. این امر سبب می‌گردد تا در اجراهای بعدی، تنها پرس‌وجوهای جدید وارد پایگاه داده شود. تصویر ۳ نمائی از نرم‌افزار فوق را نشان می‌دهد:

date	SearchWord	Row1
1:43 3/16/2015	آموزش و سواد و ...	1
1:43 3/16/2015	حقوق و بشر و و ...	2
1:43 3/16/2015	"روش پرانگیزی ...	3
1:43 3/16/2015	"دیدگاه مشترک م...	4
1:43 3/16/2015	"...Product bundlin"	5
1:43 3/16/2015	آماده‌شدن و برای...	6
1:44 3/16/2015	هوش و محیطی	7
1:44 3/16/2015	هوش و محیطی	8
1:44 3/16/2015	"اکرومگانی"	9
1:45 3/16/2015	"جمای موثر"	10
1:45 3/16/2015	مدیریت و ناهرنی‌ها	11
1:45 3/16/2015	جابکی و عاطفی	12
1:45 3/16/2015	"حوادث قلبی صر...	13
1:45 3/16/2015	جابکی و عاطفی	14
1:45 3/16/2015	"مورترین بلات"	15
1:46 3/16/2015	"معات"	16

تصویر ۳. نمایش پرس‌وجوهای پردازش شده در فاز اول نرم‌افزار فراخوان خودکار کلیدواژه (پردازش پرس‌وجوها)

برای روشن‌تر شدن موضوع و بررسی نحوه عملکرد پردازشگر پرس‌وجوها، به سه مثال زیر توجه کنید:  
 مثال ۱: پرس‌وجوی (زن یا زنان یا بانوان) و (شاغل یا کارمند) ایرانی که ترکیبی از کلمات، کمانک و عملگرهای AND و OR می‌باشد، به شرح زیر به شش پرس‌وجوی مجزا و ساده که هر کدام فراوانی یکسانی با پرس‌وجوی اصلی خواهند داشت، تفکیک می‌گردد:

• زن شاغل ایرانی، زن کارمند ایرانی، زنان شاغل ایرانی، زنان کارمند ایرانی، بانوان شاغل ایرانی، بانوان کارمند ایرانی.

مثال ۲: پرس‌وجوی مشارکت و (زن یا زنان) به دو پرس‌وجوی ساده زیر تفکیک می‌شود:

• مشارکت زن، مشارکت زنان

مثال ۳: در این مثال که حالت AND ساده را نشان می‌دهد، پرس‌وجوی کتابخانه و دیجیتالی و توسعه و وچالش و ها به

ترکیب زیر تغییر می‌یابد:

• کتابخانه دیجیتالی توسعه و چالش‌ها

مرحله ۲: طراحی و پیاده‌سازی فاز دوم نرم‌افزار فراخوان خودکار کلیدواژه (پردازش کلیدواژگان)

در این مرحله، کلیدواژگان مقالات تمام‌متن فارسی از پایگاه داده رایسست دریافت می‌شود و پس از انجام بررسی‌های لازم

به همراه اطلاعات تکمیلی در پایگاه داده نرم‌افزار درج می‌گردد (تصویر ۴).

year	mainsubject_id	keytext	no1
1386	4	تحلیل گفتمان انتقا...	2
1386	4	تغییر گفتمان	2
1386	4	جورج لوکاج	2
1386	4	روایت	2
1386	4	نقد ادبی	2
1387	4	نظامی	2
1391	4	قصه‌های عامیانه ا	2
1392	4	شاه	2
1386	4	هنر	2
1388	4	نقش گرامی هلیدی	2
1390	4	داستان معاصر فار	2
1390	4	گفتمان کاوی	2
1391	4	انسان	2
1391	4	خداینامه	2
1391	4	زبان های ایرانی	2
1391	4	ساختار	2
1391	4	سروده های کردی	2
1391	4	گردل	2

تصویر ۴. نمایش نمونه‌ای از محتوای فاز دوم نرم‌افزار فراخوان خودکار کلیدواژه (پردازش کلیدواژگان)

این اطلاعات موارد زیر را شامل می‌شود:

- فراوانی (تعداد تکرار) کلیدواژه‌ها در کلیه مقالات تمام‌متن فارسی،
  - تاریخ جدیدترین مقاله‌ای که از کلیدواژه در آن استفاده شده است (تاریخ و زمان در سیستم‌های مشابه، مقوله‌ای تاثیرگذار محسوب می‌گردد. در پایگاه اطلاعاتی مقالات تمام‌متن فارسی مرکز منطقه‌ای نیز این متغیر، متغیری مهم به شمار می‌رود.
  - شناسه موضوعی مقاله حاوی کلیدواژه (شناسه موضوعی برای مجلات حاوی مقالات در رایسست تعریف شده است که حوزه‌هایی چون علوم انسانی، فنی مهندسی و ... را شامل می‌گردد.
- این اطلاعات با گذشت زمان و درج در پایگاه داده رایسست، لازم است در پایگاه داده نرم‌افزار نیز به روز رسانی شود. نحوه اجرای این عملیات و جزئیات آن در مرحله ۳ از بخش ۴-۲ آورده شده است.

**مرحله ۳: طراحی و پیاده‌سازی فاز سوم نرم‌افزار فراخوان خودکار کلیدواژه (نگاشت پرس‌وجو به کلیدواژه‌ها)**

- بخش سوم نرم‌افزار پردازش خودکار کلیدواژه، نگاشت پرس‌وجو و کلیدواژه نام دارد که خود حاوی دو قسمت است: «پرس‌وجوهای جدید» و «پرس‌وجوهای پردازش شده» (تصویر ۵) که در ادامه هر یک به اختصار توضیح داده می‌شود:



تصویر ۵. نمایش قسمت نگاشت پرس و جو و کلیدواژه در نرم افزار

**پرس وجوهای جدید:** منظور از پرس وجوهای جدید پرس وجوهائی است که تاکنون توسط نرم افزار پردازش نشده اند (با توجه به پیشرفت علوم و نیازهای خاص کاربران همواره با پرس وجوهای جدید مواجه هستیم ضمن آنکه هر پرس وجویی که نتواند با کلیدواژه های موجود در پایگاه داده رایشست نگاشت شود جدید محسوب می شود و در لیست انتظار برای نگاشت بعدی قرار می گیرد). در این مرحله، بین پرس وجوها و کلیدواژه های که از پایگاه داده رایشست دریافت کرده ایم (کلیدواژگان پردازش شده)، نگاشت انجام می شود. نگاشت مشخص می کند که آیا یک کلیدواژه با پرس وجویی در پایگاه داده مرتبط است یا خیر؟ این ارتباط در نهایت در فرایند پیشنهاد کلیدواژه مورد استفاده قرار می گیرد. بدیهی است کلیدواژه های که توسط کاربران بیشتری پرس وجو شده اند، محبوب ترند.

**پرس وجوهای پردازش شده:** منظور از پرس وجوهای پردازش شده، پرس وجوهائی است که قبلاً حداقل یک بار ارتباط آنها با کلیدواژگان موجود، توسط نرم افزار بررسی شده باشد. در اینجا نیز ارتباط بین کلیدواژه ها و این پرس وجوها بررسی شده و در صورت وجود ارتباط، همانند مرحله قبل، در پایگاه داده نرم افزار جهت استفاده در الگوریتم پیشنهاد کلیدواژه، درج می گردد. در قسمت قبل یعنی «پرس وجوهای جدید» اطلاعات جدید نگاشت می شد. حال فرض کنید که یک سری اطلاعات در مرحله قبل پردازش شده اما نگاشت نشده باشند. به خاطر وجود این عامل، پس از مدتی در پایگاه داده یک سری کلیدواژه خواهیم داشت. این امکان نیز وجود دارد که این پرس وجوهای نگاشت نشده و باقی مانده در پایگاه داده به این کلیدواژه ها قابل نگاشت باشند. علت وجودی بخش «پرس وجوهای پردازش شده» دقیقاً انجام این مهم است. این بخش، بخشی بسیار گسترده است و همه پرس وجوهائی را که تاکنون پردازش شده ولی با کلیدواژه ها نگاشت نشده اند در خود جای می دهد. همین حجم بالای داده ها سبب می شود تا برای پردازش اطلاعات یک بازه زمانی محدود تعیین کنیم. بازه زمانی مربوطه به روز، ماه و سال میلادی ثبت می گردد (در این پروژه از بازه زمانی یک ماهه برای تحلیل داده ها استفاده به عمل آمد تا نرم افزار در فواصل زمانی کوتاه تری به روز رسانی شود).

برای مثال، بازه زمانی ۲۰۱۳-۰۱-۰۱ تا تاریخ ۲۰۱۴-۰۱-۰۱ را وارد می کنیم و بعد بر روی دکمه «دریافت اطلاعات» کلیک می کنیم. بدین ترتیب، اطلاعاتی که پیشتر تحلیل شده ولی به کلیدواژه ای نگاشت نشده، فهرست می شود. مثلاً فرض کنید کاربری پرس وجویی با محتوای «آمار کتابداری» را ۲ بار در تاریخ ۲۰۱۲/۰۸/۰۹ درخواست کرده باشد و در آن زمان هیچ کلیدواژه ای متناسب با آن پرس وجو در پایگاه داده رایشست (مربوط به مقالات تمام متن فارسی) وجود نداشته باشد. باز فرض کنید پس از مدتی سندی با این کلیدواژه در پایگاه داده رایشست و بعد از انجام مراحل پردازش توسط نرم افزار، درج شده باشد، بنابراین، نیاز است مراحل نگاشت دوباره تکرار شود. بدین صورت که با کلیک زدن بر روی دکمه «دریافت اطلاعات»، دکمه «درج در پایگاه» فعال می شود. پس از کلیک زدن بر روی این دکمه، پرس وجوها در پایگاه کلیدواژه ها جستجو می شوند. اگر این پرس وجوها در این پایگاه وجود داشتند نگاشت انجام می شود در غیر این صورت همچنان به عنوان پرس وجوی نگاشت نشده برای نگاشت های احتمالی آتی باقی می ماند. توضیحات تکمیلی در این خصوص در مرحله ۴ از بخش ۴ آورده شده است.

تمامی توضیحاتی که تا این لحظه در این فصل آورده شده است را می‌توان به نوعی در مقوله کلان پیش پردازش جای داد. در ادامه، طراحی الگوریتم پیشنهاد کلیدواژه آورده می‌شود.

#### مرحله ۴: طراحی الگوریتم پیشنهاد کلیدواژه

تمامی مراحل کار با نرم‌افزار توضیح داده شد. در ادامه، نحوه استفاده از این اطلاعات و الگوریتم پیشنهاد کلیدواژه به کاربر، توضیح داده می‌شود. این الگوریتم از ۴ مختصه استفاده می‌کند که عبارتند از:

۱. فراوانی پرس‌وجو: این مختصه بر اساس تعداد دفعاتی که کاربران پرس‌وجویی خاص را درون‌داد کرده‌اند، عمل می‌کند بدین صورت که هر چه فراوانی درون‌داد یک پرس‌وجو بیشتر باشد، پس طبیعتاً از اهمیت بیشتری نیز برخوردار بوده است.
۲. جدیدترین تاریخی که پرس‌وجو در آن تاریخ درون‌داد شده است.
۳. تاریخ خود کلیدواژه: همیشه تاریخ جدیدترین مقاله‌ای که کلیدواژه در آن ظاهر شده است توسط نرم‌افزار به عنوان تاریخ خود کلیدواژه انتخاب می‌شود.

#### ۴. فراوانی کلیدواژه در پایگاه مقالات تمام‌متن فارسی

نهایتاً بر اساس این چهار مختصه، به هر کلیدواژه وزنی داده می‌شود که بر اساس آن کلیدواژه‌های دارای بیشترین وزن، به کاربر پیشنهاد داده می‌شود. در این راستا، لازم است ابتدا مقدار هر مختصه نرمال سازی شود. برای انجام این مهم از فرمول زیر استفاده به عمل آمد (لازم است فرمول زیر از راست به چپ خوانده شود):

#### داده منهای ابتدای بازه) تقسیم بر (انتهای بازه منهای ابتدای بازه)

از طرفی تصمیم بر آن شد تا برای هر یک از این چهار مختصه ارزش یکسانی در نظر گرفته شود و بنابراین وزن‌دهی نهایی برای هر کلیدواژه، بر اساس برآیند وزن چهار مختصه انجام شد (علت تخصیص ارزش یکسان به چهار مختصه، عملکرد مثبت نرم‌افزار با این روال وزن‌دهی بود).

#### مرحله ۵: طراحی و پیاده‌سازی موتور جستجو بر اساس امکان جدید (جستجوی موضوعی با پیشنهاد

#### کلیدواژه به کاربر)

مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری دارای موتور جستجویی با عنوان «پایگاه مقالات تمام‌متن فارسی» می‌باشد که جستجوی مقالات فارسی را بر اساس کلیدواژه، نام نویسندگان، عنوان مقاله، شماره مدرک و چکیده به انجام می‌رساند. در پژوهش حاضر، این موتور جستجو با افزودن دو امکان مهم ارتقاء یافت که عبارت بودند از: (۱) جستجوی موضوعی و (۲) پیشنهاد کلیدواژه.

در جستجوی موضوعی، کاربران امکان می‌یابند در صورت تمایل در یکی از حوزه‌های موضوعی نشریات مرکز منطقه‌ای، جستجوی خود را به انجام برسانند. برای مثال، کاربری که در حوزه موضوعی فنی مهندسی کلیدواژه «جوش» را وارد کرده باشد، با پیشنهاداتی مانند «جوش ترمیت»، «جوش نفوذ کامل»، «جوش‌پذیری»، «جوش چند پاسه محیطی» و «جوش‌های غیرمشابه» مواجه می‌شود (تصویر ۶). حال، چنانچه همین کلیدواژه در حوزه موضوعی پزشکی جستجو شود، کلیدواژه‌های پیشنهادی به شرح زیر خواهد بود: «جوش شیرین»، «جوش نخوردن»، «جوش خوردن»، «جوش خوردگی» و «جوش نخوردگی استخوان اسکافوئید» (تصویر ۷). با انتخاب یک حوزه موضوعی و درون‌داد یک کلیدواژه، تمامی اسنادی که کلیدواژه مربوطه در آن‌ها وجود داشته باشد (صرفاً از میان مجلات مربوط به حوزه موضوعی انتخاب شده) بازیابی می‌شوند. بدیهی است چنانچه کاربر مایل به انتخاب یک حوزه موضوعی خاص نباشد می‌تواند از حالت «پیش فرض» سیستم استفاده کند. در حقیقت، برای استفاده از جستجوی موضوعی لازم است کاربر، نرم‌افزار را از حالت پیش فرض خارج نموده و با انتخاب لفظ «کلیدواژه» از منوی



سمت راست، یک حوزه موضوعی را برگزینند. اگرچه این نوع تفکیک موضوعی در گوگل ساجست انجام نمی‌شود، پیشنهادات نامرتبب بیشتری هم تولید می‌کند. با تفکیک موضوعی در پژوهش حاضر سعی شد، حجم پیشنهادهای نامرتببب به حداقل برسد. در مورد پیشنهاد کلیدواژه نیز پیرو درونداد حداقل دو کاراکتر توسط کاربر، نرم‌افزار به صورت خودکار فعال شده و ۵ کلیدواژه را به کاربر پیشنهاد می‌دهد. شایان ذکر است ورود دو کاراکتر برای فعال‌سازی نرم‌افزار صرفاً نقطه شروع است و با ادامه ورود کاراکترهای بعدی فهرست پیشنهادهای تغییر می‌یابد که کاربر بسته به نیاز خود می‌تواند یکی را برگزیند.

تصویر ۶. نمایش کلیدواژه‌های پیشنهادی برای «جوش» در حوزه موضوعی فنی مهندسی توسط نرم‌افزار

تصویر ۷. نمایش کلیدواژه‌های پیشنهادی برای «جوش» در حوزه موضوعی پزشکی توسط نرم‌افزار

پس از معرفی مختصات نرم‌افزار رایست ساجست در ادامه به معرفی چگونگی استفاده از این نرم‌افزار توسط کاربر سیستم پرداخته می‌شود. منظور از کاربر سیستم فردی است که مسئولیت روزآمد نگه داشتن اطلاعات و داده‌های نرم‌افزار بر

عهده‌ی او گذاشته می‌شود و با کاربر نهائی فرق دارد. به عبارت دیگر، کاربر سیستمی نرم‌افزار را در فواصل زمانی منظم (در این پژوهش بازه‌ی زمانی یک ماهه) روز آمد سازی می‌کند و کاربر نهائی فردی است که برای تأمین نیاز اطلاعاتی خود به وبگاه مرکز مراجعه نموده و پرس‌وجویی را در پایگاه مقالات تمام‌متن فارسی درون‌داد نموده و اطلاعات خود را در قالب مقاله دریافت می‌نماید.

### راهنمای استفاده از نرم‌افزار توسط کاربر سیستم

در این قسمت چگونگی استفاده از نرم‌افزار توسط کاربر سیستم به تصویر کشیده می‌شود.

#### مرحله ۱:

در گام نخست اجرای نرم‌افزار، کاربر سیستم نام کاربری و کلمه رمز عبور خود را که توسط مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری و برای روزآمد نگه داشتن و نیز نگهداری نرم‌افزار در اختیارش گذاشته شده، وارد می‌نماید (تصویر ۸).



تصویر ۸: نمایش صفحه ورود به نرم‌افزار فراخوان خودکار کلیدواژه

درخواست نام کاربری و کلمه رمز عبور به خاطر حفظ امنیت سیستم و کنترل عملیاتی نرم‌افزار صورت می‌گیرد. معمولاً فردی که کاربر نرم‌افزار محسوب می‌گردد، اجازه دسترسی به نام کاربری و رمز عبور و استفاده از این نرم‌افزار را خواهد داشت. پس از ورود نام کاربری و رمز عبور، صفحه اصلی نرم‌افزار باز می‌شود (تصویر ۹).



تصویر ۹. نمایش صفحه اصلی نرم‌افزار فراخوان خودکار کلیدواژه و گزینه پردازش پرس‌وجوها

## مرحله ۲:

در این صفحه، بخش‌های مختلفی (سه بخش) دیده می‌شود. بخش اول (نخستین گزینه از سمت راست منو)، **پردازش پرس‌وجوها** نام دارد. پس از کلیک زدن بر روی گزینه **پردازش پرس‌وجوها**، تصویر ۱۰ نشان داده می‌شود.

date	SearchWord	Row1
1:43 3/16/2015	آموزش و سواد و ا...	1
1:43 3/16/2015	حقوق و بشر و و...	2
1:43 3/16/2015	"روش برکنگنی ...	3
1:43 3/16/2015	"دیدگاه مشترک و...	4
1:43 3/16/2015	"Product bundlin"	5
1:43 3/16/2015	آماده‌شدن و برای...	6
1:44 3/16/2015	هوش و محیطی	7
1:44 3/16/2015	هوش و محیطی	8
1:44 3/16/2015	"اکرومگانی"	9
1:45 3/16/2015	"نمای موثر"	10
1:45 3/16/2015	مدیریت و ناهرنی‌ها	11
1:45 3/16/2015	جابکی و عاطفی	12
1:45 3/16/2015	"خواند قلبی عری...	13
1:45 3/16/2015	جابکی و عاطفی	14
1:45 3/16/2015	"نورترن بلات"	15
1:46 3/16/2015	"معاد"	16

## تصویر ۱۰: نمایش بخش پردازش پرس‌وجوها در نرم‌افزار فراخوان خودکار کلیدواژه

در تصویر ۱۰ در قسمت بالای صفحه، تاریخ آخرین رکورد پردازش شده به تقویم میلادی نشان داده می‌شود. در قسمت پایین صفحه هم، سه دکمه وجود دارد که به ترتیب عبارتند از: «دریافت اطلاعات»، «پردازش اطلاعات» و «درج در پایگاه داده». در ابتدا، صرفاً گزینه «دریافت اطلاعات» فعال است و دو دکمه دیگر غیرفعال‌اند. در حقیقت، با ادامه کار و تکمیل مرحله ۱، دکمه ۲ «پردازش اطلاعات» و با تکمیل کار در مرحله ۲، دکمه ۳ «درج در پایگاه داده» نیز فعال خواهد شد.

ابتدا بر روی دکمه «دریافت اطلاعات» کلیک زده می‌شود. با این کلیک زدن، تمامی پرس‌وجوهائی که یا از موتور جستجوی جامع و یا از پایگاه مقالات تمام‌متن فارسی مرکز منطقه‌ای گرفته شده‌اند، از زمان پردازش قبلی تا تاریخ جاری (پردازش فعلی) دریافت می‌شود. در تصویر ۱۰، ماحصل کار در بخش دریافت اطلاعات، به صورت پر شدن پنجره سمت راست نشان داده شده است. در سمت راست این پنجره، ردیف وجود دارد که از ۱ شروع شده و تعداد نهائی آن متناسب با تعداد پرس‌وجوها متغیر است. در ستون وسط، گزینه SearchWord وجود دارد که همان پرس‌وجوی کاربران است و ستون آخر نیز ستون Date است که در حقیقت تاریخ پرس‌وجو است. با کلیک زدن بر روی دکمه «پردازش اطلاعات»، عملیاتی به شرح زیر به انجام می‌رسد: عملیات ۱: نخست این پرس‌وجوها که می‌توانند ترکیبات منطقی ساده یا پیچیده باشند، به ترکیبات ساده (در صورتی که ترکیبی غیرساده باشند) تبدیل می‌شوند. برای مثال، ترکیب «ورزشکار و (مرد یا زن)» به دو ترکیب «ورزشکار مرد» و نیز «ورزشکار زن» تبدیل می‌شود.

عملیات ۲: سپس، عملیات فیلترینگ به انجام می‌رسد. علت اهمیت این مرحله آن است که همواره بعضی از پرس‌وجوها نامعتبرند. در بخش فیلترینگ کاری که انجام می‌شود این است که پرس‌وجوهائی که فقط نقطه، خط فاصله (Dash)، علامت سؤال و & را در ترکیب خود دارند فیلتر می‌شوند. ضمناً، اگر پرس‌وجو حاوی یکی از نشانه‌های زیر باشد نیز فیلتر خواهد شد:

- ترکیبی از حروف لاتین کوچک، بزرگ یا هر دو
- فضاهای سفید (White Spaces) (شامل Tab, Space, Newline, ...)

## • ارقام ۰ تا ۹

• علائم نشانه‌گذاری، علائم ریاضی (علائم ضرب، تقسیم، درصد، بعلاوه) و علامت \$.

بر این اساس، پرس‌وجوهای cat-pen، cat9 و cat حذف می‌شوند ولی پرس‌وجوی «رفت-آمد» حذف نمی‌شود چون تمامی حروف آن فارسی هستند. در حقیقت، علامت خط تیره/منها «-» هم فیلتر نشد چون در ترکیبات منطقی معنی‌دار است.



lastdate	no	Key	date	SearchWord	Row1
...12:27 7/21/2015	116	آموزش سواد اطلاع...	1:43 3/16/2015	آموزش و سواد و ا...	1
PM 7:58 6/2/2015	6	حقوق بشر و حقوق...	1:43 3/16/2015	حقوق و بشر و و...	2
...10:28 5/27/2015	3	روش برنامه‌نگاری و ا...	1:43 3/16/2015	"روش برنامه‌نگاری ...	3
... 6:07 7/16/2015	4	دیدگاه مشترک ما ...	1:43 3/16/2015	"دیدگاه مشترک م...	4
... 1:43 3/16/2015	1	آماده‌شدن برای مد...	1:43 3/16/2015	"...Product bundlin"	5
... 1:44 3/16/2015	2	هوش محیطی	1:43 3/16/2015	آماده‌شدن و برای...	6
... 1:08 6/12/2015	3	اکتروپیکالی	1:44 3/16/2015	هوش و محیطی	7
... 7:55 7/21/2015	214	دمای مؤثر	1:44 3/16/2015	هوش و محیطی	8
... 1:45 3/16/2015	1	مدیریت ناوبری‌ها	1:44 3/16/2015	"اکتروپیکالی"	9
... 1:45 3/16/2015	2	جابجی عاطفی	1:45 3/16/2015	"دمای مؤثر"	10
... 1:45 3/16/2015	1	حوادث قلبی عروقی	1:45 3/16/2015	مدیریت و ناوبری‌ها	11
... 9:14 7/16/2015	64	نورترن بلات	1:45 3/16/2015	جابجی و عاطفی	12
... 1:13 7/13/2015	74	مهارت	1:45 3/16/2015	"حوادث قلبی عر...	13
... 6:20 7/21/2015	202	استخراج ویژگی	1:45 3/16/2015	جابجی و عاطفی	14
... 3:32 7/21/2015	122	نقاط کوانتومی	1:45 3/16/2015	"نورترن بلات"	15
... 12:01 5/6/2015	3	نرم افزار «صرفه ...	1:46 3/16/2015	"مهارت"	16

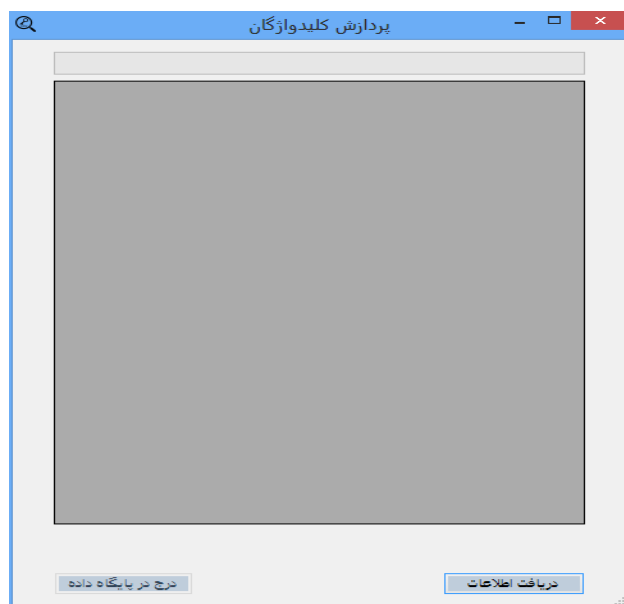
## تصویر ۱۱. نتیجه حاصل از پردازش اطلاعات (پرس‌وجوها)

در تصویر ۱۱ اطلاعات پردازش شده در پنجره سمت چپ نشان داده شده است ولی هنوز در پایگاه داده درج نشده است. بعد از انجام فیلترینگ، این اطلاعات باید در پایگاه داده نرم‌افزار درج شود. در این زمان دکمه «درج در پایگاه داده» فعال می‌شود. با کلید زدن بر روی این دکمه، اطلاعات در پایگاه داده مخصوص نرم‌افزار ذخیره می‌شود. نحوه ذخیره اطلاعات نیز بسیار حائز اهمیت است. برای روشن شدن نحوه ذخیره اطلاعات، به مثال فرضی زیر توجه کنید:

برای مثال، پرس‌وجوی «دمای مؤثر» در تاریخ ۲۰۱۵/۰۳/۱۶ به عنوان یک پرس‌وجو در سیستم درون‌داد شده است. حال کاری که «درج در پایگاه داده» انجام می‌دهد این است که اول پرس‌وجوی «دمای مؤثر» را در جدول پرس‌وجوها در پایگاه داده خودش جستجو می‌کند. چنانچه رکوردی متناسب با آن وجود داشت، فراوانی و تاریخ موجود را بررسی می‌کند. در واقع، فراوانی را اضافه کرده، تاریخ را هم به روز رسانی می‌کند. برای مثال، چنانچه پرس‌وجو آخرین بار در سال ۲۰۱۴ و با فراوانی ۵ زده شده بود، حالا در تاریخ ۲۰۱۵/۰۳/۱۶ فراوانی آن به ۶ افزایش خواهد یافت و اگر این رکورد پیشتر وجود نداشته باشد، به عنوان یک رکورد جدید با فراوانی ۱ و تاریخ درون‌داد شده، ثبت می‌شود.

## مرحله ۳:

بخش دوم منوی اصلی در تصویر ۹، «پردازش کلیدواژگان» است که دو دکمه دارد: دکمه «دریافت اطلاعات» و دکمه «درج در پایگاه داده». در ادامه، مختصات هر یک به اختصار توضیح داده می‌شود.



تصویر ۱۲. پردازش کلیدواژگان در نرم‌افزار

**دریافت اطلاعات:** کاری که در این قسمت (تصویر ۱۲) انجام می‌شود این است که آخرین سندی که کلیدواژه‌اش پردازش شده را پیدا می‌کند و سندهای بعد از آن را از پایگاه داده رابست دریافت می‌نماید که قاعدتاً این سندها، مقالات تمام‌متن فارسی هستند. اطلاعات آن‌ها در قالب Text و متن مقاله‌ها در قالب فایل PDF است. این بخش اطلاعات سندها (شامل فراوانی، کلیدواژه، شناسه موضوع اصلی (ID) و سال مربوط به سند) را به ما نشان می‌دهد (تصویر ۱۳).

year	mainsubject_id	keytext	no 1
1386	4	تحلیل گفتمان انتقا...	2
1386	4	تغییر گفتمانی	2
1386	4	جورج لوکاج	2
1386	4	روایت	2
1386	4	نقد ادبی	2
1387	4	نظامی	2
1391	4	قصه‌های عامیانه ا...	2
1392	4	شاه	2
1386	4	هنر	2
1388	4	نقش گرایی هنیدی	2
1390	4	داستان معاصر فار...	2
1390	4	گفتمان کاوی	2
1391	4	انسان	2
1391	4	خداینامه	2
1391	4	زبان های ایرانی	2
1391	4	ساختار	2
1391	4	سروده های کردی	2
1391	4	گردل	2

تصویر ۱۳. پردازش کلیدواژگان پس از دریافت اطلاعات

مثلاً همانگونه که در تصویر ۱۳ نشان داده شده است، کلیدواژه «روایت» در سندهای اخیر ۴ بار ظاهر شده و سال ۱۳۸۶ آخرین سالی بوده که از این کلیدواژه در آن استفاده شده است. پس از دریافت اطلاعات، دکمه «درج در پایگاه داده» فعال می‌شود. در این مرحله، نرم‌افزار اطلاعات را در جدول مخصوص پایگاه داده خود ذخیره می‌کند که این اطلاعات، فراوانی،

کلیدواژه، شناسه‌ی موضوع اصلی (ID) و سال مربوط به سند را شامل می‌شود. نحوه‌ی تحلیل نیز مثل حالت قبل است. یا رکوردی از قبل در جدول وجود داشته که در این صورت اطلاعات آن به روزرسانی می‌شود و فراوانی آن افزایش می‌یابد، یا نبوده که در این صورت به عنوان مدخل جدید و با فراوانی روئیت شده و سال جدید ثبت می‌شود. نکته‌ی مهم آنکه ممکن است کلیدواژه «روایت» در پایگاه داده‌ی نرم‌افزار موجود باشد ولی با موضوع اصلی دیگری ثبت شده باشد که در این صورت یک مدخل مجزا در نظر گرفته می‌شود.

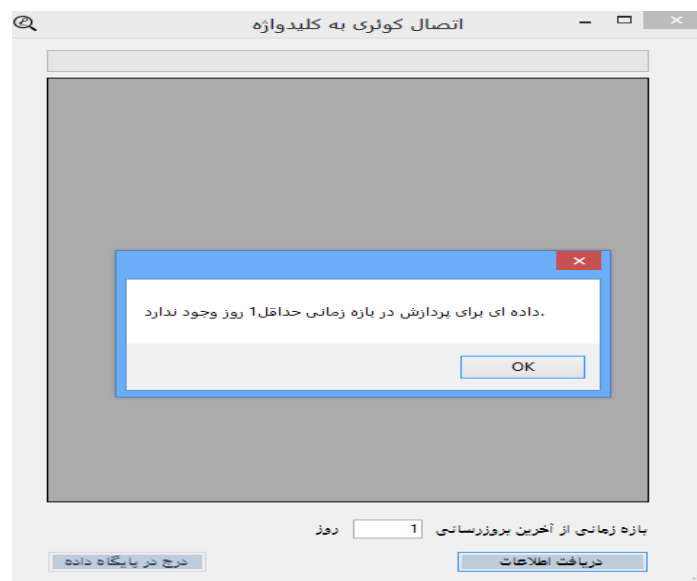
#### مرحله ۴:

بخش سوم نرم‌افزار پردازش خودکار کلیدواژه، «نگاشت پرس‌وجو و کلیدواژه» نام دارد که خود حاوی دو قسمت است «پرس‌وجوهای جدید» و «پرس‌وجوهای پردازش شده» (تصویر ۱۴) که در ادامه هر یک به اختصار توضیح داده می‌شود:



تصویر ۱۴. نمایش بخش نگاشت پرس‌وجو و کلیدواژه در نرم‌افزار

**پرس‌وجوهای جدید:** با کلید زدن بر روی دکمه «پرس‌وجوهای جدید» (تصویر ۱۴)، صفحه‌ی اتصال پرس‌وجو به کلیدواژه ظاهر می‌شود (تصویر ۱۵).



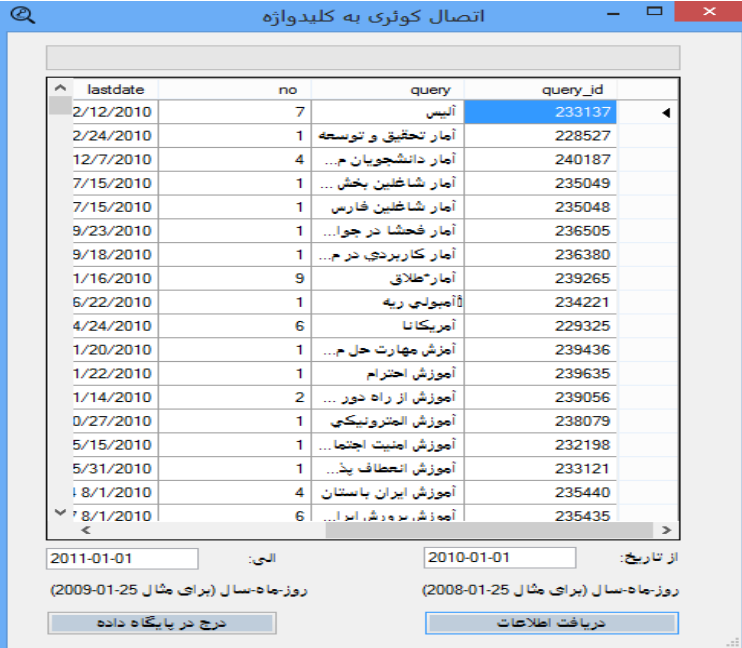
تصویر ۱۵. نمایش صفحه‌ی اتصال پرس‌وجو به کلیدواژه در نرم‌افزار (بخش اول)



هدف این مرحله آن است که بین پرس‌وجوها و کلیدواژه‌هایی که از پایگاه داده ریسست دریافت کرده‌ایم، نگاهی انجام شود. کاری که نگاهی انجام می‌دهد آن است که مشخص می‌کند آیا یک کلیدواژه با پرس‌وجویی در پایگاه داده مرتبط است یا خیر؟ برای انجام این کار لازم است بازه زمانی از تاریخ آخرین به روز رسانی تعیین شود. علت این امر آن است که در صورت عدم تعیین بازه زمانی، فرایند پردازش بسیار طولانی خواهد بود. از این رو، این بازه زمانی را در قالب تعداد روز از آخرین به روز رسانی تعیین می‌کنیم. مثلاً می‌توانیم این بروز رسانی را از آخرین تاریخ به روز رسانی تا یک روز بعد از آن انجام دهیم. سپس، بر روی دکمه «دریافت اطلاعات» کلیک زده می‌شود. آنچه ظاهر می‌شود پرس‌وجوهایی هستند که می‌خواهیم آنها را به کلیدواژه‌ها نسبت دهیم. در اینجا، پرس‌وجوها به همراه فراوانی و جدیدترین تاریخ‌شان نشان داده می‌شوند. در مرحله بعد، دکمه «درج در پایگاه داده» فعال می‌شود. پس از کلیک زدن بر روی این دکمه، کار پردازش اطلاعات آغاز می‌گردد و در پایگاه داده ثبت می‌شود که آیا این پرس‌وجو به کلیدواژه ای نگاهی شده است یا خیر. انجام این کار مزایای بسیاری دارد و به طور خلاصه اطلاعات زیر را در اختیار ما قرار می‌دهد:

- چه محتوایی در مرکز وجود ندارد که مورد مطالبه کاربران بوده است
- چه اطلاعاتی داریم که کاربران برای استفاده از آنها درخواستی نداشته‌اند یا از وجود آنها مطلع نبوده‌اند.

**پرس‌وجوهای پردازش شده:** در این قسمت نیز کاری مشابه آنچه در بخش «پرس‌وجوهای جدید» گفته شد، انجام می‌شود، تنها تفاوت با آن بخش در این است که در این مرحله با پرس‌وجوهایی سر و کار داریم که توسط این نرم‌افزار یک بار پردازش شده‌اند. تفاوت دوم در این است که اینجا بازه زمانی فرق می‌کند. در مرحله قبل بازه زمانی بر اساس تعداد روز بود اما در این مرحله چون قبلاً پردازش شده، بر اساس تاریخ شروع و پایان است (تصویر ۱۶).



lastdate	no	query	query_id
2/12/2010	7	آبیس	233137
2/24/2010	1	آمار تحقیق و توسعه	228527
12/7/2010	4	آمار دانشجویان م...	240187
7/15/2010	1	آمار شاخصین بخش ...	235049
7/15/2010	1	آمار شاخصین فارس	235048
9/23/2010	1	آمار فحشا در جوا...	236505
9/18/2010	1	آمار کاربردی در م...	236380
1/16/2010	9	آمار "طلاق"	239265
6/22/2010	1	آمبولی ریه	234221
4/24/2010	6	آمریکانا	229325
1/20/2010	1	آموزش مهارت حل م...	239436
1/22/2010	1	آموزش احترام	239635
1/14/2010	2	آموزش از راه دور ...	239056
0/27/2010	1	آموزش المکترونیکی	238079
5/15/2010	1	آموزش امنیت اجتماع...	232198
5/31/2010	1	آموزش انعطاف پذ...	233121
8/1/2010	4	آموزش ایران باستان	235440
8/1/2010	6	آموزش روش سراسر...	235435

از تاریخ: 2010-01-01  
الی: 2011-01-01  
روز-ماه-سال (برای مثال 2008-01-25)  
روز-ماه-سال (برای مثال 2009-01-25)

درج در پایگاه داده      دریافت اطلاعات

تصویر ۱۶. نمایش اتصال پرس‌وجو به کلیدواژه (بخش دوم)

### راهنمای استفاده گام به گام از نرم‌افزار توسط کاربر نهایی

در این قسمت چگونگی کار با نرم‌افزار توسط کاربر نهایی نشان داده می‌شود. در گام نخست لازم است کاربر نهایی آدرس [www.ricest.ac.ir](http://www.ricest.ac.ir) را در مرورگر سیستم خود وارد نماید (تصویر ۱۷).



تصویر ۱۷. آدرس وبگاه مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

در مرحله بعد لازم است بر روی گزینه «مقاله‌های الکترونیکی» کلیک زده شود (تصویر ۱۷). گزینه مقاله‌های الکترونیکی، خود دارای سه زیر بخش است که عبارتند از: «مقاله‌های الکترونیکی فارسی»، «مقاله‌های الکترونیکی عربی» و «مقاله‌های الکترونیکی انگلیسی».

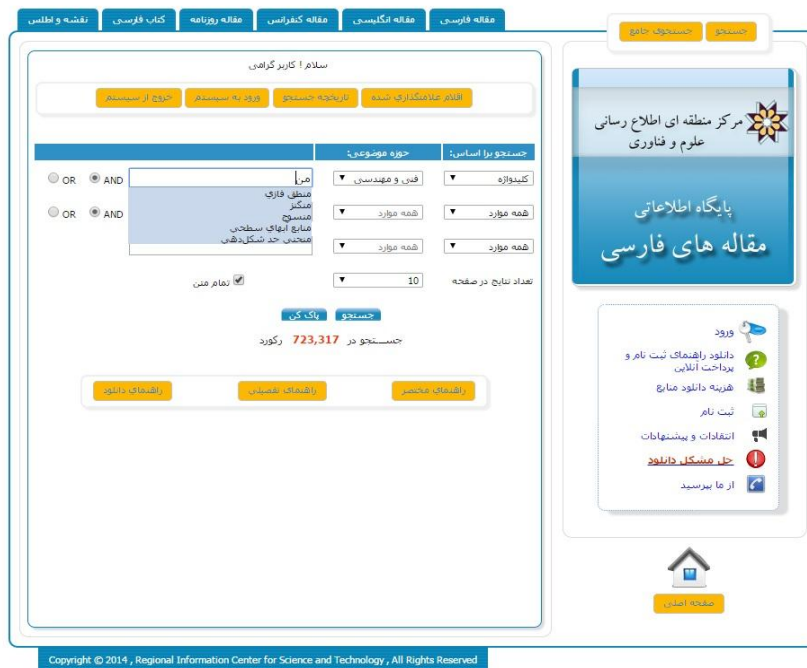


تصویر ۱۸. گزینه مقاله‌های الکترونیکی و اجزاء آن در وبگاه مرکز منطقه‌ای

در گام بعدی، بر روی گزینه «مقاله‌های الکترونیکی فارسی» کلیک زده می‌شود (تصویر ۱۸). با کلیک زدن بر روی این گزینه، صفحه درون‌داد پرس‌وجو در پایگاه اطلاعاتی مقاله‌های الکترونیکی باز خواهد شد. تصویر ۱۹، صفحه ورود پرس‌وجو را در این پایگاه پیش از بارگذاری نرم‌افزار نشان می‌دهد.



تصویر ۱۹. صفحه‌ی درون‌داد پرس‌وجو در پایگاه اطلاعاتی مقاله‌های الکترونیکی مرکز منطقه‌ای پیش از بارگذاری نرم‌افزار به همین ترتیب، تصویر ۲۰ صفحه‌ی جدید این پایگاه را نشان می‌دهد.



تصویر ۲۰. نمایش صفحه‌ی جدید طراحی شده برای پایگاه اطلاعاتی مقاله‌های فارسی

در این صفحه، چنانچه کاربر از قسمت «جستجو بر اساس» گزینه «کلیدواژه» را انتخاب نکند، «حوزه موضوعی» غیرفعال خواهد بود و سیستم مانند زمان قبل از اجرای پروژه حاضر، فرایند جستجو را انجام خواهد داد. حال چنانچه کاربر از گزینه «جستجو بر اساس» گزینه «کلیدواژه» را برگزیند، «حوزه موضوعی» فعال خواهد شد. در این حالت، کاربر می‌تواند بر اساس علائق تحقیقاتی خود یکی از حوزه‌های موضوعی (فنی مهندسی، علوم انسانی، پزشکی، علوم کشاورزی، ...) یا همه موارد را انتخاب نماید. با وارد کردن حداقل دو حرف و در صوتی که کاربر یکی از حوزه‌های موضوعی را انتخاب کرده باشد، نرم‌افزار پیشنهاد دهنده کلیدواژه فعال می‌شود و برای حوزه موضوعی انتخاب شده ۵ کلیدواژه را پیشنهاد می‌دهد. این ۵ کلیدواژه هم بر

اساس بیشترین وزن کسب شده برای متغیرها و به صورت نزولی مرتب می‌شوند. برای مثال همانگونه که در تصویر ۲۰ نشان داده شده است کاربر نهائی پس از انتخاب حوزه موضوعی «فنی مهندسی» دو حرف «من» را در جعبه پرس‌وجو وارد کرده و نرم‌افزار کلیدواژه‌های «منطق فازی»، «مگننز»، «منسوج»، «منابع آبهای سطحی» و «منحنی حد شکل‌دهی» را پیشنهاد داده است.

البته، کاربر چنانچه موارد پیشنهادی را مناسب نداند می‌تواند به تایپ پرس‌وجو ادامه دهد و کلیدواژه مورد نظر را از میان گزینه‌های جدید پیشنهاد شده انتخاب نماید. بدیهی است با تایپ هر کاراکتر یا حرف جدید، فهرست پیشنهادها نیز متناسب با آن تغییر می‌یابد. البته، این امکان هم وجود دارد که کاربر هیچ یک از موارد پیشنهاد شده را مناسب نداند و خود رأساً نسبت به درونداد کلیدواژه کامل اقدام کند.

The screenshot shows the RICEST website interface. At the top, there is a search bar and navigation buttons: 'جستجوی جدید', 'فهرست منتخب', 'صفحه خانگی', 'حذف انتخاب‌ها', and 'افزودن به اتمام علامتگذاری شده'. Below the search bar, the results for 'منطق فازی' are displayed. The first result is 'بهبود بنده خطر زمین لغزش‌های ناشی از زلزله 1375 سرعین، ایران با استفاده از روش‌های کیفی و کمی' by سیدمحمد فاطمی عقیبا, published in 'مجله مهندسی' 21728901. The second result is 'سیستم خبره فازی برای تحلیل نسبت‌های مالی (مطالعه موردی: صنعت سیمان)' by سیداحمد نبی جفالی, published in 'مجله مهندسی' 2160734. The third result is 'تحلیل عدم قطعیت در شبکه‌های جمع‌آوری فاضلاب با منطق فازی' by علی قاسمی, published in 'مجله مهندسی' 2057844.

## تصویر ۲۱. نمایشی از رکوردهای بازیابی شده از پایگاه، پیرو کلید زدن بر روی یکی از پیشنهادها

پس از انتخاب یکی از ۵ کلیدواژه پیشنهاد شده و کلید زدن بر روی گزینه «جستجو»، مقالات مرتبط و موجود در پایگاه اطلاعاتی مقالات فارسی بازیابی می‌شود (تصویر ۲۱). در این مرحله کاربر با کلید زدن بر روی گزینه «اطلاعات بیشتر» می‌تواند به اطلاعات کتابشناختی مقاله مورد نظر و چکیده (در صورت موجود بودن) دست یابد (تصویر ۲۲).



### پهنه‌بندی خطر زمین‌لغزش‌های ناشی از زلزله 1375 سرعین، ایران با استفاده از روش‌های کیفی و کمی

سیدمحمود فاطمی‌عقدا، محمدرضا مهدویفر، وحید باقری

خلاصه:

در این پژوهش یکی از روش‌های جدید پهنه‌بندی خطر زمین‌لغزش‌های ناشی از زلزله (مدل ناحیه‌ای جامع ریسک لغزش‌های ناشی از زلزله با مدل کتل [1]) برای پهنه‌بندی این نوع از زمین‌لغزش‌ها و ارزیابی آن با روش مهندسی بررسی شده است. بدین منظور محدوده تحت تأثیر زلزله سرعین (سال 1375) برای منطقه الگو انتخاب شده است. در این روش که محاسبه با واژه‌ها نامیده می‌شود، سیستم‌های منطق فازی به کار گرفته شده است. محاسبه با واژه‌ها روشی برای پیشرفت مدل‌سازی خطر زمین‌لغزش‌های ناشی از زلزله در مقیاس منطقه‌ای فراهم می‌کند که در واقع روشی حداقل برای تبدیل و استفاده از زبان گفتاری آسان به جای تعداد در محاسبات و استدلال است. ابتدا اطلاعات لازم شامل فاصله به هم ریختگی، رده زمین، رطوبت، شدت زلزله، زاویه شیب، ارتفاع نامنه، عمق خاک، ناظوری نامنه و پوشش گیاهی با استفاده از عکس‌های هوایی، تصاویر ماهواره‌ای، سنجش از دور (RS)، نقشه‌های موجود و بازدیدهای میدانی از منطقه بررسی شده، جمع‌آوری شده است. این داده‌ها با استفاده از سامانه اطلاعات جغرافیایی، رقوم و وزن دهی شده‌اند و با استفاده از برنامه کتل، میزان خطر و تراکم زمین‌لغزش‌های ناشی از زلزله محاسبه شده است و سپس مقایسه زمین‌لغزش‌های پیش‌بینی شده با استفاده از روش مذکور و زمین‌لغزش‌های رخ داده ناشی از زلزله سرعین انجام گرفته است. در نهایت برای ارزیابی روش مذکور در بیش‌ترین زمین‌لغزش‌های ناشی از زلزله، مقایسه تجربی کتل با روش مهندسی انجام شده است.

کلیدواژه‌ها:

سامانه اطلاعات جغرافیایی، منطق فازی، پهنه‌بندی خطر منطقه الگو، زلزله، زمین‌لغزش، ناشی از زلزله

عنوان نشریه:

زمین‌شناسی مهندسی، دانشگاه خوارزمی

فایل متن کامل:

PDF

شماره سند:	2172890	زبان هنرکد:	فارسی	سال انتشار:	1392	نوع هنرکد:	مقاله فارسی	تعداد صفحات:	30
ار صفحه:	1639	تا صفحه:	1668	دوره:	7	شماره:	1	مسلسل:	0

بارگشت صفحه اصلی

## تصویر ۲۲. نمایش اطلاعات کتابشناختی مقاله و چکیده در پایگاه

چنانچه کاربر مایل باشد متن کامل مقاله مربوطه را نیز دریافت کند می‌تواند با ثبت نام در نظام پرداخت الکترونیک موجود در وبگاه مرکز و کلید زدن بر روی گزینه متن کامل، به متن کامل مقاله مورد نظر دسترسی پیدا کند (تصویر ۲۳).

۱۳۳۹

نشریه زمین‌شناسی مهندسی، جلد هفتم، شماره ۱ بهار و تابستان ۱۳۹۲

### پهنه‌بندی خطر زمین‌لغزش‌های ناشی از زلزله ۱۳۷۵ سرعین، ایران با استفاده از روش‌های کیفی و کمی

\* سیدمحمود فاطمی‌عقدا؛ دانشگاه خوارزمی تهران  
محمدرضا مهدویفر؛ پژوهشگاه بین‌المللی زلزله‌شناسی و مهندسی زلزله  
وحید باقری؛ دانشگاه خوارزمی تهران  
تاریخ: فروردین ۹۰/۳۳ پذیرش: ۹۰/۱۰/۱۳

#### چکیده

در این پژوهش یکی از روش‌های جدید پهنه‌بندی خطر زمین‌لغزش‌های ناشی از زلزله (مدل ناحیه‌ای جامع زمین‌لغزش‌های ناشی از زلزله با مدل کتل [1]) برای پهنه‌بندی این نوع از زمین‌لغزش‌ها و ارزیابی آن با روش مهندسی بررسی شده است. بدین منظور محدوده تحت تأثیر زلزله سرعین (سال ۱۳۷۵) برای منطقه الگو انتخاب شده است. در این روش که محاسبه با واژه‌ها نامیده می‌شود، سیستم‌های منطق فازی به کار گرفته شده است. محاسبه با واژه‌ها روشی برای پیشرفت مدل‌سازی خطر زمین‌لغزش‌های ناشی از زلزله در مقیاس منطقه‌ای فراهم می‌کند که در واقع روشی حداقل برای تبدیل و استفاده از زبان گفتاری آسان به جای تعداد در محاسبات و استدلال است. ابتدا اطلاعات لازم شامل فاصله به هم ریختگی، رده زمین، رطوبت، شدت زلزله، زاویه شیب، ارتفاع نامنه، عمق خاک، ناظوری نامنه و پوشش گیاهی با استفاده از عکس‌های هوایی، تصاویر ماهواره‌ای، سنجش از دور (RS)، نقشه‌های موجود و بازدیدهای میدانی از منطقه بررسی شده، جمع‌آوری شده است. این داده‌ها با استفاده از سامانه اطلاعات جغرافیایی، رقوم و وزن‌دهی شده‌اند و با استفاده از برنامه کتل، میزان خطر و تراکم زمین‌لغزش‌های ناشی از زلزله محاسبه شده است و سپس مقایسه زمین‌لغزش‌های پیش‌بینی شده با استفاده از روش مذکور و زمین‌لغزش‌های رخ داده ناشی از زلزله سرعین انجام گرفته است. در نهایت برای ارزیابی روش مذکور در بیش‌ترین زمین‌لغزش‌های ناشی از زلزله، مقایسه تجربی کتل با روش مهندسی انجام شده است.

واژه‌های کلیدی: زلزله، زمین‌لغزش، ناشی از زلزله، منطق فازی، پهنه‌بندی خطر منطقه‌ای، سامانه اطلاعات جغرافیایی (GIS).

\* نویسنده مسئول: Fatemi@tmu.ac.ir

۱. CAMEL: A Comprehensive Areal Model of Earthquake induced Landslides

## تصویر ۲۳. نمایش متن کامل مقاله در پایگاه

## مقایسه جستجو در نسخه قدیمی و جدید پایگاه مقالات فارسی

در این قسمت برای اینکه بتوان به گونه‌ای شفاف‌تر مزایای دو امکان اضافه شده به پایگاه مقالات تمام‌متن فارسی (جستجوی موضوعی+پیشنهاد پرس‌وجو) را به تصویر کشید، یک جستجوی آزمایشی با یک کلیدواژه خاص در نسخه قدیمی (فاقد این دو امکان) و جدید پایگاه انجام شد که نتایج این بررسی در ادامه آورده می‌شود.



تصویر ۲۴. جستجوی کلیدواژه «گندم» در نسخه قدیمی پایگاه

همانگونه که در این تصویر مشاهده می‌شود، در نسخه قدیمی پایگاه لازم است یک کلیدواژه یا پرس‌وجو به صورت کامل تایپ و وارد شود و نمی‌توان با تایپ دو یا چند کاراکتر جستجوی موفق داشت. به هر تقدیر، پس از درونداد کلمه «گندم» و کلید زدن بر روی گزینه جستجو، ۳۱۶۰ مقاله مطابق تصویر ۲۵ بازیابی می‌شود. کاربر می‌تواند با تنظیم گزینه مربوطه در هر صفحه، ۱۰ تا ۱۰۰ مقاله را مشاهده نماید.



تصویر ۲۵. تعداد رکوردهای بازیابی شده با کلیدواژه «گندم» در نسخه قدیمی پایگاه



در تصویر ۲۶، نسخه‌ی ارتقاء یافته‌ی پایگاه نشان داده شده است. برخلاف نسخه‌ی قبلی، در این نسخه کاربر می‌تواند با وارد کردن دو حرف، ۵ پیشنهاد را ملاحظه نماید. همانگونه که در تصویر ۲۶ نشان داده شده است کاربر ابتدا از منوی سمت راست گزینه‌ی «کلیدواژه» را که برای فعال سازی نسخه‌ی جدید نرم‌افزار ضروری است، انتخاب نموده و سپس دو حرف «گن» را درون‌داد کرده است. بدین ترتیب ۵ کلیدواژه «گندم»، «گندم دوروم»، «گندم نان»، «گندم کاران» و «گندم‌های سرداری و پارسی» پیشنهاد شده است. کاربر در این آزمایش کلیدواژه «گندم» را انتخاب نموده و بر روی گزینه‌ی «مشاهده» کلیک زده است.



تصویر ۲۶. جستجوی کلیدواژه «گندم» در نسخه‌ی جدید پایگاه

در این نسخه و با طی همان مراحل قبلی این بار به جای ۳۱۶۰ مقاله، تنها ۸۹۵ مقاله بازیابی شده است (تصویر ۲۷).



تصویر ۲۷. تعداد رکوردهای بازیابی شده با کلیدواژه «گندم» در نسخه‌ی جدید پایگاه

مقایسه عملکرد نسخه قدیمی پایگاه مقالات تمام‌متن فارسی با صورت ارتقاء یافته آن نشان می‌دهد که در نسخه جدید، تعداد مدارک بازیابی شده بسیار کمتر از تعداد مدارکی است که از نسخه قدیمی به دست می‌آید (۸۹۵ مقاله در مقابل ۳۱۶۰ مقاله). دلیل این امر نیز آن است که در نسخه جدید، کاربر با انتخاب یک حوزه موضوعی خاص صرفاً مقالاتی را که به آن حوزه موضوعی مربوط بوده، جستجو کرده و بنابراین با وجود کاهش بازیابی دقت جستجو افزایش پیدا کرده است. به عبارت دیگر، کاربران در نسخه جدید راحت‌تر و سریع‌تر به رکوردهای مورد نظر خود دسترسی پیدا می‌کنند. نکته دیگر در مقایسه دو نسخه آن است که ترتیب مقالات نیز در دو نسخه با هم فرق دارد که این به خاطر محدود کردن جستجو به یک حوزه موضوعی خاص و نیز روش خاص وزن‌دهی است که در پژوهش حاضر از آن استفاده شد. به بیان دیگر، در نسخه قدیمی با درونداد کلیدواژه «گندم» مقاله «ایران چه می‌کارد؟ چه صادر می‌کند؟ و چه وارد می‌کند» به عنوان مقاله نخست بازیابی می‌شود ولی در نسخه جدید با درونداد همان کلیدواژه مقاله «مقایسه شاخص‌های اراضی اصلاح‌شده و اصلاح نشده در روش پارامتری ارزیابی تناسب اراضی» بازیابی می‌شود. نکته پایانی اینکه در نسخه ارتقاء یافته، امکان جستجوی مقاله به طور همزمان و از تمام حوزه‌های موضوعی نیز وجود دارد و بنابراین تمامی قابلیت‌های نسخه قبلی در نسخه فعلی نیز موجود است.

### ارزیابی نسخه جدید نرم‌افزار

برای ارزیابی نسخه جدید نرم‌افزار، از معیارهای کیفی استفاده شد. به عبارت دیگر، نسخه قبلی و نیز لینک نسخه ارتقاء یافته نرم‌افزار به ۳ کارشناس ارشد علم اطلاعات و دانش‌شناسی (با حداقل ۵ سال سابقه کار در مراکز اطلاع‌رسانی) و نیز ۵ کاربر نهایی از دانشجویان تحصیلات تکمیلی و مراجعه‌کننده به مرکز منطقه‌ای، داده شد و پس از معرفی مختصات کار از دو گروه خواسته شد تجربه و نظر خود را در مورد نسخه جدید اعلام دارند. درخواست شد در نگارش نظرات، به نقش خود به عنوان افرادی که خدمات اطلاع‌رسانی را به کاربران ارائه می‌دهند (گروه اول) و نیز جوینده اطلاعات (گروه دوم) توجه کنند. نظرات دو گروه به شرح زیر بود:

نظرات گروه اول:

نفر اول: این امکان بسیار خوبی است و باید مدت‌ها پیش برقرار می‌شد. با راه‌اندازی این نسخه، فرایند بازیابی اطلاعات کوتاه‌تر و آسان‌تر می‌شود و به دلیل اینکه مقالات به تفکیک حوزه موضوعی ارائه می‌شود، دقت مقالات بازیابی شده و ارتباط آن با هدف کاربر بیشتر است.

نفر دوم: جالب و مفید است، استفاده از آن کار ما و کاربر را ساده‌تر می‌کند. امکان خوبی است که به جهت ارائه خدمت جدید و کاربردی بودن آن، باعث ارتقاء نام سازمان می‌شود.

نفر سوم: اینجور امکانات در حوزه بازیابی اطلاعات بسیار مفید و مؤثر است. البته، لازم است که در بخش‌های دیگر هم پشتیبانی لازم انجام شود و تفکیک موضوعی مجلات و مقالات با دقت و به صورت کامل انجام شود.

نظرات گروه دوم:

گروه دوم (۵ کاربر نهایی) که همگی دانشجویان تحصیلات تکمیلی بودند، پس از رویت دو نسخه و کار با آن‌ها، نظر خود را در خصوص کارایی احتمالی نسخه جدید برای کاربران نهایی اعلام داشتند که متن این نظرات در ادامه آورده شده است:

نفر اول: راحت‌تر و سریع‌تر است.

نفر دوم: باعث صرفه‌جویی در وقت می‌شود و مقالات مربوط به رشته خودمان را بهتر مشاهده می‌کنیم.

نفر سوم: از نسخه قبلی خیلی بهتر است ولی در هر دو نسخه گاهی دنبال چیزی می‌گردیم که در پایگاه مرکز نیست.

نفر چهارم: به نظرم نسخه جدید خیلی بهتر است. انگار یک واژه‌نامه هم در اختیار دارم. اگر کلیدواژه‌ای را ندانم با وارد کردن چند حرف می‌توانم کلیدواژه‌های موجود در پایگاه را هم ببینم.

نفر پنجم: خیلی مفید است. باعث می‌شود اشکالات تاپی کمی‌تر شود. تازه می‌توانم بفهمم در پایگاه چه کلیدواژه‌هایی استفاده شده که می‌توانم از آن‌ها استفاده کنم. دقت بازیابی هم بیشتر می‌شود چون اطلاعات حوزه‌های دیگر که دنبالش نیستم، بازیابی نمی‌شود.

با توجه به محدودیت‌های تحقیق (و نبود نسخه‌های مشابه در زمان انجام پژوهش حاضر برای انجام مقایسه کمی)، هدف از ارزیابی، انجام ارزیابی‌هایی با طیف وسیع نبوده است. از اینرو، بررسی نظرات تعداد اندکی از افراد مورد آزمایش (کارشناسان ارشد علم اطلاعات و دانش‌شناسی و کاربران نهائی با مدرک کارشناسی ارشد و از رشته‌های مختلف) نشان داد که اولاً تقریباً همگان اذعان دارند که نسخه جدید در مقایسه با نسخه قدیمی موفق‌تر است. دوم آنکه به دلیل ارائه تعداد کمتری از رکوردها، سبب صرفه‌جویی در وقت کاربر یا اطلاع‌رسان می‌گردد. سوم آنکه به دلیل انجام جستجو در حوزه‌های موضوعی خاص، باعث افزایش دقت بازیابی می‌گردد. چهارم آنکه با ارائه پیشنهادها سبب کاهش مشکلات تایپی و نیز جهت‌دهی به انتخاب کلیدواژه‌ها توسط کاربر می‌شود.

## نتیجه‌گیری

هدف از انجام پژوهش حاضر آن بود تا نرم‌افزاری تهیه و در نظام رایسست تلفیق شود که بتواند: (۱) جستجوی موضوعی را در پایگاه مقالات تمام‌متن فارسی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری محقق سازد و (۲) پس از درونداد حداقل دو کاراکتر در جعبه پرس‌وجو، پنج پرس‌وجوی پیشنهادی را برای استفاده احتمالی کاربر ارائه نماید.

برای انجام تحقیق از مخزن کلیدواژه‌های رایسست (البته، صرفاً کلیدواژه‌های فارسی) به عنوان منبع اصلی داده تحقیق استفاده شد. این کلیدواژه‌ها بررسی و برای استفاده در پژوهش حاضر اصلاح و آماده‌سازی شدند. مدخل‌های مخزن از ترکیبات پیچیده به ترکیبات ساده تبدیل شد. سپس الگوریتم ساخت نرم‌افزار پروژه حاضر تهیه گردیده و به برنامه تبدیل گردید. این نرم‌افزار حاوی دو بخش اساسی است. بخش اول مربوط به کاربر سیستم است که به عنوان پشتیبان سیستم کار می‌کند و لازم است روزآمدسازی اطلاعات نرم‌افزار را به انجام برساند. بخش دوم که تولید نسخه جدید از صفحه مربوط به پایگاه مقالات تمام‌متن فارسی است و نرم‌افزار اصلی در آن تلفیق گردیده است. کاربران عادی در حقیقت نسخه اصلاح شده پایگاه تمام‌متن مقالات فارسی را روئیت خواهند نمود. این نرم‌افزار کاربران را قادر می‌سازد: (۱) جستجوی خود را بر اساس حوزه‌های موضوعی مورد علاقه به انجام برسانند و (۲) با درونداد کردن حداقل دو کاراکتر، کلیدواژه‌ها و پرس‌وجوهای پیشنهادی (پنج مورد) را از سیستم دریافت نمایند. پس از ساخت نرم‌افزار، نتیجه کار توسط سه متخصص علم اطلاعات و دانش‌شناسی با مدرک کارشناسی ارشد و پنج کاربر نهایی از مقطع تحصیلی کارشناسی ارشد مورد ارزیابی قرار گرفت. افراد مورد آزمایش اذعان داشتند که نسخه جدید در مقایسه با نسخه قدیمی موفق‌تر است. دوم آنکه به دلیل ارائه تعداد کمتری از رکوردها، سبب صرفه‌جویی در وقت کاربر یا اطلاع‌رسان می‌گردد. سوم آنکه به دلیل انجام جستجو در حوزه‌های موضوعی خاص، باعث افزایش دقت بازیابی می‌شود و چهارم آنکه با ارائه پیشنهادها سبب کاهش مشکلات تایپی و نیز جهت‌دهی به انتخاب کلیدواژه‌ها توسط کاربر می‌گردد.

نتایج به دست آمده از پژوهش حاضر دارای کاربردهای عملی متعددی است که در این قسمت مختصری از آن آورده می‌شود. می‌توان از این نرم‌افزار در ارتقاء زیرساخت موتورهای کاوش فارسی و به ویژه رایسست استفاده نمود. می‌توان از طریق آموزش کاربران نهایی و نیز متخصصان علم اطلاعات و دانش‌شناسی، جامعه علمی را به استفاده از امکانات ارائه شده ترغیب نمود. یافته‌های پژوهش حاضر برای دانشجویان و اساتید علم اطلاعات و دانش‌شناسی و نیز زبان‌شناسی رایانه‌ای نیز بسیار سودمند است چرا که پژوهش حاضر از تعامل صورت گرفته بین این دو حوزه موضوعی حاصل گردیده است. نتایج پژوهش حاضر همچنین برای محققان حوزه رایانه و بازیابی اطلاعات نیز مفید است چرا که می‌توانند راه‌کارهایی بهینه را برای عملیاتی سازی اینگونه پژوهش‌ها ارائه نمایند.

بر اساس نتایج به دست آمده در پژوهش حاضر می‌توان پژوهش‌های دیگری را نیز پایه‌ریزی کرد. به عنوان مثال، می‌توان این نرم‌افزار را برای پایگاه‌های فارسی دیگر مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری نیز راه‌اندازی کرد. می‌توان چنین نرم‌افزاری را با استفاده از متغیرهای دیگر یا بیشتر تولید کرد و نتایج را با نتایج بدست آمده از پژوهش حاضر مقایسه نمود. همچنین می‌توان این کار را با پایگاه‌های انگلیسی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری به انجام رساند.

## منابع و مراجع

- [۱] خسروی، ع، فتاحی، ر، پربرخ، م، و دینانی، م. ح. (۱۳۹۲). بررسی مارآمدی کلیدواژه‌ها و عبارت‌های پیشنهادی موتور کاوش گوگل در بسط جستجو و افزایش ربط از دیدگاه دانشجویان تحصیلات تکمیلی. *پژوهش‌نامه کتابداری و اطلاع‌رسانی*، ۳(۱)، ۱۳۳-۱۵۰.
- [۲] طاهریان، آ، و حیاتی، ز. (۱۳۸۸). بررسی مقایسه‌ای الگوی جستجوی کلیدواژه‌ای با دریافت کمک و الگوی جستجوی بدون دریافت کمک در موتور کاوش گوگل از دیدگاه کاربران. *فصلنامه مطالعات تربیتی و روانشناسی دانشگاه فردوسی مشهد*، ۱۰(۲)، ۹۱-۱۱۲.
- [۳] فتاحی، ر. (۱۳۸۵). شناسائی و تحلیل واژگان عمومی در منابع وب: رویکردی نو به بسط عبارت جستجو با استفاده از زبان طبیعی در موتورهای کاوش. *دوفصلنامه مطالعات علوم تربیتی و روانشناسی*، ۷(۱)، ۳۱-۵۳.
- [۴] کلانتر، س. م، و منصفی، ر. (۱۳۸۱). سیستم تطبیقی فیلترینگ اطلاعات وب بوسیله الگوریتم‌های ژنتیک. هشتمین کنفرانس سالانه انجمن کامپیوتر ایران، مشهد، دانشگاه فردوسی مشهد. بازبایی شده در ۱۲ سپتامبر ۱۳۹۴ از [http://www.civilica.com/Paper-ACCSI08-ACCSI08\\_025.html](http://www.civilica.com/Paper-ACCSI08-ACCSI08_025.html)
- [۵] مهرداد، ج. (۱۳۹۲). *آشنائی با مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری*. شیراز: تخت جمشید.
- [6] Bahatia, S., Majumdar, D., & Mitra, P. (2011). Query suggestions in the absence of query logs. In *SIGIR'11*, pp. 795-804.
- [7] Bar-Yossef, Z., & Kraus, N. (2011). Context-sensitive query auto-completion. In *WWW'11*, pp. 107-116. Retrieved Sep. 12, 2015 from [www.conference.org/proceedings/www2011/proceedings/p107.pdf](http://www.conference.org/proceedings/www2011/proceedings/p107.pdf).
- [8] Bialczak, P., Mazurczyk, W., & Szczypiorski, K. (2014). Sending hidden data via google suggest. Retrieved Nov. 22, 2014, from [www.arxiv.org/ftp/arxiv/papers/1107/1107](http://www.arxiv.org/ftp/arxiv/papers/1107/1107).
- [9] Golbandi, N., Katzir, L., Koren, Y., & Lempel, R. (2013). Expediting search trend detection via prediction of query counts. In *WSDM'13*, pp. 295-304.
- [10] Caiyz, F., Liang, Sh., & de Rijke, M. (2014). In *CIKM'14*, November 3-7, Shanghai, China. Retrieved December 17, 2014, from <https://staff.fnwi.uva.nl/m.derijke/publications/cikm2014-fp-qac.pdf>.
- [11] Cirasella, J. (2007). Google sets, google suggest, and google search history: Three more tools for the reference librarian's bag of tricks. Retrieved Jan. 26, 2015, from [www.works.gc.cuny.edu/cgi/viewcontent.cgi?article=1006&content=facworks](http://www.works.gc.cuny.edu/cgi/viewcontent.cgi?article=1006&content=facworks).
- [12] Golbandi, N. G., Katzir, L. K., Koren, Y. K., & Lempel, R. L. (2013). Expediting search trend detection via prediction of query counts. In *WSDM'13*, 295-304.
- [13] Kastrinakis, D., & Tzitzikas, Y. (2010). Advancing search query autocompletion with more and better suggestions. Retrieved Feb. 16, 2015 from [www.users.ics.forth.gr/~tzitzik/publications/Tzitzikas\\_2010\\_ICWE.pdf](http://www.users.ics.forth.gr/~tzitzik/publications/Tzitzikas_2010_ICWE.pdf).
- [14] Mohan, M. (2014). Over 151 google products and dervices you probably don't know. Retrieved December 15, 2014, from <http://www.minterest.org/google-products-services-you-probably-dont-know/>.
- [15] Sengstock, C., & Gertz, M. (2011). Conquer: A system for efficient context-aware query suggestions. In *WWW'11, ACM*, pp. 265-268.
- [16] Shokouhi, M. (2011). Detecting seasonal queries by time-series analysis. In *SIGIR'11*, pp. 1171-1172.
- [17] Shokouhi, M., & Radinsky, K. (2012). Time-sensitive query auto-completion. In *SIGIR'12*, pp. 601-610
- [18] Shokouhi, M. (2013). Learning to personalize query auto-completion. In *SIGIR'13*. Retrieved April, 27, 2015, from [www.research.microsoft.com/pubs/193319/SIGIR2013-shokouhi-personalizedQAD.pdf](http://www.research.microsoft.com/pubs/193319/SIGIR2013-shokouhi-personalizedQAD.pdf).
- [19] Strizhevskaya, A., Baytin, A., Galinskaya, I., & Serdyukov, P. (2012). Actualization of query suggestions using query logs. In *WWW'12*, pp. 611-612.

- [20] Sullivan, D. (2011). How google instant's autocomplete suggestions work? Retrieved Jan. 12, 2015, from <http://searchengineland.com/how-google-instant-autocomplete-suggestions-work-62592>.
- [21] Ward, D., Hahn, J., & Feist, K. (2012). Autocomplete as a research tool: A study on providing search suggestions. *Information Technology and Libraries*, December Issue. Retrieved Feb. 28, 2015, from [www.ejournals.bc.edu/ojs/index.php/ital./article/download/1930/pdf](http://www.ejournals.bc.edu/ojs/index.php/ital./article/download/1930/pdf).
- [22] Whiting, S., & Jose, J. M. (2014). Recent and robust query auto-completion. Retrieved Jun. 14, 2015 from [www.stewh.com/wp-content/uploads/2014/02/fp539-whiting.pdf](http://www.stewh.com/wp-content/uploads/2014/02/fp539-whiting.pdf).
- [23] Whiting, S., McMinn, J., & Jose, J. M. (2013). Exploring real-time temporal query auto-completion. In DIR 2013. Retrieved April 23, 2015 from [www.ceur-ws.org/vol-986/paper\\_14.pdf](http://www.ceur-ws.org/vol-986/paper_14.pdf).
- [24] Xiao, Ch., Qin, J., & Wang, W. (2013). Efficient error-tolerant query autocompletion. The 39<sup>th</sup> International Conference on Very large Data Bases, 26-30, August, Riva del Garda, Trento, Italy. Retrieved May 22, 2015, from [www.vldb.org/pvldb.vol6/p373-xiao.pdf](http://www.vldb.org/pvldb.vol6/p373-xiao.pdf).