Archive of SID

اولین همایش ملی جویشگر بومی
۲۷ و ۲۸ بهمن ماه سال ۱۳۹۴
پژوهشگاه ارتباطات و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)
وزارت ارتباطات و فناوری اطلاعات
ITRC

بررسی فرآیند تبدیل یک جویشگر پیشرفته جهت پشتیبانی خودکار از مجموعه سرویس‌های وب پویا

مهران علیدوست نیا[۱]، رضا ابراهیمی آتانی[۲]، هیراد اسیمی[۳]، مهرداد اسمعیل پور[۴]، امیررضا جواهردشتی[۴]، بهنام اسحاقی[۴]، آرتا خادمی نژاد[۴]، امید محمدرضایی[۴]، میلاد ابراهیم پور[۴] و فرج الله حسن پور[۵]
[۱] کارشناس ارشد نرم‌افزار، گروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه گیلان، رشت
alidoost@msc.guilan.ac.ir
[۲] استادیار، گروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه گیلان، رشت
rebrahimi@guilan.ac.ir
[۳] دانشجوی کارشناسی ارشد، دانشکده فنی، دانشگاه گیلان، رشت
assimi@msc.guilan.ac.ir
[۴] عضو تیم طراحی و پیاده‌سازی، گروه نرم‌افزاری فتاسان، شرکت فضاگستران تبادل الکترونیک، رشت
{esmaeilpour, javaherdashti, eshaghi, khademinejad, m.rezaei, ebrahimpour}@fatasan.ir
[۵] مشاور فنی و رسانه ای، گروه نرم‌افزاری فتاسان، شرکت فضاگستران تبادل الکترونیک، رشت
hasanpour@fatasan.ir

چکیده

در سال‌های اخیر توسعه سرویس‌های جانبی در جویشگرهای معتبر جهانی از اهمیت بالایی برخوردار بوده است. این اهمیت از تعداد کاربران مشترک در سرویس‌های این جویشگرها مشخص می‌شود. در این مقاله، هدف اصلی، ارائه یک معماری مبدل برای تبدیل یک جویشگر به یک واحد پشتیبانی کامل از سرویس‌های جانبی آن است. این معماری مبدل به صورت تعاملی از قابلیت‌های استفاده مجدد از کدهای جویشگر بهره می‌گیرد. تبدیل در این روش، تبدیل زیرساخت جویشگر بهینه و با کمترین هزینه زمانی و ریالی امکان‌پذیر خواهد بود. تمرکز اصلی در این پروژه بر روی توسعه سرویس و افزایش حضور کاربران است.

نتایج پیاده‌سازی‌ها نشان می‌دهد که با استفاده از معماری مبدل جویشگر، بخش‌های کلیدی آن شامل خزنده‌ها، پارسر، پایگاه داده‌ها، بخش کنترلی، تازه‌سازی، بروزرسانی و رتبه‌بندی به طور حداکثری قابلیت استفاده مجدد را دارند. بیشترین حجم تغییرات مربوط به خزنده‌ها می‌شود که در بدترین شرایط حداکثر از ۷٪ تغییرات در زیرساخت جویشگر پارسی‌یاب بیشتر نمی‌شود.

کلمات کلیدی

جویشگر سرویس‌گرا، وب سرویس، سرویس‌های پویا، خزنده‌های پویا، رتبه‌بندی تخصصی، جویشگر فارسی.

۱- مقدمه

در سال‌های اخیر، رشد بزرگی در حوزه سرویس‌های جویشگرهای بزرگ و پیشرفته در دنیا شاهد بوده‌ایم. این رشد شگرف نه تنها در زیرساخت‌های فناوری اطلاعات، که در زیرساخت‌های محتوایی نیز به چشم می‌خورد. نکته بسیار مهم در رابطه با این ابرسرویس‌ها، رضایتمندی کاربران و توجه مستقیم به نیازهای آنها بوده است. تنها دارا بودن زیرساخت یک جویشگر پیشرفته، دلیلی بر فراگیر بودن سرویس‌های آن نخواهد بود. این مسئله علاوه بر جویشگرهای جهانی و بزرگی مانند گوگل، بایدو و بینگ، شامل حال جویشگرهای ایرانی نیز می‌شود. باید به این سوال پاسخ اساسی داد که آیا جویشگرهای بومی در سال‌های اخیر توانسته‌اند در جذب کاربران ایرانی توفیق چندانی حاصل نمایند؟ پاسخ این سوال را پیش از نگاه الگوریتمیک و زیرساختی، باید در محصولات جویشگرها و توجه به نیاز کاربران جستجو کرد.

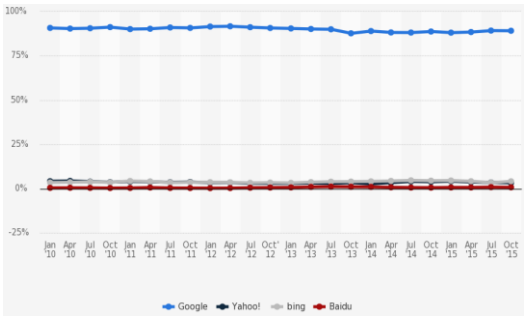
بر اساس آمارها در کشورهای دارای زیرساخت جویشگر پیشرفته، سهم جهانی رقبای بزرگی مانند گوگل، همچنان ثابت و قابل توجه است [۱]. این امر بیش از هر چیز دیگری به سرویس‌های کاربردی و جدایی ناپذیر این غول اطلاعاتی از جویشگر آن وابسته است. برای مثال سرویس اطلاعاتی جویشگر به طور همزمان به سرویس‌هایی مانند پست الکترونیک گوگل *Gmail*، ترند *Trend*، تقویم *Calendar* و *webmaster* متصل است. این اتصال خودکار علاوه بر رفع نیازهای اطلاعاتی و خدماتی کاربران، باعث بهبود عملکرد سرویس شده و در بازه‌های زمانی بزرگتر، عملکرد جویشگر را بهتر می‌کند. همانطور که در شکل۱ قابل مشاهده است، سهم گوگل در سال‌های اخیر از بازار جهانی تقریبا ثابت و بین ۸۹ تا ۹۰ درصد در حال نوسان بوده است. توجه به این نکته حائز اهمیت است که این سهم در برخی کشورها دارای کاهش و افزایش‌های چشمگیری است. در کشور روسیه این سهم در حدود ۳۲ درصد و در کشور چین کمتر از ۳ درصد پیش‌بینی می‌شود [۲]. رمز موفقیت در بازار جهانی گوگل، توجه به نیاز کاربران و ارائه سرویس‌های جانبی کاربرپسند است. البته این جویشگر از نظر الگوریتمیک و زیرساخت فناوری اطلاعات یکی از بهترین‌ها است و با دارا بودن بیش از ۱۱۰ مرکز داده بزرگ در سراسر جهان، به عنوان موفق‌ترین جویشگر لقب گرفته است [۳].

هدف اصلی از ارائه این مقاله، ارائه راهکاری برای تبدیل جویشگرهای با زیرساخت فناوری اطلاعات به هسته‌های محرک پشتیبان از مجموعه سرویس‌های وب پویا است. در این مقاله یک روش هدفمند و مرحله به مرحله همراه با پیاده‌سازی جویشگر سرویس‌گرای "*روزبه‌روز*" از پروژه موتور جستجوی "*پارسی‌یاب*" نشان داده می‌شود. تغییرات و نیازمندی‌های جدید برای تبدیل یک جویشگر واقعی به یک پشتیبان تمام عیار در سرویس‌های وب محتوا-محور بررسی می‌گردد. همچنین مشکلات پیش رو جهت تبدیل تمامی اجزای یک موتور جستجو از جمله خزنده‌ها، واحدهای کنترلی، واحدهای داده، واحدهای تجزیه، ایندکس و

www.SID.ir

Archive of SID

رتبه‌بندی مورد بررسی قرار می‌گیرد. در پایان نشان خواهیم داد که تبدیل یک جویشگر با زیرساخت اطلاعاتی پیشرفته به یک هسته محرک سرویس‌گرا نه تنها کار پر هزینه‌ای نیست، بلکه به تولید سرویس‌های خودکار و پویا منجر خواهد شد که نتیجه ای جز رضایت کاربران در بر نخواهد داشت.



شکل (۱) : سهم بازار جهانی چهار جویشگر برتر از ۲۰۱۰ تا ۲۰۱۵ [۴]

۲- انگیزه‌های مطالعاتی

در این بخش به بررسی مروری سرویس‌های وب در جویشگرهای جهانی می‌پردازیم. همچنین عملکرد سرویس‌های خارجی در جذب کاربران هدف را مورد ارزیابی قرار داده و سهم هر یک از بازار جهانی سرویس‌های اطلاعاتی را بررسی می‌کنیم. همچنین در خصوص جویشگرهای سرویس‌گرا و ویژگی‌های اساسی آنها بحث خواهیم نمود.

۲-۱- سرویس‌گرایی در جویشگرها

یک جویشگر برای رقابت با بازار جهانی باید همواره آماده ارائه سرویس‌های اطلاعاتی جدید باشد. در اصل، یک جویشگر دارای یک بانک اطلاعاتی پردازش نشده است و اینکه چطور این بانک را سازماندهی کند، به خدماتی بستگی دارد که توسط سرویس‌های وب آن قابل ارائه هستند. در واقع خروجی اصلی کار یک جویشگر می‌تواند مجموعه ای از سرویس‌های اطلاعاتی و به هم پیوسته باشد. هر چه این میزان از سرویس‌ها بیشتر و متنوع تر در اختیار کاربران قرار گیرند، میزان رضایتمندی و ترغیب کاربران به استفاده از جویشگرها بیشتر می‌شود. سرویس‌گرایی در یک جویشگر به معنای آمادگی آن جویشگر برای تولید و پشتیبانی از سرویس‌های اطلاعاتی تحت وب است. اینکه یک موتور جستجوی پیشرفته تا چه حد قابلیت سرویس‌گرایی دارد، بیش از هر چیز دیگری به معماری نرم‌افزاری و زیرساختی آن بستگی خواهد داشت. در این بین استفاده از خواص برنامه‌نویسی پویا، ماژولار و رعایت اصول مهندسی نرم‌افزار بسیار اهمیت پیدا می‌کند [۵]. البته گاهی هم پیش می‌آید که یک سرویس وظیفه پشتیبانی از سایر سرویس‌ها را بر عهده دارد. مانند آنچه که بین شبکه اجتماعی فیسبوک و جویشگر بینگ اتفاق می‌افتد [۶]. جدای از این مسائل، مهمترین نکته در سرویس‌گرایی، قابلیت تعامل و سازگاری بین محیط مبدا و مقصد است.

۲-۲- جویشگرهای بومی بین‌المللی

سهم اصلی جویشگرها در روسیه با ۶۰ درصد در اختیار *Yandex* و ۳۲ درصد در اختیار گوگل است. لازم به ذکر است که ۲ درصد از سهم جویشگرهای ترکیه نیز در اختیار *Yandex* است. دو جویشگر بومی کره جنوبی عبارتند از *Naver* و *Daum* که به ترتیب ۷۷ و ۲۰ درصد از سهم بازار را در اختیار دارند. در جمهوری چک نیز جویشگر بومی *Seznam* نیز ۴۵ درصد از سهم بازار را در این کشور در اختیار دارد. حال باید این سوال مطرح شود که دلیل افزایش استفاده روزافزون از موتور جستجوهای بومی در این کشورها چیست؟ برای پاسخ به این سوال، باید به سراغ بررسی سرویس‌های پایه و جانبی در جویشگرهای بین‌المللی برویم که در قالب جداول ۱، ۲ و ۳ با هم مقایسه می‌شوند.

جدول (۱) : خدمات مبتنی بر تولید محتوای جویشگرهای بومی

نام جویشگر	Baidu	Sogou	Qihoo	Yandex	Naver	Daum	Seznam
شبکه های اجتماعی	✓	✓	✓	-	✓	✓	-
وبلاگ	-	-	✓	-	✓	✓	-
ایمیل	-	-	✓	✓	✓	✓	✓
اشتراک گذاری فایل	✓	-	✓	✓	✓	✓	-
خرید	✓	✓	✓	✓	✓	✓	✓
چت	✓	✓	-	-	-	✓	-

جدول (۲) : مقایسه خدمات جانبی جویشگرهای بومی

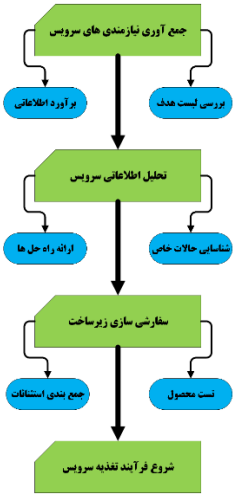
نام جویشگر	Baidu	Sogou	Qihoo	Yandex	Naver	Daum	Seznam
نقشه	✓	✓	✓	✓	✓	✓	✓
جستجو با گفتار	✓	-	-	-	-	✓	-
ترجمه	✓	✓	-	✓	✓	-	✓
تبلیغات	✓	✓	✓	✓	✓	✓	✓
ابزارهای آماری	✓	-	-	✓	-	✓	✓
فرهنگ لغت	✓	✓	-	✓	✓	✓	✓
واسط کاربری برای توسعه دهندگان	✓	✓	✓	✓	✓	✓	✓
خدمات مکان محور	✓	✓	✓	✓	✓	✓	✓

جدول (۳) : مقایسه زیرساخت ارائه خدمات در جویشگرهای بومی

نام جویشگر	Baidu	Sogou	Qihoo	Yandex	Naver	Daum	Seznam
مرورگر	✓	✓	✓	✓	-	-	✓
تبلیغات هوشمند	✓	-	-	✓	-	-	✓
سامانه دسکتاپ	✓	✓	-	-	-	-	-
سامانه موبایل	✓	✓	✓	-	-	✓	✓
خدمات مبتنی بر موبایل	✓	✓	✓	✓	✓	-	✓

۳- معماری مبدل

معماری جویشگرهای پیشرفته دارای پیچیدگی در پیاده‌سازی بوده و این امر تنها با پیروی از یک فرآیند مهندسی نرم‌افزار مطمئن قابل رهگیری است. تمامی فرآیندهای استاندارد مهندسی نرم‌افزار پیشرفته از جمله: نیازمندی‌ها، کنترل فرآیند، تحلیل، راستی آزمایی و تست، پیاده‌سازی و نگهداری باید به طور منظم و مدون انجام گرفته باشد. طبق اصول مهندسی نرم‌افزار در محیط‌های مرتبط با وب [۷و۵]، باید الگوی مورد استفاده برای تک تک ماژول‌ها و کلاس‌های زیر برنامه های مختلف جویشگر موجود باشد. حتی برای استفاده از مفاهیم وب معنایی نیز یکی از پیش‌نیازها، رعایت اصول مهندسی نرم‌افزار و داشتن اسناد و موارد مرتبط با آن است [۸].



شکل (۲) : فرآیند کلی پشتیبانی جویشگر از سرویس‌های وب پویا

در این پروژه، اسناد و مراحل مربوط به مهندسی نرم‌افزار از پیش در اختیار تیم قرار داشت و به همین دلیل فرآیند پیاده‌سازی محصول با سرعت و دقت بالاتری انجام گرفت. فرآیند مربوط به معماری مبدل جویشگر بیشتر بر روی سه مرحله از فرآیند مهندسی نرم‌افزار محصول تاثیرگذار است. اولین و مهمترین مرحله، نیازمندی‌های سرویس جدید است که به کلی با نیازمندی‌های جویشگر تفاوت دارد. این تفاوت‌ها تنها از نظر معنایی بوده و از لحاظ رفتاری، بسیار شبیه به الگوی جویشگر خواهد بود. دومین مرحله مربوط به تحلیل نیازمندی‌های سیستم است. این مرحله مهمترین بخش سیستم را شامل می‌شود. شناسایی حالات خاص و تغییراتی که باید در زیرساخت‌های نرم‌افزاری جویشگر انجام شود، از مهمترین اهداف این مرحله است. این مراحل به طور ترتیبی در شکل۲ قابل نمایش هستند. سومین مرحله، سفارشی‌سازی

www.SID.ir

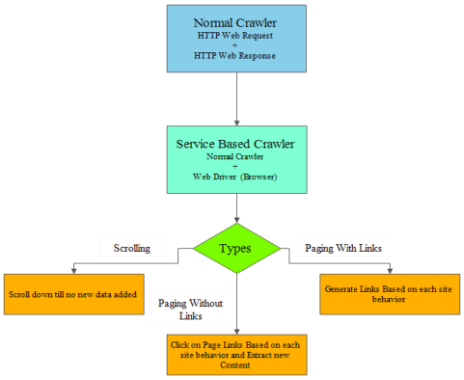
Archive of SID

محصول است که به طور خاص با مراحل مربوط به پیاده‌سازی محصول در فرآیند مهندسی نرم‌افزار منطبق می‌گردد. پس از سفارشی‌سازی محصول، نوبت به تست آن می‌رسد که فرآیندی مشابه با فرآیند مهندسی نرم‌افزار را طی خواهد کرد. همان‌طور که بیان گردید، جویشگر *پارسی‌یاب* بر اساس معماری انطباقی پیشنهادی پیاده‌سازی شده و تمامی حالات ترکیبی سرویس *روزبه‌روز*، بر اساس همین اصول بازنویسی و طراحی مجدد شده‌اند.

۳-۱- پیاده‌سازی خزنده‌های سفارشی

در پروژه پارسی‌یاب، کراول یک سایت با سه شیوه عادی، بروزرسانی و بازگشتی انجام می‌شود. روش‌های عادی و بازگشتی اساس کاری مشابهی دارند و در آنها تمام لینک‌های داخلی سایت کراول و ذخیره می‌شوند. اما در خزش به شیوه بروزرسانی، تنها لینک‌هایی که پس از یک بازه زمانی خاص به سایت اضافه شده‌اند کراول شده و سایر لینک‌های داخلی بررسی نمی‌شوند. نسخه عادی خزنده توانایی کراول کردن تمامی سایت‌هایی که از فناوری AJAX برای نمایش محتوا استفاده نمی‌کنند را دارد. درصد اطمینان خزنده در مواجهه با سایت‌های مختلف بیش از ۹۹٪ ارزیابی گردیده و موارد استثناء به صورت دوره‌ای به لیست های اختصاصی افزوده می‌شود تا عملکرد جویشگر بهبود یابد.

برای سفارشی‌سازی نسخه عادی خزنده به خزنده سرویسی که بتواند سرویس رتبه‌بندی فروشگاه‌های الکترونیکی را به طور خودکار پشتیبانی کند، ابتدا باید سایت‌هایی که نیاز به کراول دارند مشخص شوند. در این زمینه لیستی از ۱۵۷ سایت فروشگاهی که رتبه آلکسای [۹] آنها زیر ۱۰۰۰ در منطقه ایران بودند، مورد بررسی قرار گرفت. در میان این سایت‌ها ۱۴۷ سایت به وسیله نسخه عادی خزنده پارسی‌یاب قابل کراول بودند، اما ۱۰ سایت باقیمانده از فناوری *AJAX* برای نمایش اطلاعات استفاده می کردند و برای کراول تمام محتوای آنها نیاز به ایجاد تغییراتی در نسخه عادی خزنده بود که با توجه به رفتار سایت‌ها، روند مناسبی را در پیش بگیرد.



شکل (۳) : فرآیند کلی معماری مبدل خزنده‌های سرویس

برای ساخت خزنده سرویس فروشگاه‌ها از زیرساخت مرورگر فنتوم جی‌اس درون خزنده عادی استفاده شده است [۱۰]. این مرورگر بدون رابط کاربری بوده و توانایی غیرفعال کردن بارگذاری عکس‌ها برای ایجاد سرعت بیشتر را دارد. تغییرات مورد نیاز در کد اصلی برای اضافه کردن این مرورگر تقریبا برابر ۰٫۵ درصد کدهای خزنده عادی است. پس از اضافه شدن این مرورگر، باید رفتار سایت‌ها به دسته‌بندی‌های مختلف تقسیم شده و روش‌هایی برای کراول کامل آنها طراحی شود. سه شیوه کلی استفاده شده در ده سایت باقیمانده برای استفاده از فناوری AJAX، Paging With Links و Paging Without Links و Scrolling بودند. بر اساس شکل۳، در هر دو شیوه Paging محتوا در صفحات مختلف قرار می‌گیرند و با کلیک روی لینک هر صفحه، محتوای جدید به صورت غیرهمزمان تولید و بدون بارگذاری مجدد، داده می‌شود. در شیوه Scrolling نیز محتوای جدید با رسیدن نوار اسکراول به نقطه انتهایی بارگذاری می‌شود و تا پایان یافتن داده‌ها ادامه پیدا می‌کند.

۳-۲- پارسر و ذخیره‌سازی اطلاعات

روند کار پارسر به صورت آفلاین است. به خاطر بهینه‌سازی پهنای باند و صرفه‌جویی در منابع مربوط به خزنده‌های موازی، در معماری جویشگر پارسی‌یاب، توجه خاصی به پردازش‌های موازی شده است. وظیفه اصلی هماهنگی بین خزنده‌ها و پارسر به عهده کنترل‌کننده مرکزی سیستم است. این کنترلر است که خزنده‌ها را راه‌اندازی کرده و آنها را تغذیه می‌کند و همچنین وظیفه فراخوانی و اجرای پارسر نیز بر عهده آن خواهد بود. بخش کنترلر یک فایل راهنما به ازای هر مرتبه کراول ایجاد کرده و به عنوان راهنمای تجزیه اطلاعات، در اختیار پارسر قرار می‌دهد. پارسر نیز اطلاعات را پردازش کرده و با توجه به فایل‌های راهنما در پایگاه داده اولیه ذخیره می‌کند. برای ذخیره‌سازی داده‌های بزرگ از فرمت‌های *clob* در مدیریت پایگاه داده اوراکل استفاده می‌شود. سرعت و دقت بالای این نوع داده سبب شده تا دسترسی‌های راه دور نیز از سرورهای دیگر تحت پروتکل *ftp* به راحتی ایجاد شود. یکی از ویژگی‌های مهم پارسر پارسی‌یاب، توانایی پارس کردن فایل‌های با انواع استاندارد استکه از انواع فایل‌های *Word*، *PDF*، *Power Point* و *Excel* به طور صریح پشتیبانی می‌نماید.

۴- ارزیابی و نتایج پیاده‌سازی

در این بخش نتایج تجربی حاصل از ارزیابی معماری مبدل سرویس در جویشگر پارسی‌یاب مورد بررسی قرار می‌گیرد. در این پروژه از سه سرور تخصصی برای پیاده‌سازی پنج خزنده و سه پارسر به صورت همزمان استفاده شده است.

۴-۱- نتایج پیاده‌سازی خزنده‌های تخصصی

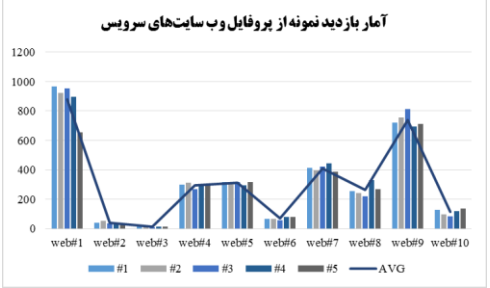
خزنده‌های تخصصی بر اساس پروفایل وب سایت‌های هدف کار می‌کنند و به طور طبیعی وابسته به رفتار آنها هستند. اما چون رفتار دسته بزرگی از وب سایت‌ها مانند روال معمول جویشگر است، با تغییرات خاص و هدفمندی روبرو هستیم. همانطور که در جدول۴ قابل مشاهده است، جزئیات تبدیل در خزنده‌های سرویس وب پویا در جویشگر پارسی‌یاب نشان داده شده است. این خزنده جهت پشتیبانی از سرویس رتبه‌بندی فروشگاه‌های الکترونیکی سفارشی‌سازی شده است. برخی از وب سایت‌ها دارای حالات خاص زیادی بوده‌اند و این امر روی زمان کراول سایت تاثیر منفی می‌گذارد. اما نکته مهم میزان کم تغییرات است که با بررسی های جامع در مرحله نیازمندی و تحلیل مهندسی نرم‌افزار محصول، تحقق می‌یابد. شکل های ۵ هم مقایسه سفارشی‌سازی در ابعاد مختلف است.

جدول (۴) : مقایسه روش‌های سفارشی‌سازی در وب سایت‌های نمونه

وب سایت نمونه	متوسط خطوط اضافه شده	تعداد متدهای اضافه شده	تغییرات کد	درصد افزایش زمان کراول	نوع سفارشی‌سازی
کافه بازار	۱۵۰	۳	٪ ۵	٪ ۳۰۰	Scrolling
دیجی‌کالا	۲۰۰	۴	٪ ۷	٪ ۴۰۰	Paging With Links
لاین کامپیوتر	۳۰	۳	٪ ۱	٪ ۱۵۰	Paging With Links
دیوار	۵۰	۱	٪ ۲	٪ ۳۰۰	Scrolling
زنبیل	۵۰	۱	٪ ۲	٪ ۲۰۰	Paging With Links
شیپور	۹۰	۱	٪ ۳	٪ ۱۰۰	Paging With Links
آل دیجیتال	۱۱۰	۱	٪ ۴	٪ ۲۵۰	Paging With Links + Paging Without Links
جی اس ام	۲۰۰	۱	٪ ۷	٪ ۴۰۰	Scrolling + Paging Without Links

۴-۲- محاسبه معیار تازه‌سازی

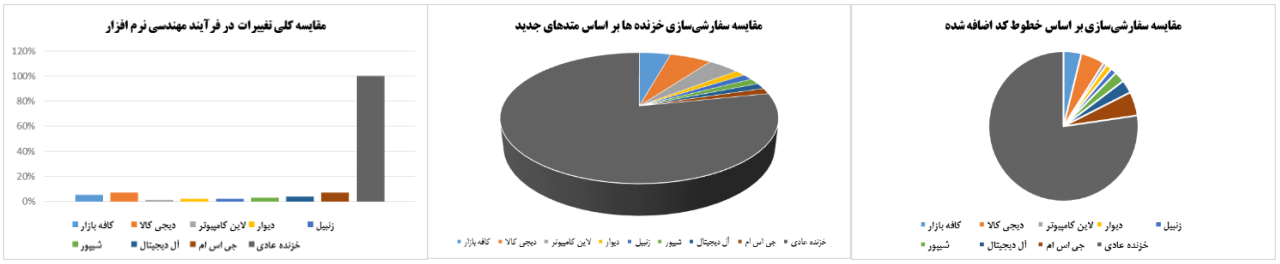
معیار *d*، پارامتری برای تعیین زمان بروزرسانی بعدی یک وب سایت هدف است. این معیار با کمی تغییر در راهبرد سرویس‌های جویشگر پارسی‌یاب تاثیر مستقیمی دارد. محاسبه این معیار با یک مقدار پیش فرض شروع می‌شود. برای محاسبه مقدار جدید پارامتر *d* لازم است تا ابتدا دو عدد تاثیرات مثبت (positive changes) و تاثیرات منفی (negative changes) محاسبه شوند. رابطه محاسبه تاثیرات مثبت تابعی از d و صفحات جدید است. در صورتی که تعداد new pages کمتر از ۱۰ باشد، این اثر منفی بوده و منجر به افزایش *d* می‌شود. همچنین مقدار به دست آمده به ازای ۱۰۰۰ صفحه جدید بیشترین مقدار قابل محاسبه با این رابطه است و تاثیر مقادیر بالاتر از این عدد محدود شده و رفتار مشابهی خواهد داشت. تاثیر این پارامتر در بروزرسانی بعدی ۱۰ وب سایت مورد بررسی در این پروژه، در شکل۴ قابل مشاهده است. همچنین محاسبه این مقدار بر اساس فرمول های نمایی انجام می گیرد.



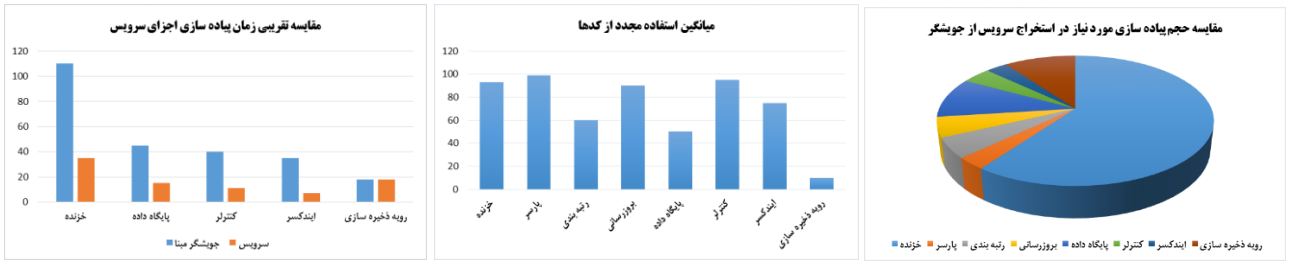
www.SID.ir

Archive of SID

شکل (۴) : تاثیر پارامتر d در محاسبه مجدد معیار تازه‌سازی



شکل (۵) : مقایسه سفارشی‌سازی خزنده‌ها با رویکرد میزان تغییرات در کدهای خزنده مبنا، متدهای جویشگر مبنا و همچنین تغییرات در فرآیند مهندسی نرم‌افزار



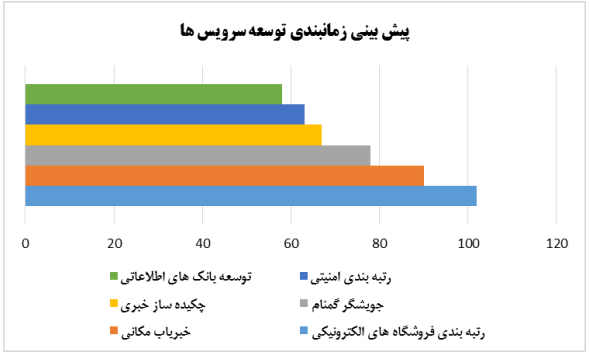
شکل (۶) : مقایسه حجم بازنویسی اجزای مختلف سرویس، میزان استفاده مجدد از کد جویشگر و مقایسه زمان پیاده‌سازی اجزای مختلف سرویس در معماری مبدل پیشنهادی

۴-۳- ارزیابی بهینه‌سازی اجزای جویشگر

تنها بخشی که به طور کامل نیاز به بازنویسی مجدد دارد، بخش پردازش اطلاعات پایگاه داده است که در جویشگر با عنوان *DtoD* شناخته می‌شد. این بار باید یک رویه کامل جهت پردازش نیازمندی‌های اطلاعاتی سرویس رتبه‌بندی فروشگاه‌های الکترونیکی نوشته شود که پردازش اطلاعات تخصصی از جمله نظرات کاربران، محصولات جدید، قیمت‌ها، پرفروش‌ترین‌ها و غیره را انجام دهد و نتایج به صورت دوره‌ای در بانک داده سرویس ذخیره شود. این رویه جدید که *DtoS* نام دارد، این وظیفه را بر عهده خواهد داشت. جمع‌بندی تغییرات با توجه به قابلیت استفاده مجدد از کدهای جویشگر به تفکیک هر بخش و به طور مقایسه‌ای در نمودارهای شکل۶ نشان داده شده است. بیشترین استفاده مجدد از کدها در فرآیند مهندسی نرم‌افزار خزنده‌ها مد نظر قرار دارد. همچنین کمترین میزان استفاده مجدد در کدها نیز مربوط به رویه‌های ذخیره‌سازی و تجزیه است که بنا به نیازمندی‌های خودکار سرویس‌ها متفاوت از جویشگر خواهد بود.

۴-۴- پیش‌بینی توسعه جویشگرهای سرویس‌گرا

ساخت و توسعه سرویس‌ها در افزایش رضایتمندی کاربران و رفع نیازهای فضای مجازی آنها اهمیت ویژه‌ای دارد. جویشگرهای بومی نیز نیازمند توسعه سرویس‌های جانبی هدفمند، در کنار توسعه زیرساخت جویشگرها هستند. این توسعه باید بر مبنای پیشینه اطلاعاتی و زیرساختی جویشگر باشد. در مجموعه شکل۶ نمودار مقایسه زمان اجرا و پیاده‌سازی هر یک از اجزای سرویس را در مقایسه با زمان توسعه اجزای جویشگر مبنا نشان می‌دهد که بسیار حائز اهمیت بوده و با یک پیاده‌سازی چابک، در کوتاه‌ترین زمان ممکن قابل اجرا خواهد بود. شکل۷ نیز پیش‌بینی توسعه سرویس‌های مبتنی بر معماری مبدل را نشان می‌دهد.



شکل (۷) : پیش‌بینی زمان توسعه سرویس‌ها با معماری مبدل

۵- نتیجه

در این مقاله، هدف اصلی ارائه یک معماری مبدل برای تبدیل جویشگر پیشرفته *پارسی‌یاب* به یک پشتیبان قدرتمند از سرویس‌های جانبی آن بود. مجموعه سرویس *روزبه‌روز* از معماری مبدل پارسی‌یاب بهره گرفته و یک نمونه از سرویس‌های پیاده‌سازی شده توسط این معماری، ارزیابی گردید. هدف ما از این پروژه توسعه چابک و هدفمند سرویس‌های جانبی جویشگرهای بومی جهت توسعه زیرساخت‌های فنی و افزایش رضایتمندی کاربران بود. با بهینه‌سازی زیرساخت یک جویشگر پیشرفته می‌توان با کمترین تغییرات در فرآیندهای مهندسی نرم‌افزار، مجموعه سرویس‌های قدرتمند تجاری را راه‌اندازی نمود. این معماری در بخش‌های خزنده، کنترلی، پایگاه داده، رتبه‌بندی و تازه‌سازی اطلاعات اعمال می‌گردد. به طور میانگین در خزنده‌ها، در بدترین شرایط حداکثر به تغییرات در ۷٪ کدها نیاز داریم. بیشتر تغییرات در بخش تامین نیازمندی‌های جدید اعمال گردیده و راه برای یک پیاده‌سازی چابک باز است. یکی از نتایج مهم، توجه به کاهش هزینه‌ها در توسعه سرویس‌های جانبی جویشگرهای پیشرفته است.

سپاسگزاری

لازم می‌دانیم تا از پارک علم و فناوری استان گیلان به خاطر حمایت‌های بی‌دریغ از پروژه "توسعه سرویس‌های جانبی موتور جستجوی پارسی‌یاب" تشکر نماییم. همچنین از سازمان ارتباطات و فناوری اطلاعات استان گیلان بخاطر حمایت‌های زیرساختی در توسعه این پروژه تقدیر و تشکر می‌کنیم. این پروژه به عنوان یکی از پروژه‌های مصوب به شماره ۹۴/۴۱۴۰ مورخ ۹۴/۰۷/۲۲ در پارک علم و فناوری استان گیلان به ثبت رسیده است.

مراجع

[1] Webcertain Team, The Webcertain Global Search and Social Report 2015, Thechnical Report, International Digital Hub, Jan 28, 2015.

[2] Incitez Pte. LTD, *China Search Engine Market Overview 2015; A Statistical Analysis of China Search Engine Market*, Thechnical Report, Published on Aug 27, 2015.

[3] Google Inc, "Google's Green Data Centers: Network POP Case Study", Google Green, Technical Report online, updated 2014.

[4] Statista Team, *Worldwide market share of leading search engines from January 2010 to October 2015*, Technical Report, 2015.

[5] J. Nijjar, I. Bocić, and T. Bultan, *Data Model Property Inference, Verification, and Repair for Web Applications*, ACM Transactions on Software Engineering and Methodology (TOSEM) - Special Issue on ISSTA 2013, Volume 24 Issue 4, Article No. 25, August 2015.

[6] G. Marchionini, *Exploratory search: from finding to understanding*, Communications of the ACM 49.4, Pages 41-46, 2006.

[7] H. Li, C. Isaac, L. Wang-Chien, and C. L. Giles, *CiteSeerx: an architecture and web service design for an academic document search engine*, In Proceedings of the 15th international conference on World Wide Web, pp. 883-884. ACM, 2006.

[8] M. Brambilla, C. Irene, S. Ceri, D. Cerizza, E. Della Valle, and F. Michele Facca, *A software engineering approach to design and development of semantic web service applications*, In proceeding of the Semantic Web-ISWC 2006, pp. 172-186. Springer Berlin Heidelberg, 2006.

www.SID.ir

Archive of SID

[9] C. Li, B. Cheng, J. Chen, P. Gu, N. Deng, and D. Li, *A Web Service Performance Evaluation Approach Based on Users Experience*, In Proceeding of the conference on Web Services (ICWS), 2011 IEEE International Conference on, pp. 734-735. IEEE, 2011.

[10] R. Angmo, and M. Sharma, *Selenium Tool: A Web based Automation Testing Framwork*, International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), 2014.

www.SID.ir