تحلیل لاگ پایگاه اطلاعات و مدارک علمی ایران (ایرانداک) و رفتار جستجوی کاربران آن

مریم خسروی[1] | **حمیدرضا جمالی مهموئی**[2]

۱. [پدیدآور رابط] کارشناسی ارشد روان‌شناسی تربیتی؛ مربی پژوهشی؛ پژوهشگاه علوم و فناوری اطلاعات ایران khosravi@irandoc.ac.ir

۲. دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛ دانشگاه خوارزمی h.jamali@gmail.com

مقالهٔ پژوهشی

دریافت: ۱۳۹۲/۰۲/۳۱
پذیرش: ۱۳۹۲/۰۹/۱۷

دورهٔ ۲۹ شماره ۴
ص‌ص. ۹۷۹-۱۰۰۶

چکیده: این پژوهش با هدف تحلیل رفتار کاربران و درک میزان و نحوهٔ استفادهٔ آنها از پایگاه اطلاعاتی ایرانداک به مطالعهٔ یک دوره ۲۵ روزه (در ماه‌های بهمن و اسفند ۹۱) لاگ جستجوی کاربرانی می‌پردازد که از بیرون ایرانداک به جستجوی اطلاعات در این پایگاه اطلاعاتی پرداخته‌اند. تحلیل‌های انجام‌شده روی لاگ شامل تحلیل کدگذاری جغرافیایی، تحلیل های آماری و تحلیل خرد رفتار اطلاع جویی است. یافته‌ها حاکی از آن است که کاربران شهرهایی چون تهران، مشهد و تبریز بیش از سایر نقاط از پایگاه استفاده کرده‌اند. ساعات بین ۹ الی ۱۳ روزهای دوشنبه و یکشنبه اوج استفاده است. پربسامدترین طول عبارت جستجوها بین ۱۰ الی ۱۵ حرف و پربسامدترین موضوع جستجو شده علوم اجتماعی است. همچنین مشخص شد که برخی از کاربران با مفاهیمی نظیر اصلاح عبارت جستجو یا نحوهٔ صحیح استفاده از عملگرهای بول آشنا نیستند و این نکات منجر به بازیابی ناکارآمد از پایگاه می شود. ضعف‌هایی از جمله ناکارآمدی الگوریتم بازیابی در مورد کلمات کوتاه نیز در خود پایگاه آشکار شد. انجام این تحقیق نشان داد که استفاده از لاگ می تواند در ارزیابی سامانهٔ بازیابی و شناخت رفتار کاربران آن بسیار مفید باشد.

کلیدواژه‌ها: تحلیل لاگ؛ رفتار اطلاع‌یابی؛ جستجوی اطلاعات؛ پایگاه اطلاعاتی ایرانداک

پژوهشنامهٔ پردازش و مدیریت اطلاعات
فصلنامه | علمی پژوهشی
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISA، ISC و Scopus
http://jipm.irandoc.ac.ir
پژوهشگاه علوم و فناوری اطلاعات ایران

۱. مقدمه

بازیابی اطلاعات مقولهٔ اصلی و ربط، مفهوم محوری علم اطلاعات است (Saracevic 2007). به این علت در حوزهٔ علم اطلاعات که شامل زیرحوزه‌هایی مثل تعامل انسان و اطلاعات است، دو جزء از اجزاء یک نظام اطلاعاتی بسیار مورد مطالعه قرار می‌گیرند: یکی، خود نظام‌های بازیابی اطلاعات و دیگری، کاربران این نظام‌ها. مطالعهٔ این دو جزء با هدف بهبود نظام‌های بازیابی و ارتقای خدمت‌دهی به کاربران نیز صورت می‌گیرد. هر دو جزء، هم کاربران و هم نظام‌ها، مدام در حال تغییر و تحول هستند و نیاز ما به شناخت کاربران و رفتار آنان همچنان پایدار می‌ماند. علاوه بر این، چگونگی عملکرد یک نظام اطلاعاتی نیز باید مورد ارزیابی قرار گیرد تا باآشکارشدن نواقص، برای رفع آنها اقدام گردد.

حوزهٔ پژوهشی مطالعهٔ رفتار کاربران، حوزه‌ای غنی است و پژوهش‌های پیمایشی متعددی دربارهٔ رفتار اطلاع‌یابی کاربران انجام شده است. با این حال، پژوهش های معدودی به تجزیه و تحلیل داده‌های عینی حاصل از فایل‌های تراکنشی پایگاه‌های اطلاعاتی پرداخته‌اند (Lambert 2010). در پیمایش‌ها به‌طور معمول آنچه آشکار می‌شود، رفتاری است که افراد تصور می‌کنند انجام می‌دهند، یا می‌گویند که انجام می‌دهند، یا حتی مایل‌اند انجام دهند. گاه این موارد با آنچه که کاربر در عمل انجام می‌دهد، تفاوت زیادی دارد و روش تحلیل لاگ، بر خلاف پیمایش، رفتاری را آشکار می‌کند که کاربر واقعاً انجام داده است.

یکی از پایگاه‌های اصلی اطلاعات علمی کشور، پایگاه اطلاعاتی ایرانداک است که محتوای اصلی آن پایان‌نامه‌های دانشگاه‌های کشور می‌باشد. الگوریتم‌های جستجوی این پایگاه و واسط گرافیکی آن تاکنون در چند مقطع دستخوش بازطراحی و اصلاح شده است؛ از جمله اینکه در سال ۱۳۸۹ نظام جستجوی پایگاه عوض شد. از آنجا که تاکنون نه دربارهٔ رفتار کسانی که در این پایگاه اقدام به جستجوی اطلاعات می‌کنند و نه بر روی عملکرد این پایگاه، هیچ پژوهشی از دید علم اطلاعات صورت نگرفته، این مقاله در صدد است تا با بررسی آثار به‌جای مانده از تعامل کاربران با پایگاه اطلاعاتی ایرانداک، علاوه بر به‌دست آوردن چشم‌اندازی کلی از چگونگی عملکرد نظام و کاربرد آن، به شناسایی راهبردهای جستجو در این پایگاه اطلاعاتی پرداخته و سپس، با استفاده از رفتار کاربر،

میزان استفاده و نحوهٔ استفادهٔ کاربران تحلیل شود. انجام این کار از این جهت ضرورت دارد که توسعهٔ الگوریتم‌های بازیابی و کل سامانهٔ پایگاه، به‌صورت بومی انجام شده و مبتنی بر نرم‌افزارهای خارجی نیست و لازم است که مورد مطالعه قرار گیرد تا نقاط قوت و ضعف آن مشخص گردد.

پایگاه اطلاعاتی ایرانداک اکنون با بیش از ۶۱۵ هزار رکورد اطلاعاتی، روزانه ۲۲ هزار نفر بازدیدکننده دارد و بر اساس گزارش شورای عالی اطلاع‌رسانی، از لحاظ حجم محتوا، مقام اول را در میان سایت‌های دولتی داراست. مهم‌ترین پایگاه‌های ایرانداک عبارتند از پایان‌نامه‌ها، طرح‌های پژوهشی، مقالات علمی نشریات، گزارش‌های دولتی و مقالات سمینارها و همایش‌ها. این پایگاه دارای یک واسط است با ویژگی‌های جستجوی ساده، پیشرفته و حرفه‌ای که تصویر صفحه جستجوی پیشرفته آن در شکل ۱ ارائه شده است.

شکل ۱. واسط گرافیکی جستجوی پیشرفتهٔ پایگاه ایرانداک

با توجه به رشد روزافزون محتوای این پایگاه و اهمیت و افزایش تعداد کاربران آن

کـه جامعـهٔ دانشـجویان تحصـیلات تکمیلـی و اعضـای هیئـت علمـی هسـتند، لازم اسـت مطالعه‌ای بر روی چگونگی استفادهٔ کاربران از آن و عملکرد خود پایگاه صورت گیرد تـا نقاط قوت و ضعف آن آشکار گردد. در همین راستا، سعی بر این است تا با تحلیـل لاگ[1] (گزارش‌های تراکنشی وب) و تجزیه و تحلیل فایل‌های تراکنشی که تعامل بین کـاربران و پایگاه را نشان می دهد، علاوه بر درک بهتـر رفتـار جسـتجوی کـاربران، نـواقص احتمـالی موجود در خود پایگاه نیز شناسایی شده و زمینهٔ ارتقای آن فراهم آید.

۲. هدف و پرسش‌های پژوهش

- ◇ هدف اصلی این پژوهش، مطالعهٔ رفتار اطلاع‌جویی کاربران پایگاه اطلاعاتی ایرانداک با استفاده از تحلیل لاگ بوده و به‌دنبال پاسخ‌گویی به سؤالات زیر است:
- ◇ روند زمانی استفاده از پایگاه نظیر ساعات و روزهای هفته چگونه است؟
- ◇ منشأ کاربران پژوهشگاه از نظر جغرافیا و دامنهٔ اینترنتی چیست؟
- ◇ چه موضوعاتی مورد جستجوی کاربران است؟
- ◇ رفتار جستجوی اطلاعات کاربران (تعدد، موارد جستجو، طول عبارات مورد جسـتجو، نتایج حاصل از جستجو) چگونه است؟

۳. پیشینهٔ پژوهش

تحلیل لاگ نظام‌های اطلاعاتی روشی است که برای ارزیـابی نظـام‌هـا و نیـز مطالعـهٔ کاربران از دههٔ ۱۹۶۰ مورد استفاده بوده است (Peters 1993). پیشینه، ویژگی‌هـا، معایـب و مزایای این روش را آثار دیگری به‌تفصـیل مـورد بحـث قـرار داده‌انـد، از جملـه: (جمـالی مهموئی ۱۳۸۳) و (جمالی مهموئی[2] و همکاران ۲۰۰۵). در اینجا ما به مرور آثار جدیدتری که با هدف بررسی رفتار جستجوی کاربران انجام شده است و نیز آثار ایرانی می‌پردازیم.

در میان محققان خارجی، اسـپینک، جانسـن[3] و همکـاران آثـار زیـادی بـا اسـتفاده از تحلیل لاگ موتورهای جستجوی وب منتشر کـرده‌انـد کـه نشـان‌دهنـدهٔ رفتـار جسـتجوی

1. log analysis
2. Jamali
3. Jansen

کاربران وب هستند. به‌عنوان مثال اوزمولتو[1]، اوزمولتو و اسپینک نشان دادند که بسیاری از کاربران، به‌خصوص در اینترنت جستجوهای کوتاهی انجام می‌دهند که نیازهای واقعی آنها را به‌طور واضح نشان نمی‌دهد (Ozmultu, Ozmultu & Spink 2003). جانسن، اسپینک و ساراسویک گزارش کردند که ۶۷ درصد از کاربران فقط یک جستجو انجام داده و ۱۹ درصد آنها تنها یک تغییر در جستجوی خود ایجاد می‌کنند (۲۰۰۰). اسپینک و جانسن در مطالعهٔ دیگری دریافتند که کاربران دو یا سه واژه را جستجو کرده و تنها اولین صفحهٔ نتایج را مشاهده می‌کنند (۲۰۰۶). لامبرت نیز در بررسی لاگ موتور جستجوی MCI نشان داد که جستجوها به‌طور معمول ساده و متشکل از دو کلمه هستند (۲۰۱۰).

تحقیقات اسپینک و همکارانش اطلاعات زیادی در مورد رفتار جستجو در وب به‌دست می‌دهد، اما رفتار جستجو در موتورهای جستجو احتمالاً با رفتار جستجو در یک پایگاه تخصصی علمی متفاوت است. برخی آثار به بررسی این‌گونه پایگاه‌ها پرداخته‌اند. یی[2] و همکاران در بررسی لاگ جستجوی کاربران پایگاه‌های داده‌های روانشناسی سایک‌اینفو[3] و دو پایگاه حوزهٔ علوم انسانی دریافتند که واژه‌های مرکب یا چندتایی، شناخت بهتری از مقوله‌بندی پرسش رفتار اطلاع‌یابی کاربران را به‌دست می‌دهند و نیز میانگین طول پرسش‌ها در سایک‌اینفو ۳/۱۶ واژه و در دو پایگاه دیگر ۳/۴۲ واژه بود (۲۰۰۶). نیکلاس[4] و همکاران به بررسی لاگ پایگاه ساینس‌دایرکت[5] پرداختند و نشان دادند که میان رفتار جستجوی دانشجویان در رشته‌های مختلف، تفاوت‌های فاحشی وجود داشت (۲۰۰۸). نتایج بررسی دیگری بر روی لاگ یک وبگاه علمی توسط پارک و لی[6] مشخص کرد که طولانی‌ترین نشست‌های کاربران کمتر از ۳۰ دقیقه، میانگین ۲۴ دقیقه و ۱۵ ثانیه، و حداقل آنها ۱ ثانیه و حداکثر ۱۰ ساعت بود. همچنین، موضوعات علوم زیست‌محیطی در سال ۲۰۰۹ و علوم زیستی در سال ۲۰۱۰ دارای بالاترین میزان جستجو بوده‌اند (۲۰۱۳).

در ایران، در حوزه علم اطلاعات و دانش‌شناسی تاکنون هیچ تحلیلی بر روی لاگ

1. Ozmultu
2. Yi
3. PsycINFO
4. Nicholas
5. ScienceDirect
6. Park and Lee

خام پایگاه‌ها یا وبگاه‌ها انجام نشده است. شاید یک علت آن در دسترس‌نبودن این لاگ‌ها برای محققان بوده، در حالی که مطالب نظری چندی در زمینهٔ تحلیل لاگ نوشته شده است؛ از جمله ستوده (۱۳۸۲)، جمالی مهموئی (۱۳۸۳)، و رجبعلی بگلو (۱۳۸۶). در یکی از معدود مطالعات پژوهشی این حوزه، زوارقی به تحلیل استفاده از وبگاه مرکز اطلاعات و مدارک علمی ایران در بازهٔ زمانی ۱۲ اسفند ۱۳۸۲ تا آبان ۱۳۸۳ پرداخت. البته وی لاگ خام را تحلیل نکرده است. وی نشان داد که روز سه‌شنبه پراستفاده‌ترین روز هفته بوده است (۱۳۸۴). همچنین، حریری و مهربان با استفاده از تحلیل‌های گوگل تحلیل راهبردهای جستجو در پایگاه‌های اطلاعاتی فناوری نانو را مطالعه کردند و دریافتند که متوسط طول پرسش‌های به‌کار برده شده از طریق موتورهای جستجو ۳/۳۶ واژه بوده و کاربران از راهبردهای جستجوی ساده برای بازیابی اطلاعات استفاده کرده‌اند (۱۳۹۲).

پیشینهٔ پژوهش نشان می‌دهد که با وجود اینکه تحلیل لاگ به‌عنوان یکی از شیوه‌های ارزیابی پایگاه‌ها و مطالعهٔ رفتار اطلاع‌جویی کاربران در خارج جایگاه خود را پیدا کرده، هنوز در ایران کار پژوهشی چندانی در این زمینه انجام نشده است و مطالعهٔ حاضر، احتمالاً نخستین مطالعهٔ گسترده‌ای است که بر روی لاگ خام یک پایگاه اطلاعاتی ایرانی در حوزهٔ علم اطلاعات صورت می‌گیرد.

٤. روش پژوهش

روش پژوهش در این مطالعه تحلیل لاگ جستجوی کاربرانی است که از بیرون ایرانداک و از طریق وب به جستجوی پایگاه‌های اطلاعاتی آن می‌پردازند. لاگ، یک فایل متنی ساده با فرمت txt است که توسط نرم‌افزاری که بر روی سرور نصب شده، هر گونه کنش کاربر در تعامل با نظام اطلاعاتی (کلیک‌کردن با موس یا فشردن کلید بر روی صفحه‌کلید) را در قالب یک خط داده ثبت می‌کند.

لاگ مورد تحلیل در این تحقیق شامل ۱۷۳۰۳۸ جستجو است که ۲۵ روز از ۱۶ فوریه ۲۰۱۳ (مطابق با ۲۸ بهمن ۱۳۹۱) تا ۱۲ مارس ۲۰۱۳ (مطابق با ۲۲ اسفند ۱۳۹۱) را دربر می‌گیرد. از این تعداد جستجوی انجام شده، ۵۰۲۱۰ جستجو شامل شناسهٔ کاربر و آدرس پروتکل اینترنت کاربر و بقیهٔ جستجوها فاقد شناسهٔ کاربر است. علت انتخاب بازهٔ ۲۶ روزه این بود که تنها در همین مدت لاگ بر روی سرور ایرانداک موجود بود، لذا

تصمیم گرفته شد که کل لاگ موجود تحلیل شود. ترکیب‌بندی لاگ که توسط واحد رایانهٔ ایرانداک در اختیار محققان قرار گرفت، به‌صورت زیر است: لاگ به‌ترتیب از چپ به راست شامل فیلدهای زیر می‌باشد: شمارهٔ جستجو (۹۷۳۵)، عبارت جستجو (مسئولیت مدنی ...)، تاریخ جستجو (۱ مارس ۲۰۱۳)، زمان جستجو (۸:۴۵:۰۰)، تاریخ روزآمدشدن جستجو و زمان روزآمدشدن جستجو که در مورد تمام جستجوها دقیقاً شبیه تاریخ و زمان جستجو است، تعداد رکوردهای بازیابی‌شده برای جستجو (۳)، و شماره آی‌پی کاربر که در مثال‌های زیر اعداد آخر آنها برای حفظ محرمانگی ناشناس شده است (XXX. ۲۶، ۱۴۶، ۲). همان‌طور که گفته شد، در مورد بیش از ۵۰ هزار جستجو شناسهٔ کاربر نیز در انتهای لاگ قرار داشت. علت تفاوت در وجود یا نبودِ شناسهٔ کاربر این بود که لاگ فراهم‌شده از دو سرور متفاوت گردآوری شده است (در ادامه نمونهٔ آنها آورده شده است).

٤-١. نمونه لاگ

(مسئولیت مدنی مدیران در شرکت های تجاری) | (مسئولیت مدنی مدیران در شرکت‌های تجاری) ("مسئولیت مدنی مدیران در شرکت‌های تجاری"/۶) ۹۷۳۵

2013/03/01 08:45:00 2013/03/01 08:45:00 3 2.146.26.XXX

(مسئولیت مدنی مدیران شرکت های سهامی در حقوق ایران) | (مسئولیت مدنی مدیران شرکت‌های سهامی در حقوق ایران) | ("مسئولیت مدنی مدیران شرکت‌های سهامی در حقوق ایران"/۶ ۵۴۵۹۷

2013/02/23 20:42:34 2013/02/23 20:42:34 20 2.146.26.XXX

(مشاوره دینی) | (مشاوره دینی) | ("مشاوره دینی"/۶) ۵۲۷۵۸

2013/02/23 16:47:27/23 16:47:27 2013/02/23 16:47:27 1637.255.XX

پیش از پردازش و تحلیل لاگ قدری پالایش و پیش‌پردازش بر روی لاگ صورت گرفت؛ از جمله: پاک‌سازی لاگ از پرسش‌های مخدوش، تبدیل زمان جستجوها از ساعت گرینویچ (UTC) به زمان تهران (UTC+3:30) و محاسبهٔ مجدد زمان و تاریخ

جستجوها، شمارش تعداد حروف و کلمات هر جستجو، تعیین وجود عملگرهای مختلف در هر جستجو، و تبدیل داده‌ها به فایل اکسل و ورود آنها به اس‌پی‌اس‌اس.

تحلیل‌های انجام‌شده روی لاگ شامل موارد زیر است که نحوهٔ انجام هر کدام توضیح داده می‌شود:

تحلیل کدگذاری جغرافیایی: با استفاده از یک برنامه نوشته‌شده به زبان برنامه‌نویسی پایتون[1] که برای انجام عملیات Reverse DNS Look UP یا برگردان معکوس نام دامنه نوشته شده، تمامی آی‌پی‌های کاربران تبدیل به نام دامنه شد. این برنامه که توسط دیوید کلارک[2] از گروه پژوهشی سایبر نوشته و اجازهٔ استفاده از آن برای انجام چنین کاری به محققان داده شده است، شماره آی‌پی را تبدیل به نام دامنه کرده و اطلاعات جغرافیایی مربوط به آن، شامل منطقه جغرافیایی، کشور، شهر، طول و عرض جغرافیایی را مشخص می‌کند.

با استفاده از فرمول‌نویسی در نرم‌افزار اکسل و استفاده از توابع خود این نرم افزار، شمارش تعداد کاراکترها و کلمات عبارت جستجو و تجزیه و تحلیل آنها صورت گرفت. تحلیل‌های آماری نظیر تعداد رکوردهای بازیابی‌شده و موارد دیگر با استفاده از نرم‌افزار اس‌پی‌اس‌اس انجام گردید.

تحلیل خرد رفتار اطلاع‌جویی: برای تحلیل خرد، چند کاربر که اقدام به انجام چندین جستجوی متوالی کرده بودند، انتخاب شدند و روند جستجوی آنها مورد تحلیل خرد قرار گرفت. برای انتخاب این چند کاربر ابتدا به‌صورت هدفمند کاربرانی که اقدام به انجام چندین جستجوی متوالی کرده بودند، جدا شده و در مرحلهٔ بعد، از بین آنها چند مورد به‌صورت تصادفی انتخاب گردید.

تحلیل موضوعی: برای بررسی موضوعات جستجوها، مطابق جداول مورگان از مجموع ۱۷۳۰۳۸ جستجو با سطح اطمینان ۹۸ درصد و حاشیه خطای ۳، تعداد ۱۵۰۰ جستجو به‌صورت تصادفی انتخاب شد. سپس، برای مشخص‌شدن پراکندگی موضوعی جستجوهای انجام‌شده، بر اساس دسته‌بندی پایگاه اسکوپوس به هر جستجو یک موضوع اختصاص داده‌شد.

---

1. Python
2. David J. Clark

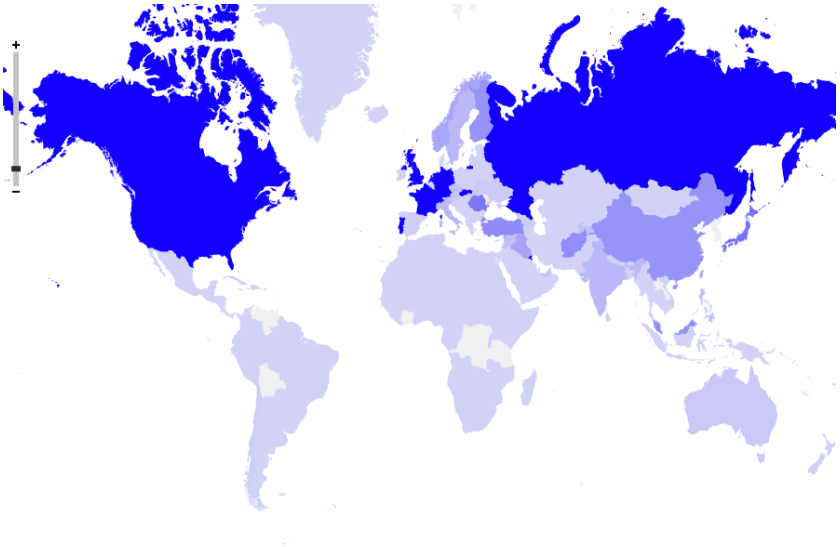
۵. یافته‌ها

۵-۱. پراکندگی جغرافیایی جستجوگران پایگاه

از مجموع ۱۷۳۰۳۸ جستجوی انجام‌شده، ۱۶۰۸۱۳ جستجو مربوط به کاربران ایرانی بود و بقیهٔ جستجوها توسط کاربرانی از چهل کشور جهان انجام شده بود. کاربران آمریکا با ۲۶۶۲ و انگلستان با ۱۰۷۲ جستجو بیشترین استفاده را از پایگاه داشته و دو کشور فارسی زبان افغانستان (با ۳۷ جستجو) و تاجیکستان (با ۱۲ جستجو) استفاده قابل توجهی از پایگاه نکرده بودند. کشورِ ۳۰۶۲ جستجو نیز به‌دلیل اینکه در فرایند تبدیل شماره‌های آی‌پی به نام یک آی‌پی با نامِ دامنه و کشور ثبت‌کننده، به نام یک پروکسی ناشناس ثبت شده بود، قابل تشخیص نبود. از کل جستجوها، حدود ۸۸۰۰ جستجو (۵ درصد) به‌زبان انگلیسی و باقی جستجوها (۹۵ درصد) فارسی بود. البته گاهی در برخی از جستجوهای فارسی کلمات تخصصی انگلیسی نیز به‌کار رفته بود.

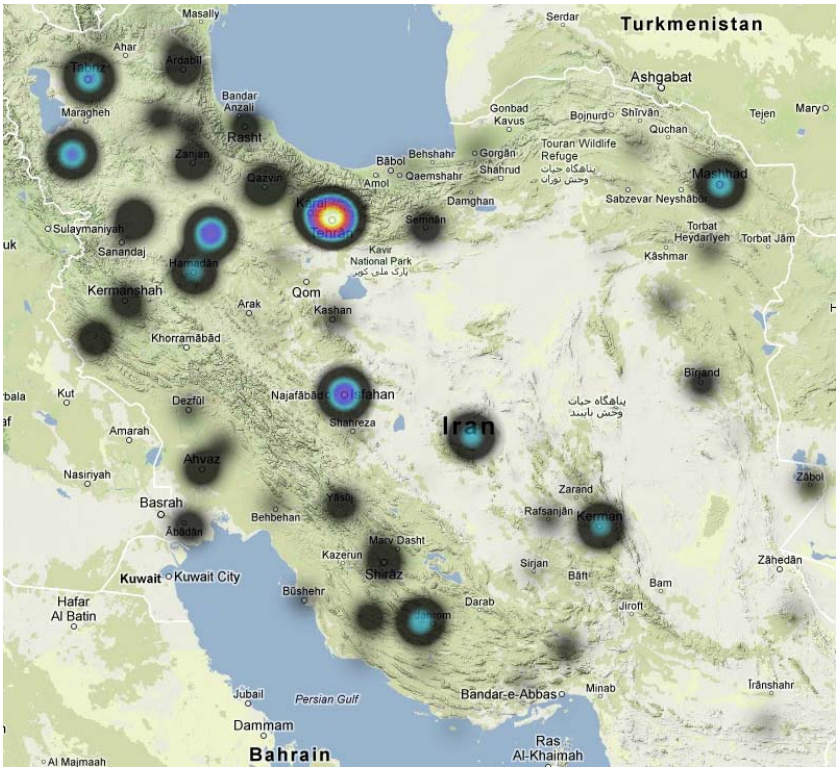
برای درک بهتر پراکندگی جغرافیایی کاربرانِ پایگاه، در شکل ۲ نقشهٔ حرارتی[۱] پراکندگی جغرافیایی کاربران در کشورهای مختلف به جز ایران نشان داده شده است. نقشهٔ حرارتی با استفاده از امکانات وبگاه OpenHeatMap.com تولید شده است. در نقشهٔ حرارتی، رنگ تیره‌تر (آبی تیره) به معنای این است که تعداد جستجوی بیشتری توسط کاربران آن کشور انجام شده است. همان‌طور که نقشه نشان می‌دهد، کشورهایی مثل آمریکا، کانادا، انگلیس، روسیه، فرانسه و آلمان دارای تعداد جستجوی بیشتری نسبت به کشورهایی مثل هند و ترکیه بوده‌اند.

۱. نقشه حرارتی یا heatmap نوعی نقشه است که برای مصورسازی تراکم یا چگالی یک متغیر در یک پهنه جغرافیایی به‌کار می‌رود. این تراکم معمولاً با رنگ متمایز می‌شود.



شکل ۲. نقشهٔ حرارتی جغرافیایی کاربران پایگاه به استثنای کاربران ایرانی

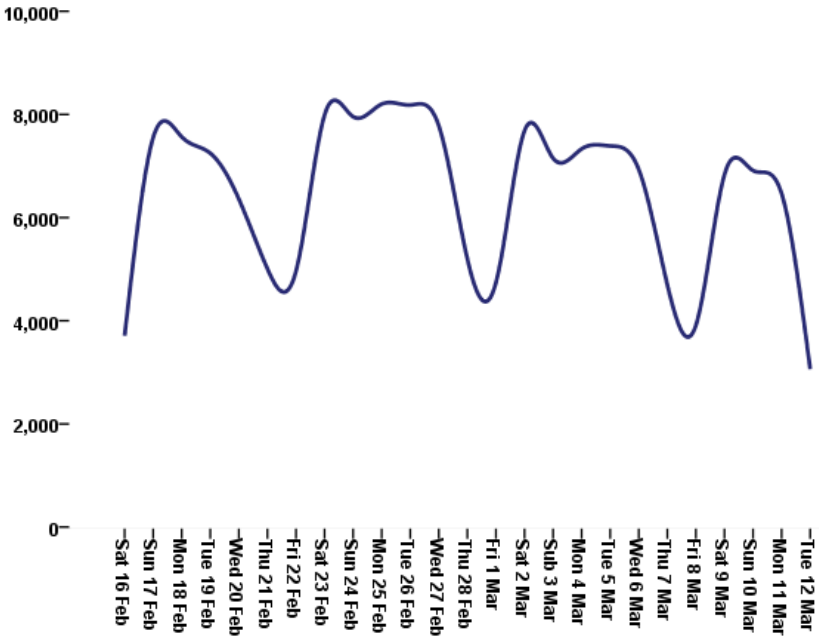
استفاده از پایگاه در داخل کشور نیز در نقاط مختلف آن پراکنده بوده و همان‌طور که انتظار می‌رفت، در تهران بیش از سایر نقاط از آن استفاده شده است. برای نشان دادن تفاوت در استفادهٔ شهرها از پایگاه، نقشهٔ حرارتی با استفاده از سرویس myheatmap.com ترسیم شده و در شکل ۳ ارائه شده است. در این نقشه نیز هر چه رنگ به سمت قرمز پیش برود، به معنای بسامد بیشتر جستجوهاست. همان‌طور که ملاحظه می‌شود، تعداد جستجوها در شهرهای بزرگِ تهران، مشهد و تبریز بیشتر بوده است.



شکل ۳. نقشهٔ حرارتی جغرافیایی کاربران پایگاه در نقاط مختلف ایران

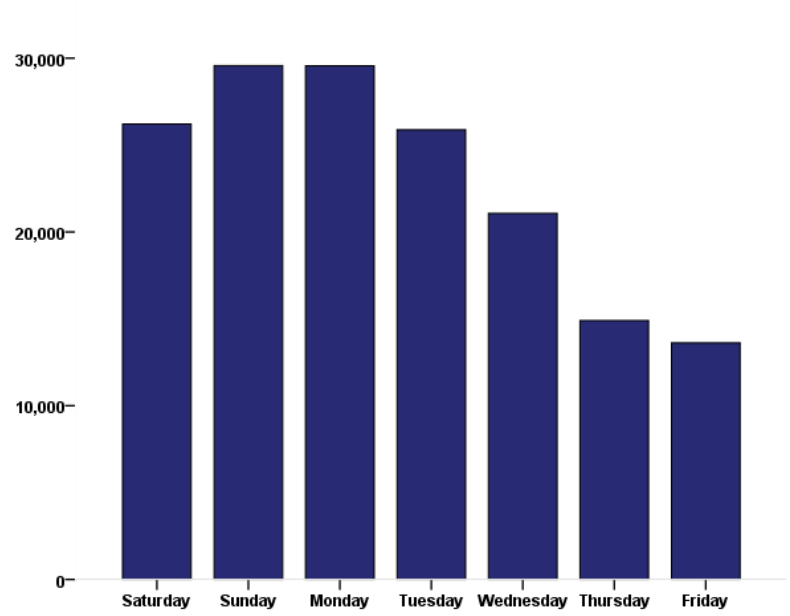
۵-۲. توزیع زمانی و تاریخی جستجوها

همان‌طور که از نمودار ۴ مشخص است، افزایش و کاهش در تعداد جستجوها با روزهای کاری هفته سازگاری نشان می‌دهد، به این‌صورت که در روزهای آخر هفته، یعنی پنجشنبه‌ها و جمعه‌ها، شاهد افت محسوس تعداد جستجوها هستیم. این نمودار شامل کل جستجوها تنها توسط کاربران ایرانی است و همان‌طور که در نمودار دیده می‌شود، در رفتار کاربران ایرانی نوعی الگو مطابق روزهای هفته مشهود است.



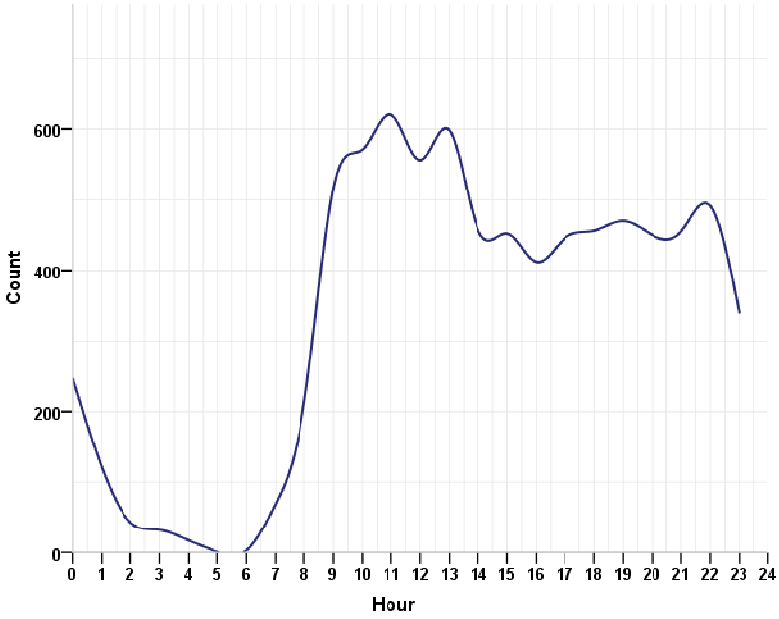
شکل ٤. تعداد جستجوهای انجام‌شده بر حسب تاریخ توسط کاربران ایرانی

شکل ۵، توزیع تعداد جستجوها در روزهای هفته را برای کاربران ایرانی نشان می‌دهد و همان‌طور که پیشتر اشاره شد، میزان استفاده از پایگاه در روزهای آخر هفته کم می‌شود، به‌طوری که جمعه‌ها کمترین تعداد جستجو و روزهای یکشنبه و دوشنبه که روزهای میانی هفته است، بیشترین تعداد جستجو صورت می‌گیرد.



شکل ۵. تعداد جستجوهای انجام‌شده در مجموع بر حسب روزهای هفته توسط کاربران ایرانی

نمودار ۶، توزیع جستجوها را برای یک روز کاری (شنبه ۲۳ فوریه مطابق با ۵ اسفند ۱۳۹۱) و تنها برای کاربران ایرانی نشان می‌دهد و همان‌طور که در این نمودار دیده می شود، بین ساعات ۹ تا ۱۳ اوج میزان استفاده از پایگاه در این تاریخ توسط کاربران ایرانی است.

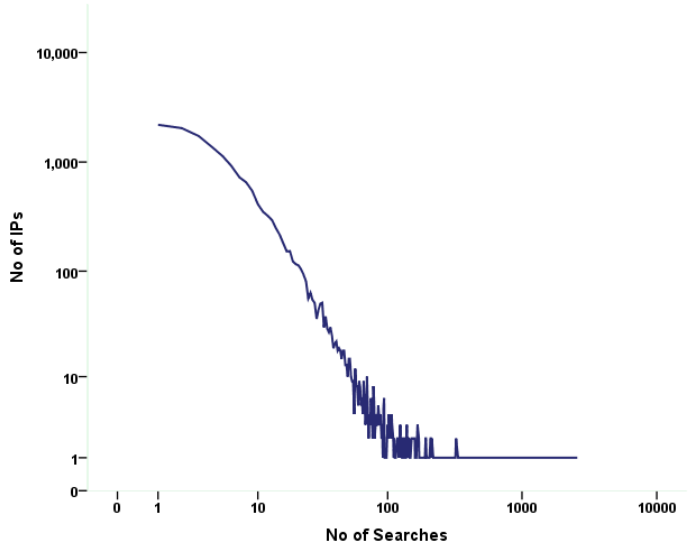


شکل ٦. تعداد جستجوهای انجام‌شده در ٢٤ ساعت در روز شنبه ٢٣ فوریه توسط کاربران ایرانی

۵-۳. وفاداری کاربران و حجم استفاده توسط آنها

دانستن این نکته که هر کاربر تا چه میزان از وبگاه استفاده کرده، می‌تواند مفید باشد؛ با این توضیح که تعریف ما از کاربر در این تحقیق مبتنی بر آدرس آی‌پی است. هر آی‌پیِ منحصربه‌فرد یک کاربر تلقی می‌شود. این تفسیر قدری با تساهل صورت می گیرد، چرا که در دنیای واقعی هر آی‌پی لزوماً به یک کاربر خاص اختصاص ندارد. هر آی‌پی، در واقع یک رایانه است و آن رایانه ممکن است یک رایانهٔ عمومی در یک آزمایشگاه باشد یا در یک کتابخانه و یا در یک دانشگاه که افراد متعددی با فواصل زمانی اندک از آن استفاده می کنند. آی‌پی حتی ممکن است متعلق به یک سرویس پروکسی باشد که چند ده یا چند صد کاربر در پشت آن پنهان باشند. به هر حال این یکی از محدودیت‌های شناخته‌شده در روش تحلیل لاگ است و راه‌حلی برای آن وجود ندارد. نمودار ۷، نمودار لوگاریتمی است که در محور افقی آن تعداد جستجو و در محور عمودی آن آی‌پی‌ها نشان داده شده است. محاسبه نشان داد که ۳۰/۳ درصد (۴۶۹۷) از آی‌پی‌ها حدود ۷۰/۷

درصد (۱۳۴۵۳۲) از جستجوها را انجام داده‌اند. این توزیع با قانون توزیع پارتو[1] یا قانون ۸۰:۲۰ مطابقت ندارد. اگر این قانون در مورد این داده‌ها صادق بود، باید ۸۰ درصد از جستجوها توسط ۲۰ درصد از کاربران صورت می‌گرفت که البته امری منطقی نیست. همان‌طور که در نمودار مشخص است، یک آی‌پی بیش از دو هزار جستجو انجام داده و تعداد زیادی از آی‌پی‌ها تنها یک جستجو انجام داده‌اند. در مجموع، ۱۵۴۷۴ آی‌پی منحصربه‌فرد (کاربر منحصربه‌فرد) در لاگ وجود داشته، و به این معناست که هر آی‌پی به‌طور متوسط ۱۱/۱۸ جستجو انجام داده است.



شکل ۷. نمودار لوگاریتمی تعداد جستجوهای انجام‌شده توسط آدرس‌های آی‌پی

۵-٤. تعداد نتایج بازیابی‌شده برای هر جستجو

بررسی آماری تعداد نتایج بازیابی‌شده برای هر جستجو وضعیت مطلوبی را برای عملکرد این پایگاه نشان نمی‌دهد. علت آن شاید کارکرد بد نظام بازیابی پایگاه و یا سطح پایین سواد جستجوی کاربران باشد. به هر حال، توزیع تعداد نتایج بازیابی‌شده برای هر

1. Pareto

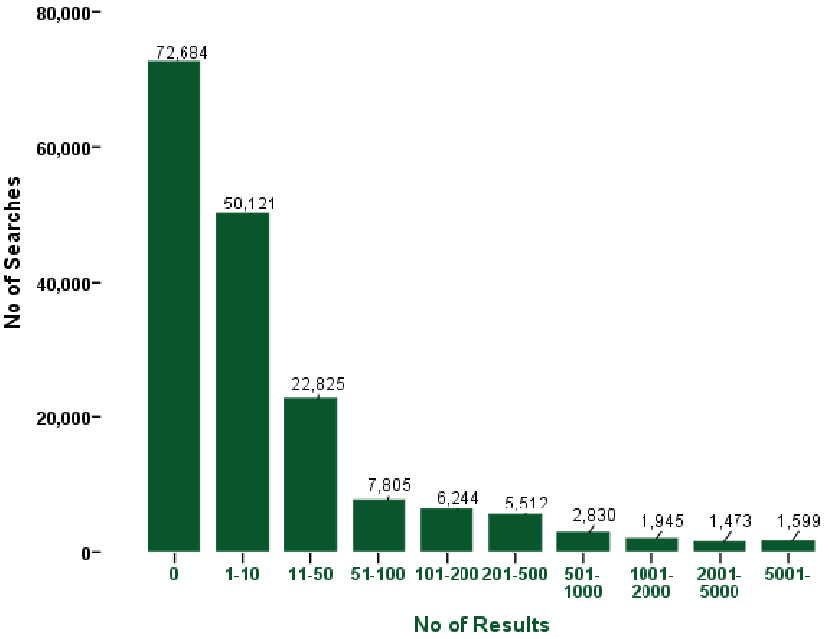
جستجو توزیع نرمال یا منطقی و مطلوب نیست. حداقل نتیجهٔ بازیابی‌شده صفر و حداکثر ۵۵۵۹۶۰ رکورد بوده است. میانهٔ نتایج ۱ و میانگین ۱۵۱۳/۰۹ است که اختلاف زیاد میان میانگین و میانه، خود نشان دهندهٔ توزیع نامطلوب و غیرمنطقیِ تعداد نتایج بازیابی شده است.

بررسی دقیق عبارت جستجو در مورد رکوردهایی که تعداد نتایج آنها بسیار زیاد بود، مشخص کرد که اکثر آنها شبیه به این است که فردی (شاید خود متولیان پایگاه) در حال آزمایش پایگاه بوده است. بررسی نام کاربر برخی از این جستجوگران نیز مشخص کرد که بعضی از جستجوها توسط خود متولیان صورت گرفته است. در برخی موارد نیز عبارت جستجوشده کاملاً بی‌معنی است که شاید فرد به اشتباه چیزی را تایپ و جستجو کرده است. پایگاه در صورت جستجوی هر گونه عبارتِ تک‌حرفی یا دوحرفی تمام رکوردهای موجود را به‌عنوان نتیجه ارائه می‌دهد و لذا تعداد رکوردهای بازیابی‌شده برای عبارت‌های جستجویی مثل آ، ب، ج، ی، آب، مه، نخ، سد، cd، hj و نظایر آنها ۵۳۲۰۸۰ رکورد برای همهٔ آنها بوده است. بر همین اساس، جدول ۱ آمار توصیفی را قبل از حذف جستجوهایی که طول عبارت جستجوی آنها کوتاه‌تر از سه حرف بوده و پس از حذف آنها نشان می‌دهد. بر اساس تعداد کلمه، میانگین طول جستجوها ۴/۵ واژه و میانهٔ آنها ۳ واژه بوده است.

جدول ۱. آمار توصیفی مربوط به تعداد نتایج ارائه‌شده برای هر جستجو

پیش از حذف جستجوهای کوتاه‌تر از ۳ حرف		پس از حذف جستجوهای کوتاه‌تر از ۳ حرف	
میانه	۱	میانه	۱
میانگین	۱۵۱۳٫۰۹	میانگین	۸۶۱٫۶۵
انحراف معیار	۲۶۵۲۴٫۹۲	انحراف معیار	۱۸۹۲۱٫۴۷
خطای استاندارد	۶۳٫۷۶	خطای استاندارد	۴۵٫۵۱
واریانس	۷۰۳۵۷۱۴۱۱٫۸۲	واریانس	۳۵۸۰۲۲۲۹۳٫۲۴
حداقل	۰	حداقل	۰
حداکثر	۵۵۵۹۶۰	حداکثر	۵۴۶۸۱۹
چولگی	۱۹٫۸	چولگی	۲۷

نمودار ۸ نشان می دهد که نتیجهٔ ۷۲۶۸۴ جستجو صفر بوده و برای ۱۵۹۹ جستجو نیز بیش از ۵۰۰۰ رکورد بازیابی شده است.



شکل ۸. نمودار ستونی تعداد جستجوهای انجام‌شده و تعداد نتایج بازیابی‌شده برای هر جستجو

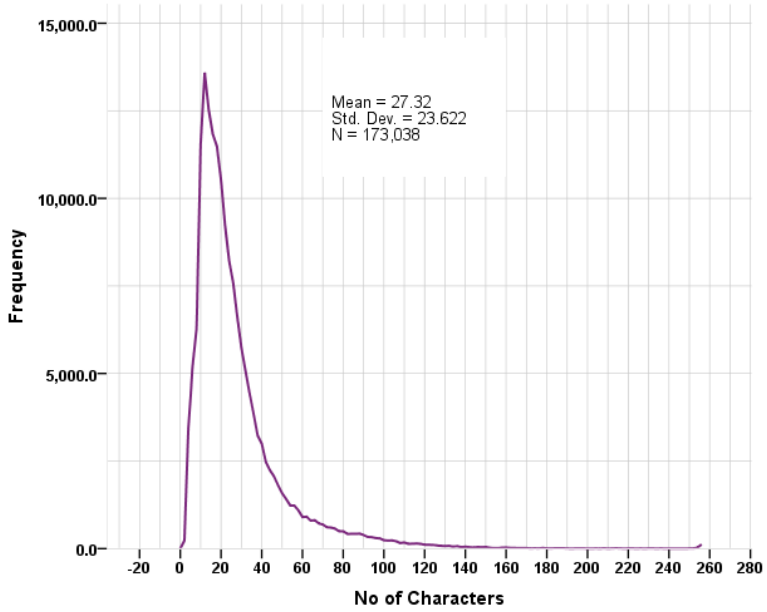
۵-۵. طول عبارات جستجوها

مطابق توضیحی که قبلاً داده شد، در جدول ۲ نیز آمار مربوط به طول کاراکترها قبل و بعد از حذف جستجوهایی با عبارت جستجوی کوتاه‌تر از سه حرف ارائه شده است. طول عبارت جستجو به‌طور متوسط حدود ۲۷/۳ کاراکتر و میانهٔ آن حدود ۲۰ کاراکتر بوده است. حداکثر طول کاراکتر در جستجوها ۲۵۵ کاراکتر بوده که ۲۵۵ کاراکتر محدودیت فیلد جستجو در پایگاه است.

جدول ۲. طول عبارات جستجو

پیش از حذف جستجوهای کوتاه‌تر از ۳ حرف		پس از حذف جستجوهای کوتاه‌تر از ۳ حرف	
میانه	۲۰	میانه	۲۱
میانگین	۲۷٫۳	میانگین	۲۷٫۳
انحراف معیار	۲۳٫۶	انحراف معیار	۲۳٫۶
خطای استاندارد	۶۳٫۷۶	خطای استاندارد	۰٫۰۵۷
واریانس	۵۵۷٫۹	واریانس	۵۵۷٫۸
حداقل	۰	حداقل	۳
حداکثر	۲۵۵	حداکثر	۲۵۵
چولگی	۱۹٫۸	چولگی	۳

شکل ۹ نیز بسامد جستجوها را بر اساس طول عبارت جستجو نشان می‌دهد که مشخص است پربسامدترین جستجوها از نظر تعداد حروف، بین ۱۰ تا ۱۵ حرف داشته‌اند.



شکل ۹. بسامد جستجوها بر حسب طول عبارت جستجو (تعداد حروف)

۵-٦. استفاده از عملگرها

امکان استفاده از عملگرها در پایگاه ایرانداک وجود دارد. این عملگرها در جدول ۳ نشان داده شده‌اند. کاربر از برخی از این عملگرها، مانند علامت به‌علاوه (+) و منها () استفاده می‌کند و برخی نظیر علامت تعجب یا خط عمودی صاف توسط خود نظام در صورت استفادهٔ کاربر از گزینه‌های موجود روی سایت به‌عبارت جستجوی ثبت‌شده در لاگ اضافه می‌شود. برخی از عملگرها نیز مربوط به الگوریتم جستجوی خود پایگاه است، به‌عنوان مثال عملگر /۶ موقعی به‌عبارت جستجو اضافه می‌شود که جستجو طولانی باشد و الگوریتم با افزودن این عملگر به‌عبارت مشخص می‌کند که پایگاه فقط ۶ کلمه از عبارت جستجو را کاوش خواهد کرد. علمگری که بیشترین استفاده را داشته، علامت & است که علت آن شاید این باشد که این عملگر، عملگر پیش‌فرض در واسط جستجوی حرفه‌ای است.

جدول ۳. تعداد و درصد عملگرهای به‌کار رفته در عبارت‌های جستجو

عملگر	کارکرد و معنا	تعداد جستجو	درصد از کل جستجوها
&	عملگر بول برای اشتراک دو فیلد در جستجوی پیشرفته	۲۸۴۴۱	۱۶,۴
!	عملگر بول برای اجتماع دو فیلد در جستجوی پیشرفته	۱۵۳	۰,۰۹
*	کوتاه‌سازی در جستجوی حرفه‌ای	۱۱۳	۰,۰۷
+	عملگر بول برای اشتراک دو کلمه عبارت در جستجوی حرفه‌ای	۱۳۲۸۹	۷,۷
-	عملگر بول برای حذف یک کلمه یاعبارت از جستجو در جستجوی حرفه‌ای	۲۳۸۳	۱,۴
|	عملگر بول برای حذف (not) یک فیلد از عبارت جستجو	۲۱۳۶	۱,۲
()	پرانتز برای فرمول‌بندی عبارت جستجو	۲۶۵۷	۱,۵
6/	کدی که خود سیستم به‌عبارت می‌افزاید و در مواردی به‌کار می‌رود که عبارت جستجو طولانی باشد؛ به این معنا که پایگاه تنها ۶ کلمه از عبارت جستجو را کاوش خواهد کرد	۱۱۲۸	۰,۶

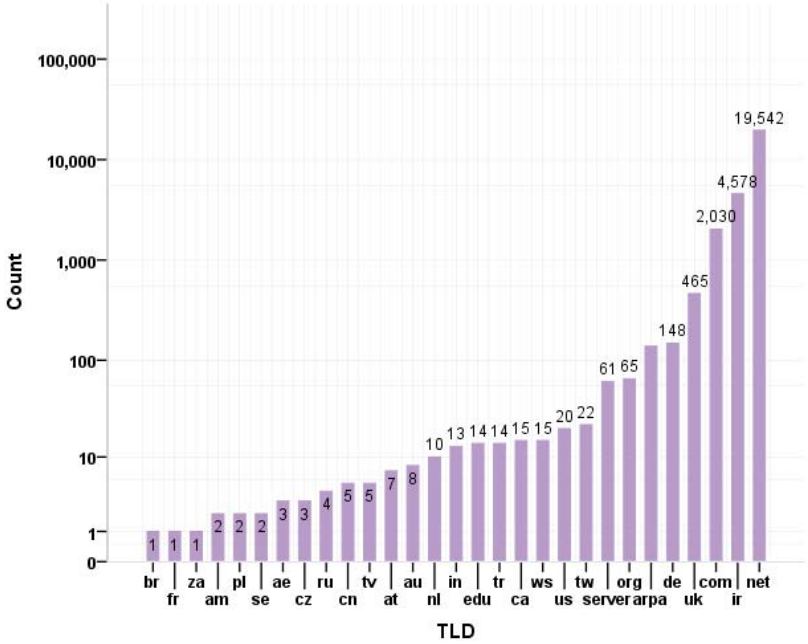
۵-۷. نشست

تحلیل نشست در مورد لاگ ایرانداک اطلاعات چندان سودمندی در اختیار ما قرار نمی‌دهد. علت اصلی آن تعداد زیاد آی‌پی‌های ناشناس و پروکسی‌هاست که مانع شناسایی کاربران واقعی می‌شود. در حقیقت، شاید هر آی‌پی معرف چند ده یا چند صد کاربر باشد. محاسبهٔ طول نشست، بدون در نظر گرفتن این نکته برای تمام آی‌پی‌ها برای هر روز به‌طور مجزا نشان می‌دهد که حداقل طول نشست صفر ثانیه، حداکثر آن نزدیک به ۲۴ ساعت و میانگین آن ۷۷ دقیقه و ۲۹ ثانیه و میانهٔ آن ۵ دقیقه و ۵۶ ثانیه است. از مجموع ۲۹۶۱۳ نشست مورد تحلیل ۶۳۰۷ نشست دارای طول صفر ثانیه (افرادی که تنها یک جستجو در روز انجام داده و ترک کرده‌اند) بوده است. در صورتی که نشست‌های طولانی‌تر از ۹۰ دقیقه را از محاسبه کنار بگذاریم، ۲۴۸۵۴ نشست باقی می‌ماند که دارای میانگین ۱۱ دقیقه و ۷ ثانیه و میانهٔ ۳ دقیقه و ۲۹ ثانیه هستند.

لازم به ذکر است که نشست تنها فاصلهٔ دو جستجو را در صورت انجام دو جستجو توسط کاربر محاسبه می‌کند و زمانی که کاربر صرف رؤیت نتایج جستجو کرده، اما جستجویی انجام نداده در سیستم ثبت نمی‌شود و راهی برای محاسبهٔ آن وجود ندارد.

۵- ۸. دامنهٔ سطح بالا

تحلیل دامنه‌های سطح بالا (شکل ۱۰) نشان می‌دهد که کاربران به‌ترتیب بیشتر از دامنه‌های دات‌نت، دات‌آی‌آر، و دات‌کام استفاده کرده‌اند. همچنین، تحلیل بیشتر نشان داد که در مجموع تنها ۷۳۱ آی‌پی متعلق به نام دامنه‌های دانشگاهی ایران (ac.ir) وجود داشت. این البته شاید دور از انتظار باشد، ولی علت آن احتمالاً این است که برخی از شماره‌های آی‌پی دانشگاهی ایران به‌عنوان آی‌پی دانشگاه ثبت نشده‌اند. در مقابل، دامنه‌های شرکت‌های سرویس‌دهندهٔ اینترنت دارای آی‌پی بیشتری در جستجوها بودند، به‌صورتی که دامنهٔ پارس‌آنلاین دارای ۳۹۲۱ و دامنهٔ رسانا دارای ۱۳۲۴۰ آی‌پی بود. تعداد ۱۴۵۸۴۴ آی‌پی نیز نامشخص بود. تنها ۱۲ آی‌پی متعلق به نام دامنه‌های دانشگاهی آمریکا (edu) بود که مربوط به دانشگاه‌های تگزاس، کلارکسون، دوک و کورنل بوده و ۱۲ آی‌پی نیز متعلق به دانشگاه‌های بریتانیا (ac.uk) بود.



شکل ۱۰. دامنه‌های سطح بالا

۵- ۹. موضوعات مورد جستجو

برای مقوله‌بندی موضوعی جستجوهای انجام‌شده، هر یک از ۱۵۰۰ جستجوی انتخاب‌شده، به‌عنوان نمونه، در پایگاه اطلاعاتی ایرانداک جستجو شدند و سپس بر اساس دسته‌بندی پایگاه اسکوپوس به اولین موضوع بازیابی‌شده، کد رده موضوعی داده شد. قابل ذکر است که تک‌واژه‌هایی مانند دیاگرام، وابستگی، ساخت، انحلال و ... که جنبهٔ عمومی دارند، در دسته عمومی، نام افراد و یا مکان‌های جغرافیایی (مثل فروغی، سحرخیز، لبنان و ...) در اسامی، و لغاتی همچون ایستگاه راه‌آهن، تئاتر، مدرسه، سینما و ... در ناشناخته‌ها قرار گرفتند. فراوانی و درصد رده‌های موضوعی جستجوهای انجام‌شده در جدول ۴ نمایش داده شده است. بیشترین جستجوهای انجام‌شده به علوم اجتماعی (۱۸/۵ درصد)، و سپس هنر و علوم انسانی (۱۳/۹ درصد) تعلق دارد.

جدول ٤. رده‌بندی موضوعی جستجوهای انجام‌شده

ردیف	موضوع	فراوانی	درصد
۱	علوم اجتماعی	۲۷۸	۱۸/۵
۲	هنر و علوم انسانی	۲۰۸	۱۳/۹
۳	مهندسی شیمی	۱۶۹	۱۱/۳
۴	ناشناخته	۱۴۹	۹/۹
۵	روانشناسی	۱۲۲	۸/۱
۶	علوم محیطی	۱۱۹	۷/۹
۷	اسامی	۱۰۱	۶/۷
۸	انرژی	۸۱	۵/۴
۹	کشاورزی و علوم زیستی	۶۰	۴
۱۰	علوم تصمیم‌گیری	۳۲	۲/۱
۱۱	علم اعصاب	۲۸	۱/۹
۱۲	علوم کامپیوتر	۲۰	۱/۳
۱۳	اقتصاد، اقتصادسنجی و مالی	۲۰	۱/۳
۱۴	حرفه‌های بهداشت	۲۰	۱/۳
۱۵	زمین‌شناسی و علم نجوم	۱۶	۱/۱
۱۶	ایمنی‌شناسی و میکروب‌شناسی	۱۵	۱
۱۷	بیوشیمی، ژنتیک و زیست مولکولی	۱۰	۰/۷
۱۸	پزشکی	۹	۰/۶
۱۹	شیمی	۷	۰/۵
۲۰	فیزیک و نجوم	۷	۰/۵
۲۱	عمومی	۶	۰/۴
۲۲	ریاضیات	۶	۰/۴
۲۳	دامپروری	۵	۰/۳۳
۲۴	علم مواد	۴	۰/۳
۲۵	داروشناسی	۴	۰/۳
۲۶	مهندسی	۳	۰/۲
۲۷	پرستاری	۱	۰/۱
۲۸	داروسازی و سم‌شناسی	۰	۰
۲۹	جمع	۱۰۰	۱۵۰۰

۵-۱۰. تحلیل خرد

در این قسمت، رفتار جستجوی دو کاربر به‌صورت جزئی و روایی مورد بررسی قرار می‌گیرد.

کاربر «الف»: کاربری از ایران با شهر نامشخص، روز جمعه ۱۱ اسفند (۱ مارس) در ساعت ۱۳:۵۸:۴۵ با استفاده از واسط جستجوی ساده، عبارت «بررسی فقهی و حقوقی شرایط ضمن عقد نکاح» را جستجو کرده و هیچ رکوردی بازیابی نمی‌کند. ۱ دقیقه و ۴ ثانیه بعد، در ساعت ۱۳:۵۹:۴۹ وی با استفاده از واسط جستجوی پیشرفته دو عبارت «بررسی فقهی و حقوقی» و «شرایط ضمن عقد» و «نکاح» را با عملگر & با هم جستجو می‌کند و ۲ رکورد بازیابی می نماید. وی ظاهراً حدود ۵ دقیقه و ۴۸ ثانیه را صرف بررسی این دو رکورد کرده است و سپس در ساعت ۱۴:۰۶:۳۷ جستجوی قبلی خود را تکرار می‌کند؛ با این تفاوت که این بار به‌جای کلمه «نکاح» از کلمه «ازدواج» استفاده می‌کند و دو رکورد دیگر بازیابی می‌نماید. وی کمتر از یک دقیقه بعد، در ساعت ۱۴:۰۷:۱۶ همان جستجوی قبلی را تکرار می‌کند و این بار به جای کلمهٔ «شرایط» از کلمه «شروط» استفاده می‌کند. در نتیجهٔ این تغییر، تعداد رکوردهای بازیابی‌شده به ۴ عدد افزایش می‌یابد. وی جستجو را رها کرده و حدود دو ساعت بعد، در ساعت ۱۶:۲۱:۱۳ مجدداً با استفاده از واسط جستجوی پیشرفتهٔ پایگاه سه کلمه «وکالت»، «زن» و «طلاق» را با عملگر & با هم ترکیب کرده و جستجو می‌کند و در نتیجهٔ این جستجو ۶ رکورد بازیابی می‌نماید.

کاربر «ب»: کاربری از ایران از منطقه پیشگامان استان گیلان در روز دوشنبه، ۷ اسفند (۲۵ فوریه) در ساعت ۱۰:۰۵ با استفاده از جستجوی سادهٔ عبارت «تأثیر حاکمیت شرکتی بر رابطه بین فرصت سرمایه‌گذاری و خط‌مشی سود» را جستجو می‌کند و نمی‌تواند رکوردی بازیابی نماید. ۵۱ ثانیه بعد، در ساعت ۱۰:۰۵:۵۱ وی جستجو را با همان عبارت، ولی با افزودن «های» به‌دنبال کلمه فرصت، تکرار کرده و ۳ رکورد بازیابی کرده است. پس از گذراندن زمانی برای وارسی رکوردهای بازیابی‌شده، در ساعت ۱۳: ۱۲: ۱۰، وی این بار با استفاده از علامت + و واسط جستجوی حرفه‌ای ادامه داده و رکوردی بازیابی نکرده است. وی بارها در ساعات متفاوت روز نیز به همین طریق تلاش نموده و نتوانسته رکوردی بازیابی نماید. در ساعت ۹ و ۲۰ ثانیهٔ روز بعد، یعنی ۸ اسفند، جستجوی عبارت مورد نظر را به زبان لاتین انجام داده و رکوردی بازیابی نکرده است.

آخرین جستجوی وی در ساعت ۱۹:۱۲:۰۲ همین روز با عبارت «خط‌مشی تقسیم سود» صورت گرفته است. جدول ۵ سیر جستجوی این کاربر را نشان می‌دهد.

جدول ۵. سیر عملکرد جستجوی کاربر «ب»

نتایج	تاریخ و زمان	عبارت جستجو
0	25/02/2013 10:05:00	تأثیر حاکمیت شرکتی بر رابطه بین فرصت سرمایه‌گذاری و خط‌مشی تقسیم سود
3	25/02/2013 10:05:51	تأثیر حاکمیت شرکتی بر رابطه بین فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 10:12:13	تأثیر + حاکمیت شرکتی بر رابطه بین فرصت سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 10:14:55	تأثیر حاکمیت + شرکتی بر رابطه بین فرصت سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 10:15:28	تأثیر حاکمیت شرکتی + بر رابطه بین فرصت سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 10:16:15	تأثیر حاکمیت شرکتی بر رابطه + بین فرصت سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 10:16:42	تأثیر حاکمیت شرکتی بر رابطه بین + فرصت سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 10:17:14	تأثیر حاکمیت شرکتی بر رابطه بین فرصت سرمایه‌گذاری + و خط‌مشی تقسیم سود
0	25/02/2013 10:17:47	تأثیر حاکمیت شرکتی بر رابطه بین فرصت سرمایه‌گذاری و خط‌مشی + تقسیم سود
7	25/02/2013 12:50:02	تأثیر حاکمیت شرکتی بر ارتباط بین فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود در شرکت‌های پذیرفته‌شده در بورس
0	25/02/2013 12:50:57	تأثیر حاکمیت شرکتی بر ارتباط بین فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود

نتایج	تاریخ و زمان	عبارت جستجو
0	25/02/2013 12:51:58	تأثیر + حاکمیت شرکتی بر ارتباط بین فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 13:06:04	تأثیر + حاکمیت شرکتی بر ارتباط بین فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 13:08:23	تأثیر حاکمیت + شرکتی بر ارتباط بین فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 13:08:44	تأثیر حاکمیت شرکتی + بر ارتباط بین فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 13:09:00	تأثیر حاکمیت شرکتی بر ارتباط + بین فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 13:09:17	تأثیر حاکمیت شرکتی بر ارتباط بین + فرصت‌های سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 13:09:51	تأثیر حاکمیت شرکتی بر ارتباط بین فرصت‌های + سرمایه‌گذاری و خط‌مشی تقسیم سود
0	25/02/2013 13:10:09	تأثیر حاکمیت شرکتی بر ارتباط بین فرصت‌های سرمایه‌گذاری + و خط‌مشی تقسیم سود
0	26/02/2013 09:00:21	The Impact of Corporate Governance on the Relationship between Investment Opportunities and Dividend Policy
0	26/02/2013 19:12:02	خط مشی تقسیم سود

٦. بحث و نتیجه‌گیری

پژوهش حاضر که به روش تحلیل لاگ روی یک پایگاه علمی ایرانی انجام گرفته، نتایجی داشته که می‌تواند برای متولیان این پایگاه و دیگر پایگاه‌های مشابه سودمند باشد. مهم‌ترین این یافته‌ها عبارتند از:

اکثر کاربران پایگاه از داخل ایران هستند، ولی در خارج از ایران نیز کاربرانی دارد که خدمت‌دهی به آنها در صورتی که جزو سیاست‌های ایرانداک باشد، قابل توسعه است. اطلاع‌رسانی بهینه در مورد پایگاه و پوشش آن و افزایش رؤیت‌پذیری خود پایگاه و محتوای آن می‌تواند کاربران بیشتری را جذب کند. در این میان خدمت‌دهی به کاربران

افغانستان و تاجیکستان می‌تواند باعث تقویت روابط علمی و نزدیک‌شدن زبان علمی این سه کشور هم‌زبان شود.

بسیاری از جستجوها به شیوه‌ای انجام‌شده که گویی عنوان یک اثر (مثلاً پایان‌نامه) عیناً در فیلد جستجو کپی شده است. اکثر این جستجوها منجر به بازیابی هیچ رکوردی نشده است. علت این امر احتمالاً این است که دانشجویان برای بررسی این که موضوع پایان‌نامهٔ خودشان تکراری هست یا خیر، عین عنوان را در پایگاه جستجو می‌کنند و احتمالاً پس از بازیابی‌نشدن هیچ رکوردی، چنین نتیجه‌گیری می‌کنند که کار آنها جدید است. بدیهی است که این شیوهٔ جستجو اشتباه می‌باشد و لازم است در این زمینه اطلاع‌رسانی و آموزش کافی صورت گیرد. این حقیقت که بیش از ۴۰ درصد جستجوها با نتیجهٔ صفر مواجه شده‌اند، نشان‌دهندهٔ ضرورت بازنگری اساسی در الگوریتم‌های جستجو و شیوهٔ خدمت‌دهی پایگاه است.

متوسط طول عبارت جستجو ۲۷ حرف و ۵/۴ واژه است. این نتیجه تقریباً با پژوهش یی و همکاران (۲۰۰۶) که میانگین طول پرسش‌ها در پایگاه سایک‌اینفو ۳/۱۶ واژه و در دو پایگاه دیگر ۳/۴۲ واژه بود، هم‌سو است.

بررسی رفتار اطلاعاتی کاربرانی که بیش از یک جستجو در یک نشست انجام داده‌اند، نشان می‌دهد که بسیاری با مفاهیم بسط یا اصلاح عبارت جستجو ناآشنا هستند و شیوهٔ صحیح استفاده از عملگرهای بولی را نیز نمی‌دانند. لازم است که ایرانداک علاوه بر تهیهٔ راهنمای مناسب برای پایگاه، برنامه‌هایی برای آموزش کاربران نیز داشته باشد.

تحلیل خرد برخی از جستجوها نشان‌دهندهٔ فایده‌ای است که استفاده از یک اصطلاحنامه به‌عنوان یک زیرسیستم در پایگاه می‌تواند داشته باشد. به‌عنوان مثال، در صورتی که کاربر کلمهٔ «نکاح» را جستجو کند و خودبه‌خود کلمهٔ «ازدواج» نیز بازیابی شود، عملکرد پایگاه بسیار بهبود خواهد یافت. با توجه به تجربهٔ غنی ایرانداک در توسعهٔ اصطلاحنامه‌ها، پیشنهاد می‌شود که نسبت به یکپارچه‌سازی آنها با موتور جستجوی پایگاه اقدام شود.

پراکندگی موضوعی جستجوهای انجام‌شده نشان می‌دهد که در برخی رشته‌ها مثل جامعه‌شناسی، روانشناسی، مهندسی، اقتصاد و مدیریت تعداد جستجوها بیش از سایر رشته‌ها، همچون علوم پایه است. علت این امر جای بررسی دارد. شاید یک علت عدم

استقبال، ناآگاهی جامعهٔ علمی برخی رشته‌ها از سودمندی پایگاه باشد، یا اینکه پایگاه واقعاً در برخی حوزه‌ها دارای محتوای غنی نیست، یا اینکه ابزار جستجوی پایگاه برای جستجوی برخی حوزه‌ها (مثل شیمی) چندان کارآمد نمی‌باشد.

تمرکز جستجوها در روزهای یکشنبه و دوشنبه و از ساعت ۱۰ الی ۱۳ است. در حالی‌که زوارقی (۱۳۸۴) نشان داده بود که سه‌شنبه‌ها پرترافیک‌ترین روز بوده است.

بررسی آی‌پی‌ها نشان‌دهندهٔ منشاء به ظاهر غیردانشگاهی بسیاری از کاربران است که احتمالاً حقیقت ندارد و علت آن شاید نابسامان‌بودن خدمات فناوری اطلاعات و اینترنت دانشگاه‌ها و نحوهٔ ثبت آی‌پی آنهاست که منجر به این می‌شود که منشأ بسیاری از رایانه‌های موجود در دانشگاه‌ها قابل انتساب به آن دانشگاه نباشد. لازم است که دانشگاه‌ها اقدام لازم را در مورد ساماندهی خدمات اینترنتی خود انجام دهند.

الگوریتم‌های جستجوی پایگاه دارای اشکالات اساسی است که به ناموفق‌بودن بسیاری از جستجوها منجر می‌شود. به‌عنوان مثال کلمات دوحرفی نادیده گرفته می‌شوند و در مورد عبارت‌های جستجوی طولانی تنها شش کلمه جستجو می‌شود. بنا به استعلامی که از ایرانداک شد، این الگوریتم هم اکنون در حال اصلاح و بهینه‌سازی است.

انجام این تحقیق نشان داد که استفاده از لاگ می‌تواند در ارزیابی و زمینه‌سازی برای بهبود یک سامانهٔ بازیابی بسیار مفید باشد. از این‌رو پیشنهاد می‌شود که ایرانداک و سایر پایگاه‌های مشابه نسبت به ثبت و نگهداری آرشیو لاگ در فواصل زمانی معقول اقدام نمایند و شیوهٔ ثبت لاگ خود را نیز بهبود بخشند. به‌عنوان مثال، در لاگ فعلی فیلد یا نوع جستجو ثبت نمی‌شود، در حالی که ثبت این اطلاعات می‌تواند زمینهٔ انجام تحلیل‌های بسیار غنی‌تری را فراهم آورد.

با وجود اینکه یافته های تحقیق، برخی از ابعاد رفتار کاربران را مشخص می‌کند، اما ماهیت این رفتارها و یا علت و چرایی آنها نیازمند انجام تحقیقات پیمایشی و کیفی بر روی کاربران است. همچنین، با توجه به اینکه تحقیق حاضر نشان‌دهندهٔ فایده احتمالی استفاده از اصطلاحنامه در پایگاه است، لازم است مطالعه‌ای برای تعیین شیوهٔ صحیح تعبیهٔ اصطلاحنامه در نظام بازیابی صورت گیرد.

۷. منابع

جمالی مهموئی، حمیدرضا. ۱۳۸۳. یادداشتی بر روش تحلیل گزارش‌های وب (وبلاگ). *اطلاع‌شناسی*. ۱ (۳): ۱۸۶-۱۷۵.

حریری، نجلا و سحر مهربان. ۱۳۹۲. زودآیند. راهبردهای جستجوی کاربران پایگاه های اطلاعاتی فناوری نانو: تحلیل گزارش تراکنش. *پژوهشنامه پردازش و مدیریت اطلاعات*. http://Jipm.irandoc.ac.ir (دسترسی در ۱۳۹۲/۵/۲).

رجبعلی بگلو، رضا. ۱۳۸۶. آشنایی با شیوه‌های تحلیل گزارش: فرایند تعامل در نظام‌های الکترونیکی. *فصلنامه کتابداری و اطلاع‌رسانی*. ۱۰ (۳۹).

زوارقی، رسول. ۱۳۸۴. تحلیل گزارش‌های وب، روشی نوین برای ارزیابی عملکرد وب‌سایت‌ها: مطالعهٔ موردی: وب‌سایت مرکز اطلاعات و مدارک علمی ایران. *اطلاع‌شناسی*. ۲ (۴-۳): ۱۱۶- ۸۸.

ستوده، هاجر. ۱۳۸۲. روش تحلیل گزارش‌های وب (وب‌لاگ). *اطلاع‌شناسی*. ۱ (۱): ۶۹-۸۴.

Jamali, H. R., D. Nicholas, and P. Huntington. 2005. Use and users of scholarly e-journals: a review of log analysis studies. *Aslib Proceedings* 57 (6): 554-571.

Jansen, B.J., A. Spink, and T. Saracevic. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing & Management*. 36 (2): 207-27.

Lambert, Frank. 2010. Web Searching to Meet Everyday Information Needs: A Comparative

Longitudinal Study of Queries. Submitted to an Online Community Information System Prato CIRN-DIAC Community Informatics Conference. 2010: Refereed Stream.

Nicholas, D, P. Huntington, and H. R. Jamali. 2008. User diversity: as demonstrated by deep log analysis. *Electronic Library*. 26 (1): 21-38.

Ozmutlu, S., H. C. Ozmutlu, and A. Spink. 2003. Are people asking questions of general Web search engines? *Online Information Review*. 27 (6): 396-406.

Park, M. and T. S. Lee. 2013. Understanding science and technology information users through transaction log analysis. *Library Hi Tech*. 31 (1): 123-140.

Peters, Thomas A. 1993. The history and development of transaction log analysis. *Library Hi Tech*. 11 (2): 41-66.

Saracevic, Tefko. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestation of relevance. *Journal of the American Society for Information Science and Technology*. 58 (13): 1915-1933.

Spink, A., and. B. J. Jansen. 2006. Searching multimedia federated content web collections. *Online Information Review*. 30 (5): 485-495.

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.

Log Analysis of the IRANDOC Database and the Analysis of Its Users' Information Seeking Behavior

Maryam Khosravi[1] | **Hamid R. Jamali Mahmuei**[2]

1. [Corresponding Author] Instructor of Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran khosravi@Irandoc.ac.ir
2. Ph. D in Knowledge & Information Science; Department of Library and Information Studies; Faculty of Psychology and Education; Kharazmi University; Iran h.jamali@gmail.com

Iranian Journal of Information Processing & Management

Iranian Research Institute for Science and Technology
ISSN 2251-8223
eISSN 2251-8231
Indexed in SCOPUS, ISC & LISA
Vol.29 | No.4 | pp: 979-1006
Summer 2014

Abstract: Log analysis or Web transactions Log are beneficial for enhancing information retrieval systems, productivity of the system for users, and understanding how users benefited from a system and their information behavior. The purpose of this paper is to analyze users' behavior, the level of their using IRANDOC database and the way they use it. In order to do this, the users log to the database from outside IRANDOC searching information was studied for a period of 24 days in Bahman and Esfand 1391. The analysis of users log includes: geographic coding analysis, statistical analysis, and microanalysis of information seeking behavior. The results indicate that users from the cities of Tehran, Mashhad and Tabriz used the database more than other cities. Sundays and Mondays between 9 to13 were the peak use days and hours. 10 to 15 search characters long were the most frequently used search items, and the most searched subject was Social Sciences. The findings showed also that some users were not familiar with concepts like search term modification and how to use Boolean operators which resulted in inefficient information retrieval. Inefficient retrieval algorithm for short words was also found in the database. It was also found that log analysis can be of use in the evaluation of an information retrieval system as well as in user behavior analysis.

Keywords: Log Analysis; Information Seeking Behavior; Information Searching; IRANDOC Database