



پیشنهاد ایجاد سامانه بومی رتبه‌بندی وبگاهها

لیلا ربیعی^۱، مریم امیری^۲، مریم باقری^۳

گروه مدیریت یکپارچه شبکه، پژوهشکده فناوری ارتباطات، پژوهشگاه ارتباطات و فناوری اطلاعات (مرکز تحقیقات مخابرات ایران)، تهران

l.rabiei@itrc.ac.ir^۱

amiri.maryam@aut.ac.ir^۲

m.bagheri@itrc.ac.ir^۳

چکیده

پژوهش حاضر با هدف تعیین ضرورت نیاز به طراحی و پیاده‌سازی سامانه بومی رتبه‌بندی وبگاهها انجام پذیرفت. در حال حاضر نرم‌افزارها و خدمات غیر بومی برای تحلیل و رتبه‌بندی وبگاهها در کشور توسط مدیران وبگاهها و کاربران اینترنت مورد استفاده قرار گرفته است و طرفداران بسیاری دارد. این در حالی است که این نرم‌افزارها و خدمات، علاوه بر به چالش انداختن امنیت داده‌های داخلی و خروج سرمایه ملی از کشور، اطلاعات دقیقی به کاربران ارائه ندادند و نقاط ضعفی برای به وجود آمدن امکان تقلب در آن‌ها وجود دارد. شناسایی و بررسی عملکرد این نرم‌افزارها، در این مقاله انجام شده است. با توجه به روی کار آمدن جویشگران بومی فرصتی برای جمع‌آوری داده‌های مورد نیاز برای ارائه هرچه دقیق‌تر رتبه‌بندی وبگاهها به وجود آمده که این موضوع به همراه شناسایی و بررسی سایر منابع داده‌ای سامانه بومی رتبه‌بندی انجام شده است. گام‌های اصلی در طراحی سامانه، شامل استخراج مقادیر شاخص‌های رتبه‌بندی از منابع داده‌ای و ذخیره آنها در بانک اطلاعاتی، پردازش روی شاخص‌ها و ارائه رتبه وبگاه، شناسایی شده و در رابطه با روش پیاده‌سازی آن به نحوی که بتوان در سریع‌ترین زمان از سامانه رتبه‌بندی بهره‌برداری نمود، بحث شده است.

کلمات کلیدی

وبگاه، جویشگر بومی، سامانه رتبه بند بومی، تحلیلگر وب، بر وبگاه، برون وبگاه، شاخص‌های رتبه‌بندی، شاخص‌های تحلیل وب، منابع داده‌ای، کسب‌وکارهای اینترنتی

۱- مقدمه

همزمان با آغاز رشد وب فارسی و به‌راه افتادن کسب‌وکارهای اینترنتی متنوع و متعدد در کشور، مدیران وبگاهها همواره به دنبال شاخصی برای سنجش ارجحیت وبگاهشان بوده‌اند. رتبه‌بندی یک وبگاه^۱ از حیث درک میزان قابلیت اعتماد و همچنین رونق کسب‌وکار اینترنتی ایجاد شده به واسطه‌ی آن، حائز اهمیت است. شرکت الکسا، شرکتی در کالیفرنیا است که داده‌ها و تحلیل‌های ترافیک وب را به‌صورت رایگان و تجاری ارائه می‌کند. این شرکت در وبگاه خود داده‌های مربوط به کاربران و بازدیدکنندگان وبگاهها را ذخیره کرده و تحلیل‌هایی را بر روی آن‌ها انجام داده و این وبگاهها را رتبه‌بندی می‌نماید.

وبگاه‌های توسعه یافته در کشور را می‌توان به دو بخش، شامل وبگاه‌های ارائه دهنده هرگونه محتوا و وبگاه‌های درآمدزا تقسیم بندی کرد. در حال حاضر رتبه الکسا برای مدیران وبگاهها و کاربران آنها از حیث اعتبار وبگاه و محتوای آن، اهمیت دارد؛ اما بیشترین دغدغه برای رتبه الکسا از آن وبگاه‌های درآمدزا و در واقع همان کسب‌وکارهای اینترنتی در داخل کشور است.

امروزه مهم‌ترین عامل رونق کسب‌وکارهای اینترنتی در ایران تبلیغات در وب است. بسیاری از شرکت‌های ایرانی، تنها به رتبه الکسای وبگاه مورد نظر که قصد تبلیغ در آن را دارند، اکتفا می‌کنند. همین موضوع باعث شده است تا الکسا و کسب رتبه‌ی بهتر در آن، برای صاحبان

وبگاهها اهمیت ویژه‌ای پیدا کند. هرچند رتبه‌ای که الکسا ارائه می‌دهد به دلیل نمونه برداری غیر دقیق از داده‌ها چندان دقیق نیست اما با این حال راه منطقی کسب رتبه بهتر در الکسا، جذب مخاطب و افزایش ترافیک از طریق تولید محتوای غنی و کاربردی است؛ اما از آنجایی که بسیاری از مردم همواره بدنال میان‌بر هستند، ده‌ها سرویس و روش غیرمنطقی شکل گرفته است که از طریق آن‌ها مدیران وبگاهها تلاش می‌کنند تا رتبه الکسای وبگاه خود را به شکل غیر واقعی بهبود بخشند. بر اساس آمار الکسا [1] حدود شش درصد از کاربران الکسا از ایران هستند و ایران در لیست برترین ملیت‌های مخاطب الکسا، در رتبه پنجم قرار دارد. بنابراین به نظر می‌رسد با راه‌اندازی سامانه بومی رتبه‌بندی وبگاهها و جایگزین نمودن آن با الکسا و امثال آن در کشور، علاوه بر ارائه آمار و مستندات دقیقتر از وضعیت ترافیکی وب وبگاه‌های داخلی و پوشش نقاط ضعف الکسا و راه‌های تقلب در آن، می‌توان از رشد کاذب و غیر اصولی وبگاهها و کسب‌وکارهای اینترنتی در ایران جلوگیری نمود؛ بدین ترتیب مبنای ارزش یک وبگاه رتبه‌ای خواهد بود که سامانه بومی رتبه‌بندی وبگاهها به آن اختصاص داده است. همچنین با حفظ امنیت داده‌های کشور و سرمایه ملی، در آمد حاصل از استفاده از سامانه بومی روانه شرکت‌های داخلی خواهد شد.

از طرفی جویشگران بومی به منظور ابلاغ موثر ارزش پیشنهادی خود به کاربران و جلب توجه آنها به کیفیت و سایر مشخصه‌های بازاری که مزیت رقابتی آنها را در مقایسه با سایر بازیگران بنام حاضر در بازار تشکیل می‌دهد، نیازمند خلق تصویری قدرتمند از خود در محیط مجازی

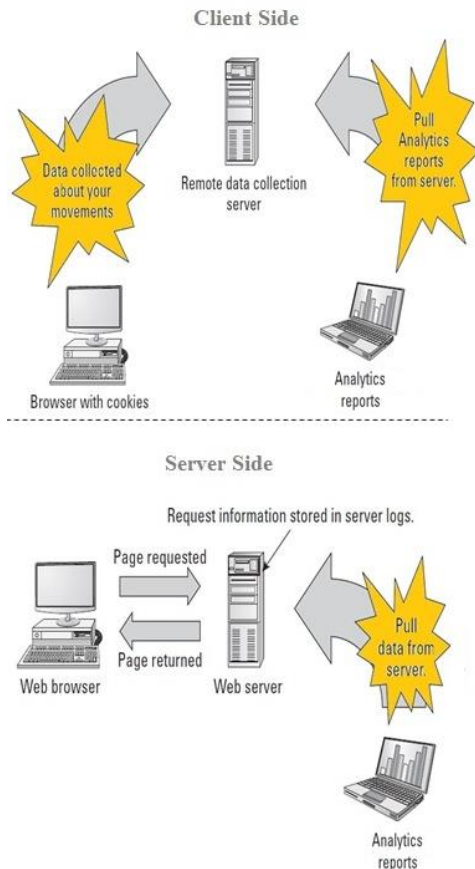
فیلتر کردن Robotها و امثال آنها که در یک دوره زمانی مشخص حداقل یک بار عمل بازدید از وبگاه را انجام دادند، بازدیدکننده‌های منحصر بفرد گفته می‌شود. طبق این تعریف بازدید چندباره یک کاربر خاص از وبگاه در دوره زمانی تعریف شده فقط یک بار شمارش می‌گردد.

شاخص‌های رتبه‌بندی بر مبنای ارجاعات صفحه مطابق [4] شامل شاخص لینک‌های خروج از صفحه (Outbound links from pages)، لینک‌های ورود به صفحه (Input links to pages)، رتبه صفحاتی که به صفحه مورد نظر لینک داده‌اند (PR) و شاخص تعدیل (damping factor) عددی بین صفر و یک است که معادل احتمال اینکه کاربر بازدیدکننده از یک صفحه، صفحه بعدی را از لینک‌های آن صفحه انتخاب کند، در نظر گرفته می‌شود.

۲-۴- منابع داده‌ای برای تحلیل و رتبه‌بندی وب

منابع داده‌ای جهت اندازه‌گیری شاخص‌های تحلیل ترافیکی وبگاه‌ها که جهت رتبه‌بندی آنها نیز بکار می‌روند، به دو دسته کلی شامل منابع داده‌ای بر وبگاه (On-Site) و منابع داده‌ای برون وبگاه (Off-Site) تقسیم‌بندی می‌شوند.

هرگونه ابزاری که بتوان مقادیر شاخص‌ها را از طریق آن استخراج نمود و این داده‌ها مستخرج از خود وبگاهی باشند که در حال تحلیل آن هستیم، منبع داده‌ای بر وبگاه خوانده می‌شود. منبع داده‌ای بر وبگاه به دو نوع سمت کاربر (Client Side) و سمت سرور (Server Side) تقسیم شده است [5][2]. ابزارهایی مانند کوکی‌ها، برچسب‌ها، اسکرپت‌ها و اطلاعات ثبت‌شده در وب سرورها (Server Logs) از جمله منابع داده‌ای بر وبگاه هستند. شکل (۱) نمایشگر نحوه استخراج داده از در هریک از انواع منابع داده‌ای بر وبگاه است.



شکل (۱): انواع منابع داده‌ای بر وبگاه

ابزارهای دیگری که بتوان داده‌های مربوط به وبگاه‌ها را بدون هیچ ارتباطی با سرور آنها، جهت مقاردهی به شاخص‌های تحلیل و رتبه‌بندی وبگاه‌ها، استخراج نمود، منابع داده‌ای برون وبگاه خوانده می‌شوند. به‌عنوان مثال دریافت اطلاعات ترافیکی وبگاه‌ها از طریق قرار دادن نوار ابزار در مرورگر وب کاربران و ذخیره‌ی آن در سرور مربوط به سامانه، استخراج داده از مراکز داده‌ای و فراهم‌کنندگان سرویس و یا استخراج اطلاعات جستجوها و بازدیدهای انجام‌شده توسط کاربران که در سرورهای جویشگر بومی ثبت شده‌اند، منابع داده‌ای برون وبگاه نامیده می‌شوند.

هستند که یکی از راه‌کارهای آن توسعه سکوها خدمات جویشگر به کاربران است؛ خدماتی که توسط رقبای بنام ایشان در دنیا ارائه می‌شود. از این رو سامانه بومی رتبه‌بندی و تحلیل وبگاه‌ها به‌عنوان یک خدمت در کنار سایر خدمات جویشگر بومی قابل ارائه است و از طرفی نیز جویشگر بومی می‌تواند جهت هرچه بهتر ارائه دادن مهم‌ترین خدمت خود یعنی "گزارش جویش" از سامانه، خدمت ارائه رتبه‌ی وبگاه‌ها را دریافت نموده و آن را در الگوریتم رتبه‌بندی خود به نحو موثری وارد سازند. همچنین استفاده از اطلاعات کاربری و بازدید از وبگاه‌ها که در جویشگر ثبت می‌شود، به‌عنوان یک منبع داده‌ای برای رتبه‌بندی وبگاه‌ها، از مهم‌ترین عوامل پیوند میان جویشگر بومی و سامانه بومی رتبه‌بندی، به‌شمار می‌آید.

بدون شک داشتن نگاه راهبردی که مبتنی بر شناخت مسأله و نقاط ضعف رقبای خارجی محبوب در کشور و همچنین نگاه بلندمدت به‌تمامی عناصر دخیل در فرآیند تحلیل و رتبه‌بندی وبگاه‌ها می‌باشد، شرط اساسی برای موفقیت بهره‌برداری از سامانه بومی رتبه‌بندی در کشور است؛ لذا مقاله حاضر با اهداف بیان ضرورت و چگونگی طراحی و پیاده‌سازی سامانه تحلیل و رتبه‌بندی وبگاه‌ها مبتنی بر ترافیک، رفع نیاز به رصد ترافیک اینترنتی ایران و تعیین سهم ترافیک وبگاه‌های بومی به‌عنوان زیرمجموعه‌ای از پژوهش‌های طرح جویشگر بومی انجام می‌پذیرد. ترتیب ارائه مطالب در این پژوهش از این قرار است که ابتدا در بخش اول به تعریف برخی مفاهیم اولیه‌ی مورد نیاز در حوزه تحلیل و رتبه‌بندی وبگاه‌ها پرداخته و سپس در بخش دوم کارهای مشابهی که در این راستا انجام گرفته‌اند را شرح می‌دهیم. در بخش سوم به طرح مسأله‌ی دلایل نیاز به طراحی و پیاده‌سازی سامانه بومی رتبه‌بندی وبگاه‌ها پرداخته و در بخش چهارم نیز راه‌حل پیشنهادی خود را برای حل مسأله تشریح شده، بیان می‌کنیم. در بخش پنجم نکاتی در رابطه با توسعه‌ی این پژوهش و کارهای جدید و قابل انجام در این راستا و نتیجه‌گیری از کل پژوهش می‌پردازیم.

۲- مروری بر مفاهیم پایه

در این بخش به تعریف برخی مفاهیم اصلی که در این مقاله به‌کاررفته است، می‌پردازیم.

۲-۱- تحلیلگر وب

تحلیلگر وب، داده‌های موجود در وب را با هدف شناسایی و بهینه کردن استفاده از وب، اندازه‌گیری، جمع‌آوری و تحلیل کرده و گزارش آن را به مدیران وبگاه‌ها که از تحلیلگر استفاده می‌کنند، ارائه می‌دهد [2].

تحلیلگر وب تنها ابزاری برای اندازه‌گیری ترافیک وب نیست بلکه به‌عنوان یک ابزار برای ارزیابی جنبه‌های مختلفی از وبگاه، از قبیل جمعیت‌شناسی کاربران، بهبود رتبه وبگاه در جویشگر (SEO) و ردیابی تبلیغات، مورد استفاده قرار می‌گیرد. اساسی‌ترین تحلیل‌هایی که روی وبگاه‌ها صورت می‌گیرد، تحلیل بر مبنای ترافیک وبگاه و آمار بازدیدها از آن است.

۲-۲- رتبه‌بندی وب

رتبه‌ی یک وبگاه میزان ارجحیت آن را به نسبت سایر وبگاه‌ها بر اساس شاخص‌های مشخصی، تعیین می‌کند. در حوزه تجاری، رتبه‌بندی به‌وسیله ارزیابی جنبه‌هایی از وبگاه، باعث سوق به اهداف کسب‌وکار می‌شود. این جنبه‌ها بر اساس نوع رتبه بند مشخص می‌شود.

معمولاً رتبه‌بندی وبگاه‌ها دو نوع اصلی را شامل می‌شود که یکی رتبه‌بندی بر مبنای تحلیل‌های ترافیکی از وب و دیگری رتبه‌بندی بر پایه ارجاعات^۲ وب است. به‌عنوان مثال رتبه بند الکسا، وبگاه‌ها را بر اساس تحلیل ترافیکی آنها شامل شاخص‌های بازدید از صفحات، رتبه‌بندی می‌کند و رتبه بند گوگل، (در بخش بعد در رابطه با هر دو رتبه بند بحث خواهد شد) وبگاه‌ها را بر اساس تحلیل لینک‌های موجود در صفحات آن یا همان ارجاعات صفحه، رتبه‌بندی می‌کند.

۲-۳- شاخص‌های رتبه‌بندی وب

شاخص‌های رتبه‌بندی با توجه به انواع رتبه‌بندی که در بند قبل ذکر شد، به دو دسته شاخص‌های تحلیل وب و شاخص‌های تحلیل لینک‌های صفحات وب، تقسیم می‌شوند.

شاخص‌های تحلیل وب در واقع همان شاخص‌هایی هستند که توسط تحلیلگران وب نیز به‌کاررفته و مطابق با [3] دارای انواع دسته شاخص‌های بلوک‌های ساختاری^۳، خصوصیات بازدید^۴، خصوصیات محتوا^۵ و معیارهای تبدیلی^۶ هستند. شاخص‌هایی که به‌عنوان شاخص‌های پایه‌ای برای رتبه‌بندی بر مبنای تحلیل ترافیکی شناخته می‌شوند، در دسته‌ی شاخص‌های بلوک‌های ساختاری قرار گرفته و شامل بازدیدهای صفحه (Page views) و بازدیدکننده‌های منحصر بفرد (Unique visitors) هستند. به تعداد دفعاتی که یک صفحه مورد بازدید قرار گرفته است «بازدیدهای صفحه» گفته می‌شود و به تعداد کاربران منحصر بفرد انسانی (با

۳- کارهای مشابه

شرکت الکسا در سال ۱۹۹۶ میلادی توسط آمازون به عنوان یک شرکت مستقل تأسیس گردید. هدف از تأسیس آن ارائه داده‌های حاصل از تحلیل از وب می‌باشد و در حال حاضر بیش از ۳۰ میلیون وب‌وبگاه را مورد بررسی قرار می‌دهد [6]. اطلاعاتی که الکسا از وب‌وبگاه‌ها در اختیار قرار می‌دهد شامل تعداد زیادی شاخص از قبیل اطلاعات جمعیت شناختی، محبوبیت، آمار بازدیدکنندگان و بسیاری معیارهای دیگر می‌باشد. به طور کلی الکسا محصولات خود را بر اساس نوع مشتری و نیاز آن ارائه می‌دهد. محصولات این شرکت برای سه دسته فروشندگان دیجیتال، مالکین وبگاه و طراحان محتوا قابل استفاده است. مطابق آنچه توسط الکسا اعلام شده است، رتبه یک وبگاه ترکیبی از بازدیدکننده‌های منحصر بفرد روزانه و تعداد صفحات بازدید شده روزانه در دوره سه ماهه می‌باشد [7]. الکسا از میانگین حسابی که روی میانگین هندسی روزانه دو معیار مذکور اعمال می‌گردد برای تعیین رتبه سه ماهه آن وبگاه استفاده می‌شود. بنابراین رابطه بین رتبه یک وبگاه و شاخص‌های ذکر شده به صورت رابطه (۱) می‌باشد.

$$\text{Alexa traffic rank} \approx 1 / \text{AM3 (GM (DUV, DPV))}$$

رابطه (۱)

تعریف هر یک از نمادها به صورت زیر است:

- GM: میانگین هندسی روزانه
- AM3: میانگین حسابی در دوره سه ماهه
- DUV: تعداد بازدیدکننده روزانه
- DPV: تعداد بازدیدهای روزانه که توسط بازدیدکنندگان منحصر بفرد انجام گرفته است (بازدیدهای چندباره از یک صفحه توسط یک بازدیدکننده در طول یک روز فقط یک بار شمرده می‌شود)

در خصوص منابع داده‌ای جهت گردآوری اطلاعات آنچه توسط الکسا و سایر منابع منتشر و تأکید شده است، نوار ابزارهایی است که توسط مشتریان و کاربران اینترنت روی مرورگرها نصب شده است. همچنین بخش دیگری از اطلاعات که دارای دقت بیشتری است، از همکاری مالکین وبگاه‌هایی به دست می‌آید که با خرید خدمت «تحلیل داخلی و ممیزی وب‌وبگاه» امکان فعالیت Crawlerها و Robotها را برای تحقق اهداف تحلیل وبگاه خود یا سایر وبگاه‌ها فراهم می‌سازند.

شرکت مطرح گوگل محصولی تحت عنوان Google Analytics را برای استفاده و بهره‌برداری صاحبان وبگاه‌ها و کسب‌وکارهای مربوطه‌شان ارائه داده است تا ایشان بتوانند به اندازه‌گیری ترافیک وبگاه خود و کاربران و مخاطبان آن، به پایش و تحلیل وبگاه در جهت بالا بردن کارایی آن و هدفمند کردن کسب‌وکار و درآمدزایی بیشتر بپردازند. بعلاوه این شرکت محصول رتبه‌بندی وبگاه‌ها را تحت عنوان «گوگل رنک» یا Google Page Rank ارائه کرده است تا کاربران وب بتوانند از طریق رتبه اختصاصی گوگل به صفحاتی دسترسی کنند که در آنها گشت‌وگذار می‌کنند، به میزان اهمیت و اعتبار آن پی ببرند و زمان خود را در استفاده از وب بهینه‌تر نمایند. بنابراین می‌توان گفت رتبه بند گوگل فرصت مناسبی برای انتخاب بهینه کاربران وب در استناد به محتوای صفحات وب بر اساس میزان ارجاعات آن صفحات در اینترنت که نشان‌دهنده اهمیت آنهاست، پیش می‌آورد. گوگل رنک همان امتیازدهی به صفحات وب مبتنی بر ارجاعات و لینک‌هاست که امتیازی بین ۰ تا ۱۰ را به آن‌ها اختصاص می‌دهد. شاخص‌هایی که گوگل رنک در تحلیل لینک‌های صفحات و اختصاص این امتیاز اندازه‌گیری و پردازش می‌کند، در واقع همان شاخص‌هایی است که در بخش ۲ به عنوان شاخص‌های رتبه‌بندی مبتنی بر ارجاعات نام برده شد. مهم‌ترین منبع داده‌ای که گوگل از آن برای مقداردهی به شاخص‌های رتبه‌بندی از آن استفاده می‌کند همان ربات خزنده (Crawler) گوگل است که توسط آن صفحات را پیمایش کرده و اطلاعات مورد نیاز از لینک‌های صفحه را به دست می‌آورد.

الگوریتم رتبه‌بندی گوگل بر اساس گرافی که هر وبگاه از صفحات خود و ارتباط با صفحات سایر وبگاه‌ها دارد، محاسبات مربوطه را انجام می‌دهد. محاسبات در الگوریتم رتبه‌بندی بر روی چهار شاخص اصلی نام برده در بخش ۲ صورت می‌پذیرد. ساده‌ترین رابطه ارائه‌شده توسط گوگل برای محاسبه رتبه وبگاه‌ها مطابق زیر است:

$$\text{PR (A)} = (1-d) + d (\text{PR (t1)/C (t1)} + \dots + \text{PR (tn)/C (tn)})$$

رابطه (۲) [4]

در رابطه (۲)، «d» همان شاخص تعدیل است که معمولاً ۰.۸۵ در نظر گرفته می‌شود. PR (t1) رتبه‌ی مربوط به صفحاتی است که به صفحه A لینک هستند (inbound links) و C (t1) تعداد لینک‌های بیرونی صفحاتی (Outbound links) که به صفحه A لینک است. T1 تا Tn تمام صفحاتی هستند که به صفحه A لینک شده‌اند.

در ابتدا برای صفحاتی که قبلاً PR آنها محاسبه نشده است مقدار پیش‌فرض PR برابر یک در نظر گرفته می‌شود. نکته‌ای که وجود دارد این است که به علت آنکه مقدار PR صفحات با هر محاسبه به‌روزی می‌شوند برای دستیابی به یک مقدار درست‌تر و کم کردن خطاها این فرمول را چند بار تکرار می‌کنیم و در هر دوره‌ی اجرا مقادیر به‌روزرسانی می‌شوند. ثابت شده است که معمولاً ۴۰ تا ۵۰ دوره‌ی اجرا برای دستیابی به یک رتبه درست لازم و کافی است [8].

۴- طرح مسأله

امروزه رتبه‌بندی از حیث درک میزان اهمیت یک صفحه وب، در گشت‌وگذار اینترنتی^۷ جهت بازدید و پیمایش صفحات وب و یا استناد به آن و در نتیجه در رونق کسب‌وکارهای اینترنتی حائز اهمیت است. همزمان با آغاز رشد وب فارسی و به راه افتادن کسب‌وکارهای اینترنتی متنوع و متعدد در کشور، مدیران سایت‌ها همواره به دنبال معیاری برای سنجش میزان ارجح بودن سایت‌ها بوده‌اند.

از سوی دیگر زمانی که یک کسب‌وکار اینترنتی به واسطه تأسیس یک وبگاه ایجاد می‌شود یکی از مهم‌ترین عواملی که باعث می‌شود تا بعد از مدتی درآمدزایی داشته باشد انجام تبلیغات و بازاریابی برای آن کسب‌وکار است. در حال حاضر وبگاه‌های تخصصی و غیرتخصصی برای انجام تبلیغات در کشور وجود دارند که از راه تقبل نمودن تبلیغات برای سایر وبگاه‌ها، کسب درآمد می‌کنند. این قبیل وبگاه‌ها بیشترین اهمیت را برای رتبه الکسای خود قائل هستند چراکه مدیران وبگاه‌های درآمدزا برای بازاریابی و تبلیغات به وبگاه‌هایی اعتماد می‌کنند که بیشترین بازدیدکننده را داشته و از محبوبیت بالایی میان مردم برخوردار باشند. این امر در حال حاضر در کشور با اتکالی به رتبه الکسای آنها قابل تشخیص و انکاست.

بنابراین می‌توان گفت یکی از مهم‌ترین عوامل رونق یافتن و اعتبار دهی به یک وبگاه و کسب‌وکار اینترنتی در ایران، رتبه‌ی الکسای سایت نمایش‌دهنده می‌باشد. اما الکسا مرجعی نه‌چندان دقیق برای رتبه‌بندی وبگاه‌ها چه در سطح محلی و چه در سطح بین‌المللی است. چراکه راه‌های زیادی برای دور زدن رتبه الکسا و ارائه آمار غیردقیق و حتی غیرواقعی به وجود آمده است.

به عنوان مثال یکی از راه‌های افزایش رتبه‌ی الکسا که متأسفانه از متداول‌ترین روش‌ها در ایران است، استفاده از روش تبلیغاتی پای‌آپ (اسپم) است. این صفحات ناخواسته بر روی ماینیور کاربران ظاهر می‌شوند و باعث رشد سریع و البته کاذب رتبه‌ی الکسا می‌شوند. این مورد یکی از نقاط ضعف الکسا است.

مسئله دیگر اینکه دستیابی به رفتار کاربر در نحوه دنبال کردن وبگاه‌های مختلف که در نهایت موجب تولید یک رتبه‌بندی صحیح از وبگاه‌ها خواهد شد با تکیه بر منابع داده‌ای داخل کشور و بومی بسیار دقیق‌تر و قابل استنادتر از سرویس غیربومی مانند الکسا خواهد بود.

راه‌حل غلبه بر چنین کاستی‌هایی در زمینه‌ی رتبه‌بندی صفحات وب راه‌اندازی یک سرویس بومی رتبه‌بندی وبگاه‌ها می‌باشد. با راه‌اندازی سرویس بومی رتبه‌بندی وبگاه‌ها نقاط ضعف الکسا شناسایی شده و راه‌کارهایی بر اساس داده‌های دقیق‌تر و مستندتر برای رتبه‌بندی، پیاده‌سازی خواهند شد. در این مقاله سعی بر آن است که با توجه به اصالت در اطلاعات، حفظ امنیت و سرمایه مصرف‌کنندگان و قابلیت اعتماد که همگی با پشتوانه بومی بودن محصول به وجود خواهند آمد، جایگزین مناسبی برای الکسا و امثال آن در کشور پیشنهاد گردد.

۵- راه‌حل پیشنهادی

با توجه به مشکل مطرح‌شده همان‌طور که در بخش قبل نیز اشاره شد، راه‌اندازی یک سرویس بومی رتبه‌بندی وبگاه‌ها راه‌حلی غلبه بر کاستی‌های موجود و ایجاد یک رتبه‌بند دقیق و قابل اعتماد خواهد بود. این امر در نهایت موجب ایجاد رقابت بیشتر بین وبگاه‌ها و در نتیجه حصول کیفیت بالاتر سرویس و یا محتوای ارائه‌شده به کاربر و از سوی دیگر با شناخت رفتار کاربر باعث رونق کسب‌وکارهای اینترنتی خواهد شد.

نمای ساده‌ای از اجزای معماری سامانه‌ی رتبه‌بند پیشنهادی در شکل (۲) آورده شده است.

همان‌طور که در شکل (۲) نشان داده شده است گام‌های اصلی در طراحی این رتبه‌بند عبارت‌اند از: شناسایی منابع داده‌ای، استخراج اطلاعات از این منابع، ایجاد انباری داده و ذخیره‌سازی اطلاعات استخراجی، استخراج شاخص‌های مورد نیاز جهت تحلیل، رتبه‌بندی و در نهایت پردازش در سامانه و به موازات آن ارائه‌ی اطلاعات حاصل از تحلیل وبگاه‌ها و رتبه‌ی آن‌ها و در نهایت ذخیره‌ی خروجی‌ها و اطلاعات حاصل از پردازش. در ادامه‌ی این بخش چالش‌های موجود و راه‌حل‌های پیشنهادی در طراحی گام‌های برشمرده شده ارائه می‌گردد. اولین چالش شناسایی منابع داده‌ای و امکان‌سنجی به‌کارگیری آن‌ها است. با توجه به بررسی‌های صورت گرفته منابع داده‌ای بالقوه مورد استفاده در این سامانه را می‌توان در دو بخش تقسیم‌بندی کرد.

۶- نتیجه‌گیری و کارهای آتی

در این مقاله بر مبنای شاخص‌ها، نیازمندی‌های استخراجی و نتایج مطالعات امکان‌سنجی در خصوص منابع داده‌ای، معماری سامانه بومی رتبه‌بندی و وبگاه‌ها طراحی شده و در کارهای آتی وارد مرحله مربوط به طراحی بخش‌های مختلف شامل ماژول‌های معماری و پیاده‌سازی و آزمون خواهیم شد. لازم به ذکر است که در طراحی و پیاده‌سازی سامانه جنبه‌های امنیتی مربوطه لحاظ خواهد شد.

در مرحله طراحی، هر آنچه مرتبط با استخراج یک مقدار معتبر برای هر یک از شاخص‌های نهایی شده در مراحل قبل می‌باشد، با جزئیات کامل طراحی خواهد گردید. طراحی شامل بررسی موارد ذیل می‌باشد:

- چگونگی جمع‌آوری داده
- انواع داده جمع‌آوری شده
- ذخیره‌سازی داده
- منابع جمع‌آوری داده
- تخمین حجم داده‌های ذخیره‌شده
- چگونگی پردازش داده‌ها
- ابزارهای جمع‌آوری داده
- روش‌های تحلیل داده
-

پیاده‌سازی ماژول‌ها این سامانه بومی با توجه به معیارهای مورد نیاز برای پوشش دهی ویژگی‌های مستخرج و تعیین منابع داده‌ای مربوطه و انتخاب راه‌حل مناسب با استفاده از متدولوژی چابک انجام خواهد پذیرفت. بدین صورت که ابتدا یکی از معیارها انتخاب، داده‌های مرتبط با آن از منابع داده مربوطه توسط ابزارهای مشخص شده در طراحی جمع‌آوری خواهد گردید، سپس پردازش روی آنها انجام گردیده و رتبه و نتایج تحلیل حاصل از آن نمایش داده خواهد شد. به همین ترتیب کلیه معیارها پیاده‌سازی و نهایی خواهد گردید. در آخرین مرحله، این سامانه شامل کلیه ماژول‌ها پس از آزمون یکپارچگی ارائه خواهد گردید.

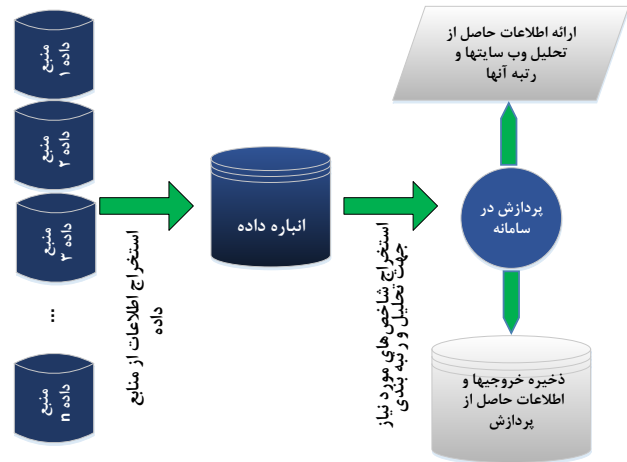
همچنین در گام بعدی پس از پیاده‌سازی و ارائه نهایی این سامانه با در نظر گرفتن نیازمندی‌های توسعه در تعامل با مشتریان و شرکت‌های داخلی می‌توان به ارائه‌ی طرح کسب‌وکار سامانه پرداخت و بستر مناسب برای تجاری‌سازی محصول و کسب درآمد از آن را فراهم نمود.

مراجع

- [1] <http://www.alexa.com/siteinfo/alexa.com>
- [2] https://en.wikipedia.org/wiki/Web_analytics
- [3] Jason Burby, Angie Brown & WAA Standards Committee, *Web Analytics definitions standard version 4*, August 16, 2007.
- [4] <https://en.wikipedia.org/wiki/PageRank>
- [5] Brian Clifton, *Advanced web metrics*, Wiley Publishing, 2010.
- [6] https://en.wikipedia.org/wiki/Alexa_Internet
- [7] <https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined->
- [8] https://en.wikipedia.org/wiki/PageRank#Distributed_algorithm_for_PageRank_computation

زیرنویس‌ها

- ^۱ Webpage Rank
- ^۲ References
- ^۳ Building Blocks
- ^۴ Visit Characterization
- ^۵ Content Characterization
- ^۶ Conversion Metrics
- ^۷ Web Surfing
- ^۸ tag
- ^۹ Agile
- ^{۱۰} scrum
- ^{۱۱} iterative



شکل (۲) اجزای سامانه رتبه‌بندی پیشنهادی

دسته اول منابع داده‌ای بر وبگاه و دسته دوم منابع داده‌ای برون وبگاه می‌باشند. از جمله منابع بر وبگاه می‌توان به کوکی‌ها و برچسب‌ها^۱ و منابع برون وبگاه به فایل‌های سابقه ذخیره‌شده بر روی سرور ارائه‌دهندگان خدمات اینترنت، شرکت ارتباطات زیرساخت و یا log کاربران در استفاده از جویشگرهای ملی اشاره کرد.

علاوه بر آن برای جمع‌آوری اطلاعات از رفتار کاربران در وب و میزان مراجعه آن‌ها به وبگاه‌ها می‌توان اقدام به ایجاد نوار ابزارهایی برای مرورگرهای مختلف نمود تا با نصب نوار ابزار توسط کاربران اینترنت امکان ردگیری بازدیدها و رفتارهای آن‌ها توسط سامانه رتبه‌بندی مهیا گردد. این بخش از اطلاعات نیز در دسته منابع داده‌ای برون وبگاه قرار می‌گیرد.

همچنین بخش دیگری از اطلاعات که دارای دقت بیشتری است، از همکاری مالکین وبگاه‌ها به دست می‌آید که با خرید خدمت «تحلیل داخلی و ممیزی وب‌وبگاه» امکان فعالیت Crawlerها و Robotها را برای تحقق اهداف تحلیل وبگاه خود یا سایر وبگاه‌ها فراهم می‌سازند.

در کنار تأمین منابع داده‌ای لازم می‌بایست از سوی دیگر الگوریتم رتبه‌بندی استخراج گردد. پس از بررسی برخی رتبه‌بندی‌های مطرح از جمله الکسا و Google Page Rank شش شاخص پایه برای رتبه‌بندی، مطابق آنچه در بخش تعاریف اشاره شد، استخراج شده‌اند. در این رتبه‌بندی شاخص‌های ترافیکی شامل شاخص‌های زیر بکار گرفته خواهد شد.

۱. بازدید صفحات
۲. تعداد بازدیدکننده‌های منحصر بفرد

پیشنهاد می‌شود سامانه بومی رتبه‌بندی و پایش وبگاه‌ها معرفی‌شده در این مقاله به روش توسعه نرم‌افزاری چابک^۲ طراحی و پیاده‌سازی و اجرا شود و در نهایت به‌عنوان یک محصول نرم‌افزاری با مدل کسب‌وکار تعریف‌شده، ارائه گردد. با الگوبرداری از رتبه‌بندی‌های معتبر و با توجه به فرآیند کردن محصول خروجی، پیشنهاد اولیه برای مدل کسب‌وکار این محصول، استفاده از مدل Ransom است تا بخشی از خدمات به‌صورت رایگان و بخش دیگری از آن‌ها که ممکن است در صورت توسعه محصول اولیه قابل حصول باشد، با دریافت هزینه از کاربران محصول در اختیار قرار گیرد. لیکن مدل کسب‌وکار این محصول، در تعامل با ذینفعان، نهایی خواهد شد.

برای تولید و توسعه نرم‌افزار، از متدولوژی اسکرام^۳ که یک مدل تکراری^{۱۱} از چارچوب چابک برای حل مسائل است استفاده خواهد شد. در واقع اسکرام یک فرایند و یا تکنیک تولید محصول نیست بلکه روشی است که به‌وسیله آن می‌توان مدیریت تولید محصول به‌صورت چابک را بهینه نمود.

با توجه به این متدولوژی، برای چرخه اول تولید محصول دو شاخص اصلی بازدید صفحات و تعداد بازدیدکننده‌های منحصر بفرد در نظر گرفته شده است اما می‌توان در چرخه‌های بعدی شاخص‌های بیشتری به سامانه رتبه‌بندی اضافه نمود.

پس از مشخص شدن منابع داده‌ای و شاخص‌های لازم برای الگوریتم رتبه‌بندی باید اطلاعات مورد نیاز از منابع داده‌ای استخراج و در انبار داده ذخیره شود. در مرحله بعد شاخص‌های ذکرشده، مقداردهی شده و برای انجام پردازش وارد سامانه رتبه‌بندی می‌شوند. در نهایت پس از اعمال الگوریتم رتبه‌بندی اطلاعات حاصل از تحلیل و رتبه‌بندی وبگاه‌ها ارائه خواهد شد.