

## بسط جستجو و واکاوی قضاوت ربط در پایگاههای اطلاعاتی

### مروری بر پیشینه پژوهش

با گسترش وب و بویژه توسعه موتورهای کاوش، دسترسی کاربران به اطلاعات افزایش چشمگیری یافته است. در این زمینه، هرچند که کاربران تلاش می کنند با استفاده از انواع قابلیتهای موتورهای کاوش به جستجو و بازیابی اطلاعات مورد نیاز خود بپردازند اما، در درجه نخست این وظیفه طراحان موتورهای کاوش است که از پیش راههای مناسب کمک به کاربران را برای جستجوی کارآمد را در نظر گرفته و پیاده سازی کنند. با توجه به اهمیت چنین امری، حوزه پژوهش در زمینه موتورهای کاوش در دهه اخیر بسیار پویا شده و ادبیات آن بسیار غنی است. برای شناسایی و بازیابی مناسب ترین و مرتبط ترین منابع اطلاعاتی در زمینه های مورد نیاز، پایگاههای اطلاعاتی و موتورهای کاوش امکانات و قابلیتهایی را در اختیار کاربران قرار می دهند. یکی از این امکانات، استفاده از راهبرد بسط جستجو (Query expansion) است. بازنگری و پالایش عبارت جستجو می تواند احتمال بازیابی منابع مرتبط تر را افزایش دهد. در عین حال، میزان مرتبط بودن منابع بازیابی شده به قضاوت کاربر بستگی دارد.

پژوهشهای فراوانی در سه دهه گذشته در زمینه بسط جستجو و قضاوت ربط (Relevance judgement) در پایگاههای اطلاعاتی انجام شده است که یافته های ارزشمندی را نه تنها برای بهسازی عملکرد پایگاههای اطلاعاتی به دست داده بلکه زمینه های جدیدتری را برای انجام پژوهشهای بیشتر فراهم آورده است. آنچه در

زیر ارائه شده، مروری بر ادبیات تحقیق یا پیشینه پژوهش در این زمینه است که برپایه موضوع تخصصی، دسته بندی شده است.

#### ۴-۲-۲. مطالعات حوزه عبارت‌های پیشنهادی، گزینش و قضاوت ربط

شاید بتوان گفت استفاده از کلیدواژه و عبارت‌های پیشنهادی در موتورهای کاوش تخصصی، رویکرد استفاده از هستی‌شناسی‌ها در اصلاح عبارت‌های جستجوی کاربران در ارائه لیست‌های پیشنهادی در موتورهای کاوش از دهه ۱۹۹۰مورد مطالعه جدی قرار گرفته است. در این راستا، ویدیانتورو (Widyantoro and Yen, ۲۰۰۱) در پژوهشی که در موتور کاوشی که چکیده مقالات ارایه می‌داد، به اصلاح عبارت جستجوی کاربران با استفاده از هستی‌شناسی فازی پرداخت. نتایج پژوهش نشان داد که اصلاح عبارت جستجو از طریق کلیدواژه‌های پیشنهادی هستی‌شناسی به طور موثری مورد توجه کاربران قرار گرفته و رضایت کاربران را در پی داشته است. استفاده از این ویژگی در نظام‌های بازیابی اطلاعات می‌تواند به عنوان یک ویژگی مهم مورد توجه قرار گیرد.

یکی دیگر از این پژوهش‌ها، مطالعه انیک و کانتامنی (Anick and Kantamneni, ۲۰۰۳) بود که به منظور بررسی تفاوت بین استفاده از کلیدواژه و عبارت‌های پیشنهادی و عدم استفاده از آنها در جستجو در موتورهای کاوش انجام شد. این بررسی به منظور تحلیل داد و گرفت نشست‌های جستجوی دو گروه کاربران انجام گرفت. نتایج نشان داد که هر چند از نظر میزان «موفقیت» در گروه مورد بررسی در انجام جستجوهای مورد نظر تفاوتی وجود ندارد، ولی کاربران استفاده-

کننده از این و کلیدواژه و عبارت‌های پیشنهادی، در صورت تداوم جستجو و تکرار آن در آینده و به دلیل آشنایی با این امکان، از آن به صورتی کارآمدتر در جستجوی خود استفاده خواهند کرد.

یکی از زمینه‌های مطالعاتی این حوزه، گزینش کلیدواژه و عبارت‌های جستجو با استفاده از ابزارهای موجود نظیر اصطلاحنامه‌ها و برخی منابع دیگر بوده است. در این راستا، فیدل (Fidel, ۱۹۹۱) مطالعه خود را بر دو نوع منبع مهم به منظور گزینش کلیدواژه‌های جستجو متمرکز می‌کند: متن آزاد و توصیف‌گرها. به دنبال آن (Spink and saracevic, ۱۹۹۷) دو رویکرد عمده انتخاب کلیدواژه جستجو را شناسائی کردند. رویکرد نخست مبتنی بر الگوریتم و رویکرد دوم مبتنی بر انسان بود. همچنین اسپینک و ساراسویک (Spink and saracevic, ۱۹۹۷) پنج منبع عمده را برای گزینش کلیدواژه شناسائی نمود: بیان پرسش، تعامل کاربر با رابط، اصطلاحنامه، رابط انسانی و بازخورد ربط کلیدواژه از عمده منابع گزینش کلیدواژه‌ها بودند.

همانطور که اشاره شد، در برخی از مطالعه‌ها گزینش کلیدواژه‌ها مبتنی بر رویکرد کاربرمدار بوده است. در این زمینه شیری (Shiri and Crawford, ۲۰۰۶) پژوهشی مروری به منظور بررسی مطالعات مبتنی بر رویکردهای کاربرمدار جهت شناسائی منابع گزینش کلیدواژه‌های جستجو انجام داد. بررسی وی نشان داد که عمده روش‌های گزینش کلیدواژه‌های جستجو با کمک اصطلاحنامه‌ها و کلیدواژه‌هایی که از بسط جستجوی حاصل از استفاده اصطلاحنامه‌ها توسط کاربر نهائی انجام می‌شود، بدست می‌آید.

گزینش کلیدواژه از میان مجموعه‌ای از کلیدواژه‌های پیشنهادی نیز یکی دیگر از زمینه‌های پژوهشی این حوزه را تشکیل می‌دهد. در این مورد روتون (Ruthven, ۲۰۰۳) رفتار واژه‌گزینی کاربران را در مجموعه‌ای از واژگان بسط یافته که از بیست و پنج مورد مدرک با بیشترین مورد بازیابی مطالعه کرد. نتایج پژوهش او نشان داد که کاربران روابط معنایی بین واژگان عبارت پرسش و واژگان بسط یافته پیشنهادی را درک نمی‌کنند.

یکی دیگر از این پژوهش‌ها توسط هانگ (Huang et al., ۲۰۰۳) و به منظور بررسی شیوه‌های کارآمد در پیشنهاد کلیدواژه و عبارت‌های مرتبط با استفاده از سیاهه‌های حاصل از نشست‌های مختلف جستجو و با روش داد و گرفت، انجام شد. نتایج وی نشان داد که ربط بین کلیدواژه‌های جستجو و کلیدواژه‌های استخراجی در حد انتظار نبود.

از جمله پژوهش‌های دیگر می‌توان به پژوهش ليو (Liu and Belkin, ۲۰۰۸) که به منظور نشان دادن اهمیت تدوین عبارت جستجو با استفاده از قابلیت عبارت پیشنهادی مستخرج از بازخورد یا قضاوت ربط انجام داد، اشاره کرد. به طور کلی، یافته‌های حاصل از مطالعات تدوین تعاملی عبارت جستجو (بسط جستجو توسط کاربر براساس نتایجی که نظام اطلاعاتی به وی ارائه می‌دهد نشان داده است که کاربرانی که برای جستجوی اطلاعات آموزش ندیده‌اند و یا کاربرانی که دانش کمی درباره مجموعه دارند قادر به تشخیص واژگان مناسب برای بازیابی و درک مرتبط‌تر نیستند. اگر بستر و حوزه موضوعی مطرح در عبارت جستجوی کاربر

تعیین شود و در کمک به کاربر برای تدوین مجدد عبارت جستجوی خود به کار برود، نظام تعاملی بالقوه به هدفش رسیده است.

در مطالعه‌ای دیگر ایتی‌میادیس (Efthimiadis, ۲۰۰۰) بسط جستجوی تعاملی مبتنی بر ارزیابی کاربر در محیط بازخورد ربط را مورد مطالعه قرار داد. داده‌ها از ۲۵ جستجوی انجام شده در پایگاه INSPEC جمع‌آوری شد. این پژوهش بر انجام مطالعه‌های بیشتر برای بسط جستجو و انجام مطالعات کاربران در نظام‌های ارزیابی عمومی تأکید کرد.

#### ۴-۲-۴. مطالعات حوزه بار شناختی<sup>۱</sup> و جستجو

یکی از رویکردهای پژوهشی حوزه‌های جستجو، یافتن راهی برای کاهش بار شناختی کاربران هنگام ارزیابی اطلاعات است. در این راستا تائو (Tao, ۲۰۰۶) با هدف بررسی افزایش بار شناختی در نتایج زیاد و نامرتب، پژوهشی انجام داد. نتایج وی نشان داد که ظرفیت حافظه کاری (کوتاه مدت) محدود است و حداکثرش عنصر غیر مرتبط را می‌تواند بطور همزمان در خود نگه دارد و تنها یک چهارم از عناصری که در ذهن کاربران باقی مانده است آنهایی است که ارتباطی با هدف جستجو ندارد. به عبارت دیگر، حافظه کاری کاربران تمایل ندارد عناصر غیر مرتبط را هنگام جستجو در خود نگهدارند.

نتایج پژوهشی که گویزکا (Gwizdka, ۲۰۰۸) به منظور ارزیابی تأثیر بار شناختی بر روی جستجوگران وب انجام داد، نشان داد که بین مؤلفه‌های عملکرد

---

<sup>۱</sup> Cognitive load

فعالیت اولیه جستجو<sup>۲</sup> (حقیقت‌یابی<sup>۳</sup>) با سختی فعالیت ذهنی<sup>۴</sup> ارتباط معنی‌داری وجود دارد. اما بین عملکرد فعالیت ثانویه جستجو (گردآوری اطلاعات) با سختی فعالیت ذهنی ارتباط معناداری یافت نشد.<sup>۵</sup>

#### ۴-۲-۲. مطالعات داخل کشور

در ایران در حوزه بسط جستجو به جز فتاحی که در پیشینه پژوهش مطالعه‌های بسط جستجو در خارج از کشور مطرح گردید، پژوهش جدی انجام نگرفته است. برخی از پژوهش‌ها نظیر کشاورز (۱۳۸۳)؛ (Tabatabai & shore, ۲۰۰۵)؛ آزادی (۱۳۸۴) در نتایج خود به مشکلات کاربران در تدوین عبارت جستجو و برخی از مشکلات این حوزه اشاره‌ای داشته‌اند، اما به چگونگی جستجو به زبان طبیعی و با بسط جستجو توجهی نکردند. به طور کلی، هیچ پژوهشی تاکنون در زمینه بسط جستجو بر اساس عبارتهای پیشنهادی موتورهای کاوش، به شیوه‌ای که مورد نظر پژوهش حاضر است، انجام نگرفته است.

#### ۴-۳. استنتاج از پیشینه پژوهش

بررسی متون نشان می‌دهد که موضوع بسط جستجو از مدتها پیش به موازات توسعه نظام‌های بازیابی اطلاعات مورد توجه جدی پژوهشگران قرار گرفته است. از دهه‌های ۹۰ به بعد هم‌زمان با رشد اینترنت و توسعه موتورهای کاوش این

<sup>۲</sup> Primary task performance

<sup>۳</sup> Fact finding

<sup>۴</sup> Subjective task difficulty

<sup>۵</sup> همانگونه که در بخش روش شناسی مطرح خواهد شد، در پژوهش حاضر نیز نظیر همین رویکرد مورد استفاده قرار خواهد گرفت. در ابتدا، جستجوی اولیه بر اساس عبارتهای پیشنهادی انجام خواهد گرفت و دیدگاه کاربران در مورد مرتبط بودن آنها با نیاز آنان مورد ارزیابی قرار خواهد گرفت. سپس، بعد از آنکه کاربران روی یک یا چند مورد از عبارتهای پیشنهادی کلیک کرده و نتایج حاصل را مورد بررسی قرار دادند، از آنها در باره میزان مرتبط بودن نتایج در مقایسه با ارزیابی اولیه آنها بر اساس جستجوی اولیه سوال خواهد شد و مشکلات مرتبط با این فرایند نیز مورد مطالعه قرار خواهد گرفت.

راهبرد در وب هم مورد توجه قرار گرفته است. بدین منظور، عمده پژوهش‌ها تاکنون به دلیل نوپا بودن این راهبرد در نظام‌های جدید بازیابی اطلاعات از دیدگاه نظام‌مدار بوده است.

بررسی متون نشان داد که به دلیل نو بودن موضوع، اغلب پژوهش‌ها با روش‌های پیمایشی و تجربی و رویکرد کمی و با استفاده از ابزارهایی نظیر پرسشنامه، مشاهده، پروتکل بلنداندیشی، ثبت وقایع و سیاهه‌های داد و گرفت و جامعه مورد پژوهش اغلب آنها کلیدواژه‌های پیشنهادی و متخصصان کامپیوتر بوده است. در حوزه بسط جستجو به زبان طبیعی، دیدگاه سایر کاربران به ویژه دانشجویان مورد بررسی قرار نگرفته است. در این راستا بررسی دیدگاه‌های کاربران موتورهای کاوش به منزله یکی از شکاف‌های پژوهشی در این حوزه است که می‌تواند سودمند باشد. همچنان‌که برخی پژوهش‌گران، به این نوع بررسی و کارآمدی از جنبه کاربرمداری تأکید نموده بودند.