



### چکیده

متن حاضر، یادداشتی تکمیلی است بر مقاله‌ای تحت عنوان 'روش تحلیل گزارش‌های وب (وبلاگ)' به قلم هاجر ستوده که در صفحات ۸۴۶۹ شماره دوم از جلد اول نشریه اطلاع‌شناسی منتشر شده و ضمن معرفی روش تحلیل گزارش‌های وب به تشریح و ویژگی‌های روش شناختی آن پرداخته است. این یادداشت با تمرکز بر مسائل فنی به شرح برخی مشکلات موجود در این روش از جمله تأثیر حافظه پنهان و رویت‌ها، مشکل حجم عظیم داده‌ها، پیچیدگی شناسایی کاربران، و طول نشست و نقص نرم‌افزارهای تجاری موجود برای تحلیل گزارش‌ها می‌پردازد. همچنین به برخی شیوه‌های ممکن برای بهبود کیفیت این روش تحقیق برای سنجش رفتار اطلاعاتی کاربران اشاره می‌کند.

کلیدواژه‌ها: روش‌شناسی، وب، تحلیل گزارش، وبلاگ.

## یادداشتی بر روش تحلیل گزارش‌های وب (وبلاگ)

حمیدرضا جمالی مهموئی



پروہشگاہ علوم انسانی و مطالعات فرہنگی  
پرتال جامع علوم انسانی

## یادداشتی بر روش تحلیل گزارش‌های وب (وبلاگ)

حمیدرضا جمالی مهموثی<sup>۱</sup>

### مقدمه

در صفحات ۶۹-۸۴ شماره دوم از جلد اول نشریه اطلاع‌شناسی مقاله‌ای تحت عنوان "روش تحلیل گزارش‌های وب (وبلاگ)"<sup>۲</sup> به قلم هاجر ستوده منتشر شد که نخستین مقاله در این موضوع در متون فارسی حوزه کتابداری و اطلاع‌رسانی محسوب می‌شود. اگرچه ممکن است متخصصان حوزه علوم رایانه‌ای در ایران تا به حال پژوهش‌هایی را بر روی گزارش‌های وب انجام داده باشند که اغلب با رویکرد فنی خاص این متخصصان و جهت سنجش کارایی وب سایت‌ها انجام می‌شود، اما تا به حال پژوهشی با این روش و با رویکرد علم اطلاع‌رسانی در ایران انجام نشده است. صرف‌نظر از دلایل فنی که این عدم پژوهش می‌تواند داشته باشد از جمله مشکل دسترسی به فایل‌های لاگ که در اختیار صاحبان خدمت‌دهنده‌های اینترنت قرار دارند و یا فقدان همکاری علمی لازم میان متخصصان رایانه و کتابداران، یقیناً یکی از دلایل این نقیصه نبود دانش و آشنایی کافی با این روش در میان کتابداران و اطلاع‌رسانان ایرانی است. از این رو، مقاله مورد اشاره می‌تواند گام بسیار مثبتی در جهت معرفی این روش باشد. از آنجاکه هدف مؤلف بیشتر بررسی نقاط قوت و ضعف این روش به لحاظ روش‌شناسی

۱. دانشجوی دکترای اطلاع‌رسانی  
دانشگاه سیتی لندن

hrjamali@hotmail.com

2. Web Log

بوده است (ص ۷۰)، برخی نکات، به‌ویژه مسائل فنی، تنها به‌صورت گذرا اشاره شده و برخی نواقص از قلم افتاده است. متن حاضر سعی دارد تا جهت تکمیل مقاله ارزشمند مذکور نکاتی را به‌ویژه با تکیه بر تجربیات گروه تحقیقاتی سایبر<sup>۱</sup> که از گروه‌های تحقیقاتی فعال در به‌کارگیری این روش است به‌صورت تکمیلی مطرح کند.

فایل‌های گزارش وب یا به‌زعم محققان گروه سایبر "تلویزیون‌های مداربسته فضای سایبر"<sup>۲</sup> (نیکلاس<sup>۳</sup> و دیگران، ۱۹۹۹، ص ۲۶۳) فایل‌های متنی هستند که در حالت پیش‌فرض در شاخه اصلی یک سایت، یعنی همان شاخه‌ای که صفحه خانگی یک سایت ذخیره شده، ذخیره می‌شوند (گوتزمن<sup>۴</sup>، ۱۹۹۹). استفاده از این فایل‌ها در حقیقت در ابتدا به‌عنوان یک محصول فرعی پروتکل انتقال فرامتن<sup>۵</sup> ایجاد شدند (فیبیر<sup>۶</sup>). استفاده از این فایل‌ها برای بررسی رفتار اطلاعاتی کاربران و کارایی وب‌سایت‌ها که به نوعی تداوم تحقیقات مشابه دهه‌های قبل روی گزارش‌های تراکنشی<sup>۷</sup> فهرست‌های رایانه‌ای و نظام‌های ذخیره و بازیابی محسوب می‌شود به لحاظ مشکلات و ابهاماتی که به همراه دارد با دیدگاه‌های ضد و نقیضی همراه بوده است. برخی چون گوتزمن آنها را به "گنجینه اطلاعات ارزشمند" تشبیه کرده، برخی چون فولر و دوگراف<sup>۸</sup> با تردید آنها را منشأ نتایج گمراه‌کننده در مورد فعالیت کاربران دانسته و برخی دیگر نظیر اودل<sup>۹</sup> از این هم فراتر رفته و آنها را "دروغ‌های لعنت‌شده" نامیده‌اند (نیکلاس و هانتینگتون<sup>۱۰</sup>، ۲۰۰۳، ص ۳۹۳).

## مشکلات روش تحلیل گزارش‌های وب

نویسنده مقاله در صفحه ۷۳ به اجمال به برخی مشکلات روش از جمله وجود حافظه پنهان<sup>۱۱</sup>، عدم شناسایی کاربر، حجم زیاد داده‌ها و عدم برخی وقایع نظیر استفاده از کلیدهای Back و Forward در فایل‌های گزارش اشاره کرده است. در اینجا ضمن تشریح بیشتر این موارد به برخی موارد نیز دیگر اشاره می‌شود.

**الف. حافظه پنهان.** وجود حافظه پنهان تأثیر چشمگیری بر تحلیل آماری نتایج حاصل از تحلیل گزارش‌ها دارد و موجب افت آمار مربوط به استفاده از صفحات وب می‌شود. این مشکل چند منشأ دارد. بخشی از آن ناشی از عملکرد نرم‌افزار مرورگر مورد استفاده توسط کاربر است. این نرم‌افزارها به‌طور مثال، اینترنت اکسپلورر<sup>۱۲</sup> جهت افزایش کارایی خود و تسریع فرایند مرور صفحات توسط کاربر، اقدام به ذخیره صفحات رؤیت شده طی یک دوره زمانی خاص بر روی فضای محدودی از رایانه کاربر

1. Centre for Information Behaviour and the Evaluation of Research: Cyber (City University London)
2. CCTV of Cyberspace
3. Nicholas
4. Gutzman
5. HyperText Transfer Protocol: HTTP
6. Fieber
7. Transactional Log
8. Fuller & De Graaf
9. Udell
10. Huntington
11. Cache
12. Internet Explorer

می‌کنند. در طول این دوره اگر کاربر صفحه‌ای را بازگذاری کند، برای بار دوم نیازمند به بازگذاری آن از روی اینترنت نیست، زیرا مرورگر در دفعات بعد آن را از آرشیو موقت خود فرامی‌خواند نه از روی اینترنت. عدم ثبت استفاده از کلیدهای Back و Forward نیز به دلیل همین ویژگی است. به همین دلیل است که گاه برای اطمینان از روزآمد بودن صفحات وب، کاربر نیاز پیدا می‌کند تا از کلید بازگذاری مجدد یا Refresh استفاده کند. البته کاربر می‌تواند این ویژگی مرورگر را غیرفعال و یا به دلخواه خود تنظیم کند. بنابر تحقیق فیبر ۳۵ تا ۵۵ درصد از عملکرد کاربر به سبب همین ویژگی مرورگرها در فایل‌های گزارش ثبت نمی‌شود. ذکر این نکته نیز لازم است که بنا بر تحقیق نیکلاس و همکارانش (۲۰۰۲، ص ۶۳) میزان ذخیره صفحات در حافظه پنهان رایانه کاربر تا حدودی متأثر از معماری و ساختار وب سایت است.

نوع دیگر حافظه پنهان، خدمت‌دهنده‌های پروکسی<sup>۱</sup> هستند. پروکسی‌ها همانند واسطه‌ای میان کاربر و خدمت‌دهنده در جهت افزایش کارایی و فیلتر کردن اطلاعات عمل می‌کنند. پروکسی‌ها گاه جهت افزایش سرعت شبکه کاربران خود اقدام به ذخیره صفحات وب می‌کنند. در نتیجه زمانی که کاربر صفحه‌ای را درخواست می‌کند، صفحه از حافظه پروکسی به کاربر منتقل می‌شود نه از خدمت‌دهنده اصلی وب، و طبعاً این درخواست کاربر در گزارش وب خدمت‌دهنده اصلی ثبت نمی‌شود.

وجود حافظه پنهان نه تنها بر ثبت آمار استفاده از سایت اثر دارد، بلکه در تحلیل زمان رویت هر صفحه توسط کاربر نیز تأثیر منفی دارد (نیکلاس و دیگران، ۲۰۰۲، ص ۶۳). به این دلیل که کاربر ممکن است چندین صفحه را بدون اینکه در فایل گزارش ثبت شود رویت کند و در نتیجه محقق در زمان تحلیل گزارش چنین تصور کند که کاربر تمام این مدت را صرف مشاهده یک صفحه یعنی آخرین صفحه رویت شده توسط وی که در گزارش ثبت شده کرده است.

**ب. روبات‌ها و خزنده‌های اینترنت<sup>۳</sup>.** یکی دیگر از عوامل گمراه‌کننده در این روش، وجود خزنده‌ها و روبات‌هایی است که به صورت مرتب به کاوش وب می‌پردازند. نمونه بارز آنها روبات‌هایی هستند که توسط موتورهای جست‌وجو جهت بررسی محتوای صفحات وب و نمایه‌سازی آنها به خدمت گرفته می‌شوند. هر بار که چنین روباتی یک سایت را کاوش می‌کند تمامی عملکرد آن همانند عملکرد یک کاربر ثبت می‌شود در حالی که در حقیقت هیچ انسانی از سایت استفاده نکرده است. به‌زعم

1. Proxy Server
2. Robot
3. Internet Crawler

گوتزمن (۱۹۹۹) حتی تا یک سوم از ترافیک ثبت شده یک وب سایت ممکن است ناشی از عملکرد روبات‌ها باشد. البته در مواردی کاوش سایت توسط روبات‌ها به دلیل اینکه در فایلی به نام "Robot.txt" روی خدمت‌دهنده وب ثبت می‌شود، قابل شناسایی است و می‌توان با رجوع به آنها این نوع اطلاعات را از گزارش‌های وب حذف کرد. در عین حال، تعداد رو به افزایش روبات‌های ناشناس وب نیز مشکل دیگری است (نیکلاس و دیگران، ۲۰۰۲، ص ۶۴).

**ج. حجم زیاد داده‌ها.** از آنجا که تمام وقایع با جزئیات در فایل گزارش ثبت می‌شوند، حجم داده‌های موجود در این فایل‌ها بسیار زیاد است. به عنوان نمونه استفاده از سایت روزنامه‌های Times و Sunday Times روزانه منجر به ثبت حدود ۱۷۴ میلیون سطر داده در فایل گزارش آنها می‌شود. (نیکلاس و دیگران، ۲۰۰۰، ص ۴۰۲). طبیعتاً تحلیل صحیح این حجم از داده ممکن است با مشکلاتی مواجه شود. اما، در حقیقت، تمامی این داده‌ها مفید نیستند. بسته به نوع وب سایت و محتوای آن بخش قابل توجهی از این داده‌ها می‌تواند ناشی از ثبت انتقال اجزاء یک صفحه (نظیر تصاویر و آیکون‌ها) به کاربر باشد که ارزش اطلاعاتی ندارند. در تحقیقی بر روی دو سایت مربوط به اطلاعات بهداشتی حدود ۸۵ درصد از حجم فایل‌های گزارش، مربوط به این نوع داده‌های زائد تشخیص داده شده و حذف شد (نیکلاس و دیگران، ۲۰۰۲، ص ۶۶). بسته به اهداف تحقیق ممکن است لازم باشد محقق بخش قابل توجه دیگری از داده‌ها را نیز از تحقیق حذف کند. به طور مثال، در تحلیل گزارش وب یک نشریه الکترونیکی که با فرمت PDF عرضه می‌شود، می‌توان تنها به سطوروی از داده‌ها پرداخت که مربوط به این نو فایل‌ها هستند.

**د. شناسایی کاربر.** مطالعه رفتار کاربر بدون آنکه شناخت دقیقی از هویت، منشأ، و ویژگی‌های کاربر داشته باشیم حاصل چندانی نخواهد داشت. یکی از بزرگ‌ترین چالش‌ها در روش تحلیل گزارش‌های وب مشکل شناسایی کاربر است. این مشکل دلایل متعددی دارد. گزارش‌های وب از پروتکل اینترنت<sup>۱</sup> یا آی.پی برای شناسایی کاربر استفاده می‌کنند. هر رایانه‌ای بر روی اینترنت دارای یک آدرس منحصر به فرد است که آدرس آی.پی نام دارد. این آدرس (به عنوان مثال ۹۰، ۳، ۱۵۱، ۱۳۲) از چهار عدد تشکیل شده که هر یک در محدوده صفر تا ۲۵۵ قرار دارند (آرمس<sup>۲</sup>، ۲۰۰۰، ص ۲۳). زمانی که کاربری از وب استفاده می‌کند آدرس آی.پی رایانه وی در گزارش ثبت می‌شود.



اولین مشکل از همین جا ناشی می‌شود، زیرا در حقیقت، فایل گزارش عملکرد یک رایانه را ثبت می‌کند نه یک کاربر مشخص واحد را. چه بسا چندین نفر از یک رایانه به صورت مشترک استفاده کنند نظیر رایانه‌های عمومی داخل کتابخانه‌ها، خدمات عمومی، کافه‌های اینترنتی، یا حتی رایانه‌های مشترک اعضای یک خانواده. مشکل دوم این است که دو نوع آدرس آی.پی وجود دارد: ثابت و پویا یا شناور. به سبب محدود بودن آدرس‌های آی.پی، بسیاری از فراهم‌کنندگان خدمات اینترنتی<sup>۱</sup> یا آی.اس.پی‌ها از آدرس‌های آی.پی شناور برای مشترکان خود استفاده می‌کنند. به این صورت که هر بار که مشترک اقدام به اتصال به اینترنت از طریق آی.اس.پی می‌کند، آی.اس.پی یک آدرس شناور به وی اختصاص می‌دهد که به احتمال بسیار زیاد متفاوت از آدرس قبلی همان رایانه است. در نتیجه شناسایی کاربر از طریق آدرس آی.پی ثبت شده در گزارش‌های فایل می‌تواند گمراه‌کننده باشد. با این حال، تحلیل بر اساس آی.پی خود می‌تواند به عنوان یک روش مستقل برای مطالعه استفاده از وب به کار رود (برای نمونه نگاه کنید به دیویس، ۲۰۰۳). گاه محققان سعی می‌کنند تا به جای شناسایی انفرادی کاربران در تحلیل گزارش وب به دسته‌بندی آنها پردازند. از این روش در تحلیل گزارش وب سایت روزنامه‌های *The Times / The Sunday Times* استفاده شد (نیکلاس و دیگران، ۲۰۰۰). به این صورت که بر اساس حوزه آدرس یو.آر.آل<sup>۲</sup> (نظیر *.com .org .co.uk .edu*) و آدرس آی.پی مشخص شد که چه میزان از استفاده‌کنندگان از نهادهای دولتی (*.org*)، چه میزان از دانشگاه‌ها (مثلاً *ut.ac.ir* دانشگاه تهران)، چه میزان از مؤسسات تجاری و بخش خصوصی (*.com*) و یا بر اساس حوزه جغرافیایی چه میزان از یک کشور به خصوص (مثلاً *.uk* بریتانیا) بوده‌اند. در این مورد نیز مشکلاتی وجود دارد؛ به‌طور مثال، بسیاری از آی.اس.پی‌ها در کشورهای مختلف آدرس‌های آی.پی خود را در آمریکا به ثبت می‌رسانند. در نتیجه در هنگام تحلیل، تمامی کاربرانی که از آن آی.اس.پی استفاده کرده‌اند به اشتباه امریکایی تلقی می‌شوند؛ حال آنکه ممکن است در یونان یا در جای دیگری باشند. نمونه این اشتباه در مطالعه روزنامه‌های فوق‌الذکر رخ داده است. بر اساس آدرس‌های آی.پی، حدود ۶۰ درصد خوانندگان وب سایت این روزنامه‌ها از آمریکا بوده‌اند، حال اینکه بر اساس بانک اطلاعاتی مشترکان که نزد ناشر قرار دارد (خوانندگان این روزنامه‌ها جهت مطالعه شکل پیوسته آنها نیاز به ثبت مشخصات خود دارند)، تنها ۲۶ درصد از خوانندگان، خود را مقیم آمریکا معرفی کرده‌اند (نیکلاس و دیگران، ۲۰۰۰، ص ۴۰۷).

1. Internet Service Provider: ISP
2. Uniform Resource Locator: URL

در مورد سایت‌هایی که از شناسه کاربری و کلمه عبور و یا فن‌آوری کوکی<sup>۱</sup> برای شناسایی کاربران تکراری و یا صدور مجوز دسترسی به محتوای خود استفاده می‌کنند، شناسایی کاربر قدری مطمئن‌تر است. با این حال در این مورد زیر وجود سرورهای پروکسی و دیواره‌های آتش<sup>۲</sup> که جهت افزایش کارایی و امنیت شبکه‌ها استفاده می‌شوند مانع از شناسایی دقیق و مطمئن کاربر می‌شود. این دو فن‌آوری در حقیقت آدرس‌های آی.پی. حقیقی کاربران را پنهان می‌کنند. بنابراین شناسایی دقیق کاربر زمانی به نحو نسبتاً مطلوب میسر می‌شود که کاربر با استفاده از شناسه کاربری و رمز عبور و با اختصاص یک آدرس آی.پی. ثابت و بدون وجود پروکسی و دیواره آتش از یک رایانه غیر عمومی به استفاده از وب پردازد.

ه. طول زمان یک نشست<sup>۳</sup>. یکی از متغیرهای مورد استفاده در روش تحلیل گزارش‌های وب، تداوم یا طول زمان یک نشست است. تمام فعالیت‌های (کلیک‌های) یک کاربر از یک وب سایت به‌خصوص طی یک فاصله خاص زمانی یک نشست تلقی می‌شود (زاویتز، ۱۹۹۸). از آنجا که کاربران معمولاً پس از استفاده از یک وب سایت خروج خود را رسماً ثبت نمی‌کنند (log off نمی‌کنند)، در نتیجه سنجش این متغیر معمولاً سؤال برانگیز است. برای رفع این نقیصه معمولاً به‌صورت پیش فرض، زمان مشخصی به عنوان حداکثر زمان مجاز برای غیرفعال بودن کاربر در نظر گرفته می‌شود و پس از آن فعالیت کاربر به عنوان یک کاربر جدید یا نشست جدید در نظر گرفته می‌شود. به این معنی که اگر کاربری به مدت فرضاً ۳۰ دقیقه غیرفعال باشد (روی هیچ فرآیندی کلیک نکند)، سیستم چنین در نظر می‌گیرد که کاربر در حال استفاده از وب نیست.

## نرم‌افزارهای تحلیل گزارش‌های وب

همان‌طور که نویسنده مقاله متذکر شده (ص ۷۴) نرم‌افزارهای تجاری متعددی برای تحلیل گزارش‌های وب وجود دارد؛ اما نکته این است که اکثر این نرم‌افزارها به منظور گزارش‌گیری برای صاحبان وب طراحی شده‌اند و چندان جوابگوی نیازهای متخصصان اطلاع‌رسانی نیستند. در حقیقت، این نرم‌افزارها بعضاً متغیرها و ملاک‌های از پیش تعریف شده را به پژوهشگر تحمیل می‌کنند. گروه سایبر به جهت نامناسب بودن این نرم‌افزارها از شیوه تجزیه داده‌ها<sup>۴</sup> و نرم‌افزار اس.پی.اس.اس<sup>۵</sup> جهت استخراج آمار استفاده می‌کند. این شیوه اگر چه زمان، منابع، و تخصص بیشتری می‌طلبد اما از دقت بیشتری

1. Cooky
2. Firewall
3. Session
4. Parsing
5. Statistical Package for Social Sciences: SPSS



برخوردار است و بسته به اهداف تحقیق می‌تواند جوابگویی محقق باشد.

## ارتقاء کیفی روش تحلیل گزارش‌های وب

صرف‌نظر از اینکه می‌توان در کنار روش تحلیل گزارش وب از ترکیبی از روش‌های مختلف از جمله مصاحبه، پرسشنامه، و مشاهده به صورت توأم جهت دستیابی به نتایج مطمئن‌تر استفاده کرد، با استفاده از فنون و ترفندهای خاص می‌توان کیفیت خود روش تحلیل گزارش وب را نیز ارتقا بخشید.

یکی از فنون به کار گرفته شده توسط گروه سایبر روش تحلیل خرد<sup>۱</sup> است (نیکلاس و هانتینگتون، ۲۰۰۳، ص ۳۹۲). منظور از تحلیل خرد، تحلیل بخش مطمئنی از داده‌های گزارش وب (نه تمام داده‌ها) و ردگیری تعداد نسبتاً کمی از کاربران است. به صورت خاص می‌توان از سه شیوه برای تحلیل خرد استفاده کرد.

۱. جداسازی یک گروه خاص از کاربران که می‌توان در مورد منشا جغرافیایی آنها مطمئن بود. این شیوه به عنوان مثال در مورد کاربران بریتانیایی که حوزه آدرس یو.آر.آل آنها با "uk" کاملاً مشخص است کاربرد دارد.

۲. تحلیل یک گروه جزئی از کاربران که احتمال اینکه آدرس‌های آی.پی. آنها منعکس‌کننده استفاده افراد مشخصی باشد بیشتر است. به این معنی که آدرس‌های آی.پی. از نوع ثابت بوده و مشخص هستند.

۳. ردگیری و بررسی استفاده‌های صورت گرفته توسط افراد خاص و نه گروه‌ها. در تحلیل گزارش‌های وب می‌توان داده‌های مربوط به یک کاربر منفرد را بر اساس اطلاعات مربوط به آدرس آی.پی، مشخصات مرورگر و شناسه کاربری جدا کرد و به صورت موردی بررسی نمود (برای نمونه نگاه کنید به نیکلاس و هانتینگتون، ۲۰۰۳).

تدابیر دیگری نیز می‌توان بر حسب مورد تحقیق و ویژگی‌های سایت مورد بررسی به کار برد، از جمله در مواردی که اطلاعات مربوط به کاربران در یک بانک اطلاعاتی موجود است (نظیر مشترکان یک نشریه یا روزنامه الکترونیک) می‌توان آنها را به صورت اطلاعات تکمیلی برای شناخت کاربران و نیز سنجش میزان صحت نتایج حاصل از تحلیل به کار برد.

باید توجه داشت که محتوا و ساختار داده‌های ذخیره شده در گزارش‌های وب بسته به نوع سایت، نرم‌افزار به کار رفته برای ثبت داده‌ها، و حتی سیاست صاحبان سایت‌ها

تفاوت می‌کند. به‌طور مثال، ساختار و نوع داده‌ها در فایل گزارش سایت یک ناشر نشریات الکترونیکی با گزارش‌های یک وب سایت عمومی تفاوت زیادی دارد (برای نمونه نگاه کنید به نیکلاس و دیگران، ۲۰۰۱؛ ۲۰۰۲؛ ۲۰۰۳). همان‌طور که مشکلات ذکر شده در این نوشته را نمی‌توان در مورد تمام انواع فایل‌های گزارش وب تعمیم داد، در هنگام به‌کارگیری شیوه تحلیل نیز باید به ویژگی‌های خاص آن سایت و گزارش‌های آن اندیشید.

## خاتمه

در خاتمه ذکر این نکته نیز لازم است که اگر چه روش تحلیل گزارش‌های وب افق‌های تازه‌ای در مطالعه خیل کثیر کاربران اطلاعات الکترونیکی گشوده و تکامل این روش چه در مرحله ثبت داده‌ها در فایل‌های گزارش و چه در مرحله تحلیل همچنان ادامه دارد، اما شیوه منطقی آن است که همواره با احتیاط به نتایج این‌گونه پژوهش‌ها نگریست. چه بسا در مواردی استفاده از این روش برای طرح سؤال مناسب‌تر است تا رسیدن به پاسخ سؤالاتی خاص. مطالعه رفتار کاربر با استفاده از این روش سؤالات زیادی در مورد رفتار آنها برمی‌انگیزد که پاسخ آنها را باید با شیوه‌های کیفی جست. هنگام استفاده از این روش باید به‌خاطر داشت که وب در حقیقت یک نظام برای توزیع صفحات وب است و نه یک نظام ذخیره و بازبازی اطلاعات. آنچه که در داده‌های فایل گزارش وب نهفته است بیانگر چیزی است که کاربران (با واسطه یک رایانه) انجام داده‌اند، نه آنچه که فکر می‌کنند انجام داده‌اند یا تصور می‌کنند بایستی انجام می‌دادند.

پرتال جامع علوم انسانی

مأخذ

Arms, W. Y. (2000). *Digital Libraries*. Cambridge: MIT press.

Davis, P. M., and Solla, L. R. (2003). An IP-Level analysis of usage statistics for electronic journals in chemistry: making inferences about user behavior. *Journal of the American Society for Information Science and Technology*, 54(11): 1062-1068.

Fieber, John (2004). "Browser caching and Web log analysis". [Online] Available: <http://ella.slis.indiana.edu/~jfieber/papers/bcwla/bcwla.html>. last visited 18/02/2004

- Gutzman, A. (1999). "Analysing traffic on your E-commerce site". [Online] Available: [http://ecommerce.internet.com/news/insights/ectech/article/0,9561\\_00.html](http://ecommerce.internet.com/news/insights/ectech/article/0,9561_00.html).last visited 18/02/2004
- Nicholas, D., and Huntington, P. (2003). "Micro-mining and segmented log file analysis: a method for enriching the data field from Internet log files". *Journal of Information Science*, 29(5): 391-404.
- Nicholas, D., Huntington, P., Lievesley, N., and Wasti, A. (2000). "Evaluating consumer website logs: a case study of The Times/The Sunday Times Website". *Journal of Information Science*, 26(6): 399-411.
- Nicholas, D., Huntington, P., Lievesley, N., and Withey, R. (1999). "Cracking the code: web log analysis". *Online & CD-ROM Review*, 23(5): 263-265.
- Nicholas, D., Huntington, P., and Watkinson, Anthony. (2003). "Digital journals, Big Deal and online searching behaviour: a pilot study". *Aslib Proceedings: New Information Perspectives*. 55(1-2): 84-109.
- Nicholas, D., Huntington, P., and Williams, P. (2001). "Establishing metrics for the evaluation of touch screen kiosks". *Journal of Information Science*, 27(2): 61-71
- Nicholas, D., Huntington, P., and Williams, P. (2002). "Evaluating metrics for comparing the use of web sites: a case study of two consumer health web sites". *Journal of Information Science*, 28(1): 63-75.
- Zawitz, M. W. (1998). "Web statistics- measuring user activity". [Online] Available: <http://www.ojp.usdoj.gov/bjs/pub/ascii/wsmua.txt>.last visited 18/02/2004



پروہشگاہ علوم انسانی و مطالعات فرہنگی  
پرتال جامع علوم انسانی