ارزیابی ساختار وب‌سایت با استفاده از مدل‌سازی رفتار بازدیدکنندگان: یک مطالعه موردی

بابک سهرابی[1*]، بابک عابدین[2]

۱- استادیار گروه مدیریت فناوری اطلاعات، دانشگاه تهران، تهران، ایران
۲- کارشناس ارشد مدیریت فناوری اطلاعات، دانشگاه تهران، تهران، ایران

دریافت: ۸٤/۹/۲۸ پذیرش: ۸٦/٤/۲٤

چکیده

رشد رو به گسترش اینترنت و به تبع آن، توسعه کسب‌وکارهای الکترونیک در جهان باعث شده تا وب‌سایتها از اهمیت بالایی برخوردار شده، نقش غیرقابل‌انکاری در برقراری ارتباط الکترونیکی بین سازمانها و مؤسسات با مشتریانشان پیدا کنند. هرچند در سالهای ابتدایی توسعه خدمات الکترونیکی، توسعه وب‌سایت از نظر کمی و اندازه مورد توجه بود، اما اکنون توسعه وب‌سایتها از نظر کیفی و کارایی مد نظر است؛ به این معنا که چگونه و با چه ابزارهایی می‌توان عملکرد وب‌سایتها را بهبود داد تا مشتریان زودتر و مؤثرتر به مقاصد خود در وب‌سایت دست‌یافته، از سرگشتگی آنها در شبکه عظیم صفحات وب جلوگیری گردد. در این راستا، ساختار وب‌سایت، بیانگر نحوه سازمان‌دهی پیوندها و ارتباطات صفحات وب‌سایت بوده، ارزیابی آن نیازمند استفاده از روش و شاخصی مناسب است. این مقاله، علاوه بر مرور روشهای فعلی مدل‌سازی وب‌سایت و ارزیابی ساختار آن، روشی را برای اندازه‌گیری وضعیت فعلی ساختار ارتباطات صفحات سایت ارائه کرده و نتایج استفاده از آن را در سایت یکی از دانشگاههای کشور ارائه می‌کند.

کلید واژه‌ها: وبکاوی، مدل‌سازی وب‌سایت، گراف، تحلیل رفتار مشتری.

---

* نویسنده مسوؤل مقاله: E-mail:bsohrabi@ut.ac.ir

Archive of SID

www.SID.ir

۱- مقدمه

رشد رو به گسترش اینترنت و به تبع آن، توسعه کسب‌وکارهای الکترونیک در جهان باعث شده تا وب‌سایتها از اهمیت بالایی برخوردار شده، نقش غیرقابل انکاری در برقراری ارتباط الکترونیکی بین سازمان‌ها و مؤسسات با مشتریانشان پیدا کنند؛ چراکه در چنین محیط الکترونیکی‌ای، وب‌سایت، پل ارتباطی بین سازمانها و مشتریانشان بوده، مشتریان از طریق صفحات مختلف وب‌سایت به خدمات سازمان دسترسی پیدا می‌کنند. در سالهای ابتدای توسعه اینترنت و کسب‌وکارهای الکترونیک، سازمانها و شرکتها درصدد توسعه هرچه بیشتر وب‌سایت، افزایش صفحات وب و جذب هرچه بیشتر بازدیدکنندگان به سایت خود بودند، لکن در سالهای اخیر، بهبود کیفیت وب‌سایت، اولویت بالاتری پیدا کرده‌است[۱]. در توجیه این مسأله به طور ساده می‌توان گفت که افزایش صفحات وب‌سایت یک سازمان، یک کلاف بزرگ و پیچ‌درپیچ می‌سازد که حرکت در آن و یافتن خدمات یا محصولات مورد نظر را کاری وقت‌گیر، خسته‌کننده و حتی ناموفق می‌کند. این مسأله باعث می‌شود تا سازمان نتواند تمام محصولات و خدماتش را به نمایش بگذارد و در نتیجه احتمالاً مشتری خود را از دست بدهد.

بنابراین باید بتوان ساختار وب‌سایت را مدل‌سازی کرده، با استفاده از شاخص یا شاخصهای مناسب وضعیت فعلی آن را ارزیابی کرد و در نهایت با ارائه راهکارهایی، آن را بهبود بخشید. قدم اول در چنین مسیری، تعیین روشی است که بتواند ساختار ارتباطات وب-سایت را مدل‌سازی کرده، شاخصهایی را برای اندازه‌گیری آن معرفی کند [۳،۲]. در ادبیات موضوع، از روشها وتئوریهای گوناگونی برای مدل‌سازی وب‌سایت و تشریح رفتار مشتری در آن استفاده شده که هرکدام از دیدگاه خاص خود به مسأله پرداخته‌اند. به عنوان مثال، در حالی که برخی روشها روی خواص و ویژگیهای ظاهری و نحوه طراحی و رنگ‌آمیزی در سایت متمرکز شده‌اند، برخی دیگر به ارتباط مفهومی و منطقی بین صفحات مختلف اشاره می‌کنند [٤]. اگرچه هر کدام از روشها و تئوریهای مختلف به کارگرفته شده در این حوزه دارای معایب، مزایا و نیز محدودیتها و قابلیتهای خاص است، اما اینکه از کدامیک استفاده شود بستگی به رویکرد تحلیلگر و نوع مسأله دارد. به عبارت دیگر، در جایی ممکن است تحلیلگر، فرایند تصمیم‌گیری مشتری و ارتباط منطقی بین صفحات وب‌سایت و اجزای مختلف ساختار سایت را مورد تجزیه و تحلیل قرار دهد و تعیین کند که کدام صفحات دارای ارتباط

Archive of SID

www.SID.ir

معنایی بیشتری با یکدیگرند و باید در کنار همدیگر قرار گیرند؛ اما در جایی دیگر ممکن است تحلیلگر بخواهد فقط روی رویدادهای رخ داده روی وب‌سایت مطالعه کند؛ بدین معنا که آنچه روی وب‌سایت اتفاق افتاده را جمع‌آوری کرده، به مطالعه روابط بین اجزای مختلف وب‌سایت می‌پردازد [٦،٥].

در این مقاله برای مدل‌سازی ساختار وب‌سایت، روشی برای تحلیل رفتار مشتری معرفی و به‌کارگیری می‌شود. از آنجا که این روش بر مبنای نتایجی که از تعدادی فرایند وب‌کاوی به دست آمده استفاده می‌کند، در بخش سوم از مقاله به معرفی مختصر استخراج دانش و وب‌کاوی پرداخته می‌شود. سپس در بخش چهارم، انواع روشهای مدل‌سازی در وب‌سایت مرور شده، روش انتخاب شده در این مقاله در بخش پنجم تشریح می‌گردد. بخش ششم، نتایج مدل شرح داده‌شده را بیان می‌کند و در نهایت از مطالب مقاله نتیجه‌گیری می‌شود.

۲- روش تحقیق

از نظر هدف، سه نوع «تحقیق بنیادی»[1]، «تحقیق کاربردی»[2] و «تحقیق عملی»[3] وجود دارد [۸،۷] که پژوهش حاضر از نوع کاربردی است، چرا که هدف از آن، توسعه دانشی خاص در زمینه مدیریت ساختار و محتوای وب‌سایت است که به وسیله آن، نیازی مشخص شناسایی و برطرف می‌گردد. از نظر ماهیت روش، سه نوع تحقیق «توصیفی»[4]، « آزمایشی»[5] و «تاریخی»[6] وجود دارد. با توجه به اینکه در این تحقیق، هدف توصیف عینی، واقعی و منظم خصوصیات یک موقعیت یا یک موضوع است، از نوع توصیفی قابل‌طبقه‌بندی است. همچنین تحقیق توصیفی دارای انواع گوناگونی است که این پژوهش از نوع توصیفی- مطالعه موردی است، چرا که به بررسی جنبه خاصی از پدیده معینی می‌پردازد. این مورد، سیستم وب‌سایت یک سازمان با حد و مرز مشخص و متشکل از عناصر و عوامل مرتبط به هم است.

در پژوهش حاضر به دنبال مدل‌سازی مجموعه‌ای از رویدادها و با استفاده از یک روش شناخته شده معین هستیم تا بتوان کاربرد این روش را در حوزه مورد مطالعه مشاهده کرد.

1. Fundamental Research
2. Applied Research
3. Action Research
4. Descriptive Research
5. Experimental Research
6. Historical Research

Archive of SID

www.SID.ir

از این رو، این تحقیق به لحاظ ماهیت، فاقد فرضیه است. همچنین از آنجا که محیط مورد مطالعه در این مقاله وب‌سایت یک دانشگاه است، برای تحلیل آن از داده‌های ذخیره شده توسط کامپیوتر سرور استفاده می‌شود. بنابراین که کل جامعه برای تحلیل به‌کارگرفته شده و این پژوهش نمونه آماری نخواهد داشت. داده‌های استفاده شده در این تحقیق از لاگ‌فایلها استخراج شده که در بخشهای بعدی مقاله شرح داده شده‌است.

مسأله و انگیزه اصلی از این پژوهش، مدل‌سازی وب‌سایت و معرفی شاخصی برای اندازه‌گیری وب‌سایت دانشگاه تهران است. از آنجا که این وب‌سایت روزانه مورد استفاده بسیاری از دانشجویان و بازدیدکنندگان مختلف قرار می‌گیرد و نیز تلاشهای زیادی برای توسعه و به‌روزرسانی آن صورت می‌گیرد، نیاز به داشتن شاخصی برای تعیین عملکرد آن از اهمیت زیادی برخوردار است؛ اما چون وضعیت فعلی یک وب‌سایت، مقوله‌ای کیفی بوده، اندازه‌گیری آن با دشواری و پیچیدگیهای خاص مواجه است، نیاز به تعیین یک شاخص کمّی و اندازه‌گیری آن، هدف اصلی این مقاله را تشکیل داده‌است.

۳- مروری بر مفاهیم و فنون استخراج دانش و وب‌کاوی

"استخراج دانش و داده‌کاوی[1]"(KDD)، فرایند شناسایی الگوهای قابل‌درک، مفید، جدید و معتبر در انبوهی از داده‌ها است. «داده»، یعنی عنوان مجموعه‌ای از حقایق و «الگو[2]»، نمودی است که به شرح زیرمجموعه داده‌ها و مدلی عملی برای زیرمجموعه‌ها می‌پردازد. واژه «فرایند»، بیانگر آن است که KDD متشکل از مراحلی است که شامل فراهم‌سازی داده‌ها، شناسایی الگوها، برآورد اطلاعات و اصلاح آنها می‌شود. داده‌کاوی، گامی از فرایند KDD است که متشکل از کاربرد تحلیل داده‌ها و الگوریتم‌های اکتشاف بوده، تحت محدودیتهای قابل‌قبول محاسبه‌ای، باعث ارائه الگو و مدل خاص داده‌ها می‌شود [۹].

وب‌کاوی، شاخه‌ای از داده‌کاوی است که با استخراج دانش مفید از شبکه گسترده وب سر و کار دارد[۱۰]. وب‌کاوی را می‌توان به سه نوع تقسیم کرد. اول وب‌کاوی محتوایی[3] که بر اطلاعات موجود در صفحات وب تمرکز می‌کند و بیشتر روی داده‌هایی از نوع متن در

1. Knowledge Discovery and Data mining
2. Pattern
3.Web content mining

Archive of SID

www.SID.ir

صفحات وب تمرکز دارد و از جمله کاربردهای متداول آن، طبقه‌بندی محتوایی و رتبه‌بندی محتوایی صفحات وب است. وب‌کاوی ساختاری[1] نوع دیگری از وب‌کاوی است که روی ساختار وب‌سایت تمرکز کرده، بیشتر اطلاعات ساختاری صفحات وب (مانند پیوندهای صفحات) را مورد بررسی قرار می‌دهد و از جمله متداولترین کاربردهای آن، طبقه‌بندی صفحات وب بر اساس پیوندها، رتبه‌بندی صفحات وب‌سایت بر اساس ترکیبی از ساختار و محتوای آنها و مهندسی معکوس مدلهای وب‌سایت است. نوع سوم وب‌کاوی، وب‌کاوی کاربری[2] نام دارد که با استخراج دانش از لاگ فایل سر و کار دارد و داده‌های مرجع آن، بیشتر شامل لاگ‌فایلهایی است که توسط سرور و در قالبهای استاندارد ذخیره می‌شود. از جمله کاربردهای متداول این مورد، می‌توان به فنون مدل‌سازی کاربران وب‌سایت، مانند سازگارسازی[3] و شخصی‌سازی[4] وب‌سایت اشاره کرد.

متداولترین و غنی‌ترین منبع داده برای وب‌کاوی، وب سرور است [۱۱]. سرورها حجم زیادی از اطلاعات را در لاگ‌فایلها ذخیره می‌کنند. این لاگ‌فایلها شامل داده‌ای پایه‌ای مانند نام کاربر و IP، زمان و تاریخ دسترسی به صفحات وب، لیست تمام صفحات مورد بازدید قرار گرفته و ... هستند و معمولاً به شکل و قالب استاندارد ذخیره می‌گردند. در برخی موارد، این داده‌ها در فایلهای متنی و در برخی موارد در فایلهای پایگاه داده‌ای نگهداری می‌شوند. نمونه‌ای از لاگ‌فایل در شکل ۱ نشان داده شده‌است.

1. Web structure mining
2. Web usage mining
3. Adaptive
4. Personalization

Archive of SID

www.SID.ir

```
#Software: Microsoft Internet Information Services ۶.۰
#Version: ۱.۰
#Date: ۲۰۰۴-۰۷-۰۲ ۱۳:۲۷:۵۱
#Fields: date time s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip cs(User-Agent) sc-status sc-substatus sc-win۳۲-status
۲۰۰۴-۰۷-۰۲ ۱۳:۲۷:۵۱ ۶۷.۱۹.۱۲۲.۱۲۲ GET / - ۸۰ - ۲۱۷.۲۱۸.۱۰۲.۳
Mozilla/۴.۰+(compatible;+MSIE+۶.۰;+Windows+NT+۵.۲;+.NET+CLR+۱.۱.۴۳۲۲) ۴۰۳ ۱۴۵
۲۰۰۴-۰۷-۰۲ ۱۳:۲۹:۰۵ ۶۷.۱۹.۱۲۲.۱۲۲ GET /index.asp - ۸۰ - ۲۱۷.۲۱۸.۱۰۲.۳
Mozilla/۴.۰+(compatible;+MSIE+۶.۰;+Windows+NT+۵.۲;+.NET+CLR+۱.۱.۴۳۲۲) ۴۰۴ ۲۰
```

شکل ۱ نمونه‌ای از یک لاگ‌فایل

مطابق شکل، لاگ‌فایل شامل تعدادی اقلام اطلاعاتی است که هرکدام معنای خاصی، مانند زمان و تاریخ بازدید از وب‌سایت، صفحه بازدید شده، خطاهای روی‌داده و غیره را شامل می‌شود. حجم این فایلها گاه به میلیونها خط می‌رسد. بنابراین، استخراج نتایج مفید از آنها نیازمند استفاده از روشها و فنون خاص استخراج دانش و داده‌کاوی است. از جمله مهمترین فنون استخراج دانش، قواعد انجمنی[1]، تحلیل مسیر[2]، الگوهای ترتیبی[3]، خوشه‌بندی[4] و دسته‌بندی[5] است. از میان این روشها، در این مقاله از روش تحلیل مسیر استفاده شده تا به وسیله آن بتوان مسیرها و مجموعه صفحاتی را که مشتریان بازدید می‌کنند شناسایی کرده، مورد تجزیه و تحلیل قرار داد.

۴- روشهای مدل‌سازی وب‌سایت و تحلیل رفتار مشتری

در ادبیات موضوع، از روشها وتئوریهای گوناگونی برای تشریح رفتار مشتری در اینترنت استفاده شده که هرکدام از دیدگاه خاص خود به مسأله پرداخته‌اند. به عنوان مثال، در حالی که برخی روشها روی خواص و ویژگیهای ظاهری و نحوه طراحی و رنگ‌آمیزی در سایت

1. Association rules
2. Path analysis
3. Sequential patterns
4. Clustering
5. Classification

Archive of SID

www.SID.ir

متمرکز شده‌اند، برخی دیگر به ارتباط مفهومی و منطقی بین صفحات مختلف اشاره دارند.

اگرچه هر کدام از روشها و تئوریهای مختلف به کار گرفته شده در این حوزه دارای معایب، مزایا و نیز محدودیتها و قابلیتهای خاص هستند، اما اینکه از کدامیک استفاده شود بستگی به رویکرد تحلیلگر و نوع مسأله دارد. در هرکدام از رویکردهای مختلف، روشهای گوناگونی معرفی و به‌کار گرفته شده‌است. نگاهی اجمالی به روشهای مختلف تحلیل و مدل‌سازی رفتار مشتری در وب‌سایت نشان می‌دهد که بسیاری از روشها مبتنی بر فنون آمار بوده، از قابلیتها و محبوبیت زیادی برخوردارند. در ادامه، به تقسیم‌بندی و شرح مختصری هر کدام از روشهای به‌کاررفته، خواهیم پرداخت. اما قبل از آن لازم است تعریفی از «کاربردپذیری»[1] ارائه گردد. این واژه به‌کرات در ادبیات موضوع مدل‌سازی و تحلیل رفتار مشتری در وب‌سایت استفاده شده و مفهوم و مقصود اصلی از آن «سهولت استفاده از وب‌سایت و حرکت در آن» است. به عبارت دیگر، هرچه مشتری و بازدیدکننده بهتر، راحت‌تر و سریعتر بتواند از صفحه‌ای به صفحه دیگر حرکت کرده، به مقاصد خود برسد، میزان کاربردپذیری وب‌سایت بالاتر خواهد بود.

۴-۱- زنجیره‌های مارکوف: زنجیره‌های مارکوف، شاخه‌ای از تئوریهای احتمالات هستند که به علت ماهیت احتمالی و امکان تحلیل روی رویدادهای احتمالی مرتبط، از محبوبیت زیادی در مدل‌سازی رفتار کاربران در وب‌سایت و اندازه‌گیری کاربردپذیری وب برخوردارند. همچنین از جمله دیگر دلایل کاربرد این روش، امکان اندازه‌گیری کاربردپذیری وب‌سایت به صورت کمّی است [۱۳]. مدلهای مارکوف، روشی شناخته‌شده برای مدل‌سازی فرایندهای تصادفی با ساختاری با تعداد حالتهای محدود است. به همین دلیل استفاده از این روش برای مدل‌سازی رفتار کاربران در وب‌سایت از محبوبیت خوبی برخوردار است [۱٤].

به طور کلی، ورودی این نوع مسائل عبارت از دنباله‌ای از صفحات وب است که توسط بازدیدکننده‌ای بازدید شده‌اند و هدف، ساختن مدلهای مارکوفی است که می‌تواند رفتار وی را در وب‌سایت مدل کرده، پیش‌بینی کند که او چه صفحاتی را در مراحل بعدی بازدید خواهدکرد.

---

1. Usability

Archive of SID

www.SID.ir

۴-۲- تحلیل پوششی داده[1]: DEA یک روش غیرپارامتریک است که می‌تواند ارزیابی واحدها را از یک فضای یک بعدی به یک دامنه چند بعدی و با حفظ یکپارچگی بین آنها، گسترش دهد. هر چند معیار کارایی مفهومی، کلی است، اما DEA برای به دست آوردن یک معیار کارایی از آوردن چندین فرض غیر دقیق استفاده نمی‌کند. این روش حتی اگر چند ورودی و خروجی داشته باشد، فقط یک معیار کارایی ارائه می‌دهد. DEA کارایی واحدهای تصمیم‌گیری (DMU)[2] را بدون اینکه هیچ فرضی درباره ارتباط کاری بین ورودیها و خروجیهای هر واحد در نظر بگیرد، ارزیابی می‌کند [۱۵]. در مطالعات انجام شده در ارزیابی عملکرد وب‌سایت با استفاده از روش تحلیل‌پوشش داده، محققان با ارائه کاربردهای مختلف از واحدهای تصمیم‌گیری در محیط وب‌سایت، سعی در اندازه‌گیری عملکرد کلی وب‌سایت دارند[۱۶؛ ۱۷].

۴-۳- روشهای مبتنی بر نظرخواهی

گروهی دیگر از روشها وجود دارند که عمدتاً به شناسایی کاستیها و مشکلات طراحی وب-سایت و تعیین سطحی کیفی آن می‌پردازند. دو روش اصلی در این حوزه عبارتند از: ارزیابی ابتکاری[3] و سنجش کاربران[4] [۱۸]. در روش ارزیابی ابتکاری، جامعه‌ای از خبرگان شناسایی و تشکیل می‌شود. سپس این جامعه به مطالعه و بررسی وب‌سایت پرداخته، اشکالات و ایرادات آن را شناسایی و استخراج می‌کنند و کاستیهای آن را از دید کاربر نهایی تعیین می‌کند. روش سنجش کاربران، کیفیت وب‌سایت را به طورمستقیم از دید بازدیدکنندگان وب‌سایت ارزیابی می‌کند؛ بدین ترتیب که از تعدادی از بازدیدکنندگان خواسته می‌شود از وب‌سایت (یا بخشی از آن) بازدیدکرده، خدمات مختلف آن را مورد بهره‌برداری قرارداده، در نهایت اشکالات و کاستیهای مشاهده شده را گزارش دهند.

۴-۴- گرافهای جهت‌دار

یکی از روشهایی که به‌تازگی مورد توجه محققان حوزه تحلیل وب‌سایتها قرارگرفته، استفاده

1. Data Envelopment Analysis (DEA)
2. Decision Markin Unit (DMU)
3. Heuristic evaluation
4. User testing

Archive of SID

www.SID.ir

از گرافها و خواص جالب توجه آنها است. در این روش، وب‌سایت را می‌توان به عنوان یک گراف ریاضی، بدین ترتیب مدل‌سازی کرد که صفحات وب‌سایت، نقش «گره‌ها[1]»و پیوندهای بین صفحات، نقش «کمان[2]» را در گراف برعهده خواهند داشت. استفاده از این روش مدل-سازی وب‌سایت، علاوه بر تحلیل کاربردپذیری وب، در موتورهای جستجوگر و الگوریتمهای جستجو بسیار مورد توجه بوده‌است [۱۹]. همچنین از گرافها در طراحی و توسعه سیستمهای مشتری‌مدار در وب‌سایتها استفاده شده [۲۰] که خصوصاً در سایتهای تجارت الکترونیک که ارتباط نزدیک با مشتری، از جمله فاکتورهای حیاتی آنها است، بسیار مورد توجه قرار دارد. علاوه بر اینها، تحقیقات زیادی در تحلیل وب‌سایت با استفاده کامل و انحصاری از تئوری گرافها صورت گرفته که شاید بتوان گفت این تحقیقات ریشه در تئوری «دنیای کوچک[3]» پدیده‌ها در شبکه داشته‌اند [۲۱].

استفاده از گرافها در مدل‌سازی وب‌سایت دارای مزایای قابل‌توجهی است که از جمله مهمترین آنها می‌توان به قابلیت نمایش و قدرت بالای آن در به تصویر درآوردن وب‌سایت در قالب گراف اشاره کرد. گرافها دارای انواع گوناگونی هستند که هرکدام ویژگیها و کاربردهای خاص دارد[۲۲]. از این میان، گرافهای جهت‌دار دارای خصوصیات ویژه‌ای هستند که باعث شده مورد توجه محققان در حوزه مدل‌سازی وب‌سایت و تحلیل ساختار آن قرار بگیرند [۲۳]. در بخشهای بعدی این مقاله به این روش بیشتر پرداخته خواهدشد.

۵- معرفی روش انتخاب شده برای مدل‌سازی وب‌سایت

در بخش گذشته به انواع روشهای به‌کارگرفته شده در ادبیات موضوع برای مدل‌سازی رفتار مشتری و وب‌سایت اشاره شد که هرکدام از روشها، مزایا و محدودیتهای خود را دارند. از میان این روشها، در این تحقیق از روش «گراف» به منظور مدل‌سازی وب‌سایت و رفتار مشتری استفاده می‌شود. برخی از مهمترین مزایای استفاده و انتخاب این روش عبارتند از:

- گرافها دارای قدرت بالای نمایشی هستند.
- گراف دارای پشتوانه قوی در علوم ریاضی است. این قابلیت به ما امکان می‌دهد از

1. Nodes
2. Arcs
3. Theory of "small world"

Archive of SID

www.SID.ir

شاخصهای عینی و عددی برای مدل‌سازی و بهبود مدل استفاده کنیم.

- استفاده از گراف در علوم کامپیوتری دارای سابقه طولانی و مطلوب است.
- استفاده از گراف در حوزه تحقیقاتی کاربردپذیری وب‌سایت رو به رشد بوده و بسیار مورد توجه قرارگرفته‌است.

۵-۱- معرفی گراف

گراف G، یک مجموعه ناتهی متناهی V(G) از عناصری با نام رأس (گره) و یک مجموعه E(G) (که می‌تواند تهی باشد) از زیر مجموعه‌های دو عضوی V(G) است که هر عضو آن، یک یال (ضلع) گراف نامیده می‌شود. مجموعه V(G) مجموعه رأسهای گراف G و E(G) مجموعه یالهای گراف G نامیده می‌شوند. برای هر گره از گراف، درجه v یعنی deg v برابر تعداد رئوسی است که در مجاورت v هستند. اگر G از مرتبه p داشته باشد و v رأسی از G باشد آنگاه ۱- p ≥ deg v ≥ ۰ [۲٤].

یک گشت در گراف G یک دنباله متناوب از رئوس و یالها است که ابتدا و انتهای آن رئوس هستند، به طوری که i = ۱، ۲، ۳، ...، n ، e(i) = v(i-۱)v(i) و W: v(۰)، e(۱) ، v(۱)، e(۲)، v(۳)، ...، e(n)، v(n)

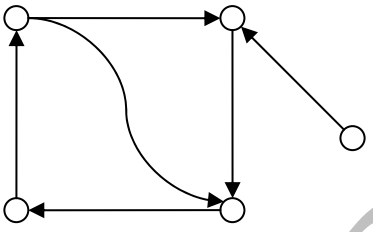
چون W از v(۰) شروع می‌شود و به v(n) پایان می‌یابد، W را به عنوان یک v(۰)-v(n) گشت مطرح می‌کنیم.

همچنین طول W برابر n تعریف می‌شود، چرا که W در رابطه با n یال است. گشت با طول صفر، یک گشت بدیهی نامیده می‌شود. بنابراین تعریف، یک مسیر، یک گشت است که رأس تکراری ندارد.

یک گراف جهت‌دار D، شکل۲، مجموعه متناهی و ناتهی V(D) از رئوس و یک مجموعه (شاید تهی) شامل زوجهای «مرتب» رأسهای مجزا است. اعضای E(D) کمان نامیده می‌شوند. رئوس گرافهای جهت‌دار D با دایره‌های کوچک و یک کمان u)، (v از D با رسم یک منحنی یا پاره خط جهت‌دار از رأس u به v نشان داده می‌شوند. از آنجا که u)، (v و v)، (u کمانهایی مجزا هستند، دو رأسی می‌توانند با دو کمان به یکدیگر متصل شوند که دارای جهتهای مخالف باشند.

Archive of SID

www.SID.ir



شکل۲ یک نمونه گراف جهت‌دار

۵-۲- مروری بر ادبیات مدل‌سازی وب‌سایت با استفاده از گراف

استفاده از گراف در حوزه مطالعه و تحلیل وب‌سایت بسیار مورد توجه بوده، کاربرد آن رو به گسترش است.

تل‌وال و ویلکینسون [۲۵] در تحقیق خود، وب‌سایتهای دانشگاههای سه کشور استرالیا، نیوزیلند و انگلستان را مورد تجزیه و تحلیل قرار داده و با استفاده از ویژگیهای گرافها، آنها را مدل‌سازی کرده‌اند. هدف آنها در این مطالعه، مقایسه کارایی و عملکرد سایتهای مختلف در ارائه محتویات و مطالب به دانشجویان و دیگر بازدیدکنندگان وب‌سایت بوده‌است. آنها با بهره‌گیری از مفاهیم گراف، مدلی را برای ارائه بهتر محتویات سایتها به مراجعه‌کنندگان وب-سایتهای دانشگاهی پیشنهاد داده‌اند.

هانگ و همکارانش [۲٦] از گرافها در مدل‌سازی سیستمهای پیشنهاد دهنده در تجارت الکترونیک[1] استفاده کرده‌اند. آنها در این مطالعه، سیستمهایی را بررسی می‌کنند که به مشتریان در پیدا کردن نیازهای خود در وب‌سایتهای تجارت الکترونیک کمک می‌کند. در این مطالعه از مزیت گرافها در نمایش داده‌ها استفاده کرده و سیستم مورد نظر را بر اساس مدل گراف خود توسعه داده‌اند.

ژو و همکارانش [۲۷] از گراف برای مدل‌سازی ساختار وب‌سایت استفاده کرده‌اند. ابتکار آنها در استفاده از روابط ریاضی حاکم بر گرافها به منظور اندازه‌گیری عملکرد وضعیت فعلی وب‌سایت و سپس ارائه یک استراتژی بهبود برای ساختار وب‌سایت مورد مطالعه خود است. ژو به همراه چن[۲۸] در تحقیقی دیگر، پژوهش خود را گسترش داده، به شرح اجزای بیشتری

1. E-Commerce recommender system

Archive of SID

www.SID.ir

از رویکرد خود مبنی بر استفاده از گرافها در حوزه مدل‌سازی ساختار وب‌سایت پرداخته و نتایج تحقیق خود را به طور عملی پیاده‌سازی کرده‌اند.

۶- مدل‌سازی وب‌سایت و اندازه‌گیری وضعیت موجود

در بخش قبلی، مزایای استفاده از گراف در مدل‌سازی وب‌سایت شرح داده‌شد. همچنین مروری بر تعاریف و خواص گرافها ارائه گردید. در این بخش، مصادیق مفاهیم گراف در وب‌سایت تشریح شده، شاخصها و معیارهایی معرفی و در نتیجه، عملکرد وب‌سایت اندازه-گیری می‌گردد.

درصورتی که یک وب‌سایت به شکل یک گراف مدل‌سازی گردد، هر صفحه یک «گره» و لینک بین صفحات، یک «کمان» تعریف می‌شود. برای اینکه گراف وب‌سایت تشکیل شود باید آمار بازدید از هرکدام از صفحات و نیز ارتباطات بین صفحات مشخص گردد. تعیین چنین ارتباطاتی، کلید اصلی مدل‌سازی در قالب یک گراف است.

برای به دست آوردن این اطلاعات و تعیین ارتباطات بین صفحات می‌توان از نرم-افزارهای تحلیل وب‌سایت استفاده کرد. این نرم‌افزارها، گزارشها و اطلاعات کلی درباره وب-سایت ارائه می‌کنند. ورودی به این نرم‌افزارها، لاگ‌فایل وب‌سایت بوده، خروجیهای اصلی آنها، گزارشهای

تحلیل وب‌سایت است که از جمله مهمترین این گزارشها می‌توان گزارش تعداد بازدید از صفحات وب، گزارش زمان بازدید از صفحات، گزارش مسیرهای طی شده در سایت، گزارش موتورهای جستجوگر و گزارش صفحات ورودی به/خروجی از سایت را نام برد. در حال حاضر، انواع گوناگونی از این نوع نرم‌افزارها در بازار وجود دارد که در این مقاله از نرم‌افزار Mach5 Analyzer استفاده شده‌است [۲۹].

قدم اول در ترسیم گراف وب‌سایت، استخراج مسیرها در آن است. لذا برای تحلیل وقایع آن، وب‌لاگ سایت دانشگاه تهران، که حاوی داده‌های این وب‌سایت از ابتدا تا انتهای پاییز سال ۱۳۸٤ بوده، به نرم‌افزار داده شده و آندسته از مسیرهایی که بیش از ۷۰۰ مرتبه بازدیدشده‌اند، در جدول ۱ نشان داده شده‌اند.

Archive of SID

www.SID.ir

جدول ۱ مسیرهای بازدید شده در وب‌سایت www.ut.ac.ir (ابتدای مهر ۸٤ تا انتهای آذر ۸٤)

ردیف	مسیر	تعداد بازدید	ردیف	مسیر	تعداد بازدید
A	(۱)/index (۲)/mail	٥۳٤۰۰	E	/index /main-links (۸)/administration	۱٥۲٥
B	/index (۳)/news	۳۸۸۹	F	/index (۹)/manues (۱۰)/addresses	۷۷۷
C	/index (٤)/faculties (٥)/engineering	۱۸٦۹	G	/index /main-links (۱۱)/research	۱۷۰۰
D	/index (٦)/main-links (۷)/overview	۱۸۱٥	H	/index /manues (۱۲)/sitemap	۷۱۹

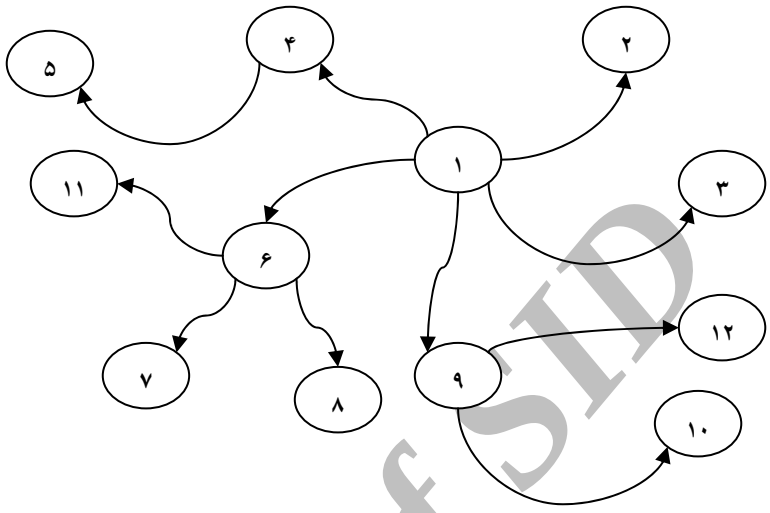
این جدول نشان می‌دهد که چه صفحاتی به‌طور متوالی بازدید شده‌اند. هر سطر جدول، بیانگر آن است که چه صفحاتی توسط چند نفر بازدید شده‌است[۱]. اعداد داخل پرانتز، کد مربوط به هر صفحه است.

با استفاده از اطلاعات به دست آمده در این جدول، شکل ۳ گراف وب‌سایت را نشان می‌دهد. عدد مربوط به گره در گراف زیر، همان کد تعریف شده در جدول قبلی است.

۱. لازم به توضیح است که در این مطالعه تنها آندسته از صفحاتی که مستقیما تحت آدرس ut.ac.ir قابل دسترسی هستند مورد بررسی قرار گرفته‌اند. به علت گسستگی محل‌های نگهداری لاگ فایل‌های سایت‌های مشتق از این آدرس، دیگر سایت‌ها مانند سایت کتابخانه‌ها و پایگاه‌های علمی، در حوزه مطالعه نمی‌باشند.

Archive of SID

www.SID.ir



شکل ۳ گراف وب‌سایت دانشگاه تهران

با استفاده از گراف ترسیم شده، شمایی از ارتباطات بین صفحات وب نمایش داده شده‌است. این شکل، بیانگر ارتباطات بین صفحات است و حجم این ارتباطات و میزان بازدید از هر کدام از مسیرها از جدول ۱ قابل‌مشاهده است. آنچه در این گراف از اهمیت فوق‌العاده برخوردار است این است که این نمودار، حاصل رویدادهایی است که واقعاً اتفاق افتاده، نه آنچه قرار بوده اتفاق بیفتد. برای اندازه‌گیری وضعیت فعلی، نیاز به شاخصهایی است که در ادامه به تعریف آنها پرداخته می‌شود. این شاخصها بر اساس نتایج تحقیقات ژو و همکارانش [۲۷، ۲۸] به دست آمده و در ادامه در تعاریف ۱ تا ۹ به آنها پرداخته شده‌است.

تعریف ۱

مدل ساختار وب‌سایت را می‌توان به صورت یک گراف وزن‌دار با تابع $G = (P, L, W)$، تعریف کرد، به طوری که: مجموعه تمام گره‌ها در گراف G: $P = \{ Pi \mid i \varepsilon [۱ \text{ و } n] \}$، $P=\{ P۱ ,P۲ ,P۳,... ,P۱۲ \}$ مجموعه تمام پیوندها در گراف G: $L= \{L(i,j) \mid i \neq j , j \varepsilon [۱ \text{ و } n]\}$ به طوری که L یک پیوند از Ni به Nj است

Archive of SID

www.SID.ir

L= {L(۱،۲) L(۱,۳) ، L(۱،٤) ، L(۲،۳) ، … ، L(۹،۱۲)}

W مجموعه وزن‌های کمان‌های گراف G :{Wij | i ≠ j و ijε [۱ و n]} = W

Wij احتمال آن است که کاربری که صفحه Ni را بازدید کرده، پیوندرا L(i,j) انتخاب کند که مقدار آن برابر است با $w_{i,j} = R_{ij} / \sum_{k=1}^{OD(i)} R_{ikk}$

(Rij در تعریف ۷ آمده)، W={W۱۲, W۱۳, W۱٤, W۲۳, …, W۹۱۲}

تعریف ۲

الگوی بازدید کاربران را می‌توان از وب‌لاگها استخراج کرد، بدین گونه که داده‌های لاگ‌فایل به گروههای بازدید کاربران[1] (UVS) تقسیم شود. هر رکورد در هر گروه بازدید کاربران به یک «دسته» از کاربرانی اختصاص دارد که همگی مسیر یکسانی را بازدید و طی کرده‌اند. بنابراین، هر سطر از جدول۱ یک گروه بازدید کاربران(UVS)است. با استفاده از این تعریف، گروههای بازدید کاربران را می‌توان به شکل ماتریسی نمایش داد که ستونهای آن نشان‌دهنده گروههای بازدید(M) و سطرهای آن، نشانگر صفحات وب‌سایت (N) است. با توجه به دو جدول قبلی، مقادیر N و M به ترتیب عبارت است از:

تعریف۳

در ماتریس تعریف شده، هر عضو در سطرi(i ε [۱ ,M])، از ماتریسV، موقعیت هر صفحه را در هر گروه بازدید کاربران ذخیره می‌کند. به همین ترتیب نیز هر عضو در ماتریس j (j ε [۱, n]) از ماتریس مذکور، مقدار صفحه j (یعنی P(j)) را در تمام گروههای بازدید ذخیره می‌کند. در نتیجه، مقدار عنصر *vij* برابر ۱ خواهد بود، اگر صفحه j در گروه بازدید i بازدید شده باشد و در غیر این صورت، مقدار آن، صفر است.

براساس این تعریف، ماتریس V برای سایت دانشگاه تهران، بر اساس جدول قبلی، به شکل زیر است که در آن، تعداد ستونها برابر تعداد صفحات و تعداد سطرها بیانگر تعداد مسیرهای جدول ۱ است.

1. User visiting sessions

Archive of SID

www.SID.ir

	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
A	۱	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
B	۱	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰
C	۱	۰	۰	۱	۱	۰	۰	۰	۰	۰	۰	۰
D	۱	۰	۰	۰	۰	۱	۱	۰	۰	۰	۰	۰
E	۱	۰	۰	۰	۰	۱	۰	۱	۰	۰	۰	۰
F	۱	۰	۰	۰	۰	۰	۰	۰	۱	۱	۰	۰
G	۱	۰	۰	۰	۰	۱	۰	۰	۰	۰	۱	۰
H	۱	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۱

تعریف۴

با تشکیل ماتریس بالا، $P.j$ عبارت از ستون jاُم از ماتریس V است که بر این اساس، تعداد عناصر این ستون با $|P.j|$ نشان داده می‌شود که مقدار آن برابر است با مجموع تمام مقادیر آن ستون. مشابه این تعریف برای $P.i$ برقرار است. همچنین $|P(i)|$، j بیانگر تعداد گروههای بازدیدی (UVS) است که هم Pi و هم Pj بازدید شده باشند. بنا بر این تعریف، برای سایت مورد مطالعه داریم:

- P.۳ = ستون سوم از ماتریس بالا و |P.۳| = ۱
- P۲. = سطر دوم از ماتریس بالا و |P۲.|= ۲

تعریف ۵

چند گره به دنبال هم که یک مسیر را نشان می‌دهند به صورت $Ni\rightarrow\ldots\rightarrow Nj$ نشان می‌دهیم که طول این مسیر برابر است با تعداد پیوندهای موجود بین آنها که با $|Ni\rightarrow\ldots\rightarrow Nj|$ نشان داده می‌شود.

تعریف ۶

مجموعه تمام مقاصد Pi ، تمام صفحاتی که از Pi به آنها می‌رود، با Di نشان داده می-شوند.

$\{P۱۲, P۱۱\} = D۹$

Archive of SID

www.SID.ir

تعریف ۷

درجه انجمنی[1] بین دو صفحه Pi و Pj با Rij نشان داده شده و عبارت است از احتمال شرطی اینکه اگر کسی Pi را بازدید کند، با چه احتمالی Pj را نیز بازدید می‌کند، و

i,j ε [۱, N] و Rhjε [۰,۱]ο

ο اگر |P.i|≠۰، آنگاه Rij= |P (i,j | / |P.i|: در غیر اینصورت Rij=۰

تعریف ۸

درجه اتصال[2] بین دو صفحه i و j با Cij نشان داده می‌شود، به طوری که هر چه مقدار آن بیشتر باشد، ارتباط بین دو صفحه بیشتر ارزیابی می‌شود و از رابطه زیر محاسبه می‌گردد:

به طوری که m تعداد مسیرهای موجود بین دو صفحه i و j است و ُ وزن هر مسیر است که طبق تعریف ۱ محاسبه می‌گردد:

Cij= Wi۱+W۱۳ +…+ W۵j

به‌عنوان مثال:

C۱ ۱۱= W۱۶+W۶

تعریف ۹

برای ارزیابی و اندازه‌گیری ساختار وب‌سایت، از شاخصی به نام میانگین اتصال[3] استفاده شده که با E نشان داده می‌شود و طبق رابطه زیر محاسبه می‌گردد:

$$E = \sum_{i \neq j} \frac{Cij}{[N \times (N - ۱)]}$$

مقدار N*(N-۱) برابر تعداد جفت صفحات سایت که N تعداد صفحه مورد بررسی است. طبق این تعریف، هر چه مقدار E بالاتر باشد، سایت از ساختار بهتری برخوردار خواهد بود، به این معنا که هر چه مقدار E بالاتر باشد، کاربران راحت‌تر می‌توانند صفحه مورد نظر خود را پیدا کنند.

با توجه به داده‌های جدول تعریف ۳، درجه انجمنی بین صفحات، طبق تعریف۷ اندازه‌گیری می‌شود. این درجه، بیانگر احتمال شرطی است که اگر صفحه i بازدید شود، به

1. Associate Degree
2. Connectivity Degree
3. Average Connectivity

Archive of SID

www.SID.ir

چه احتمالی صفحه j بعد از آن بازدید می‌گردد.

$\Rightarrow R_{12} = |P(2,1)| / |P_1| = 1/8 = 0/125$

$\Rightarrow \cdots$

$\Rightarrow R_{68} = |P(8,6)| / |P_6| = 1/8 = 0/125$

$\Rightarrow \cdots$

$\Rightarrow R_{9\ 12} = |P(12,9)| / |P_9| = 1/8 = 0/125$

لازم به توضیح است صفحاتی که درجه خروجی آنها، همان‌طور که در بخش معرفی گرافها آمده بود، صفر باشد، درجه انجمنی آنها هم صفر خواهد بود.

پس از محاسبه درجات انجمنی تمام صفحات می‌توان مطابق تعریف شماره ۱، وزن پیوندها یا کمانها را به ترتیب زیر به دست آورد:

$\Rightarrow W_{12} = R_{12} / (R_{12} + R_{13} + R_{14} + R_{16} + R_{19}) = 0/11$

$\Rightarrow \cdots$

$\Rightarrow W_{9\ 12} = R_{9\ 12} / (R_{9\ 10} + R_{9\ 12}) = 0/5$

این وزنهای استخراج شده بیانگر وزن نزدیکی و ارتباط بین دو صفحه‌ای است که به یکدیگر متصلند. بنابراین، هر چه مقدار آن بالاتر باشد، ارتباط بین دو صفحه بیشتر خواهد بود.

پس از محاسبه وزن بین صفحات می‌توان طبق تعریف ۸، درجه اتصال بین هر دو صفحه را محاسبه کرد. مفهوم درجه اتصال این است که ارتباط بین تمام صفحاتی که به طور مستقیم یا غیرمستقیم به یکدیگر ارتباط دارند در نظر گرفته نشده و درجه این اتصال اندازه‌گیری می‌شود. توضیح بیشتر آنکه در محاسبه وزن بین دو صفحه، اگر ارتباط مستقیم بین دو صفحه وجود نداشت، وزن آن صفر بود؛ اما در محاسبه درجه اتصال، تمام مسیرهایی که از یک صفحه خاص شروع و به صفحه دیگری ختم می‌شود مد نظر هستند.با محاسبه این شاخص،میزان ارتباط بین هر دو صفحه اندازه‌گیری می‌شود.مقدار درجه اتصال دو صفحه، بزرگتر یا مساوی وزن آن‌ها است.

مقادیر این درجه اتصال برای صفحات مختلف عبارت است از:

Archive of SID

www.SID.ir

$\Rightarrow C_{12} = W_{12} = ۰/۱۱$

$\Rightarrow ۰۰۰$

$\Rightarrow C_{15} = W_{14} * W_{45} = ۰/۲۲ * ۱ = ۰/۲۲$

$\Rightarrow ۰۰۰$

$\Rightarrow C_{1\,11} = W_{16} * W_{6\,11} = ۰/۱۱ * ۰/۳۳ = ۰/۰۳٦۳$

$\Rightarrow ۰۰۰$

$\Rightarrow C_{9\,12} = W_{9\,12} = ۰/۵$

باید توجه داشت در صورتی که یکی از مقادیر درجه اتصال صفر باشد، که برخی از مقادیر صفر در بالا نوشته نشده‌اند، لزوماً به این معنا نیست که هیچ پیوندی بین آن دو صفحه وجود ندارد، بلکه یعنی هیچ بازدیدی از هر دو آنها در قالب یک مسیر وجود نداشته‌است. به عنوان مثال، مقدار C_{42} صفر بوده، یعنی هیچ مسیری از صفحه ٤ به صفحه ۲ طی نشده، اما ممکن است این مسیر در وب‌سایت وجود داشته باشد.

به طور کلی، اگر وب‌سایت به صورت یک گراف کامل در نظر گرفته شود، آنگاه باید بین همه صفحات در حالت ایده‌آل یک پیوند وجود داشته باشد. این بدان معنا است که کل تعداد ارتباطات بین صفحات(N) برابر $N*(N-۱)$ خواهد بود. بنابراین، طبق تعریف۹، دراین مرحله، آخرین شاخص اندازه‌گیری می‌شود. بر اساس درجات اتصال به دست آمده، میانگین درجه اتصال سایت برابر است با:

$N = ۱۲ \; ; \; N*(N-۱) = ۱۳۲$

$$E = \sum_{i \neq j} \frac{Cij}{[N \times (N - ۱)]} = ٤/۲۳/\ ۱۳۲ = ۰/۰۳۲۰$$

عدد به دست آمده در بالا، بیانگر شاخص اندازه‌گیرنده وضعیت موجود ساختار وب-سایت برای دوازده صفحه‌ای از آن است که بیشترین بازدید را به خود اختصاص داده بودند. طبق مطالعات ژو و همکارانش، هرچه مقدار Eبالاتر باشد وب‌سایت از ساختار بهتری برخوردار است.

آنچه در مطالعات ژو و همکارانش و نیز در دیگر مطالعات مشابه به آن اشاره‌ای نشده، این است که هیچ حد بالا و پایینی برای E تعریف نشده‌است. این مسأله باعث می‌شود که عدد

Archive of SID

www.SID.ir

به دست آمده میزان خوبی یا بدی را به طور واضح نشان ندهد. بنابراین در ادامه این تحقیق، روش جدید و ساده‌ای برای تعیین حد بالا و پایین برای E ارائه می‌گردد.

۷- تعیین حد بالا و پایین برای شاخص اندازه‌گیر عملکرد وب‌سایت

همان طور که در بخش قبل اشاره شد، هرچه مقدار E بالاتر باشد، وب‌سایت از ساختار و عملکرد بهتری برخوردار است. بدین منظور باید حد بالایی تعریف شود تا مشخص گردد که تا کجا امکان بهبود آن وجود دارد. همچنین تعریف حد بالا و پایین کمک می‌کند تا وب-سایتهای گوناگون با یکدیگر قابل‌مقایسه بوده، وب‌سایتی که عملکرد و ساختار بهتری دارد، قابل‌شناسایی باشد.

از آنجا که مخرج کسر E عددی ثابت است، به علت ثابت بودن تعداد صفحات وب‌یات، حد بالای آن وقتی است که صورت آن به حداکثر خود برسد. صورت کسر، بیانگر میزان اتصال صفحات به یکدیگر است. هر چه ارتباط بین صفحات و لینکهای بین «صفحات مرتبط» بیشتر باشد، مقدار صورت کسر افزایش می‌یابد. زمان مقدار هر کدام از Cها بیشینه می‌گردد که مقدار Wها و نیز Rها حداکثر شود.

لازم به توضیح است که تمام صفحات را نمی‌توان به یکدیگر متصل کرد و بین آنها لینک گذاشت. تنها آن صفحاتی که به هم مرتبط هستند می‌توانند لینک به یکدیگر داشته باشند. مثلاً بین صفحه Administration و Campus-map می‌تواند لینک برقرار باشد، اما بین صفحات Administration و News پیوندی ضروری نیست.

ارتباط صفحات را می‌توان از روی نقشه‌سایت و یا همان Sitemap پیدا کرد. این نقشه توسط مدیر سایت تهیه شده، ارتباطات بین تمام صفحات وب‌سایت را نشان می‌دهد؛ اما در حالت واقعی، تمام لینکهای یک سایت توسط بازدیدکنندگان بازدید نمی‌شود، چرا که ممکن است بازدیدکنندگان نتوانسته باشند این لینکها را پیدا کنند و یا علاقه‌ای به آن صفحه خاص نداشته‌اند.

ماتریس تعریف ۲ ارتباط بین لینکهای صفحاتی که واقعاً بازدید شده بودند را نشان می‌دهد. اما اگر تمام لینکهای یک وب‌سایت مطابق نقشه سایت آن بازدید شوند، ماتریسی مشابه ماتریس زیر به وجود می‌آید، به طوری که مسیرهای جدیدی اکتشاف می‌شوند که هرچند در سایت وجود داشته‌اند، اما بازدید نشده‌اند. شش مسیر جدید در ماتریس زیر نشان داده شده‌است. تغییرات این ماتریس با ماتریس قبلی، پررنگ شده‌است.

Archive of SID

www.SID.ir

	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
A	۱	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
B	۱	۰	۱	۱	۱	۰	۰	۰	۰	۰	۰	۰
C	۱	۰	۱	۱	۱	۰	۰	۰	۰	۰	۰	۰
D	۱	۰	۰	۰	۰	۱	۱	۰	۰	۰	۰	۰
E	۱	۰	۰	۰	۰	۱	۰	۱	۰	۰	۰	۰
F	۱	۰	۰	۰	۰	۰	۰	۰	۱	۱	۰	۰
G	۱	۰	۰	۰	۰	۱	۰	۰	۰	۰	۱	۰
H	۱	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۱
I	۱	۰	۰	۱	۱	۰	۱	۰	۰	۰	۰	۰
J	۱	۰	۰	۱	۱	۰	۰	۰	۰	۰	۰	۱
K	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۱
L	۱	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱
M	۱	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۱
N	۱	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۱
O	۱	۰	۰	۱	۰	۰	۰	۰	۰	۰	۱	۰

اگر کلیه عملیات تشریح شده در بخش قبلی دوباره برای جدول بالا محاسبه گردد، نتیجه به قرار زیر تغییر می‌کند:

$$E' = \sum i \neq j Cij / [N*(N-1)] = ٦/٨/ ١٣٢ = ٠/٠٥١٠$$

بدین ترتیب، سطح بالای E به دست می‌آید. سطح پایین آن وقتی است که هیچ بازدیدی از هیچ صفحه‌ای انجام نشود که مقدار آن صفر خواهد بود. بنابراین خواهیم داشت: $0 < E < E'$ و بدین ترتیب، ضریب عملکرد فعلی سایت نسبت به حالت بیشینه ممکن، بر اساس نقشه وب-سایت، ٦٢ درصد است.

Archive of SID

www.SID.ir

۸- نتیجه‌گیری

توسعه روزافزون اینترنت و کسب‌وکارهای الکترونیک در سراسر جهان در سالهای اخیر، رویدادی است که نه تنها نمی‌توان آن را انکار کرد، بلکه تأثیرات آن بر زندگی بشر هر روز پررنگ‌تر می‌شود. این رشد فزاینده اینترنت، باعث شده تا اهمیت وب‌سایتها برای کسب-وکارها، سازمانها و حتی شخصیتهای حقیقی روبه‌روز افزایش یابد. بنابراین، اینکه یک وب-سایت خوب چه وب‌سایتی است و چگونه می‌توان میزان عملکرد آن را اندازه‌گیری کرد، از جمله موضوعاتی است که در سالهای اخیر توجه بسیاری از محققان را به خود جلب کرده-است. پژوهشگران به دنبال آن هستند که با ارائه مدلها، چارچوبها و شاخصهای گوناگون، وب‌سایت را تحلیل کرده، بر اساس نتایج به دست آمده، ساختار مناسبی را پیشنهاد دهند.

در ادبیات موضوع، روشهای گوناگونی برای ارزیابی وب‌سایت معرفی شده‌است. تمام این روشها به گونه‌ای سعی دارند تا با ارائه و معرفی شاخصهایی، رفتار بازدید و ارتباطات ساختار وب‌سایت را مدل ساخته، اندازه‌گیری کنند. نظریه گرافها یکی از این روشها است که کاربرد آن در علوم کامپیوتری در سالهای گذشته بسیار افزایش یافته‌است. اگر وب‌سایت به صورت یک گراف مدل شود، آنگاه هر صفحه، معادل یک گره و هر پیوند بین صفحات، معادل یک کمان خواهد بود. بدین ترتیب می‌توان وب‌سایت را به شکل یک گراف جهت‌دار نشان داد که مسیرها در آن، مجموعه صفحاتی است که هر کاربر در هر بار بازدید از وب-سایت طی می‌کند.

در ادامه این مقاله، با استفاده از تعریف وب‌سایت در قالب یک گراف جهت‌دار، شاخصهایی برای اندازه‌گیری ساختار وب معرفی گردید. در قدم اول، مسیرهای طی شده در وب‌سایت با استفاده از یک نرم‌افزار تحلیل وب‌سایت استخراج شدند. با استفاده از این اطلاعات، ماتریسی تشکیل گردید تا صفحات وب را که در هر نوبت توسط گروهی از کاربران بازدید شده‌اند، طبقه‌بندی کند. سپس درجه انجمنی صفحات که بیانگر میزان همبستگی هر جفت صفحه به هم متصل است، اندازه‌گیری شد و از این طریق، وزن هر پیوند تعیین گردید. درجه اتصال بین دو صفحه، شاخص دیگری بود که میزان اتصال هر دو صفحه، متصل یا غیرمتصل، را از طریق مجموع وزنهای کلیه مسیرهای ارتباط‌دهنده این دو صفحه اندازه‌گیری می‌کرد. میانگین این درجات اتصال، شاخص نهایی برای سنجش ساختار وب‌سایت بود. بدین ترتیب با استفاده از شاخص میانگین اتصال صفحات وب‌سایت، وضعیت فعلی ساختار وب سنجیده

Archive of SID

www.SID.ir

شد. در قسمت آخر مقاله، روش جدیدی برای تعیین سطح بالا و پایین مقدار E معرفی شد. با توجه به روش تعریف شده، محدوده این شاخص تعریف و مشخص گردید که ضریب عملکرد سایت چند درصد به عدد ماکزیمم امکان‌پذیر نزدیک بوده‌است.

۹- منابع

[1] Zaiane،Osmar R. Building virtual web views. Data & knowledge engineering. Vol. 39. pp:143-163, 2001.

[2] Vrazalic Lejla. Website usability in context: an activity theory based usability testing method. Proceedings of the national conference on Transformational tools for 21th centry minds, 2003.

[3] Hughes Blackmon Marilyn, Kitajima Muneo,G. Polson Peter. Repairing Usability Problems Identified by the Cognitive Walkthrough for the Web . ACM CHI Vol. No. 5 ، Issue No. 1. pp: 497-504, 2003.

[4] Y. Ivory. Melody R. Sinha. Rashmi ، A. Hearst. Marti. Empirically Validated Web Page Design Metrics. Volume No. 3 ، Issue No. 1. pp; 53-60, 2001.

[5] Huang ،ing-Hui. Web performance scale. Information & Management. (Article in press). 2004

[6] Wang. qing ،Makaroff Dwight J ،Edwards H. keith .Characterizing customer groups for an e-commerce website. Proceedings of the 5th ACM conference on Electronic commerce. USA. pp:218 – 227, 2004.

[۷] خواجه نوری، عباسقلی. روش تحقیق. مؤسسه انتشارات و چاپ دانشگاه تهران. ۱۳۷۲

[۸] ایران نژاد پاریزی، مهدی. روشهای تحقیق در علوم اجتماعی. نشر مدیران. ۱۳۷۸

[9] Piatetsky-Shapiro G. Knowledge Discovery in Real Databases: A Report on theIJCAI-89 Workshop. AI Magazine 11(5): 68–70, 1991.

[10] R.Kosala،H.Blockeel،Web mining research: a survey,SIGKDD: SIGKDD

Archive of SID

www.SID.ir

explorations: newsletter of the special interest group (SIG) on knowledge discovery & data mining ، ACM 2 (1) 1–15, 2000.

[11] A.Nanopoulos, M. Zakrzewicz،T. Morzy,Y. Manolopoulos, Indexing web access-logs for pattern queries، in: fourth ACM CIKM International Workshop on Web Information and Data Management (WIDM_02) ، 2002.

[12] K.P. Joshi ، A. Joshi ، Y. Yesha ، On using a warehouse to analyze web logs Distributed and Parallel Databases 13 (2) 161–180, 2003.

[13] Gunduz Sule ، Recommendation models for web users: user interest model and click stream tree. Unpublished Ph.D. Thesis. 2003. Department of computer engineering. Istanbul Technical University.

[١٤] راس، شلدون. ترجمه پارسیان، احمد و همدانی، علی. مبانی احتمال. نشر شیخ بهایی چاپ دوم،١٣٧٦.

[15] W.Cooper1William,M.Seiford Lawrence,Zhu Joe.DATA

[16] ENVELOPMENT ANALYSIS History ، Models and Interpretations

[17] Hahn Jungpil,J. Kauffman Robert,A METHODOLOGY FOR BUSINESS VALUE-DRIVEN WEBSITE EVALUATION: A DATA ENVELOPMENT ANALYSIS APPROACH.

[18] lo Storto Corrado, Ponsiglione Cristina A Method to Assess WEB Site Usability Based on DEA (Data Envelopment Analysis) Brajnik Giorgio. Automatic web usability evaluation:what needs to be done?, http://www.dimi.uniud.it/~giorgio/papers/hfweb00.html (Access date : 2005/1/1

[19] Bjo¨rneborn ، L. Small-world linkage and co-linkage. In Proceedings of the 12th ACM conference on hypertext and hypermedia (pp. 133–134).

Archive of SID

www.SID.ir

New York: ACM Press, 2001.

[20] Thelwall Mike ، Wilkinson David. Graph Structure in Three National Academic Webs: Power Laws with Anomalies. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY*، 54(8):706–712 ، 2003.

[21] Huang Zan ، Chung Wingyan ، Chen Hsinchun. A Graph Model for E-Commerce Recommender Systems. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY*, 55(3):259–274, 2004.

[۲۲] جی.ای.باندی و یو.اس.ار.مورتی، ترجمه حمید ضرابی زاده، نظریه گرافها وکاربردهای آن، انتشارات مؤسسه فرهنگی هنری دیباگران، سال ۱۳۷۸

[23] Bla˙zewicza, Jacek. Peschb, Erwin. Sternaa, Ma lgorzata. A novel representation of graph structures in web mining and data analysis. OMEGA، *International Journal of Management science.* 2004

[۲٤] گری چارتراند و آرتور اولرمن، ترجمه سید مهدی تشکری هاشمی، نظریه الگوریتمی و کاربردی گرافها، انتشارات دانشگاه امیرکبیر- ۱۳۸۳.

[۲۵] جی.ای.باندی و یو.اس.ار.مورتی، ترجمه حمید ضرابی زاده، نظریه گرافها و کاربردهای آن، انتشارات مؤسسه فرهنگی هنری دیباگران، سال ۱۳۷۸

[26] Thelwall Mike ، Wilkinson David. Graph Structure in Three National AcademicWebs: Power Laws with Anomalies. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY*, 54(8):706–712 ، 2003.

[27] Huang Zan ، Chung Wingyan ،Chen Hsinchun. A Graph Model for E-Commerce Recommender Systems. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND*

Archive of SID

www.SID.ir

TECHNOLOGY, 55(3):259–274 ، 2004.

[28] Zhou Baoyao ، Chen Jinlin ، Shi Jin ، Zhang Hongjiang ، Wu Qiufeng. website Link Structure Evaluation and Improvement Based on User Visiting Patterns. ACM. pp: 241-242, 2001.

[29] Zhou Baoyao ، Chen Jinlin. USER BEHAVIOR BASED WEBSITE LINK STRUCTURE EVALUATION AND IMPROVEMENT

[30] http://www.mach5.com: Mach5 Analyzer v4.1.5

Archive of SID

www.SID.ir