

## تحلیل رفتار اصلاح پرس و جوی کاربر بر اساس ویژگی‌های معنایی: تحلیل لاگ کاربران پایگاه اطلاعات گنج (ایراندک)

\* **المیرا کریمی:** دانشجوی دکتری علم اطلاعات و دانش‌شناسی، پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)، تهران، ایران (نویسنده مسئول)

karim@students.irandoc.ac.ir

**محمود بابایی:** استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)، تهران، ایران

**ملوک‌السادات حسینی بهشتی:** استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)، تهران، ایران

### چکیده

دریافت: ۱۳۹۷/۶/۱۱

پذیرش: ۱۳۹۷/۱۰/۱۷

**زمینه و هدف:** نظام‌های اطلاع‌رسانی را بدون یافتن درکی روشن از مقوله‌هایی چون نیازهای استفاده‌کنندگان، چگونگی جستجوی اطلاعات توسط آنها، و چگونگی ارزیابی اطلاعاتی که دریافت می‌کنند، نمی‌توان به نحو مطلوبی طراحی کرد یا توسعه داد. این پژوهش به منظور بررسی رفتار اطلاع‌یابی کاربران، به تحلیل لاگ کاربران پایگاه اطلاعات گنج (ایراندک) پرداخته است. آن چه در این پژوهش برای بررسی در نظر گرفته شده است، ویژگی‌های معنایی جستجو و اصلاح جستجو در میان کاربران گنج است.

**روش پژوهش:** روش این پژوهش، تحلیل گزارش‌های تراکنش است که یکی از روش‌های تجزیه و تحلیل پایگاه‌های اطلاعاتی به شمار می‌رود. در این پژوهش، لاگ کاربران پایگاه اطلاعات گنج در بازه زمانی سه ماهه (از اردیبهشت تا مرداد ۱۳۹۵) از نظر ویژگی‌های معنایی تحلیل شد. ابزار این پژوهش سیاهه واری بود که ویژگی‌های معنایی را مدنظر قرار داده و به روش مشاهده انجام شد. تعداد کل جستجوهای دریافتی ده میلیون رکورد بود که آنها به جستجوهای حوزه علم اطلاعات (با محدودسازی با اصطلاحات اصطلاحنامه در این حوزه) محدود شد. تعداد کل جستجوهای حوزه علم اطلاعات ۱۰۶۶۴۱ بود که تحلیل بر روی آنها انجام گرفت.

**یافته‌ها:** تحلیل معنایی جستجوها، وجود رابطه‌های معنایی را بین جفت‌های اصطلاحات مبتنی بر اصطلاحنامه در جستجوهای کاربر نشان داد. نتایج نشان داد که کاربران برای رسیدن به پاسخ مطلوب خود در پایگاه گنج جستجوهای خود را براساس روابط معنایی اصلاح می‌کنند.

**نتیجه‌گیری:** نتایج این پژوهش می‌تواند در بهبود سیستم گنج برای پاسخگویی به کاربر و برای پیشنهاد برای جستجو به کاربر به کار رود. به گونه‌ای که با آگاهی از این مسئله که کاربران اغلب تاکتیک‌های خاصی را در اصلاح پرس‌وجوی خود به کار می‌برند و از روابط معنایی خاصی در آن استفاده می‌کنند می‌توان پیشنهادات مناسبی را به آنها ارائه داد. به طوری که با جستجوی یک مفهوم خاص در موارد مشابه، مفهوم دارای رابطه معنایی که اغلب مورد استفاده کاربران قرار می‌گیرد به آنها پیشنهاد شود.

**کلیدواژه‌ها:** رفتار اطلاعاتی کاربر، مطالعه کاربر، تحلیل لاگ، پایگاه اطلاعات گنج، روابط معنایی، هستی‌شناسی

تعارض منافع: گزارش نشده است.

منبع حمایت‌کننده: حامی مالی نداشته است.

**شیوه استناد به این مقاله**

**APA:** Karimi, E., Babaei, M., Hosseini Beheshti, M (2018). Analysis of user query refinement behavior based on semantic features: User log analysis of Ganj database (IranDoc). Human Information Interaction. 5(3);1-14. (Persian)

**Vancouver:** Karimi E, Babaei M, Hosseini Beheshti M. Analysis of user query refinement behavior based on semantic features: User log analysis of Ganj database (IranDoc). Human Information Interaction. 2018;5(3):1-14. (Persian)



انتشار مجله تعامل انسان و اطلاعات با حمایت مالی دانشگاه فوارزمی انجام می‌شود.

انتشار این مقاله به صورت دسترسی آزاد مطابق با CC BY-NC-SA 3.0 صورت گرفته است.

## Analysis of user query refinement behavior based on semantic features: User log analysis of Ganj database (IranDoc)

**\*Elmira Karimi:** PhD Student of Knowledge and Information Studies, Iranian Research Institute for Information Science and Technology, Tehran, Iran. (Corresponding author) [karim@students.irandoc.ac.ir](mailto:karim@students.irandoc.ac.ir)

**Mahmoud Babaei:** Assistant Professor of Knowledge and Information Studies, Iranian Research Institute for Information Science and Technology, Tehran, Iran

**Molookosadat Hosseini Beheshti:** Assistant Professor of Knowledge and Information Studies, Iranian Research Institute for Information Science and Technology, Tehran, Iran

Received: 09/08/2018

Accepted: 02/10/2018

### Abstract

**Background and Aim:** Information systems cannot be well designed or developed without a clear understanding of needs of users, manner of their information seeking and evaluating. This research has been designed to analyze the Ganj (Iranian research institute of science and technology database) users' query refinement behaviors via log analysis.

**Methods:** The method of this research is log analysis. In this research, the Ganj users' query refinement behaviors were analyzed. User's logs during 3 months between May and June 2016 were analyzed for semantic features. The study tool was a researcher-made checklist of semantic features. The total searches were 10 million records which were limited to information science domain. About 106641 records were selected for analysis.

**Results:** The semantic relationships (based on thesaurus relationships) were revealed between pair's terms in user's searches. The results showed that users refined their searches based on some semantic relationships.

**Conclusion:** The results of this research can be used to improve the Ganj results and for term suggestion for users, so that they be able to choose proper terms while there are several related terms.

**Keywords:** User information behavior, User study, Log analysis, Semantic relationships, Pntology

*Conflicts of Interest: None*

*Funding: None.*

### How to cite this article

**APA:** Karimi, E., Babaei, M., Hosseini Beheshti, M., (2018). Analysis of user query refinement behavior based on semantic features: User log analysis of Ganj database (IranDoc). Human Information Interaction. 5(3);1-14. (Persian)

**Vancouver:** Karimi E, Babaei M, Hosseini Beheshti M. Analysis of user query refinement behavior based on semantic features: User log analysis of Ganj database (IranDoc). Human Information Interaction. 2018;5(3):1-14. (Persian)



The journal of *Human Information Interaction* is supported by Kharazmi University, Tehran, Iran.  
This work is published under [CC BY-NC-SA 3.0](https://creativecommons.org/licenses/by-nc-sa/3.0/) licence.

## مقدمه

آن را تغییر دهد، و خواه از نتایج راضی باشد یا نه، آن را کاملاً از نو شروع کند یا به کلی متوقف سازد. هر چند نظام بازیابی اطلاعات از نظر صورت ظاهری زبان پرس و جو، روند بازخورد و فرمت‌های رکوردها تفاوت قابل ملاحظه‌ای دارند، ماهیت اطلاعاتی که از هر دو سو بین کاربر و نظام ردوبدل می‌شود، در نظام‌های مختلف تا حد زیادی مشابه است. (میدو و دیگران، ۱۳۹۰، ص ۲۶۴ و ۲۶۵). اغلب نظام‌های بازیابی اطلاعات تاکنون تفسیر و گرفتن بازخورد از کاربر را براساس کلیدواژه‌های مورد نظر وی و اسناد مرتبط به آن انجام می‌دادند. امروزه، برخی از نظام‌ها و به‌ویژه موتورهای جستجو، مانند گوگل، سعی در معنایی کردن این بازخوردها دارند. این کار در گوگل، براساس پروفایل‌های کاربر و نحوه‌ی جستجو و توجه به جستجوی پیشین کاربر و بسامد توجه سایر کاربران به سندی خاص در حوزه‌ی مورد جستجو انجام می‌گیرد.

از طرفی، سه نفر از پیشگامان در پیشرفت دیدگاه تعاملی در بازیابی اطلاعات، مدل‌های بازیابی اطلاعات کاملی پیشنهاد دادند: اینگورسن<sup>۴</sup> (۱۹۹۲، ۱۹۹۶) مدل شناختی بازیابی اطلاعات، بلکین<sup>۵</sup> (۱۹۸۴ و ۱۹۹۶) مدل بازیابی اطلاعات به عنوان پشتیبانی برای تعامل با اطلاعات و ساراسویک<sup>۶</sup> (۱۹۹۶، ۱۹۹۷) مدل تعاملی طبقه بندی شده. گرچه عناصر اصلی آنها متفاوت است، هر سه آنها در یک چیز مشترک هستند: کاربران دریافت کنندگان منفعل متون بازیابی شده نیستند بلکه جستجوگران فعال اطلاعات هستند. (هوانگ و ایتیمیدیس<sup>۷</sup>، ۲۰۰۹). براساس مدل‌های تعاملی بازیابی اطلاعات، فرض شده که اصلاح پرس و جو بین کاربر و سیستم تعامل ایجاد می‌کند. اصلاح پرس و جو نیز تعامل بین سطوح بیرونی و عمیق‌تری از تعامل با کاربر است. با وجود این مسئله، اغلب فرمول‌بندی مجدد پرس و جو را آسان‌تر از فرمول‌بندی اولیه پرس و جو نمی‌دانند. کاربران کلیدواژه‌هایی که می‌شناسند را در پرس و جو اولیه وارد می‌کنند. اگر پرس و جو اولیه نتایج جستجوی مورد انتظار را بازیابی نکرد، کاربران باید دومین کلیدواژه‌های خود را وارد کنند. به این دلیل که کاربران در بیشتر مواقع در ترکیب اطلاعات از اسناد قبلی بازیابی شده به پرس و جوهایشان تجربه‌ی خوبی ندارند، فرایند فرمول‌بندی مجدد می‌تواند خسته کننده‌تر و پیچیده‌تر از فرمول‌بندی اولیه باشد. (فرنچ، براون و کیم<sup>۸</sup>، ۱۹۹۷). انتظار می‌رود همان‌طور که کاربران در فرمول‌بندی پرس و جو از تاکتیک‌ها و تکنیک‌های متفاوتی استفاده می‌کنند، در فرمول‌بندی مجدد نیز تاکتیک‌ها را به کار

واژه‌ی پرس و جو برای بیان نیاز از سوی کاربر یا واسطه‌ی اطلاعاتی به نظام بازیابی اطلاعات، به کار می‌رود. در واقع کاربران به صورت مداوم پرسش جستجوی قبلی خود را به امید بازیابی بهتر اصلاح می‌کنند. این اصلاحات، فرمول بندی مجدد<sup>۱</sup> یا اصلاح پرس و جو<sup>۲</sup> نامیده می‌شوند. (Huang and Eftimiadis, 2009). در واقع پرس و جو، عبارت<sup>۳</sup> یا مجموعه‌ای از عبارات است که از سوی کاربر به نظام بازیابی اطلاعات ارائه می‌شود تا اطلاعاتی بازیابی شود. (میدو و دیگران، ۱۳۹۰، ص ۲۶۴). تمرکز این تعریف بر «عبارت» نشان‌دهنده این موضوع است که نیاز ممکن است به درستی از سوی کاربر به درستی شناخته شود یا برعکس، این تشخیص از سوی او انجام نگیرد و علاوه بر این، ممکن است نیاز شناخته شده در قالب کلمات و کلیدواژه‌های نادرست و نامناسب بیان شود.

«بیان نیاز اطلاعاتی» به صورت‌های مختلفی انجام می‌گیرد. یکی از این صورت‌ها، می‌تواند بیان نیاز برای یک واسطه اطلاعاتی یا برای شخصی باشد که حرفه‌ی وی کمک به انجام جستجوهای اطلاعاتی است. حالت دوم، می‌تواند شکلی از نیاز باشد که پس از مشورت با واسطه اطلاعاتی بیان می‌شود. این دو نوع بیان، چون با مشورت شخص حرفه‌ای انجام می‌شود، تا حد زیادی، اشتباهات را در بیان نیاز کاهش می‌دهد. نوع سوم بیان، که تنها نوع بیانی است که غالباً در جستجوی کاربر نهایی لزوم پیدا می‌کند، پرس و جویی است که به نظام بازیابی اطلاعات ارائه می‌شود. این آخرین نوع بیان نیاز اطلاعاتی، بیانی است که نظام بازیابی اطلاعات آن را تفسیر می‌کند. در اغلب نظام‌های بازیابی اطلاعات امروزه نیاز از سوی خود کاربر ارائه می‌شود. از آن‌جا که بیان به صورت مستقیم از سوی کاربر انجام می‌شود، ممکن است به این دلیل که آشنایی با نظام و خطمشی‌های آن در تمام کاربران یکسان نیست و از طرفی به این دلیل که ممکن است کاربر در تشخیص و بیان نیاز خود اشتباه کند، نیاز به تفسیر درست و گرفتن بازخورد از کاربر در مراحل مختلف بازیابی اطلاعات وجود داشته باشد.

در این مورد، تقریباً کلیه نظام‌های بازیابی اطلاعات، اطلاعات خاصی درباره‌ی چگونگی پیشرفت جستجو مانند تعداد رکوردهای یافته شده براساس ملاحظات خاص را به عنوان بازخورد به کاربر برمی‌گردانند. بر اساس این اطلاعات، کاربر تصمیم می‌گیرد که اگر می‌خواهد پرس و جویی مطرح کند، این پرس و جو چه باشد، آیا کاوش را به همان صورتی که از ابتدا تصور کرده بود ادامه دهد یا

<sup>4</sup> Ingwersen

<sup>5</sup> Belkin

<sup>6</sup> Saracevic

<sup>7</sup> Huang and Eftimiadis

<sup>8</sup> French, Brown, and Kim

<sup>1</sup> Reformulation

<sup>2</sup> Query refinement

<sup>3</sup> statement

برند.

پژوهش‌ها زیادی الگوهای فرمول‌بندی مجدد پرس و جو را در وب بررسی کرده‌اند (مثل بروزا و دنیس<sup>۱</sup>، ۱۹۹۷؛ لاو و هورویتز<sup>۲</sup>، ۱۹۹۹؛ اسپینک، جانسن و اوزمولتو<sup>۳</sup>، ۲۰۰۰). با این حال، هیچ یک از آنها، به بررسی فرمول‌بندی مجدد فراتر از سطح تحلیل توصیفی نمی‌پردازند. از سوی دیگر، تحقیقات کمی تا امروز تاکتیک‌ها و استراتژی‌های جستجو در وب را بررسی کرده‌اند. بیشتر مطالعات پرس‌وجوهای وب، از تحلیل لاگ ثبت شده برای موتورهای جستجو یا وبسایت‌ها استفاده کرده‌اند. (اسپینک و جانسن، ۲۰۰۴؛ اسپینک، وولفرم<sup>۴</sup> و ساراسویک، ۲۰۰۲؛ اسپینک، جانسن و ساراسویک، ۲۰۰۱). پژوهش‌هایی که این مسئله را در پایگاه‌های علمی بررسی کرده‌اند، ناچیز هستند. تعداد بسیار زیادی از این تحلیل‌ها، تحلیل‌های کمی را انجام می‌دهند و تحلیل‌های معنایی به جز در موارد نادر، مورد توجه پژوهش‌گران تحلیل لاگ قرار نگرفته‌اند. (بروزا و دنیس، ۱۹۹۷؛ کائو<sup>۵</sup> و هورویتز، ۱۹۹۹)

نظام‌های بازبازی اطلاعات بیشتر اوقات، عبارت جستجوی کاربران را به عنوان نقطه شروع کشف نیاز اطلاعاتی وی در نظر می‌گیرند، اما بسیاری اوقات کاربران بدون توجه به روابط خاص میان اصطلاحات، آنها را به صورت جداگانه در یک عبارت جستجو قرار می‌دهند. با استفاده از هستی‌شناسی، نظام می‌تواند اطلاعات دقیقی را که مورد نیاز کاربر است درک و با رفع ابهام از پرسش، آن را به صورت دقیق بیان کند. (میرزاییگی، ۱۳۹۰) علاوه‌براین، بسط جستجو که از فرایندهای مهم و گاهی مفقوده نظام‌های بازبازی کنونی است، از طریق هستی‌شناسی‌ها می‌تواند به خوبی صورت گیرد. در این رویکرد اصطلاحاتی که به عنوان پرسش وارد نظام می‌شوند با مفاهیم مرتبط جابه‌جا و به کاربر پیشنهاد می‌شوند (آندریو<sup>۶</sup>، ۲۰۰۵؛ به نقل از میرزاییگی، ۱۳۹۰). در حال حاضر، ابزارهای موجود تنها از روابط تاکسونومیک<sup>۷</sup> بین اصطلاحات استفاده می‌کنند و معنای مفاهیم دامنه‌ای را استفاده نمی‌کنند و نتایج آنها ساختار ضعیف تاکسونومی دارد و فاقد ارتباطات غیر تاکسونومیک است. (بن مصطفا<sup>۸</sup> و دیگران، ۲۰۱۳).

در واقع کاربران در جستجوهای خود، برای گرفتن پاسخ مرتبط به نیاز خود، بارها و بارها پرس و جوی خود را اصلاح می‌کنند. اما این که پرس‌وجوهای اصلاح شده آنها از نظر معنایی به آن چه ابتدا

مورد جستجو قرار داده‌اند مرتبط است یا خیر مسئله مبهمی است که این پژوهش به دنبال یافتن پاسخی برای آن است. پایگاه اطلاعات علمی ایران (گنج) به عنوان یک نظام بازبازی اطلاعات حاوی منابع اطلاعاتی با ارزش (پایان‌نامه‌ها، مقالات و طرح‌های پژوهشی) با مشکلاتی در بازبازی اطلاعات مواجه است. برخی از این مشکلات مربوط به مسائل لغات و مفاهیم مورد استفاده است. این پایگاه دارای اصطلاحنامه‌هایی است که برای بازبازی اطلاعات استفاده نمی‌شوند بلکه تنها کاربران قادر به استفاده از آنها در جستجوهای خود هستند. نتایج پژوهش‌هایی همچون محمدی و دخت عصمتی (۱۳۹۰) نشان داد که اصطلاحنامه‌های پژوهشگاه علوم و فناوری اطلاعات ساختار مطلوبی دارند اما برای کاربردپذیری مبتنی بر وب مناسب نیستند و قابلیت‌های آنها برای جستجو و دسترسی به منابع اطلاعاتی در مقایسه با اصطلاحنامه‌های معادل خارجی بسیار اندک است و روابط کاربری آنها برای برقراری ارتباط با کاربر اصطلاحنامه بسیار ضعیف است. در واقع مشکل از سازمان‌دهی اطلاعات نیست بلکه این مجموعه روابط محدود بین اصطلاحات هستند که مشکل‌ساز شده‌اند و در نتیجه زمینه‌ی ظهور هستی‌شناسی‌ها به صورت مجموعه مقالات یا اصطلاحنامه‌های تکمیل‌شده یا گسترش‌یافته را فراهم کرده‌اند. برای بهبود پایگاه گنج در استفاده از هستی‌شناسی نیاز است که رفتار اطلاعاتی کاربران آن براساس ویژگی‌های معنایی که می‌تواند به کار رود بررسی شود تا بتوان طبق آن چه گفته شد با استفاده از آن به بهبود سیستم کمک نمود.

با توجه به آن چه گفته شد و این که مطالعات مختلفی در زمینه بازبازی اطلاعات تأیید کردند که موتورهای بازبازی اطلاعات مبتنی بر هستی‌شناسی مسئله بافت و اصلاح پرس و جو را با استفاده از هستی‌شناسی دامنه‌ای حل می‌کنند (بولگالا<sup>۹</sup> و دیگران، ۲۰۰۷؛ بن مصطفا و دیگران، ۲۰۱۱؛ بن مصطفا و دیگران، ۲۰۱۲). این پژوهش به دنبال بررسی رفتار اصلاح پرس‌وجوی کاربران پایگاه اطلاعاتی گنج برای استفاده بیشتر در کاربرد هستی‌شناسی برای پیشنهاد اصلاح پرس‌وجو است. در واقع، این پژوهش تلاش کرده است با تحلیل لاگ کاربران براساس ویژگی‌های معنایی و هستی‌شناسی، رفتار کاربران پایگاه اطلاعات گنج را در این حوزه به تصویر بکشد.

#### اهداف پژوهش

هدف اصلی این پژوهش، شناخت رفتار جستجوی کاربران پایگاه اطلاعات گنج از نظر استفاده از ویژگی‌های معنایی در پرس‌وجو است. اهداف فرعی آن شامل موارد زیر است:

<sup>1</sup> Bruza and Dennis

<sup>2</sup> Lau and Horvitz

<sup>3</sup> Spink, Jansen, Ozmultu

<sup>4</sup> Wolfram

<sup>5</sup> Kau

<sup>6</sup> Andreou

<sup>7</sup> Taxonomic relationships

<sup>8</sup> Ben Mustapha

<sup>9</sup> Bollegala

لاگ‌های داگ‌پایل<sup>۸</sup> (جانسن و دیگران، ۲۰۰۹)، تومبا<sup>۹</sup> (کاستا و سکو<sup>۱۰</sup>، ۲۰۰۸)، AOL (هوانگ و افتمیادیس، ۲۰۰۹)، فست<sup>۱۱</sup> (اوزموتلو، ۲۰۰۹) و یاهوی آمریکا و یاهوی انگلیس (بولدی<sup>۱۲</sup> و دیگران، ۲۰۰۹) را بررسی کرده‌اند. جونز و فین<sup>۱۳</sup> (۲۰۰۳) نرم افزار تحلیل لاگ بخش کاربر را برای جمع آوری لاگ‌های گروهی از کاربران در دسترس انواع موتورهای جستجو استفاده کرده‌اند. هی<sup>۱۴</sup> و دیگران (۲۰۰۲) لاگ‌های جستجو را از اینترنت بررسی کردند. پروزا و دنیس (۱۹۹۷) از لاگ‌های نمونه اولیه پی بی اوایل<sup>۱۵</sup> که دسترسی به ۷ پایگاه داده شامل انواع مختلف اسناد را می‌داد استفاده کردند. در افتمیادیس (۲۰۰۰) اصلاحات پرس‌وجو از لاگ‌های جستجو مطالعه نشد بلکه در یک محیط آزمایشی شامل کاربران اجراکننده جستجو انجام داد. (هولینک، تسیکریکا، و وریس<sup>۱۶</sup>، ۲۰۱۰).

تعدادی از پژوهش‌های مربوط به تحلیل لاگ در زمینه‌ی اصلاحات پرس‌وجوها، پرس‌وجوهای مشاهده شده را به روش دستی (پروزا و دنیس، ۱۹۹۷؛ لو و هورویتز، ۱۹۹۹؛ جورگسن<sup>۱۷</sup> و جورگسن، ۲۰۰۵؛ ری و شی، ۲۰۰۶؛ جونز و فین، ۲۰۰۳؛ افتمیادیس، ۲۰۰۰) و تعدادی دیگری از این پژوهش‌ها، پرس‌وجوها را به روش خودکار (هی و دیگران، ۲۰۰۲؛ ویتل و دیگران، ۲۰۰۷؛ جانسن و دیگران، ۲۰۰۹؛ بولدی و دیگران، ۲۰۰۹؛ کاستا و سکو، ۲۰۰۸؛ بوزن و دیگران، ۲۰۰۷؛ هوانگ و افتمیادیس، ۲۰۰۹؛ اوزموتلو، ۲۰۰۹) طبقه‌بندی کرده‌اند. بعضی از مطالعات به روش دستی بررسی کردند که کاربران در مقایسه با پرس‌وجوی قبلی خود، چه اصلاحاتی را اضافه، حذف یا جایگزین کرده‌اند (جانسن و دیگران، ۲۰۰۹؛ وایتل و دیگران، ۲۰۰۷؛ جورگسن و جورگسن، ۲۰۰۵؛ هی و دیگران، ۲۰۰۲؛ پروزا و دنیس، ۱۹۹۷؛ جونز و فین، ۲۰۰۳؛ کوستا و سکو، ۲۰۰۸). در این نوع طبقه‌بندی اصلاح پرس‌وجوها، وقتی اصطلاحی اضافه می‌شود اصلاح در طبقه‌ی خاص کردن (مثل پرس‌وجوی بکهام به پرس‌وجوی بکهام میلان) قرار می‌گیرد، وقتی اصطلاحات حذف می‌شوند در دسته‌ی عام کردن دسته بندی می‌شود (مثل از بکهام میلان به بکهام) و وقتی اصطلاحات جایگزین می‌شوند به شکل فرمول‌بندی مجدد دسته‌بندی می‌شوند (مثل از بکهام میلان به

• شناخت وضعیت هم‌رخدادی جفت‌های اصطلاح دارای ارتباط معنایی (براساس اصطلاحنامه) در جستجوهای کاربران پایگاه اطلاعات گنج

• شناخت وضعیت حرکت و تاکتیک‌های حامی روابط معنایی (ارائه شده در گارسیا و سیسیلیا، ۲۰۰۳) در جستجوهای کاربران پایگاه اطلاعات گنج

• تحلیل اصلاحات جستجو براساس روابط معنایی (براساس اصطلاحنامه) در جستجوی کاربران پایگاه اطلاعات گنج

پرسش اصلی این پژوهش، این است که، رفتار کاربران پایگاه اطلاعات گنج از نظر استفاده از روابط معنایی در جستجوها چگونه است؟ پرسش‌های فرعی آن شامل موارد زیر است:

- وضعیت هم‌رخدادی جفت‌های اصطلاح دارای ارتباط معنایی (براساس اصطلاحنامه) در میان جستجوهای کاربران گنج چگونه است؟

- وضعیت استفاده از حرکت و تاکتیک‌های روابط معنایی (ارائه شده در گارسیا و سیسیلیا، ۲۰۰۳) در جستجوی کاربران پایگاه گنج چگونه است؟

- وضعیت اصلاحات جستجوی کاربران گنج براساس روابط معنایی (براساس اصطلاحنامه) چگونه است؟

### پیشینه پژوهش

تحلیل لاگ نظام‌های اطلاعاتی، روشی است که برای ارزیابی نظام‌ها و مطالعه کاربران از دهه ۱۹۶۰ مورد استفاده بوده است. (پیترز<sup>۱</sup>، ۱۹۹۳؛ به نقل از خسروی و جمالی مهموئی، ۱۳۹۳) بیشتر مطالعات در این حوزه برای سنجش ویژگی‌های عمومی جستجو و ویژگی‌هایی به جز ویژگی‌های معنایی انجام شده است. در واقع مطالعه‌ای با هدف مشابه با پژوهش حاضر تا کنون انجام نگرفته است. افزون بر این، بیشتر مطالعات بر روی لاگ موتورهای جستجو انجام گرفته است و کمتر لاگ یک پایگاه علمی را به عنوان منبع، برای تحلیل استفاده کرده‌اند. در اینجا مروری بر مطالعاتی که در حوزه‌ی اصلاح پرس و جو و دسته‌بندی انواع این اصلاحات انجام شده، ارائه شده است.

در بسیاری از پژوهش‌های اصلاح پرس‌وجو، از تحلیل لاگ‌های جستجوی موتورهای کاوش، استفاده شده است. بیشتر لاگ‌های جستجوی مطالعه شده موتور جستجوی اکسایت<sup>۳</sup> است (ری و شی<sup>۴</sup>، ۲۰۰۶؛ ویتل<sup>۵</sup> و دیگران، ۲۰۰۷؛ لو و هورویتز، ۱۹۹۹؛ بوزن<sup>۶</sup> و دیگران، ۲۰۰۷؛ اوزموتلو<sup>۷</sup>، ۲۰۰۹). مطالعات دیگر

<sup>7</sup> Ozmutlu

<sup>8</sup> Dogpile

<sup>9</sup> Tumba

<sup>10</sup> Costa and Seco

<sup>11</sup> Fast

<sup>12</sup> Boldi

<sup>13</sup> Jones and Fain

<sup>14</sup> He

<sup>15</sup> BabyOIL

<sup>16</sup> Hollink, Tsikrika, & Vries

<sup>17</sup> Joorgensen

<sup>1</sup> García & Sicilia

<sup>2</sup> Peters

<sup>3</sup> Excite

<sup>4</sup> Rieh and Xie

<sup>5</sup> Whittle

<sup>6</sup> Bozzon

آنها به این نتیجه رسیدند که خاص کردن معمولا با عام کردن و عام کردن با خاص کردن دنبال می‌شود. جانس و دیگران (۲۰۰۹) نتیجه‌ی پژوهش بولدی و دیگران را تأیید کردند.

علاوه بر سه نوع اصلی اصلاح، گاهی دیگر انواع اصلاحات نیز بررسی شده‌اند. به عنوان مثال، تعدادی از پژوهش‌ها، تعداد دفعاتی که کاربران پرس‌وجوی خود را اصلاح می‌کنند را بررسی کرده‌اند. (ری و شی، ۲۰۰۶؛ بروزا و دنیس، ۱۹۹۷؛ جونز و فین، ۲۰۰۳). این بررسی‌ها اصلاحات از شکل‌های مفرد به جمع یا برعکس را نیز در نظر می‌گیرند. در نمونه‌ی دیگر، بوزن و دیگران (۲۰۰۷) تفاوت‌های بین دسته‌های لغوی (تگ‌های Part-of-speech) از پرس‌وجوها و پرس‌وجوهای اصلاح شده بررسی کردند. آنها دریافتند که بیشتر پرس‌وجوهای اسمی به پرس‌وجوهای اسمی و پرس‌وجوهای فعل - اسم به پرس‌وجوهای فعل - اسم اصلاح می‌شوند. (هولینک، تسیکریکا، و وریس، ۲۰۱۰).

در این بین، همانطور که پیش از این اشاره شد، تعداد پژوهش‌هایی که به مطالعه لاگ کاربران پایگاه‌های علمی پرداختند بسیار اندک است و نمونه‌ای از این پژوهش‌ها شامل این موارد هستند: پارک و لی<sup>۶</sup> (۲۰۱۳)، نیکلاس<sup>۷</sup> و همکاران (۲۰۰۸) و یی<sup>۸</sup> و همکاران (۲۰۰۶) که هر یک به نوعی در پایگاه‌های اطلاعات علمی به بررسی لاگ کاربران پرداخته‌اند.

ری و شی (۲۰۰۶) لاگ‌های کاربران موتور جستجوی اکسایت را براساس آی دی کاربرانی که چندین بازدید از موتور جستجو داشتند بررسی کردند. واحد تحلیل لاگ در مطالعه‌ی آنها، جلسات جستجویی بود که کاربران شش یا تعداد بیشتری پرس‌وجوی منحصر به فرد وارد کرده‌اند. محققان این پژوهش، داده‌ها را در طی دو مرحله‌ی زمانی نمونه‌گیری کردند: مجموعه داده‌ی اولیه ۱۸۳ جلسه جستجوی تحلیل شده بود (Rieh and Xie, 2001) و مجموعه داده‌ی بعدی شامل ۱۳۰ جلسه جستجوی دیگر برای تحلیل بیشتر بود. مجموعه داده نهایی برای این مطالعه مورد بررسی قرار گرفت ۳۱۳ جلسه جستجو بود. سه دسته اصلی فرمول‌بندی مجدد از تحلیل داده‌ها مشخص شد، محتوا، قالب و منبع. اصلاح محتوا به نمونه‌هایی اشاره داشت که کاربران در معنای پرس و جو، اصلاح انجام داده بودند. اصلاح قالب شامل مواردی بود که تغییر را بدون جایگزینی معنای پرس‌وجوها با اضافه کردن عملگرها، چیدن دوباره ترتیب چند اصطلاح، املائی مجدد اصطلاحات پرس و جو، فاصله‌گذاری مجدد اصطلاحات و غیره، انجام داده بودند. اصلاح منبع در مورد نمونه‌هایی بود که

بکهم مادرید). بعضی از مطالعات با روش دستی نه تنها اصطلاحات را مورد بررسی قرار دادند بلکه اصلاحات را براساس معنای پرس‌وجوها دسته‌بندی کردند (ری و شی، ۲۰۰۶؛ جورگنس و جورگنس، ۲۰۰۵؛ لو و هورویتز، ۱۹۹۹؛ جونز و فین، ۲۰۰۳؛ افتمیادیس، ۲۰۰۰). هوانگ و افتمیادیس (۲۰۰۹) برای طبقه‌بندی تعاریف مبتنی بر اصطلاح را برای خاص کردن و عام کردن استفاده کردند، اما برای در مورد دسته‌ی فرمول‌بندی مجدد که پیش از این در پژوهش‌ها به کار رفته بود، به اصطلاحات جایگزین اشاره کرده‌اند که در وردنت تعاریف معنایی برای آنها ذکر شده است، مخصوصا زمانی که اصطلاح اضافه شده مترادف، هایپونیم<sup>۱</sup>، هایپرینیم<sup>۲</sup>، مرونیم<sup>۳</sup> یا هلونیم<sup>۴</sup> هستند. بولدی و دیگران (۲۰۰۹) دیدگاه حد واسط دیگری ارائه کردند که در آن هدف دسته‌بندی معنایی اصلاح پرس‌وجو به خاص کردن، عام کردن و فرمول‌بندی مجدد با جستجوی همپوشانی در اصطلاحات پرس‌وجو، فاصله زمانی بین پرس‌وجوها و ویژگی‌های جلسه کاربر به طور کلی است. همانطور که مشخص است، در بیشتر این مطالعات، انواع اصلاح پرس‌وجوی شناسایی شده، خاص کردن، کلی سازی و فرمول‌بندی مجدد بوده است.

آنیک<sup>۵</sup>، نمونه‌ی تصادفی از ۱۰۰ فرمول‌بندی مجدد را به روش دستی به ۱۱ دسته تقسیم کرد. لو و هورویتز، جانسن و دیگران و هی و دیگران، همان دسته‌های فرمول‌بندی مجدد گرفته شده از زبان شناسی را استفاده کردند. تیوان و دیگران (۲۰۰۷) طبقه‌بندی با جستجوی لاگ‌های پرس و جو ایجاد کردند و الگوریتم‌هایی برای شناسایی زیرمجموعه استراتژی‌های فرمول‌بندی مجدد به کار بردند. ویتل و دیگران، تعدادی از استراتژی‌های فرمول‌بندی مجدد را با استفاده از شبکه گرافیکی مدلسازی کردند. بروزا و دنیس، به صورت دستی ۱۰۴۰ پرس و جو را به طبقه‌بندی خودشان دسته‌بندی کردند. جو و دیگران (۲۰۰۸) نیز طبقه بندی کوچکی ایجاد کردند و از مدل زمینه تصادفی مشروط برای پیش بینی اصلاح پرس و جوها استفاده کردند. ری و شی دسته‌های فرمول بندی مفهومی مثل محتوا، فرمت و منابع. ایجاد کردند

ترتیب اصلاحات پرس‌وجوها در ویتل و دیگران (۲۰۰۷)، جانسن و دیگران (۲۰۰۹) و بولدی و دیگران (۲۰۰۹) بررسی شد. ویتل و دیگران (۲۰۰۷) دریافتند که کاربران تمایل به تکرار همان نوع اصلاح دارند (مثلا خاص کردن اغلب با خاص سازی دیگری دنبال می‌شود). این موضوع از سوی بولدی و دیگران (۲۰۰۹) تأیید نشد،

<sup>1</sup> hyponym

<sup>2</sup> heypemym

<sup>3</sup> Meronym

<sup>4</sup> helonym

<sup>5</sup> Anick

<sup>6</sup> Park and Lee

<sup>7</sup> Nicholas

<sup>8</sup> Yi

کاربر را در بردارند (حریری و مهربان، ۱۳۹۲، ۲۳۸). در حوزه‌ی مطالعات جستجوی وب، تحلیل لاگ وب یکی از ابزارهای تحقیق اولیه‌ای است که تعاملات کاربران را با سیستم جستجوی وب می‌گیرد. گرچه این روش معمولاً در کتابخانه و سیستم‌های بازیابی اطلاعات آزمایشی استفاده می‌شود، تبدیل به ابزار منطقی و غیرقابل نفوذ تعامل سیستم کاربر از تعداد زیادی از کاربران وب شده است (اسپینک و جانسن، ۲۰۰۴). لاگ‌های سرور وب معمولاً شامل آدرس آی پی (یا نام‌های دامنه) جستجوگران، تاریخ و زمان جستجو، جستجوی انجام شده و نتایج جستجو است. تحلیل لاگ، در ایجاد آمارهای کاربرد وب سایت به طور کلی تأیید می‌شود و در آینده می‌تواند استنتاجی در مورد طراحی وب سایت همچنین بررسی فرضیه‌هایی در مورد تأثیر متغیرهای طراحی مختلف بر اساس داده‌های رفتاری کاربر باشد (بورتون و والتر، ۲۰۰۱). این پژوهش از تحلیل لاگ استفاده کرده است زیرا این روش، مجموعه داده‌ای که به صورت طبیعی رخ داده و دامنه گسترده‌ای دارد از اصلاح پرس و جوا فراهم می‌کند. در واقع، در این روش، کلیه فعالیت‌های کاربران مورد پیگیری قرار می‌گیرد تا الگوی استفاده از منابع اطلاعاتی کشف و ترسیم شود. (رجبعلی بگلو، ۱۳۸۷).

در این پژوهش، لاگ کاربران پایگاه اطلاعات گنج در بازه زمانی سه ماهه (از اردیبهشت تا مرداد ۱۳۹۵) تحلیل شد. تعداد کل جستجوهای دریافتی ده میلیون رکورد است که آنها به جستجوهای حوزه علم اطلاعات (با محدودسازی با اصطلاحات اصطلاحنامه Inspec در این حوزه) محدود شدند. تعداد کل جستجوهای حوزه اطلاع رسانی ۱۰۶۶۴۱ بود که تحلیل بر روی آنها انجام گرفت. در این پژوهش، واحد اصلی تحلیل، اصلاح پرس و جو است. به این دلیل که تمرکز این مطالعه بر فرمول‌بندی مجدد پرس و جو است، داده‌های منابع مورد بررسی قرار گرفته از سوی کاربر، در تحلیل شامل نمی‌شود. برای تعریف جلسات پرس‌وجو، پژوهشگران ویژگی‌های مختلفی را در نظر می‌گیرند. جانسن و دیگران یک جلسه را یک سری از تعاملات کاربر برای رسیدن به نیاز اطلاعاتی خاص تعریف می‌کنند. سیلوراشتان<sup>۳</sup> و دیگران (۱۹۹۹) معتقدند که یک جلسه به معنای گرفتن تلاش کاربر برای در نظر گرفتن یک پرس‌وجوی خاص است که با تعدادی از پرس‌وجوهای فرمول‌بندی شده دنبال می‌شود. هی و دیگران و اوزموتلو از زمان و کلمات مشترک برای تعیین پایان جلسه استفاده می‌کنند. علاوه بر این، آرلیت<sup>۴</sup> (۲۰۰۰) محدوده‌های جلسات را با محدوده زمان محاسبه شده در نظر می‌گیرد.

کاربران تمایل به تغییر در نوع منبع اطلاعاتی داشتند (مثل اخبار، تصویر، و فایل‌های موسیقی). نتایج در دو بخش گزارش شده است، بخش اول، عناصر فرمول‌بندی مجدد را ارائه کرده است، و بخش دوم ترتیب چندین فرمول‌بندی مجدد را نشان داده است. پژوهشگران در این پژوهش، مدلی مبتنی بر مدل طبقه‌بندی ساراسویک ایجاد کردند. فرمول‌بندی پرس‌وجو محصول تعامل کاربر با سطوح شناختی، تأثیر و موقعیت است. مهم‌تر از همه، مسائل کاربران در فرایند بازیابی اطلاعات می‌تواند با بازنمون در فرایند فرمول‌بندی مجدد پرس و جو گسترش یابد. (مثل عناصر فرمول‌بندی مجدد، الگوهای فرمول‌بندی مجدد پرس و جو، بسامد فرمول‌بندی مجدد و غیره). نتایج این مطالعه نشان داد که ایجاد ابزارهای جستجویی که رفتارهای فرمول‌بندی مجدد پرس‌وجوی پیچیده را که چندین بار در فرایند تعامل بازیابی اطلاعات اتفاق می‌افتد، مهم است.

در ایران مطالعه‌ی چندانی در این زمینه انجام نشده است. مطالعه اخیر خسروی و جمالی مهموثی (۱۳۹۳) که جدیدترین مطالعه در این حوزه‌ی تحلیل لاگ در پایگاه علمی است به بررسی ویژگی‌های لاگ کاربران گنج پرداخته است و روند زمانی استفاده از پایگاه، پراکندگی جغرافیایی و دامنه اینترنتی، موضوعات مورد جستجوی کاربران و تعداد و موارد جستجو، طول عبارات مورد جستجو و نتایج حاصل از جستجو را به تصویر کشیده است. حریری و مهربان (۱۳۹۲) نیز در پژوهشی به تجزیه و تحلیل پرسش‌های جستجو و پیگیری مراحل مختلف جستجوهای انجام شده توسط کاربران پرداختند. آنها در پژوهش خود پرسش‌ها و راهبردهای جستجوی مورد استفاده کاربر، طول پرسش‌ها و تعداد آنها، تعداد پرسش‌ها براساس نوع منبع و کارایی پرسش‌ها را با توجه به تعداد بازدید، نرخ پرسش و طول زمان بازدید از هر صفحه را مورد بررسی قرار دادند.

با توجه به مروری که بر پیشینه‌ها انجام شد و بی‌توجهی به حوزه اصلاح پرس و جو براساس روابط معنایی به طور کلی در تمامی دنیا و به طور خاص در ایران، نیاز به این پژوهش برای استفاده از آن برای بهبود نظام‌های بازیابی اطلاعات احساس می‌شود.

## روش پژوهش

روش این پژوهش، تحلیل لاگ است که یکی از روش‌های تجزیه و تحلیل پایگاه‌های اطلاعاتی به شمار می‌رود. فایل‌های تراکنشی<sup>۱</sup> فایل‌هایی هستند که در سرور سایت یا پایگاه اطلاعاتی ذخیره می‌شوند و اطلاعات استفاده از آن سایت یا پایگاه اطلاعاتی توسط

<sup>۲</sup> Burton and Walther

<sup>۳</sup> Silverstein

<sup>۴</sup> Arlitt

<sup>۱</sup> Log analysis

جدول ۱- وضعیت هم‌رخدادی واژگان دارای ارتباط معنایی اصطلاحنامه

Inspec در میان کاربران پایگاه اطلاعات گنج		
مفهوم ۱	مفهوم ۲	تعداد هم‌رخدادی
اطلاعات	داده	۱۵۱
اشاعه اطلاعات	اطلاعات	۱۱۲
تبادل داده	داده	۲۸
متخصص	اطلاعات	۳
اطلاعات	تبادل اطلاعات	۱۰۶
مجموعه داده	داده	۳
خدمات اطلاعات	اطلاعات	۹۰
ارائه اطلاعات	اطلاعات	۶

برای بررسی این بخش از تحلیل لاگ، اصطلاحات مربوط به حوزه علم اطلاعات که در اصطلاحنامه Inspec فارسی با یکدیگر در ارتباط بودند انتخاب شدند و هم‌رخدادی آنها در جستجوهای کاربران بررسی شد. از میان جفت‌های اصطلاحات مورد بررسی، تعداد کمی از آنها که در جدول آمده است با یکدیگر در جستجوهای کاربر هم‌رخداد بودند.

همان‌طور که در جدول ۱ آمده است، بیشترین هم‌رخدادی را واژه‌های اطلاعات و داده داشته‌اند که از نظر معنایی به یکدیگر مرتبط هستند و رابطه بین آنها براساس روابط تقویت شده اصطلاحنامه رابطه «پردازش» است. واژه‌های اطلاعات و اشاعه اطلاعات نیز به تعداد زیادی در جستجوهای کاربران با یکدیگر به کار رفته‌اند که رابطه بین آنها «با استفاده از ابزار» و «بزار ... است» تعریف شده است.

### - وضعیت استفاده از حرکت و تاکتیک‌های روابط معنایی (ارائه شده در گارسیا و سیسیلیا، ۲۰۰۳) در جستجوی کاربران پایگاه گنج چگونه است؟

بیتس (۱۹۹۰) اشاره می‌کند، گاهی کاربر ترجیح می‌دهد که فرایند جستجو را تاحدی، براساس استراتژی‌ها و تاکتیک‌های مشخصی که بتواند به طور مؤثر در رابط کاربر حمایت شود، کنترل کند. بیتس تاکتیک‌ها را به عنوان «یک یا تعداد انگشت‌شماری از حرکاتی که جستجو را پیش می‌برند» تعریف می‌کند و به عنوان مثال، آنها را شامل تغییر یک اصطلاح به اصطلاحات خاص‌تر یا اضافه کردن اصطلاحات مشابه می‌داند. گارسیا و سیسیلیا (۲۰۰۳) معتقدند که هستی‌شناسی‌ها (به عنوان ابزارهای معنایی) می‌توانند برای حمایت مستقیم تاکتیک‌ها استفاده شوند.

در جدول ۲ خلاصه‌ای از تاکتیک‌هایی اشاره شده در بیتس (۱۹۹۰) که می‌توانند به طور مؤثری از طریق هستی‌شناسی‌ها حمایت شود، آمده است. در جدول حروف بزرگ برای مشخص کردن اصطلاحات (A,B,C) نشانه زیرمجموعه برای نشان دادن توابع استفاده شده است. (گارسیا و سیسیلیا، ۲۰۰۳).

از این رو، با توجه به ویژگی‌های عنوان شده در پژوهش‌های پیشین، در این پژوهش ترکیبی از ویژگی‌های بهینه از پژوهش‌های پیشین، که کلمات مشترک و زمان هستند، استفاده شد.

علاوه براین، چن و دار<sup>۱</sup> (۱۹۹۰) مفهوم عملگرهای معنایی را ارائه کردند، که به عنوان «حرکت‌ها و فعالیت‌هایی است که محتوای پرس و جو را تغییر می‌دهد» (ص ۱۲۱)، بنابراین و براساس این تعریف، ما آن دسته از جلساتی را انتخاب کردیم که پرس‌وجو و محتوای آن تغییر کرده است.

اطلاعاتی که در لاگ‌های دریافت شده از کاربر ارائه شده بود شامل: شماره جستجو، عبارت جستجو، تاریخ جستجو، زمان جستجو، تعداد رکوردهای بازیابی شده و شماره آی پی کاربر بود.

در مرحله پس از انتخاب جستجوهای حوزه علم اطلاعات، لاگ‌ها مورد پاک‌سازی قرار گرفت و مواردی که در آن پرس‌وجوها مخدوش بود، حذف شد.

برای شناخت وضعیت هم‌رخدادی جفت‌های اصطلاح دارای ارتباط معنایی در جستجوهای کاربران، ابتدا اصطلاحات مربوط به حوزه‌ی علم اطلاعات در اصطلاحنامه فارسی Inspec براساس نظر متخصصان انتخاب شدند. طبق ارتباطی که این اصطلاحات در اصطلاحنامه با یکدیگر داشتند، هم‌رخدادی آنها در بین جستجوهای کاربران بررسی شد.

برای شناخت وضعیت حرکت و تاکتیک‌های حامی روابط معنایی (ارائه شده در گارسیا و سیسیلیا، ۲۰۰۳) در جستجوهای کاربران که با بررسی اولیه جستجوهای کاربران، دسته‌بندی براساس این تاکتیک‌ها انجام شد و سپس میزان استفاده آنها در جستجوهای کاربران بررسی شد.

برای تحلیل اصلاحات جستجو براساس روابط معنایی (براساس اصطلاحنامه) در جستجوی کاربران، در مرحله اول، روابط معنایی موجود در اصطلاحنامه Inspec فارسی تقویت شد و انواع این روابط شناسایی شد، سپس آن دسته از جستجوهای کاربران که از نظر محتوایی تغییر کرده بودند انتخاب شدند و با انطباق آنها با روابط تقویت شده، نوع روابطی که آنها در اصلاح پرس‌وجوی خود به کار می‌برند شناسایی شد.

### یافته‌ها

در این بخش پاسخ پرسش‌های پژوهش ارائه می‌شود. وضعیت هم‌رخدادی جفت‌های اصطلاح دارای ارتباط معنایی (براساس اصطلاحنامه) در میان جستجوهای کاربران گنج چگونه است؟

<sup>۱</sup> Chen and Dhar



در این بخش به هدف بررسی این موضوع که کدام یک از این تاکتیک‌ها بیشتر مورد استفاده کاربران گنچ قرار می‌گیرد، به تحلیل لاگ‌های کاربران پرداخته شد. به این منظور، کاربران، با شماره نشست‌هایشان تفکیک شده و مجموعه‌ی جستجوهای یک کاربر با یک شماره نشست خاص، به عنوان جستجوهای اصلاح شده کاربر در نظر گرفته شد و مورد مطالعه قرار گرفت. ۵۹۶ مورد از کل جستجوها، اصلاحی در جستجوی خود در نظر گرفته بودند. بعضی از افراد با یک شماره نشست چند نوع اصلاح را بر روی پرس و جوی خود انجام داده بودند. تعداد کل تاکتیک‌های مورد استفاده کاربران ۸۰۰ مورد بود که بعضی از کاربران چند تاکتیک را

در این بخش به هدف بررسی این موضوع که کدام یک از این تاکتیک‌ها بیشتر مورد استفاده کاربران گنچ قرار می‌گیرد، به تحلیل لاگ‌های کاربران پرداخته شد. به این منظور، کاربران، با شماره نشست‌هایشان تفکیک شده و مجموعه‌ی جستجوهای یک کاربر با یک شماره نشست خاص، به عنوان جستجوهای اصلاح شده کاربر در نظر گرفته شد و مورد مطالعه قرار گرفت. ۵۹۶ مورد از کل جستجوها، اصلاحی در جستجوی خود در نظر گرفته بودند. بعضی از افراد با یک شماره نشست چند نوع اصلاح را بر روی پرس و جوی خود انجام داده بودند. تعداد کل تاکتیک‌های مورد استفاده کاربران ۸۰۰ مورد بود که بعضی از کاربران چند تاکتیک را

جدول ۲- تاکتیک‌های جستجوی حمایت شونده از سوی هستی‌شناسی (García & Sicilia, 2003)

تاکتیک	توصیف بیهتس	پیاپی‌سازی مبتنی بر هستی‌شناسی	مثال
SUPER	حرکت سلسله‌مراتبی به اصطلاح اعم (بالا تر)	تغییر B با A، B، زیرمجموعه A	حرکت از گربه چشم آبی به گربه
SUB	حرکت سلسله‌مراتبی به اصطلاح اخص (فرعی)	تغییر A با B، B، زیرمجموعه A است	حرکت از گربه به گربه چشم آبی
RELATE	حرکت یک طرفه سلسله‌مراتبی به هم پایه	تغییر B با C جایی که مفهوم A وجود دارد به طوری که B زیرمجموعه A، C، زیرمجموعه A	حرکت از گربه چشم آبی به گربه براق از طریق گربه
CONTRARY	جستجوی اصطلاحی که صورت منطقی مخالف توصیف موردنظر است	تغییر B با C جایی که اشتراک B و C تهی است (حداقل)	حرکت از محصول نهایی به محصول اولیه (به عنوان دسته‌بندی مشروبات)
SELECT	تجزیه پرس‌وجوی جستجوی پیچیده به مشکلات فرعی و کار بر روی یک مشکل در یک زمان	تجزیه جستجوی مداوم، ویرایش بعضی از شاخه‌هایی که می‌توانند بعد اضافه شوند	وقتی که جستجو برای اطلاعاتی در مورد ابزارهای آزمایش نرم افزار انجام می‌شود، تجزیه فنون مختلف (واحد، یکپارچه و ...) وقتی جستجو برای اطلاعات در مورد بازرسی نرم‌افزار شامل نقد نرم افزار نیز می‌شود
PARALLEL	گسترش فرمول‌بندی جستجو (یا گسترده‌تر) با شامل کردن مترادف‌ها یا دیگر اصطلاحات برابر مفهومی	افزودن اصطلاح C وقتی که C مشابه به هر اصطلاحی که بخشی از جهت‌یابی جاری را تشکیل می‌دهد	-
PINPOINT	ایجاد فرمول‌بندی جستجوی دقیق با حداقل سازی (یا کاهش) تعداد اصطلاحات برابر، حفظ اصطلاحات توصیفی تر	مخالف PARALLEL	-
SPECIFY	جستجوی اصطلاحاتی که به اندازه اطلاعات مطلوب خاص هستند	اجازه اصلاح قدم به قدم با پیش بردن به سوی خاص تر شدن در سلسله مراتب تابعی، ترکیب چند اصطلاح با «و» شبیه معناها	-
EXHAUST	شامل بیشتر یا تمام عناصر پرس‌وجو در فرمول بندی جستجوی اولیه؛ برای افزودن یک یا چند عنصر پرس‌وجو به فرمول جستجوی آماده	اجازه به انتخاب چند اصطلاح و «یا»- شبیه معناها	-

جدول ۳- وضعیت استفاده کاربران پایگاه اطلاعات گنچ از تاکتیک‌های حامی هستی‌شناسی

	PARALLEL	PINPOINT	SELECT	CONTRARY	RELATE	SUB	SUPER	
تعداد	۷۲	۷	۳۴۱	۳	۳۳۴	۴۶	۹۷	
درصد	۱۲	۱/۱۷	۵۷	۰/۰۵	۵۶	۸	۱۶/۲۸	

تأثیر کتاب الکترونیکی بر یادگیری ۲۰۱۶/۱۳/۰۵ ۰۸:۲۰:۴۷ مورد نظر نمی‌رسد، تاکتیکی که برای گرفتن پاسخ اجرا می‌کند  
تأثیر کتاب الکترونیکی ۲۰۱۶/۱۳/۰۵ ۰۸:۲۳:۴۴ تجزیه موارد مورد جستجو به اجزای آن است.  
در این جستجو، پس از این که کاربر در جستجوی خود به پاسخ کتابخانه کودک ۲۰۱۶/۰۹/۰۵ ۱۳:۳۱:۲۳

جدول ۴- وضعیت استفاده از روابط معنایی در اصلاح جستجوهای کاربران پایگاه اطلاعات گنج

عنوان رابطه معنایی	جستجوی اولیه	جستجوی اصلاح شده
	مطالعه موردی	مصاحبه
	کتاب خوانی	کتابداری
	مدیریت دانش	مدیریت فناوری
	تبادل داده	تبادل اطلاعات
مرتبط بودن	کتابخانه عمومی	کتابخانه دیجیتال
	شبکه اجتماعی مرجع	شبکه اجتماعی مجازی
	اطلاعات	دانش
	آموزش الکترونیکی	آموزش از راه دور
	آموزش الکترونیکی	آموزش رایانه
	فهرست موضوعی	فهرست توصیفی
	کتابخانه دانشگاهی	کتابخانه تخصصی
	ارزیابی کتاب	بررسی محتوای کتاب
	کتابخانه عمومی	کتابخانه ملی
	رابط و رتبه بندی اطلاعات	بازیابی اطلاعات
	کتابخانه عمومی	کتابخانه رقمی
	کتابخانه دانشگاهی	کتابخانه عمومی
	کتابخانه مرکزی دانشگاه	کتابخانه عمومی
مترادف	شبکه های اجتماعی	شبکه های مجازی
	مالکیت معنوی	مالکیت فکری
	ارزیابی بیرونی	اعتباربخشی
	شبکه اجتماعی	شبکه مجازی
	کتابخانه آموزشی	کتابخانه مدارس
	شبکه مجازی	شبکه اجتماعی
	کتابخانه آموزشی	کتابخانه مدارس
	آرشیو الکترونیکی	آرشیو دیجیتال
	مدیریت فناوری اطلاعات	مدیریت IT
	سیستم اطلاعات جغرافیایی	Gis
	وب معنایی	وب سه
	شبکه خصوصی مجازی	شبکه مجازی VPN
	پژوهش	تحقیق
	حریم خصوصی	حریم شخصی
	دیداری سازی	بصری سازی
	نشریات	مجلات
	فناوری اطلاعات	آی تی
	وب سایت	وبگاه
	نشریه الکترونیکی	مجله الکترونیکی
	موتور جستجو	موتور کاوش
	آموزش الکترونیکی	آموزش مجازی
	مرجع دیجیتال	مرجع مجازی
	آموزش الکترونیک	آموزش مجازی
	استفاده اطلاعات	کاربرد اطلاعات

جدول ۴- ادامه

کتابدار کودک	کتابخانه کودک	شاغل بودن
کتابدار	کتابخانه	
کتابداران دانشگاهی	کتابخانه دانشگاهی	
خدمات مرجع	منابع مرجع	با استفاده از ابزار
ذخیره سازی اطلاعات	کتابخانه دیجیتال	
بانک اطلاعاتی	بازیابی اطلاعات	
اینترنت	شبکه اجتماعی	
شبکه	اشتراگذاری اطلاعات	
شبکه اجتماعی	وب ۲	
کامپیوتر	داده کاوی	
تحلیل پیوند	وب سنجی	
کتابخانه	مطالعه	
اینترنت	وب سایت	
شبکه	نرم افزار	فراهم آوری
خدمات اطلاعاتی	دسترسی به منابع اطلاعاتی	دسترسی
کتابداری	فهرست نویسی	وظیفه داشتن
کتابخانه دیجیتال	خدمات مرجع	ارائه خدمات
مراجعین	کتابخانه	
کتابخانه دیجیتال	خدمات مرجع	بخشی است
بازیابی اطلاعات	رابط و رتبه بندی اطلاعات	
گرافیک کامپیوتری	نرم افزار کامپیوتری	
طراحی کامپیوتری	گرافیک کامپیوتری	
نرم افزار	کامپیوتر	
کتابداران آموزشی	نیازسنجی سواد اطلاعاتی	مطالعه کردن
تحلیل محتوا	کتاب	ارزیابی کردن
دانشگاه	آموزش عالی	مشابه بودن
جرايم رایانه ای	کلاهدرداری رایانه ای	
سازمان	پایگاه	استفاده کردن
مراجعین	کتابخانه	
کتابخانه عمومی	کتابخانه های ایران	نوعی است
شبکه بی سیم	شبکه	
تحلیل محتوای کیفی	تحلیل محتوا	
کتاب های تالیفی	اطلاعات کتابشناختی	شامل بودن

پیش از این به عنوان جستجوی اصلاح شده در نظر گرفته شده بود) مورد بررسی قرار گرفت. در مطالعه اولیه، روابط معنایی در حوزه علم اطلاعات در اصطلاحنامه‌ی Inspec طی فرایندی تقویت شده است. سپس روابط بین مفاهیم مورد جستجو و اصلاح شده کاربران، با روابط معنایی تقویت شده، مطابقت داده شد. بخشی از روابط معنایی که در این اصطلاحنامه وجود داشت، استخراج شد. مواردی که این انطباق در آنها مشاهده شد در جدول آمده است.

در حالی که بعضی از اصلاحات کاربران، مطابق با روابط معنایی در اصطلاحنامه تقویت شده نبود، میزان انطباق در جدول ۵ آمده است.

کتابدار کودک ۲۰۱۶/۰۹/۰۵ ۱۳:۳۴:۵۵

در این جستجو، کاربر پس از بازیابی تعداد زیادی از نتایج، برای گرفتن پاسخ مورد نظر، از اصطلاح مرتبط برای رسیدن به پاسخ استفاده کرده است.

### - وضعیت اصلاحات جستجوی کاربران گنج براساس روابط معنایی (براساس اصطلاحنامه) چگونه است؟

کاربران در جستجوهای خود، برای گرفتن پاسخ مرتبط به نیاز خود، بارها و بارها پرس و جوی خود را اصلاح می‌کنند. اما این که پرس و جوی اصلاح شده آنها از نظر معنایی به آن چه ابتدا مورد جستجو قرار داده‌اند مرتبط است یا خیر مسئله مبهمی است. برای پاسخ به این سؤال، جستجوهای اصلاح شده کاربران (طبق آنچه

جدول ۵- درصد مطابقت روابط معنایی مورد استفاده در اصلاح جستجوها با روابط معنایی در اصطلاحنامه تقویت شده

عنوان	تعداد
جستجوی اصلاح شده	۳۱۷
=جستجوهای دارای رابطه معنایی متناسب با روابط استخراج شده	۸۸
درصد مطابقت	۲۸

گونه‌ای که وقتی کاربر در جستجوی اول خود به پاسخی که مورد نظر خود نمی‌رسد، به اصلاح پرس و جوی خود دست می‌زند تا آن جا که پرس و جوی مورد نظر، او را به پاسخ مطلوب برساند. در واقع مطالعه لاگ‌های پرس و جوی پایگاه‌های اطلاعاتی دیگر ممکن است بر فرمول‌بندی مجدد استفاده شده تأثیر بگذارد، بنابراین، نتایج متفاوتی نمایش داده می‌شود یا ممکن است پرس و جوها جابه جا شوند. بنابراین، در صورتی که سیستم در پرس و جوهایی اول و دوم پاسخ مورد نظر کاربر را فراهم کند، نیازی به اصلاح پرس و جوی در مراحل بعدی نیست. با این وجود، با استفاده از تحلیل رفتاری که در این پژوهش انجام شد، می‌توان به سوی ارائه پیشنهاد جستجوی مناسب برای یافتن پاسخ مطلوب کاربر پیش رفت و سیستم را به گونه‌ای بهبود داد که امکان دسترسی سریع و مطلوب کاربر به پاسخ مورد نظر فراهم گردد.

### تعارض منافع

گزارش نشده است.

### منبع حمایت کننده

گزارش نشده است.

### References

- Anick, P. (2003). Human interaction. Using terminological feedback for web search refinement: a log-based study, Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 88-95.
- Arlitt, M. (2000). Characterizing Web user sessions. ACM SIGMETRICS Performance Eval Review, 28(2), 50-63.
- BahmanAbadi, A. (2001). Use of thesaurus in information retrieval. Book quarterly, 77-103. (Persian)
- Bates, M.J. (1990). Where should the person stop and the information search interface start? Information Processing & Management 26(5), 575-591.
- Belkin, N.J. (1984). Cognitive models and information transfer. Social Science Information Studies, 4, 111-129.
- Belkin, N.J. (1993). Interaction with texts: Information retrieval as information-seeking behavior. In: Information retrieval '93. Von der Modellierung zur Anwendung. Konstanz: Universitaetsverlag Konstanz, 55-66.

میزان انطباق اصلاحات با روابط معنایی بین جفت اصطلاحات ۲۸ درصد است.

### بحث و نتیجه‌گیری

پژوهش حاضر به روش تحلیل لاگ انجام گرفته و نتایج آن هم‌رخدادی اصطلاحات در پرس‌وجوی تعداد کمی از کاربران را نشان داد. این نشان‌دهنده این موضوع است که با وجود این که اصطلاحنامه‌ها بر روی پایگاه گنج قرار گرفته‌اند و امکان استفاده از آن برای کاربران فراهم شده است، کاربران کمتر به این اصطلاحنامه‌ها مراجعه می‌کنند و اغلب اصطلاحات را بر اساس آنچه در ذهن خود دارند ترکیب می‌کنند. دلیل این امر را باید در بررسی اصطلاحنامه‌ها یافت و این که آیا آنها به سادگی قابل استفاده هستند یا خیر.

این پژوهش دسته‌بندی جدیدی از اصلاحات کاربران ارائه نمود. نتایج این دسته‌بندی که براساس تاکتیک‌های حامی روابط معنایی و هستی‌شناسی مطرح شده از سوی گارسیا و سیسیلیا (۲۰۰۳) انجام شد نشان داد که تاکتیک‌های Relate و Select بیشتر از سایر تاکتیک‌ها مورد استفاده قرار گرفته‌اند. این نشان‌دهنده این موضوع است که کاربران تلاش می‌کنند پرس و جوی خود را در درجه اول به سمت تجزیه پرس و جوی به اجزای تشکیل دهنده آن و سپس استفاده از اصطلاح مرتبط به اصطلاح اولیه جستجوی خود پیش ببرند. این دسته‌بندی متفاوت با نتایج سایر پژوهش‌هاست که تنها به سه دسته از اصلاح پرس و جوی کاربر شامل خاص کردن، کلی‌سازی و فرمول‌بندی شناسایی کردند. علاوه بر این، نتایج این پژوهش نشان داد، تعداد زیادی از جستجوهای کاربر با اصطلاحات مترادف اصلاح شده است. در موارد دیگر، کاربران انواع مختلفی از روابط معنایی را برای اصلاح استفاده کرده‌اند که رابطه «مرتبط بودن» «با استفاده از ابزار»، «نوعی است»، «بخشی است»، «شاغل بودن»، «ارائه خدمات»، «دریافت خدمات»، «استفاده کردن»، «ارزیابی کردن»، «مشابه بودن» و «فراهم آوری دسترسی» جزو این انواع است. این انواع روابط معنایی شناسایی شده می‌توانند در بهبود نظام گنج به ویژه در پیشنهاد برای جستجو برای کاربران استفاده شود و کاربران را به سمت گرفتن پاسخی مطابق با درخواست خود سوق دهد. می‌توان این رفتارهای کاربر را برگرفته از رفتار سیستم دانست به

- Ben Mustapha, N., Aufaure M-A., Baazaoui, H. and Ben Guezala, H.(2011) 'Contextual ontology module learning from web snippets and past user queries', 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, KES 2011, 538–547.
- Ben Mustapha, N., Aufaure, M., Baazaoui Zghal, H. and Ben Ghézala, H. (2012) 'Modular ontological warehouse for adaptative information search', MEDI , 79–90.
- Ben Mustapha, N., Baazaoui-Zghal, H., Moreno, A. and Ben Ghezala, H. (2013) . A dynamic composition of ontology modules approach: application to web query reformulation, Int. J. Metadata, Semantics and Ontologies, 8(4), 309–321.
- Boldi, P., Bonchi, F., Castillo, C., and Vigna, S. (2009). From 'dango' to 'japanese cakes': Query reformulation models and patterns. In Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Milan, Italy, pages 183–190.
- Bollegala, D., Matsuo, Y. and Ishizuka, M. (2007) 'Measuring semantic similarity between words using web search engines', WWW'07: Proceedings of the 16th International Conference on World Wide Web, 757–766.
- Bozzon, A., Chirita, P. A., Firan, C. S., and Nejd, W. (2007). Lexical analysis for modeling web query reformulation. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, pages 739–740.
- Bruza, P. and Dennis, S. (1997). Query reformulation on the internet: Empirical data and the hyperindex search engine. In Proceedings of the RIAO'97 Conference on Computer-Assisted Searching on the Internet, Montreal, Canada, pages 488–499.
- Burton, M.C. & Walther, J.B. (2001). The value of Web log data in use-based design and testing. Journal of Computer Mediated Communication, 6(1), <http://www.ascusc.org/jcmc/vol6/issue3/burton.html>
- Chen, H. & Dhar, V. (1990). Online query refinement on information retrieval systems: A process model of searcher/system interactions. Proceedings of the 13th Annual-International ACM SIGIR Conference, 115-132.
- Costa, R. P. and Seco, N. (2008). Hyponymy extraction and web search behavior analysis based on query reformulation. In Proceedings of the 11<sup>th</sup> Ibero-American Conference on AI, Lisbon, Portugal.
- Efthimiadis, E. N. (2000). Interactive query expansion: a user-based evaluation in a relevance feedback environment. Journal of the American Society for Information Science, 51(11):989–1003.
- Fellbaum, C., editor (1998). WordNet: An electronic lexical database. MIT press, Cambridge, MA, USA.
- French J.C. Brown, D.E., Kim N.-H.(1997). A classification approach to Boolean query reformulation. Journal of the American Society for Information Science, 48 (8), 694-706.
- García , E., & Sicilia, M.-Á. (2003). User Interface Tactics in Ontology-Based Information Seeking. PsychNology Journal, 1( 3), 242 – 255.
- Guo, J., Xu, G., Li, H., and Cheng, X. (2008). A unified and discriminative model for query refinement. In SIGIR '08,379-386.
- Hariri, N and Haratizade, S (2015). Evaluation of user satisfaction from Islamic science thesaurus as information retrieval tool. Library and information organization national quarterly, 2(26): 141-160. (Persian)
- Hariri, N and Mehrban, S (2013). Nanotechnology Database Search Strategies: Transaction Report Analysis. Information processing and managing. 29(1), 233-252. (Persian)
- He, D., G"oker, A., and Harper, D. J. (2002). Combining evidence for automatic web session identification. Information Processing and Management, 38(5),727–742.
- Hollink, V., Tsikrika, T., & Vries, A. P. (2010). Semantic search log analysis: a method and a study on professional image search. Journal of the American Society for Information Science and Technology.
- Huang, J. and Efthimiadis, E. N. (2009). Analyzing and evaluating query reformulation strategies in web search logs. In Proceeding of the 18th ACM Conference on Information and Knowledge Management, Hong Kong, China, pages 77–86.
- Ingwersen, P. (1992). Information Retrieval Interaction, Taylor Graham, London. Online: [https://curis.ku.dk/ws/files/47050396/Ingwersen\\_IRI.pdf](https://curis.ku.dk/ws/files/47050396/Ingwersen_IRI.pdf)
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory, Journal of Documentation, 25(1), 3-50.
- J"orgensen, C. and J"orgensen, P. (2005). Image querying by image professionals. Journal of the American Society for Information Science and Technology, 56(12),1346–1359.
- Jansen, B. J., Booth, D. L., and Spink, A. (2009). Patterns of query reformulation during web searching. Journal of the American Society for Information Science and Technology, 60(7),1358–1371
- Jones, R. and Fain, D. C. (2003). Query word deletion prediction. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada, 435–436.
- Khosravi, M. and Jamali M, H,R. (2014). Log analysis of Iranian information and scientific document database (IranDoc) its user search behavior. Information management and processing 29(4): 979-1106.
- Lau, T. and Horvitz, E. (1999). Patterns of search: analyzing and modeling web query refinement. In

- Proceedings of the Seventh International Conference on User Modeling, Banff, Canada, 119-128.
- Mirzabaigi, M. (2011). Ontology usage in information retrieval: a review on current research and presenting a conceptual model. *Information processing and management research journal*, (Special for information storage and retrieval), 237-253. (Persian)
- Mohammadi, F and Dokht Esmati, M, (2011). Evaluation of web based thesauri of Iranian research institute of science and technology: A descriptive approach. *Information processing and management*, 26(3), 675-694.
- Nicholas, D, P. Huntington, and Jamali. H. R. (2008). User diversity: as demonstrated by deep log analysis. *Electronic Library*. 26 (1), 21-38.
- Ozmutlu, H. C. (2009). Markovian analysis for automatic new topic identification in search engine transaction logs. *Applied Stochastic Models in Business and Industry*, 25(6), 737-768
- Park, M. and T. S. Lee. 2013. Understanding science and technology information users through transaction log analysis. *Library Hi Tech*. 31 (1), 123-140.
- Rajabali begloo, R (2008). Methods of log process analyzing (interaction) in information systems. *Library and information science* 3(39): 181-204.(Persian)
- Rieh, S. Y. and Xie, H. (2006). Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing and Management*, 42(3),751-768.
- Saracevic, T. (1996). Modeling interaction in information retrieval (IR): A review and proposal. *Proceedings of the American Society for Information Science*, 33, 3-9, Online: <http://www.scils.rutgers.edu/~tefko/ProcASIS1996.doc>
- Saracevic, T. (1997a). The stratified model of information retrieval interaction: Extension and applications, *Proceedings of the American Society for Information Science*, 34, 313-327. Online: <http://www.scils.rutgers.edu/~tefko/ProcASIS1997.doc>.
- Saracevic, T. (1997b). "Users lost: Reflections on the past, future, and limits of information science", *SIGIR Forum*, 31(2), 16-27. Online: [http://www.scils.rutgers.edu/~tefko/SIGIR\\_Forum\\_97.doc](http://www.scils.rutgers.edu/~tefko/SIGIR_Forum_97.doc)
- Silverstein, C., Marais, H., Henzinger, M., and Moricz, M. (1999). Analysis of a very large web search engine query log. *SIGIR Forum* 33(1), 6-12.
- Spink, A., Jansen, B.J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *IEEE Computer*, 35(3), 107-111
- Spink, A., Wolfram, D., Jansen, B.J., & Saracevic, T. (2001). The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3), 226-234
- Spink, Amanda Jansen, Bernard J. Ozmutlu, H. Cenk (2000) Use of query reformulation and relevance feedback by Excite users, *Internet Research*, 10 (4),317-328
- Taheri, M, Shapuri, D, Razmi Shandi, M and Noruzi, E. (2013). Applicablty of offline Islamic database based on thesarus based on ISO 9241 and Iso 16982. *Information systems and services*, 8 (4): 81-92. (Persian)
- Teevan, J., Adar, E., Jones, R., and Potts, M.A. (2007). Information re-retrieval: repeat queries in Yahoo's logs. *SIGIR '07*, 151-158.
- Valinejad, A and Pasyar, P. (2007). Semantic retrieval challenges and usage of new thesarus. *Proceedings of the national conference on thesarus and its usage in electironic environment (15 November 2006 Qum): 491-520: Tehran, Librarian.*
- Whittle, M., Eaglestone, B., Ford, N., Gillet, V. J., and Madden, A. (2007). Data mining of search engine logs. *Journal of the American Society for Information Science and Technology*, 58(14):2382-2400.
- Yi, K., J. Behesht, J. E. Leide, and A. Large. (2006). User search behavior of domain- specific information retrieval systems: an analysis of the query logs from psycINFO and ABC-Clio's Historical Abstracts/America: history and life. *American Society for Information Science and Technology*. 57 (7), 1208.
- ZakerShahrak, M. (2008). Search: user information behavior and information retrieval systems. *Information and information seeking*, (15), 50-60. (Persian)