



## دائمه‌سنجی کاربران خدمات جویش چندرسانه‌ای بومی

مریم محمودی، معصومه عظیم زاده، محمد مهدی اثنی عشری، مژگان فرهودی

گروه سکویای فناوری اطلاعات، پژوهشکده فناوری اطلاعات، مرکز تحقیقات مخابرات ایران، تهران  
{Mahmoudy, Azim\_ma, Esnaashari, Farhoodi}@itrc.ac.ir

### چکیده

یکی از ابزارهای مهم برای یافتن اطلاعات مورد نیاز کاربران در وب، جویشگرها می‌باشند. جذابیت محتوای چندرسانه‌ای و تولید و انتشار رو به رشد آن در سطح وب، سبب شده است که ارائه خدمات جویش مبتنی بر این دسته از محتوا از اهمیت ویژه‌ای برخوردار باشد. در این مقاله سعی شده است به تحلیل و دائمه‌سنجی کاربران استفاده کننده از جویشگرهای چندرسانه‌ای بومی پرداخته شود. برای این منظور رفتار کاربران ایرانی که از جویشگرهای تصویر، صوت و ویدئوی پارسی-جو استفاده نموده‌اند به کمک تحلیل فایل ثبت وقایع (log) این جویشگر بررسی شده است. تحلیل‌ها در سطح پرس‌وجوی کاربران صورت گرفته است. نتایج به دست آمده حاکی از آن است که کاربران در هنگام به کارگیری جویشگرهای چندرسانه‌ای، علاقمندی بیشتری به موضوعات تفریحی و سرگرم کننده دارند.

### کلمات کلیدی

جویشگرهای چندرسانه‌ای، تحلیل رفتار کاربران، فایل ثبت وقایع

### ۱- مقدمه

ارتقاء روز افزون توانمندی‌های سخت‌افزارها و افزایش مداوم پهنای باند دسترسی به اینترنت سبب گردیده است که گرایش کاربران به سمت استفاده از خدمات چندرسانه‌ای برخط روند رو به رشد، همراه با شتاب فراوان، داشته باشد [۵]. همچنین، حجم محتویات چندرسانه‌ای که به صورت برخط قابل دسترسی هستند نیز به صورت مداوم و به شدت در حال افزایش است. همین مسأله موجب شده است که جویشگرها در سطح جهان به سمت ارائه خدمات ویژه جستجوی چندرسانه‌ای، نظیر جستجوی تصویر، صوت و ویدئو حرکت کنند.

از سوی دیگر، با توجه به حجم بالای محتویات چندرسانه‌ای، ارائه خدمت جویش برای این دسته از محتویات با محدودیت‌هایی، عمدتاً از نوع سخت‌افزاری و پهنای باند، مواجه است [۶]. از این رو، جویشگرها باید بتوانند با دائمه‌سنجی کاربران خدمات جویش چندرسانه‌ای خود، نمایه‌سازی محتویات چندرسانه‌ای موجود در سطح وب را بر این اساس اولویت‌بندی نمایند تا ضمن غلبه بر محدودیت‌های سخت‌افزاری و پهنای باند خود، پاسخگویی مناسبی را نیز به پرس‌وجوهای کاربران داشته باشند.

دائمه‌سنجی کاربران خدمات جویش تا کنون در مقالات متعددی مورد بررسی قرار گرفته است [۷-۱۴]. همچنین، برای جویشگر بومی پارسی‌جو نیز، تا کنون تحقیقاتی متعددی در راستای رفتارشناسی و دائمه‌سنجی کاربران این جویشگر صورت پذیرفته است [۱-۴]. در تمامی این تحقیقات، یکی از دو روش زیر برای بررسی رفتار کاربران جویشگرها و دائمه‌سنجی آنها مورد استفاده قرار گرفته است: (۱) استفاده از فایل ثبت وقایع جویشگر [۲-۴] و [۷-۱۱] و (۲) بررسی رفتار ضمنی کاربر [۱] [۱۲-۱۴]. نکته مشترک در تمامی این مقالات آن است که دائمه‌سنجی‌ها در آنها بر اساس استفاده کاربران از جویشگرهای متنی تحلیل شده است.

در زمینه دائمه‌سنجی کاربران خدمات جویش چندرسانه‌ای تحقیقات چندان زیادی صورت نپذیرفته است. به عنوان نمونه، محققین در [۱۵] نشان دادند که بیش از ۲۵ درصد واژگان

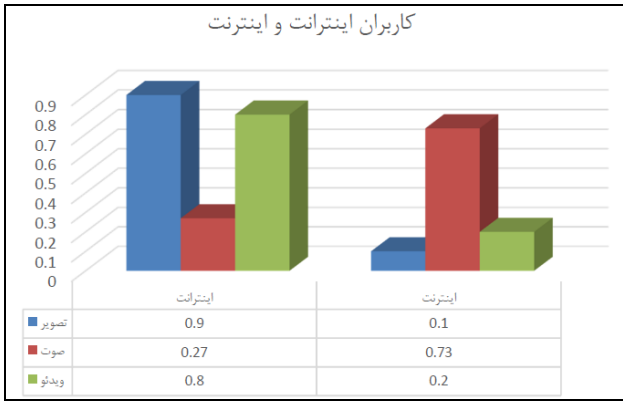
استفاده شده توسط کاربران جویشگر تصویری Excite، در دایره لغات غیراخلاقی با تنوع بسیار بالا هستند. همچنین، در [۱۶] نشان داده شده که کاربران برای پرس‌وجوهای غیراخلاقی عموماً صفحات بیشتری از نتایج را بررسی می‌کنند. در مقاله [۱۷] نویسندگان به بررسی رفتار کاربران فراجویشگر چندرسانه‌ای Dogpile در سال ۲۰۰۶ میلادی پرداخته‌اند. نتایج این بررسی نشان داده که بیش از ۵۰ درصد جستجوهای چندرسانه‌ای، جستجوی تصویری، ۲۸ درصد ویدئویی و ۲۲ درصد صوتی هستند. تمامی پرس‌وجوهای پرکاربرد تصویری و ویدئویی از نوع غیراخلاقی بوده‌اند و بیشتر پرس‌وجوهای پرکاربرد صوتی شامل نام آهنگ یا خواننده آهنگ بوده است. در مقاله [۱۸] با بررسی جویشگر یاهو نشان داده شده که نوجوانان در رده سنی ۱۰ تا ۱۹ سال بیش از ۲/۴ برابر بزرگسالان از خدمات جویش چندرسانه‌ای استفاده می‌کنند.

تا جایی که نگارندگان این مقاله اطلاع دارند، در زمینه دائمه‌سنجی کاربران خدمات جویش چندرسانه‌ای بومی تا کنون تحقیقی انجام نپذیرفته و مقاله حاضر، اولین مورد در این زمینه خواهد بود. در این مقاله، با استفاده از فایل ثبت وقایع جویشگر بومی پارسی‌جو، برای یک بازه زمانی سه ماهه، سعی شده است تحلیلی بر رفتار کاربران خدمات جویش چندرسانه‌ای جویشگر پارسی‌جو صورت پذیرد. تحلیل‌های صورت پذیرفته می‌تواند به این جویشگر در اولویت‌بندی خزش و نمایه‌گذاری محتویات چندرسانه‌ای موجود در سطح اینترنت کمک نماید.

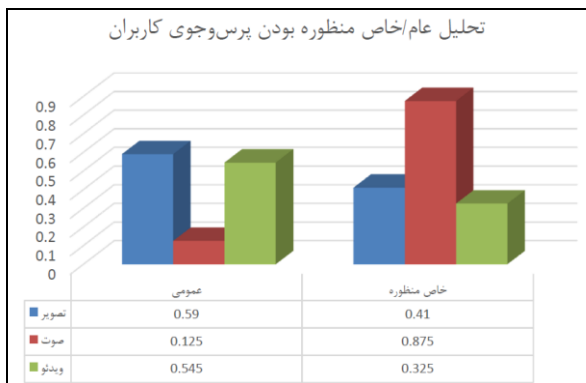
ادامه این مقاله به صورت زیر سازماندهی شده است. در بخش دوم مراحل مورد استفاده جهت انجام تحلیل‌ها شرح داده خواهد شد. بخش سوم به ارائه و بررسی نتایج حاصل از تحلیل‌های انجام شده خواهیم پرداخت. بخش چهارم شامل جمع‌بندی مقاله خواهد بود.

### ۲- تحلیل نتایج

در این بخش تحلیل نتایج ارزیابی جویشگر فارسی چندرسانه‌ای در بازه زمانی مرداد ۹۴ تا آبان ۹۴ ارائه می‌گردد.



از جمله مباحث دیگری که در این پژوهش به دنبال درک آن بودیم، نحوه ساخت پرس‌وجو توسط کاربران در خاص نمودن و دقیق نمودن نیاز اطلاعاتی خود بود. طبق بررسی که انجام گرفت مشخص شد نحوه برخورد کاربران در خدمات جویش مختلف متفاوت است به طوری که کاربران در هنگام استفاده از خدمت جویش صوتی نسبت دو خدمت دیگر یعنی تصویری و ویدئویی پرس‌وجوهای دقیق‌تری را بکار می‌گیرند. اما در هنگام کاربری خدمات جویش ویدئو و تصویری از واژه‌های عام‌منظوره بیشتر استفاده می‌کنند. شکل ۲-۴ نتایج این تحلیل را نشان می‌دهد.



شکل ۲-۴: نمودار تحلیل نحوه بیان دقیق نیاز اطلاعاتی کاربران

## ۲-۲- تصویر

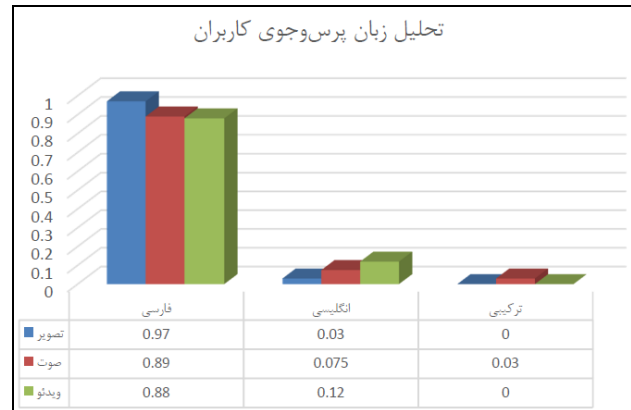
سرویس جستجوی تصویری امکان جستجوی تصاویری که عمدتاً از صفحات فارسی استخراج شده‌اند را فراهم می‌کند.

تحلیل موضوعی پرس‌وجوهای مورد استفاده توسط کاربران جویشگر تصویری پارسی‌جو، که نتایج آن در شکل ۲-۵ دیده می‌شود، نشان می‌دهد که سه دسته موضوعی که به صورت ویژه مورد علاقه کاربران این خدمت چندرسانه‌ای قرار گرفته است به ترتیب «مد و فشن»، «موضوعات هنری» و «اشخاص سرشناس» است. در میان پرس‌وجوهای دسته موضوعی اول، یعنی «مد و فشن»، بیشترین فراوانی مربوط به پرس‌وجوهای بانوان، نظیر مانتو، لباس مجلسی و مواردی از این دست اختصاص دارد. پرس‌وجوهای هنری نیز در ظاهر بیشتر مورد علاقه بانوان است. از جمله پرس‌وجوهایی از قبیل آموزش بافتنی، گل‌های کریستال و ... در این دسته فراوان دیده می‌شوند. در دسته موضوعی اشخاص سرشناس نیز بیشترین پرس‌وجوها برای یافتن تصاویر بازیگران، خصوصاً بازیگران زن ایرانی بوده است.

از سوی دیگر، کم کاربردترین دسته‌های موضوعی مشاهده شده به ترتیب سه دسته «ورزشی»، «وسایل نقلیه» و «اماکن» هستند. با توجه به آنکه موضوعات ورزشی و وسایل نقلیه بیشتر مورد توجه مردان هستند تا بانوان، می‌توان این احتمال را مطرح نمود که درصد بیشتری از کاربران جویشگر تصویری پارسی‌جو را بانوان تشکیل می‌دهند. البته این تنها یک احتمال است که اثبات آن نیازمند تحلیل‌های بیشتر و دقیق‌تری است. پایین بودن تعداد پرس‌وجوهای مربوط به اماکن، خصوصاً اماکن تاریخی و دیدنی هم، با توجه به آنکه بسیاری از پرس‌وجوهای تصویری به منظور یافتن تصاویری از چنین مکان‌هایی است، تا حدی دور از انتظار بوده است. شاید هم دلیل این مسأله مناسب‌تر بودن، یا غنی‌تر بودن نتایج حاصل از این دسته از

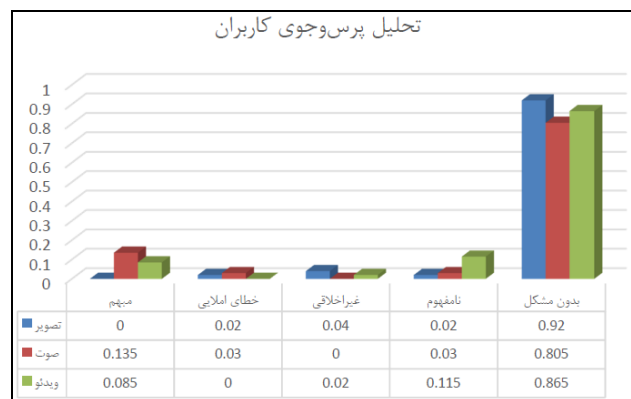
## ۱-۲- تحلیل‌های عمومی

در شکل ۱-۲ درصد پرس و جویها به تفکیک زبان فارسی و انگلیسی -به ازای هر نوع خدمت جویش آورده شده است. مطابق شکل می‌توان دریافت که کاربران بومی در هنگام ساخت پرس‌وجوی خود بیشتر از زبان فارسی استفاده می‌کنند و کمتر از ۱۰ درصد موارد از زبان انگلیسی استفاده می‌کنند. نکته جالب توجه این است که با بررسی که بر روی ۲۰۰ پرس و جوی پرفرکانس خدمات تصویر، صوت و ویدئو انجام گرفت کاربران فقط در استفاده از خدمت جویش صوتی از دو زبان فارسی و انگلیسی به صورت ترکیبی بهره می‌گیرند. همچنین با بررسی که بر این گروه از پرس‌وجوها انجام گرفت نشان می‌دهد که کاربران در واقع در هنگام جستجوی فایل‌های صوتی غیرفارسی به این نحو عمل می‌کنند.



شکل ۱-۲: تحلیل زبان پرس‌وجوی کاربران

در این پژوهش بررسی بر روی میزان کیفیت پرس‌وجوهای کاربران انجام گرفت. به همین منظور، مشکلات پرس‌وجوها از جنبه مبهم بودن، خطای املائی، غیراخلاقی بودن، نامفهوم بودن پرس‌وجوها تعیین شد و در نهایت میزان پرس و جویهای بدون مشکل مشخص شد. همان‌گونه که در شکل ۲-۲ نشان داده شده در هر سه نوع خدمت جویش کمتر از ۲۰ درصد پرس‌وجوها مشکل‌دارند که بیشتر آنها دارای مشکل ابهام و عدم شفافیت می‌باشند. بعد از آن پرس و جویهای نامفهوم بیشترین آمار را داشتند که بیانگر بی‌هدفی کاربر در رفع نیاز اطلاعاتی خاصی است و ممکن است به دنبال مقاصدی غیر از جویش همچون حمله باشد.



شکل ۲-۲: نمودار تحلیل مشکل‌دار بودن یا نبودن پرس‌وجوی کاربران

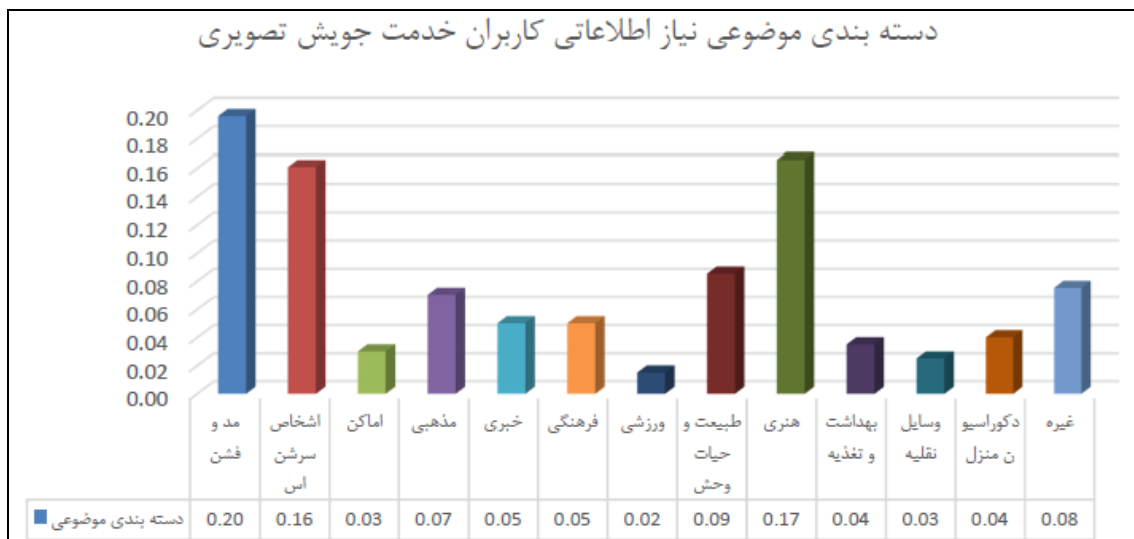
یکی از نتایج دیگری که در این بررسی ملموس گردید میزان استقبال کاربران اینترنت و اینترنت از هر یک از انواع خدمات جویش چندرسانه‌ای است. همانطور که در نمودار نمایش داده شده است (شکل ۲-۳) خدمات جویش تصویری و ویدئویی بیشتر برای کاربران اینترنت کاربردپذیری داشته اما کاربران اینترنت استقبال بیشتری در بکارگیری خدمت جویش صوتی داشتند. شاید دلیل استقبال از خدمت جویش صوتی توسط کاربران اینترنت عدم وجود سرویس مشابه داخلی در این زمینه باشد.

۳-۲- صوتی

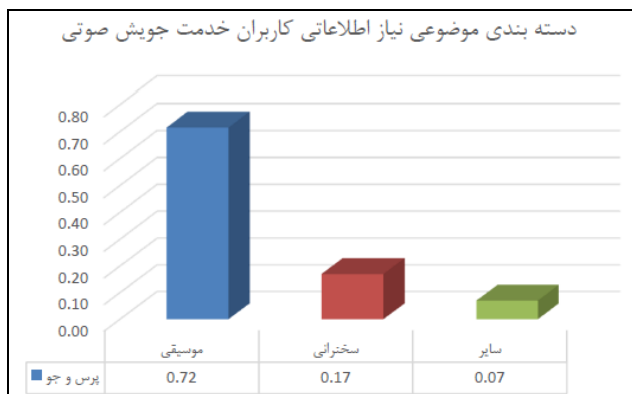
این خدمت امکان جستجو، دانلود، و پخش صوت و موسیقی را در اختیار کاربران قرار می‌دهد. یکی از تحلیل‌های صورت گرفته در این پژوهش بررسی نحوه بیان نیاز اطلاعاتی کاربران در هنگام بکارگیری خدمت جویش صوتی می‌باشد. هدف از این بررسی درک این مسئله است که کاربران در زمان ساخت پرس‌وجوی خود چه نوع اطلاعاتی را به خاطر دارند و به عنوان کلیدواژه در پرس‌وجوی خود به کار می‌گیرند. همان‌گونه که در شکل ۲-۶ نشان داده شده است، اکثر کاربران از نام هنرمند در پرس‌وجوی خود استفاده می‌کنند. بعد از آن نام آهنگ بیشترین کاربرد را دارد. همچنین کاربران در هنگام بیان نیاز اطلاعاتی خود از نام آلبوم، سبک موسیقی و حتی عباراتی از متن شعر موسیقی استفاده می‌کنند. در شکل ۲-۶ زیر میزان کاربردپذیری هر یک از این دسته‌های اطلاعاتی نمایش داده شده است.

پرس‌وجوها روی جویشگرهایی نظیر گوگل و بینگ، نسبت به جویشگر تصویری پارسی‌جو، بوده باشد.

در نهایت، تعداد نسبتاً قابل توجهی از پرس‌وجوها نیز در دسته «سایر موارد» یا «غیره» قرار گرفته‌اند. این پرس‌وجوها مواردی بوده‌اند که بیش از حد عمومی بوده‌اند و امکان قرار دادن آنها در دسته خاصی وجود نداشته است، نظیر «عکس‌های خنده‌دار»، «عکس فان‌تزی خوشگل» و «عکس‌های جالب و دیدنی». این دسته از پرس‌وجوها بیانگر آن هستند که کاربر، خود نیز اطمینان نداشته که دقیقاً به دنبال چه تصویری می‌گردد و تنها سعی نموده با استفاده از برخی واژگان عمومی، خروجی‌های جویشگر را بررسی نماید. بالا بودن تعداد پرس‌وجوها در این دسته می‌تواند نشانگر اعتماد نسبتاً پایین کاربران به جویشگر باشد که نه برای رفع نیاز اطلاعاتی خود، بلکه بیشتر با هدف بررسی نتایج حاصل از جویشگر، اقدام به ارائه پرس‌وجوهایی بدون هدف مشخص نموده‌اند.



شکل ۲-۵: دسته‌بندی موضوعی نیاز اطلاعاتی کاربران خدمت جویش تصویری

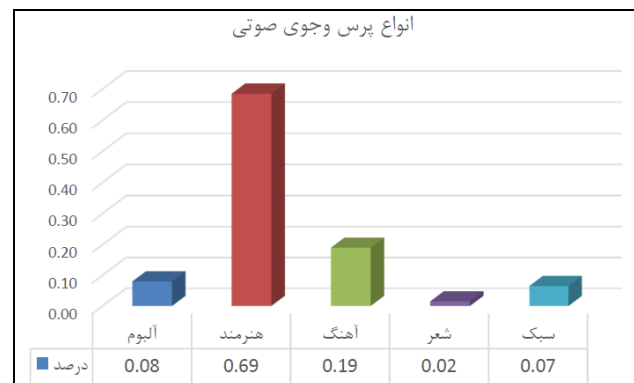


شکل ۲-۷: دسته‌بندی موضوعی نیاز اطلاعاتی کاربران خدمت جویش صوتی

۴-۲- ویدئو

خدمت جویش ویدئو، امکان جستجوی فایل‌های ویدئویی که عمدتاً از صفحات فارسی استخراج شده‌اند را فراهم می‌کند.

بررسی‌های انجام شده در بخش جویشگر ویدئو، که نتایج آن در شکل ۲-۸ مشاهده می‌شود، نشان می‌دهد که کاربران در زمینه‌های متنوعی به جستجو می‌پردازند. اما بیشترین نیاز اطلاعاتی کاربران در این حوزه موضوعات سرگرمی نظیر فیلم، کنسرت یا اجرای زنده موسیقی، کلیپ‌های طنز و خنده‌دار و غیره می‌باشد که حدوداً ۳۱ درصد از پرس‌وجوها را در برمی‌گیرد. دسته موضوعی بعدی که بعد از موضوعات سرگرمی مورد توجه قرار گرفته است، حوزه عمومی با سهم ۱۵ درصد است که به نظر می‌رسد بیشتر کاربران به دنبال رفتارسنجی جویشگر بوده‌اند تا نیاز اطلاعاتی خاصی را مطرح کرده باشند که در این

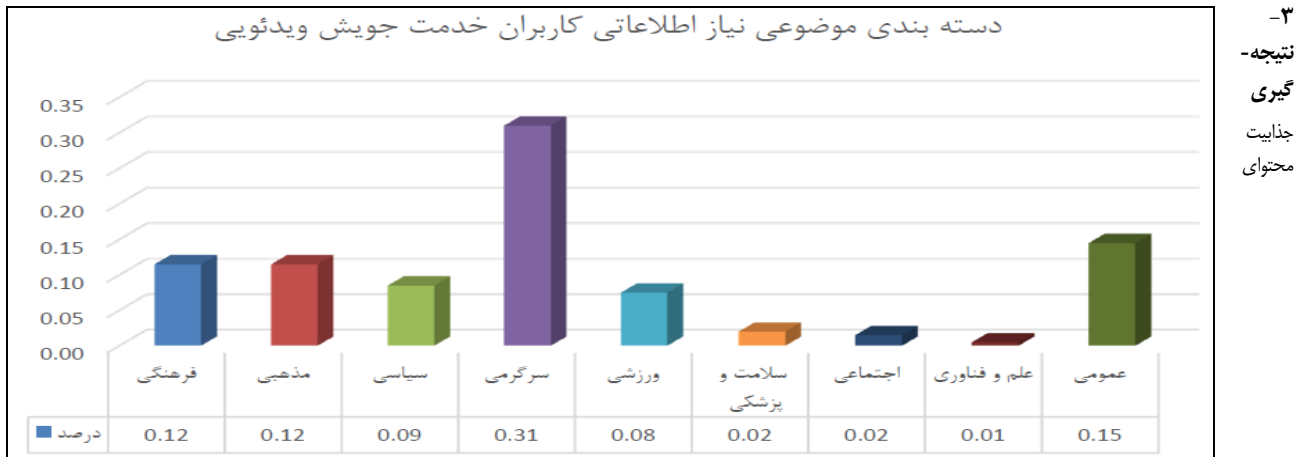


شکل ۲-۶: انواع پرس‌وجوهای خدمت جویش صوتی

از دیگر اهداف بررسی نیاز اطلاعاتی کاربران خدمت جویش صوتی شناسایی دسته‌های موضوعی مورد علاقه آنها شامل موسیقی، سخنرانی، و غیره می‌باشد. همانطور که شکل ۲-۷ نشان می‌دهد می‌توان دریافت که فایل‌های صوتی موسیقی بیشتر مورد درخواست کاربران بوده است تا دیگر انواع فایل‌های صوتی.

چندسازمانی برای کاربران وب سبب شده تا ارائه خدمات جویش مبتنی بر این دسته از محتوای الکترونیکی مورد توجه ارائه‌دهندگان خدمات قرار گیرد. اما از طرفی حجم بالای محتویات چندسازمانی از جمله چالش‌های این حوزه تلقی می‌شود. بنابراین یکی از اقدامات مهم توسط ارائه‌دهندگان خدمات جویش چندسازمانی شناسایی نیازمندی‌های اطلاعاتی کاربران و ذائقه‌سنجی آنهاست. بدین ترتیب، این خدمات

دسته پرس و جوهای نظیر فیلم، ویدئو جدید و غیره می‌باشد. بعد از آن موضوعات مذهبی و فرهنگی با سهم حدود ۱۲ درصد از موضوعات پرس و جوها مورد استقبال کاربران جویشگر بومی قرار گرفته است. موضوعات سیاسی و «سلامت و پزشکی» نیز با سهم ۹ و ۸ درصد در رده‌های بعدی استقبال کاربران قرار گرفته است. نکته قابل توجهی که در بخش تحلیل موضوعی جویشگر ویدئویی وجود دارد عدم تمایل کاربران در استفاده از جویشگر ویدئویی بومی برای اهداف علم و فناوری می‌باشد.



شکل ۳-۱: دسته‌بندی موضوعی نیاز اطلاعاتی کاربران خدمت جویش ویدئویی

- S. Park, J. H. Lee, and H. J. Bae, "End User Searching: A Weblog Analysis of NAVER, A Korean Web Search Engine," *Library and Information Science Research*, Vol. 27, No. 2, 2005, pp. 203-221.
- C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a Very Large Web Search Engine Query Log," *ACM SIGIR Forum*, Vol. 33, No. 1, 1999, pp. 6-12.
- M. Costa and J. S. Mario, "A Search Log Analysis of a Portuguese Web Search Engine," *2<sup>nd</sup> INForum-Simposio de Informatica*, pp. 525-536, 2010.
- A. Spink, S. Ozmutlu, H. C. Ozmutlu, and B. J. Jansen, "U.S. vs European Web Searching Trends," *ACM SIGIR Forum*, Vol. 36, No. 2, 2002, pp. 32-38.
- T. Joachims, "Optimizing Search Engines using Click-through Data," *ACM Conf. on Knowledge Discovery and Data Mining*, 2002.
- T. Joachims, L. Granka, B. Pang, H. Hembrooke, F. Radlinski, and G. Gay, "Accurately Interpreting Click-through Data as Implicit Feedback," *ACM Conf. on Research and Development on Information Retrieval*, 2005.
- T. Joachims, L. Granka, B. Pang, H. Hembrooke, F. Radlinski, and G. Gay, "Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search," *ACM Trans. on Information Systems*, Vol. 25, No. 2, Article 7, 2007.
- B. J. Jansen, A. Spink, and T. Saracevic, "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web," *Information Processing and Management*, Vol. 36, No. 2, 2000, pp. 207-227.
- A. Spink, H. C. Ozmutlu, and D. P. Lorence, "Web Searching for Sexual Information: An Exploratory Study," *Information Processing and Management*, Vol. 40, No. 1, 2004, pp. 113-124.
- D. Tjondronegoro, A. Spink, B. J. Jansen, "Multimedia Web Searching on a Meta-Search Engine," *12<sup>th</sup> Australasian Document Computing Symposium*, Australia, 2007.
- S. D. Torres, I. Weber, and D. Hiemstra, "Analysis of Search and Browsing of Young Users on the Web," *ACM Trans. on the Web*, Vol. 8, No. 2, Article 7, 2014.

دهندگان قادر خواهند بود ضمن استفاده بهینه از منابع سخت‌افزاری، با شناسایی محتوای مورد نظر کاربران و اولویت‌بندی مناسب آنها، در جلب رضایت کاربران خود موفق‌تر گردند. در این پژوهش نیز سعی شد گرایشات کاربران در هنگام استفاده از خدمات جویش چندسازمانی، نظیر جستجوی تصویر، صوت و ویدئو مورد بررسی قرار گیرد. نتایج این بررسی‌ها بیانگر این موضوع است که کاربران فارسی‌زبان معمولاً در بیان نیاز اطلاعاتی خود از زبان فارسی استفاده می‌کنند. همچنین بررسی‌هایی که بر روی پرس‌وجوی کاربران انجام گرفت نشان داد که اکثر کاربران با هدف سرگرمی و تفریح به جویشگرهای چندسازمانی مراجعه می‌کنند.

#### مراجع

- معصومه عظیم‌زاده، نوید فرهادی، محمد مهدی اثنی‌عشری، "تحلیل رضایت کاربر مبتنی بر رفتار ضمنی در حین جستجو"، کنفرانس بین‌المللی وب پژوهی، فروردین ۹۴.
- محمد صادق زاهدی، معصومه عظیم‌زاده، مریم محمودی، رضا بدیع، "ذائقه‌سنجی و تحلیل رفتار کاربران سیار در موتور جستجوی بومی مبتنی بر تحلیل فایل ثبت وقایع"، چهارمین کنفرانس بین‌المللی مدیریت فناوری اطلاعات، ارتباطات و کامپیوتر، تهران، تیر ۹۳.
- معصومه عظیم‌زاده، مریم محمودی، محمد مهدی اثنی‌عشری، "تحلیل داده‌های عظیم ثبت وقایع موتور جستجوی بومی با هدف رفتارشناسی و ذائقه‌سنجی کاربران فارسی‌زبان"، همایش داده‌های عظیم، مرکز تحقیقات مخابرات ایران، تهران، دی ۹۳.
- محمد صادق زاهدی، معصومه عظیم‌زاده، نوید فرهادی، علی محمد زارع بیدکی، "تحلیل رفتار کاربران فارسی‌زبان در موتور جستجوی بومی"، نوزدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، اسفند ۹۲.
- P. Daras, S. Manolopoulou, and A. Axenopoulos, "Search and Retrieval of Rich Media Objects Supporting Multiple Multimodal Queries," *IEEE Trans. on Multimedia*, Vol. 14, No. 3, 2012, pp. 734-746.
- M. J. Swain, "Image and Video Searching on the World Wide Web," *2<sup>nd</sup> UK Conf. on Image Retrieval*, 1999.
- K. Markey, "Twenty-five Years of End-User Searching, Part 1: Research Findings," *American Society for Information Science and Technology*, Vol. 58, No. 8, 2007, pp. 1071-1081.