

اینترنت پنهان و منابع اطلاعاتی نهفته در اعماق نامرئی شبکه جهان گستر وب

یزدان منصوریان¹

چکیده

مقاله حاضر به تعریف مفهوم اینترنت پنهان یا وب نامرئی و همچنین عوامل مؤثر در ایجاد آن پرداخته است. اجزای مختلف قلمرو تاریک اینترنت نظیر وب عمیق، وب مات، وب شخصی، وب ملکی و وب مرده به تفکیک تشریح شده‌اند و چگونگی ایجاد و اهمیت هر یک مورد بحث قرار گرفته است. این مقاله ضمن اشاره به حجم و کیفیت قابل توجه منابع اطلاعاتی نهفته در اینترنت پنهان، بر اهمیت مهارت‌های اطلاع‌یابی در محدوده خارج از حوزه جستجوی موتورهای کاوش تأکید می‌کند. از آنجا که اطلاعات نهفته در دل اینترنت پنهان معمولاً از کیفیت بالایی برخوردار است، این بخش از اینترنت اهمیت قابل توجهی دارد و نمی‌توان آن را نادیده گرفت.

نقش کتابداران و متخصصان اطلاع‌رسانی در آگاه‌سازی کاربران اینترنت از وجود وب نامرئی و در نتیجه ارتقای سطح دسترس‌پذیری اطلاعات برای آنان و همچنین نقشی که کتابداران می‌توانند در ساخت دروازه‌های اطلاعاتی ایفا کنند، از دیگر مباحث مطرح شده در این نوشتار است.

کلیدواژه‌ها: اینترنت پنهان، وب پنهان، وب نامرئی، وب عمیق، وب مات، بازیابی اطلاعات در وب

نگاهی کوتاه بر مشکلات جستجو و بازیابی در اینترنت

جستجو و بازیابی اطلاعات در اینترنت فرایندی پیچیده و چندبُعدی است که می‌توان آن را از زوایای مختلف مورد توجه و بررسی قرار داد. تاکنون نیز پژوهش‌های متعددی درباره جنبه‌های مختلف کاوش و بازیابی اطلاعات در این محیط صورت گرفته است. در سال‌های اخیر این موضوع همچنان جزء زمینه‌های مهم پژوهشی در حوزه بازیابی اطلاعات و مطالعات اطلاع‌یابی محسوب شده. به‌رغم تفاوت‌های موجود در رویکرد این پژوهش‌ها، محور اصلی همه آن‌ها یافتن راهکارهایی عملی برای ارتقای سطح دسترس‌پذیری اطلاعات در این شبکه و سهولت بازیابی است. (e.g. Jansen, 2000; Jansen & Spink, 2000; Wany et al, 2000; Palmquist, 2000; Kim, 2001, Cothey, 20002; Hargittai, 2002; Ford et al, 2002; Ford et al, 2003)

موتورهای جستجو اصلی‌ترین و معمولی‌ترین ابزار کاوش در اینترنت به شمار می‌آیند و معمولاً اغلب کاربران از این موتورها برای بازیابی اطلاعات اینترنتی استفاده می‌کنند. (Gordon & Pathak, 1999) اهمیت و جایگاه این ابزارهای کاوش به حدی است که بدون آن‌ها عملاً جستجو در این شبکه عظیم اگر ناممکن نباشد، قطعاً بسیار دشوار خواهد بود.

با این حال و با تمام ویژگی‌های ارزنده این موتورها و کاربرد گسترده آن‌ها در فرایند بازیابی اطلاعات، مشکلات متعددی در پیوند با کارآیی و سودمندی آن‌ها وجود دارد. پژوهش‌های گذشته حاکی از آن است

۱. دانشجوی دکتری کتابداری و اطلاع‌رسانی دانشگاه شفیلد انگلستان

که کاربران اینترنت با دشواری‌هایی در استفاده از موتورهای جستجو مواجه هستند. (Gorden & Pathak, 1999; Chu & Rojenthal, 1996) آن‌ها گاهی برای یافتن اطلاعات مورد نیاز خود ساعت‌ها وقت صرف می‌کنند و ممکن است این تلاش حاصلی هم در پی نداشته باشد. دلایل متعددی برای دشواری‌های جستجو در اینترنت وجود دارند که تاکنون موضوع پژوهش‌های فراوانی بوده‌اند. در یک طبقه‌بندی کلی، بخشی از این دشواری‌ها به حجم عظیم و روزافزون اطلاعات موجود در اینترنت مربوط می‌شود که معمولاً نتایج بازیابی موتورهای جستجو را با ریزش کاذب فراوان همراه می‌سازد. بخش دیگر به میزان مهارت‌های اطلاع‌یابی و تجربه کاربران در جستجو بازمی‌گردد. گاهی نیز ناکامی در جستجو ریشه در نوع نیاز اطلاعاتی کاربران دارد که ممکن است گسترده و مبهم باشد و کاربر نتواند آن را به شکل مناسبی در قالب یک راهبرد جستجوی کارآمد تعریف کند.

هریک از این احتمالات خود زمینه گسترده‌ای برای پژوهش و مطالعه به‌شمار می‌آیند و تشریح و تبیین آن‌ها نیازمند مطالعات مستقل و مفصلی است که این مقاله قصد پرداختن به آن‌ها را ندارد. آنچه که محور اصلی این نوشته را تشکیل می‌دهد موضوع دیگری است که تاکنون چندان مورد توجه کاربران اینترنت و پژوهشگران حوزه بازیابی اطلاعات در وب قرار نگرفته و بی‌توجهی به آن می‌تواند به طور بالقوه در ناکامی جستجو در وب مؤثر باشد. محور اصلی این مقاله پدیده‌ای تحت عنوان اینترنت پنهان² یا وب نامرئی³ می‌باشد که در ادامه معنا و مفهوم آن تشریح می‌شود و به عوامل ایجادکننده آن و اهمیت آن در فرایند بازیابی اطلاعات در اینترنت اشاره می‌گردد.

اینترنت پنهان چیست

واژه اینترنت پنهان یا وب نامرئی اشاره به یک مفهوم چندبعدی و گسترده دارد، به شکلی که ارائه تعریفی گویا و کوتاه از آن را کمی دشوار می‌کند. به بیان دیگر، ارائه یک تعریف جامع و مانع از اینترنت پنهان مستلزم تبیین بخش‌های مختلف آن است. با این حال در ساده‌ترین شکل می‌توان این بخش پنهان از اینترنت را به‌طور خلاصه این چنین تعریف نمود:

اینترنت پنهان آن بخش از فضای اینترنت است که به هر دلیل، خارج از حوزه جستجوی موتورهای کاوش قرار دارد و بازیابی اطلاعات موجود در آن از طریق استفاده مستقیم از این موتورها میسر نیست (Pedley, 2001; Price & Sherman, 2001). افزودن قید «به‌هردلیل» در این تعریف حاکی از آن است که ایجاد اینترنت پنهان ناشی از وجود فقط یک عامل نیست و به عوامل متعددی مربوط می‌شود. تأثیر همین عوامل متعدد منجر به ایجاد انواع مختلف از پنهان‌شدگی اطلاعات در اینترنت گردیده که در بخش‌های بعدی مقاله تشریح خواهند شد.

ذکر این نکته ضروری است که واژه‌های موازی دیگری نظیر وب نامرئی، وب پنهان، وب عمیق و وب تاریک یا وب مات⁴ برای این بخش از اینترنت در متون مختلف مشاهده می‌شود. گرچه این اصطلاحات گاهی به‌صورت مترادف به کار برده می‌شوند، اما در حقیقت معادل یکدیگر نیستند و هر یک اشاره به نوع خاصی از ناپیدایی⁵ در وب دارند.

از آنجا که ارتباط مستقیمی بین شیوه کار موتورهای جستجو و ایجاد اینترنت پنهان، یا به بیان دقیقتر

². Hidden Internet

³. Invisible Web

⁴. Opaque Web

⁵. Olvisibility

بروز پنهانی اطلاعات در اینترنت وجود دارد، به نظر می‌رسد قبل از توصیف بخش‌های مختلف اینترنت پنهان لازم باشد به نحوه عملکرد موتورهای جستجو اشاره شود.

اجزای تشکیل دهنده موتورهای جستجو و چگونگی عملکرد آن‌ها

به‌رغم وجود گوناگونی و تنوع فراوان در موتورهای جستجو، اغلب آن‌ها از ویژگی‌های مشترکی برخوردارند. ساختار اصلی موتور جستجو معمولاً از سه بخش عمده تشکیل شده. این قسمت‌ها عبارت‌اند از:

1. **نرم‌افزار خزنده یا روبات جستجوگر:**⁶ این قسمت از موتور جستجو در واقع نرم‌افزاری است که با دنبال کردن پیوند⁷های موجود در صفحات وب، صفحات جدید را شناسایی می‌کند و این فرایند، پیوسته تکرار می‌شود. انتخاب نام «خزنده» برای این نوع از نرم‌افزارها به دلیل نوع کاری است که انجام می‌دهند. عملکرد آن‌ها در یافتن صفحات وب مثل خزیدن در فضای وب است. هر موتور جستجو، نرم‌افزار خزنده مخصوص به خود را دارد و توانایی بازیابی اطلاعات در هر موتور، کاملاً وابسته به قدرت «خزنده» آن است. منظور از قدرت خزنده، سرعت عمل و دامنه فعالیت آن در دنبال کردن پیوندهای موجود می‌باشد. گستردگی و عمق دسترسی اطلاعات در هر موتور جستجو بیش از هر چیز به ویژگی‌های خزنده آن بستگی دارد و نمایه⁸ موتور جستجو، حاصل کار خزنده است. البته علاوه بر خزنده‌ها در هر موتور جستجو، بخشی برای معرفی سایت‌های جدید نیز وجود دارد که کاربران می‌توانند با ارسال نشانی⁹ سایت مورد نظر خود، آن را به موتور جستجو تحویل¹⁰ دهند.

2. **نمایه:** صفحات بازیابی شده در وب از طریق عملکرد خزنده، در یک پایگاه اطلاعاتی تحت عنوان «نمایه موتور جستجو» (و در واقع یک فایل بزرگ) ذخیره می‌شوند. بزرگی و روزآمدی اطلاعات موجود در نمایه هر موتور جستجو وابسته به عملکرد خزنده آن است. ساختار نمایه، اندازه و حجم آن در موتورهای جستجوی مختلف متفاوت است. به همین دلیل جستجو با کلیدواژه‌های یکسان، نتایج نسبتاً متفاوتی در موتورهای گوناگون در پی خواهد داشت.

3. **نرم‌افزار جستجو در نمایه:**¹¹ سومین جزء یک موتور جستجو، نرم‌افزار بازیابی است که به جستجو و بازیابی اطلاعات موجود در فایل نمایه براساس کلیدواژه‌های دریافتی از طرف کاربران می‌پردازد و نتایج بازیابی را نمایش می‌دهد. به بیان دیگر، هنگامی که کاربران کلیدواژه‌های مورد نظر خود را در بخش جستجوی موتور کاوش وارد می‌کنند، آنچه اتفاق می‌افتد جستجوی نمایه موتور بر اساس کلیدواژه‌های مذکور است نه جستجوی فضای اینترنت. آنچه که کاربران مشاهده می‌کنند بیرونی‌ترین لایه موتورهای جستجو است. وقتی آن‌ها به کمک کلیدواژه‌های انتخابی خود به جستجو در اینترنت می‌پردازند، در واقع آنچه را جستجو می‌کنند پایگاه اطلاعاتی (نمایه) موتور جستجو است که پیوند آن‌ها را با سایر سایت‌های اینترنتی - سایت‌ها و صفحات وب که قبلاً توسط خزنده شناسایی شده‌اند - برقرار می‌کنند.

بسته به این که خزنده‌های موتور جستجو با چه تناوب¹² و چه عمقی¹³ عمل کنند نتیجه جستجو توسط کاربر متفاوت خواهد بود، چرا که به دلیل تفاوت تناوب و عمق عملکرد خزنده موتورهای کاوش گوناگون،

⁶ . Crawler or Spider

⁷ . Links

⁸ . Index

⁹ . URL (Universal Resource Locator)

¹⁰ . Submit

¹¹ . Query Processor

¹² . Frequency of Crawling

¹³ . Depth of Crawling

حجم و اطلاعات ذخیره شده در نمایه آن‌ها با هم متفاوت می‌باشد و در نتیجه حاصل جستجو در هر یک از آن‌ها متفاوت خواهد بود. هر چند همواره درصدی از همپوشانی بین نتایج بازیابی وجود دارد.

قسمت‌های مختلف اینترنت پنهان

همانطور که گفته شد موتورهای جستجو بر اساس عملکرد نرم‌افزارهای خزنده خود و همچنین سایت‌هایی که به آن‌ها معرفی می‌شوند فایل یا پایگاه اطلاعاتی خود را می‌سازند و از این طریق امکان جستجو در اینترنت را برای کاربران فراهم می‌سازند. پایگاه اطلاعاتی موتورهای جستجو مرتباً روزآمد می‌گردند و سایت‌های جدید به آن‌ها اضافه می‌شوند و مرتباً بر حجم آن‌ها افزوده می‌گردد. بسته به فاصله روزآمدسازی این پایگاه‌ها، اندازه آن‌ها در موتورهای مختلف متفاوت است. به دلیل گستردگی و پویایی محیط وب عملاً هیچ موتور جستجویی قادر به پوشاندن تمام وب نیست و همواره بخش‌های مختلف یا به بیان بهتر منابع اطلاعاتی متعددی در اینترنت وجود دارند که در این پایگاه‌ها گنجانده نمی‌شوند. پس از بیان این مقدمه کوتاه اکنون می‌توان بخش‌های مختلف اینترنت پنهان را به شرح زیر خلاصه کرد:

1. وب تاریک یا وب مات

این بخش از اینترنت پنهان، بزرگترین قسمت این فضای گسترده را تشکیل می‌دهد. انتخاب عنوان «وب تاریک» برای این قسمت به خاطر این است که گویی منابع اطلاعاتی موجود در آن در زیر سایه‌ای قرار گرفته‌اند و از دسترسی کاربران خارج شده‌اند. به بیان دیگر، اطلاعات موجود در وب تاریک می‌تواند مورد استفاده عموم قرار گیرد و هیچ کس تعمداً در مخفی کردن آن نداشته، اما به دلیل تأثیر عواملی که در ادامه ذکر می‌گردد، عملاً اطلاعاتی که می‌توانسته توسط موتورهای جستجو بازیابی شود از حوزه و قلمرو موتورهای جستجو دور مانده است. اجزای تشکیل دهنده وب تاریک عبارت‌اند از:

1-1. **وب مرده یا نشانی‌های قطع شده:**¹⁴ محیط وب فضایی کاملاً پویا، متغیر یا به بیان بهتر ناپایدار است. هر روز صفحات و منابع مختلفی به این مجموعه عظیم افزوده می‌شوند و منابع دیگری از آن حذف می‌گردند. از سویی دیگر، صفحاتی در وب وجود دارند که هیچ پیوندی با منابع دیگر برقرار نکرده‌اند و به همین دلیل خزنده‌های موتورهای جستجو قادر به یافتن آن‌ها نیستند. در صورتی که این صفحات به موتورهای جستجو معرفی نگردند امکان شناسایی آن‌ها برای موتور جستجو وجود ندارد. طبق پژوهشی که در سال 2000 انجام شده این مجموعه جدا مانده از محیط جستجوپذیر وب حدود 20 درصد کل فضایی است که می‌تواند توسط موتورهای جستجو نمایه شود. (Broder, et al. 2000) البته از آنجا که به دلیل گستردگی بی‌وقفه اطلاعات موجود در شبکه جهانگستر وب هر گونه برآوردی - حتی نسبی - از حد و مرز این محیط خیلی زود کهنه و قدیمی می‌شود، این رقم می‌تواند بیشتر یا کمتر از 20 درصد باشد. آنچه که در اینجا بیش از صحت یا نادرستی این رقم اهمیت دارد، واقعیتی است که هم‌اکنون در وب به وجود آمده و حجم عظیمی از اطلاعات را برای کاربران، دسترس‌ناپذیر ساخته است.

2-1. **تناوب روزآمد سازی نمایه موتورهای جستجو:** همانطور که گفته شد نرم‌افزارهای خزنده مسئولیت شناسایی صفحات جدید و افزودن آن‌ها به نمایه موتورهای جستجو را برعهده دارند. توان این نرم‌افزارها محدودتر از آن است که بتوانند تمام محیط وب را پوشش دهند و تناوب روزآمدسازی آن‌ها کندتر از سرعت افزودن اطلاعات جدید به محیط وب است. به بیان ساده‌تر، نرم‌افزارهای خزنده فرصت کافی برای حرکت هماهنگ با صفحات جدید را ندارند و بنابراین همواره صفحات جدیدی به محیط وب افزوده می‌شوند

¹⁴ . Dead Web or Disconnected URLs

که هنوز مورد شناسایی این خزنده‌ها واقع نشده‌اند و به‌همین دلیل از حوزه جستجوی موتورهای دور می‌مانند. بنابراین، حتی به فرض آن که یک موتور جستجو قادر به گسترش حوزه کاوش خود به تمام محیط وب باشد (که البته هنوز هیچ موتوری این توانایی یا ادعا را ندارد)، همواره فاصله‌ای بین اطلاعات واقعی موجود در وب و آنچه که موتور مذکور جستجو می‌کند وجود خواهد داشت. در واقع موتورهای جستجو هیچ‌گاه نمی‌توانند پایه‌های سرعت افزایش اطلاعات به محیط وب حرکت کنند و همیشه چندین گام عقب‌تر هستند. بنابراین اختلاف سرعت بین افزودن اطلاعات به محیط وب و شناسایی اطلاعات جدید توسط خزنده‌ها، همواره بخشی از اطلاعات موجود در وب را در تاریکی وب پنهان قرار می‌دهد.

3-1. عمق نمایه‌سازی:¹⁵ در متون مرتبط با موتورهای جستجو اصطلاح دیگری تحت عنوان «عمق نمایه‌سازی» یا عمق عمل خزنده‌ها وجود دارد که بیانگر چگونگی و عمق نمایه‌سازی صفحات وب توسط موتورهای جستجو است. توان مالی و فناوری بسیاری از موتورهای جستجو هنوز محدودتر از آن است که بتوانند تمام صفحات و لایه‌های موجود در وب‌سایت‌ها را نمایه‌سازی کنند، چرا که این کار فرایندی هزینه‌بر و دشوار است. بنابراین هر یک از موتورهای جستجو بسته به میزان توان خود و سیاستی که برای نمایه‌سازی اتخاذ کرده، تنها بخشی از وب‌سایت‌های شناسایی شده را نمایه می‌کند. منظور از «بخشی از وب‌سایت‌ها» در واقع صفحات درونی و لایه‌های زیرین هر وب‌سایت است. بنابراین، همواره بخشی از اطلاعات نهفته در پایین‌ترین لایه‌های وب‌سایت‌ها از حوزه جستجوی کاربران دور می‌مانند. پس این مسئله نیز دلیل دیگری است که منجر به ایجاد بخشی از اینترنت پنهان می‌شود.

4-1. حداکثر صفحات قابل مرور در نتایج بازیابی:¹⁶ به‌دلیل انبوهی فزاینده اطلاعات وب معمولاً اغلب جستجوها با نتایج فراوان در بازیابی همراه‌اند. موتورهای جستجو اطلاعات بازیابی شده را در صفحات متعدد نمایش می‌دهند و هر صفحه بین 10، 20 یا 30 مورد را دربرمی‌گیرد. البته کاربران می‌توانند بسته به سلیقه یا نیاز خود این تعداد را در تنظیمات موتورهای جستجو کم یا زیاد کنند.

پژوهش‌های گذشته نشان می‌دهند که اکثر کاربران تنها صفحات اول یا دوم نتایج بازیابی موتورهای جستجو را مرور می‌کنند و به‌دلیل به‌ندرت به صفحات بعدی مراجعه می‌کنند (Jansen et al. 2000; Spink et al, 2001). گرچه موتورهای جستجو براساس نظام رتبه‌بندی¹⁷ خود سعی می‌کنند اطلاعات مرتبط‌تر را در ابتدای فهرست بازیابی نمایش دهند اما موضوع ربط¹⁸ یک مفهوم ساده نیست و نمی‌توان به راحتی نسبت به مربوط بودن یا نبودن اطلاعات مندرج در یک مدرک قضاوت کرد. بویژه آن‌که همواره قضاوت نهایی درباره‌ی ربط برعهده‌ی کاربر است و سیستم نمی‌تواند براساس معیارهای تعریف‌شده، اطلاعاتی را صددرصد مرتبط یا نامرتب بداند. موضوع ربط بحث بسیار مفصل و پیچیده‌ای است که پرداختن به جزئیات آن خارج از محدوده‌ی این مقاله است. آنچه در حوزه بحث این مقاله می‌گنجد بیان این واقعیت است که حتی در بهترین شرایط همواره بخشی از اطلاعات بازیابی شده که ممکن است مرتبط با نیاز کاربر نیز باشد در لایه‌های زیرین صفحات بازیابی شده توسط موتورهای جستجو مدفون می‌ماند و جزء وب پنهان محسوب می‌شود.

2. وب عمیق

پس از گسترش و فراگیری شبکه‌ی جهانگستر وب به‌عنوان یکی از گسترده‌ترین و مؤثرترین محمل‌های

¹⁵ . Indexing Depth

¹⁶ . Maximum Number of Viewable Results

¹⁷ . Ranking Policy

¹⁸ . Relevance

اطلاع‌رسانی، بسیاری از پایگاه‌های اطلاعات الکترونیکی پیوسته¹⁹ منابع خود را از طریق این شبکه دسترس‌پذیر ساختند. این منابع اطلاعاتی شامل پایگاه‌هایی بودند که قبل از پیدایش وب نیز وجود داشتند و خدمات اطلاع‌رسانی ارزنده‌ای ارائه می‌کردند. اکنون هزاران مورد از این پایگاه‌های کوچک و بزرگ، از جمله فهرست پیوسته عمومی²⁰ بسیاری از کتابخانه‌های دنیا از طریق شبکه وب دسترس‌پذیر هستند. هر یک از آن‌ها نظام بازیابی اطلاعات خاص خود را دارند و از طریق صفحات جستجوی مبتنی بر وب، اطلاعات و مدارک موجود در پایگاه خود را به کاربران عرضه می‌کنند. بعضی از این پایگاه‌ها و فهرست‌های پیوسته، خدمات خود را رایگان عرضه می‌دارند و برخی دیگر در برابر ارائه خدمات، هزینه اشتراک دریافت می‌کنند. مثلاً اطلاعات موجود در پایگاه اطلاعات علوم تربیتی «اریک»²¹ به رایگان دسترس‌پذیر است، ولی در مقابل، بسیاری از مجله‌های الکترونیکی و پایگاه‌های اطلاعاتی (مثل «دیالوگ»²²) هزینه‌های اشتراک قابل توجهی دریافت می‌کنند.

فارغ از رایگان بودن یا نبودن خدمات این پایگاه‌ها، همه آن‌ها در یک موضوع مشترک‌اند و آن، خارج بودن مندرجات آن‌ها از حوزه جستجوی موتورهای جستجو است. هر یک از این پایگاه‌ها یک صفحه جستجوی مبتنی بر وب دارند که امکان جستجو در آن‌ها را برای کاربران فراهم می‌کند. نتیجه جستجوها به صورت صفحات دسترس‌پذیر از طریق شبکه وب به نمایش درمی‌آید. اما هر یک از این صفحات، نتیجه جستجوی مشخص با کلیدواژه‌هایی هستند که توسط کاربران به نظام بازیابی وارد شده‌اند. بنابراین خزنده‌های موتورهای جستجو توانایی نمایه کردن مندرجات این پایگاه‌ها را ندارند و در واقع همواره در پشت مجموعه‌ای عظیم از اطلاعات متوقف شده‌اند و به درون پایگاه راهی ندارند. (Snow, 2000)

به بیان دیگر، در فضای گسترده وب، منابع اطلاعاتی ارزشمندی در عمق این دریای اطلاعات و درون پایگاه‌های اطلاعاتی وجود دارند که از دسترس کاربران موتورهای جستجو دور می‌مانند. مثلاً اگر فردی در جستجوی اطلاعاتی درباره یک موضوع تخصصی در زمینه کشاورزی باشد و جستجوی خود را تنها به موتورهای جستجو محدود کند، در واقع بخشی عظیم از اطلاعات مرتبط را از دست داده است. بی‌تردید یا به بیان محتاطانه‌تر، به احتمال زیاد، اطلاعات مرتبط و مفیدی در زمینه مذکور در اینترنت وجود دارد اما نهفته در دل پایگاه‌های اطلاعاتی متعددی می‌باشد که دسترسی به مندرجات آن‌ها فقط با مراجعه مستقیم به پایگاه‌های اطلاعاتی مذکور میسر است نه با جستجو از طریق موتورهای جستجوی عمومی.²³

کاربری که در صدد بازیابی چنین اطلاعاتی است باید ابتدا از طریق موتورهای جستجو پایگاه‌های مرتبط با موضوع مورد نظر خود را شناسایی کند و سپس جداگانه به جستجو در آن‌ها پردازد. در غیر این صورت، با جستجو از طریق موتورهای جستجو حوزه کاوش خود را به وب سطحی²⁴ محدود کرده و از دسترسی به وب عمیق بازمانده است.

3. وب خصوصی²⁵ و وب ملکی²⁶

گرچه این بخش از اینترنت پنهان تفاوت اساسی با بخش‌های دیگر دارد و به نوعی خارج از بحث

¹⁹ . On-Line

²⁰ . OPACs

²¹ . ERIC

²² . DIALOG

²³ . General –purpose search engines

²⁴ . Shallow Web

²⁵ . Private Web

²⁶ . Proprietary Web

اصلی این مقاله است، اما اشاره اجمالی به آن ضروری به نظر می‌رسد. اشاره به وب خصوصی و وب ملکی از این جهت ضروری است که انواع دیگری از پنهان‌شدگی اطلاعات در محیط اینترنت نیز معرفی شود و تفاوت‌های موجود تشریح گردد.

همان‌طور که گفته شد، تفاوتی اساسی بین این قسمت از اینترنت پنهان با قسمت‌های دیگر وجود دارد. این تفاوت اساسی از آنجا ناشی می‌شود که مخفی‌بودن اطلاعات در این بخش، کاملاً تعمّدی می‌باشد و اطلاعات مذکور جزء دارایی‌های شخصی و خصوصی افراد یا سازمان‌ها محسوب می‌شود. بی‌تردید هر شخص حقیقی یا حقوقی این حق را دارد که اطلاعات مربوط به خود را که خارج از مسائل عمومی است، به نحوی محافظت نماید.

مثلاً سازمان‌های مختلف صفحات و اطلاعاتی در اینترنت دارند که تنها به مسائل کاری آن سازمان و کارکنان آن مربوط می‌شود و حفاظت از آن ضرورتی مسلم است. این بخش از اینترنت پنهان از حوزه دسترسی موتورهای جستجو خارج است و طبیعتاً هم باید خارج باشد. منابع اطلاعاتی موجود در وب را که مربوط به مسائل شخصی و خصوصی افراد حقیقی یا حقوقی می‌باشد و به روش‌های مختلفی که در ادامه ذکر می‌شود از دسترس سایر کاربران خارج است، وب خصوصی یا وب شخصی می‌نامند.

همچنین بعضی از منابع اطلاعاتی مثل انواع نشریه‌های الکترونیکی و بانک‌های اطلاعاتی مبتنی بر وب را که دسترسی به آن‌ها مستلزم پرداخت حق اشتراک و در واقع خرید اطلاعات است و جزء محصولات شرکت‌های مختلف می‌باشد، را وب ملکی می‌نامند.

در اینجا فقط به انواع سازوکارهایی که می‌تواند یک صفحه یا منبع اطلاعاتی را از حوزه وب عمومی خارج کند و جزء وب شخصی یا وب ملکی قرار دهد اشاره می‌گردد. این بخش معمولاً به یکی از سه روش زیر ایجاد می‌شود:

3-1. استفاده از اسم کاربر²⁷ و گذرواژه²⁸: صفحاتی که دسترسی به آن‌ها مستلزم استفاده از اسم کاربر و گذرواژه باشد از حوزه جستجوی موتورهای کاوش خارج هستند.

3-2. استفاده از فایل Robots.txt: وجود این فایل، خزنده‌های موتورهای جستجو را از نمایه‌سازی وب سایت مذکور بازمی‌دارد.

3-3. استفاده از کد noindex: افزودن کد noindex به مجموعه کدهای یک صفحه وب، پیامی برای خزنده موتور جستجو به‌شمار می‌رود که آن را از نمایه‌سازی صفحه مذکور منع می‌کند و معمولاً موتورهای جستجو این مسئله را رعایت می‌کنند و از افزودن صفحات محتوی کد noindex به نمایه خود خودداری می‌کنند. بنابراین، این کد نیز می‌تواند صفحات معینی را از حوزه جستجوی موتورهای جستجو خارج کند.

در اینجا باز هم تأکید می‌شود که این بخش خاص از اینترنت پنهان، تفاوت بنیادی با بخش‌های دیگر قلمرو تاریک اینترنت دارد، چرا که این قسمت آگاهانه از دسترسی موتورهای جستجو خارج شده؛ در حالی که بخش‌های دیگر اینترنت پنهان می‌توانسته جزء قلمرو بازیابی موتورهای جستجو باشد و عوامل ناخواسته‌ای، آن‌ها را خارج از دسترسی کاربران قرار داده است.

4. اینترنت واقعاً پنهان²⁹

آنچه تاکنون درباره اینترنت ذکر شد اشاره به بخش‌هایی از اینترنت داشت که یا به دلیل عوامل غیرفنی

²⁷ . User name

²⁸ . Password

²⁹ . Truly Invisible Web

ایجاد شده بود یا بنا به خواست و تمایل صاحبان منابع اطلاعات. منظور از عوامل غیر فنی عواملی است که مثلاً در وب مات به آن‌ها اشاره شد (مثل حداکثر صفحات قابل رویت در نتایج بازیابی). اما بخش‌های دیگری نیز در قلمرو تاریک وب وجود دارند که در اثر مسائل فنی و تکنیکی، یا به بیان دیگر در اثر ناکارآمدی ابزارهای جستجو، از دسترسی کاربران دور مانده‌اند.

همان‌طور که قبلاً گفته شد موتورهای جستجو اغلب قادر به یافتن اطلاعات متنی «اچ‌تی‌ام‌ال»³⁰ هستند و توانایی بازیابی فایل‌هایی از انواع دیگر نظیر pdf, swf, ppt یا pdf را ندارند یا به دلیل کمبود منابع مالی و فنی از جستجوی فایل‌های غیرمتنی صرف‌نظر کرده‌اند. مثلاً در حال حاضر موتور جستجوی معروف «گوگل»³¹ قادر به بازیابی فایل‌های با پسوند ppt و pdf می‌باشد، اما این امکان در بسیاری دیگر از موتورهای جستجو هنوز وجود ندارد و چند سال پیش نیز «گوگل» به این امکان مجهز نبود.

بنابراین منابع اطلاعاتی متنوعی در وب وجود دارند که تنها به دلیل محدودیت تکنولوژیکی یا مالی موتورهای جستجو، از حوزه کاوش آن‌ها و در نتیجه از دسترسی کاربران دور مانده‌اند.

اهمیت اینترنت پنهان

بخش پنهان اینترنت از دو جهت اهمیت دارد. نخست آن که از نظر کمی این بخش چیزی بیش از نیمه پنهان اینترنت است. به بیان دیگر، حجم اطلاعات نهفته در اینترنت پنهان خیلی بیشتر از بخش سطحی³² یا آشکار³³ است. دوم آن که از نظر کیفی نیز این قسمت بسیار دارای اهمیت است. اطلاعات موجود در بخش‌های مختلف این مجموعه بویژه منابع اطلاعاتی موجود در وب عمیق معمولاً منابع ارزشمند و مفیدی هستند و می‌توانند در بسیاری از موارد پاسخگوی نیاز کاربران باشند. بنابراین، نمی‌توان بخش پنهان اینترنت را نادیده گرفت، چرا که کاربران می‌توانند با روش‌های مختلف سطح ناپیدایی در وب را در جستجوهای خود کاهش دهند و از این طریق، بهره‌وری کاوش خود را ارتقا بخشند.

یکی از زمینه‌های فعالیت کتابداران و متخصصان اطلاع‌رسانی، آگاه‌نمودن کاربران از وجود اینترنت پنهان است تا به این ترتیب، آنان فرایند جستجوی خود در محیط اینترنت را فقط به نتایج بازیابی موتورهای جستجو محدود نسازند و نتایج بهتری کسب کنند.

اطلاع‌یابی در اینترنت پنهان

راهکارهای مختلفی برای ارتقای سطح دسترسی پذیری اطلاعات در اینترنت پنهان وجود دارد. مهم‌ترین گام در این زمینه، آگاهی‌رسانی در مورد وجود قلمرو تاریک وب است. متأسفانه بسیاری از کاربران از وجود این بخش از اینترنت بی‌اطلاع هستند. «شرمن» و «پرایس» در بیان بخشی از مشکلات کاربران در هنگام جستجو در وب می‌نویسند: «بسیاری از کاربران بر این باورند که تمام اطلاعات موجود در وب در حوزه دسترسی موتورهای جستجو قرار دارد و همه این موتورها نیز مجموعه یکسانی از وب را کاوش می‌کنند. در ضمن معمولاً از این نکته مهم غافل‌اند که شیوه عمل موتورهای جستجو با هم متفاوت است و نیز آنچه که این موتورها جستجو می‌کنند با آنچه که در زمان جستجو در وب وجود دارد متفاوت است» (Price & Sherman, 2001 (b)).

³⁰ . HTML

³¹ . Google

³² . Surface

³³ . Visible

اگر کاربران به این مسائل آگاهی داشته باشند، می توانند جستجوهای دقیق تر و مطمئن تری انجام دهند. تاکنون تدابیر متعددی برای بهبود سطح بهره‌وری جستجو در اینترنت پیش‌بینی شده. استفاده از این تدابیر می تواند به‌عنوان مکمل جستجو از طریق موتورهای کاوش، میزان بازیابی اطلاعات مرتبط مورد نیاز را به نحو قابل توجهی ارتقا بخشد. از جمله این امکانات می توان به دروازه‌های اطلاعاتی موضوعی³⁴ اشاره کرد. تاکنون در بسیاری از رشته‌های علمی، دروازه‌های اطلاعاتی متنوعی طراحی و ساخته شده‌اند. از آنجا که انتخاب منابع موجود در این سایت‌ها توسط متخصصان موضوعی یا کتابداران انجام می‌شود، معمولاً منابع برگزیده از کیفیت و اعتبار خوبی برخوردار هستند. روش دیگر، استفاده از ابرموتورهای کاوش³⁵ است. این ابرموتورها خود موتور جستجوی واقعی نیستند بلکه به کاربران این امکان را می‌دهند که کلیدواژه‌های خود را همزمان توسط چند موتور جستجو مورد کاوش قرار دهند و نتایج جستجوی تمام آن‌ها را یکجا در اختیار داشته باشند. امکان دیگری که اخیراً مورد توجه متخصصان بازیابی اطلاعات در وب قرار گرفته، استفاده از عوامل هوشمند³⁶ است. این عوامل هوشمند که توانایی جستجو، مقایسه و انتخاب منابع اطلاعاتی بر اساس نیاز مطرح شده توسط کاربر را دارند، به‌عنوان کارگزار کاربران در این فضای گسترده به جستجو می‌پردازند و نتایج جستجو را بر این اساس فیلتر می‌کنند و به آن‌ها تحویل می‌دهند. به‌طور کلی پیش‌بینی می‌شود با پیشرفت ابزارهای بازیابی اطلاعات در وب و بهبود سواد اطلاعاتی کاربران، به تدریج از سطح ناپیدایی³⁷ اطلاعات در محیط وب کاسته شود.

نقش آموزش کتابداران در کاهش سطح پنهانی اطلاعات

همان‌طور که گفته شد بخش عمده‌ای از اطلاعات مخفی مانده در وب پنهان را می‌توان به راحتی برای کاربران این شبکه دسترس پذیر نمود. مهمترین راهکار برای نفوذ به قلمرو تاریک وب، آگاهی از وجود آن و آگاهی از روش‌های اطلاع‌یابی در آن است. بی‌تردید اگر کاربران اینترنت تصویر روشنی از نقاط قوت و ضعف موتورهای جستجو و توانایی‌ها و محدودیت‌های آن‌ها داشته باشند، بهتر می‌توانند از این ابزارهای کاوش استفاده کنند.

پنهانی اطلاعات در اینترنت یک موضوع با حد و مرزهای مشخص و معلوم نیست. در واقع نمی‌توان این شبکه را به دو بخش نهان و آشکار یا مرئی و نامرئی تقسیم نمود و محدوده و مرز مشخصی بین این دو قسمت قائل شد. ممکن است یک منبع اطلاعاتی برای یک فرد یا در یک تجربه جستجوی مشخص، جزء قلمرو وب پنهان قرار گیرد و در همان زمان برای کاربر دیگری کاملاً دسترس پذیر باشد. بنابراین تأکید بر این نکته ضروری است که نمی‌توان شبکه وب را به دو بخش پنهان و آشکار یا روشن و تاریک یا سطحی و عمیق تفکیک کرد. آنچه اهمیت دارد آگاه‌نمودن کاربران به این واقعیت است که جستجو در اینترنت فرایندی فراتر از به‌کارگیری از موتورهای جستجو می‌باشد. حد و مرز و پیچیدگی منابع اطلاعاتی موجود در این شبکه، فعلاً بسی فراتر از حوزه دسترسی موتورهای مذکور است. بازیابی بهینه اطلاعات از اینترنت نیازمند ابزارهای متعدد و شناخت کافی از این شبکه است. محدود کردن جستجو به حوزه کاوش موتورهای جستجو منجر به عدم دسترسی به اطلاعات نهفته در اینترنت پنهان خواهد شد.

³⁴ . Subject Gateway

³⁵ . Meta-Search Engine

³⁶ . Intelligent Agents

³⁷ . Invisibility Level

بنابراین، آگاهی از وجود اینترنت پنهان نخستین و مهم‌ترین گام در جهت نفوذ به آن است و این آگاهی را کتابداران می‌توانند به کاربران منتقل کنند.

نتیجه

وب پنهان یکی از واقعیت‌های موجود این شبکه جهانگستر است. پنهانی اطلاعات در محیط وب به هر یک از دلایلی که ذکر شد ایجاد شود، سطح دسترس‌پذیری اطلاعات را برای کاربران کاهش می‌دهد. از آنجا که از نظر کمی و کیفی بخش قابل توجهی از منابع اطلاعاتی مبتنی بر وب جزء قلمرو تاریک آن قرار دارد آگاهی از وجود آن و یافتن راهکارهایی برای بازیابی اطلاعات از درون آن کاملاً ضروری است. خوشبختانه قسمت عمده‌ای از این منابع را می‌توان برای کاربران دسترس‌پذیر و جستجوپذیر کرد. با توجه به تعاریفی که در این مقاله ارائه شد سهم عوامل فنی و تکنیکی در پنهانی اطلاعات در محیط اینترنت به مراتب کمتر از عوامل غیر فنی است. به بیان دیگر، نفوذ به درون منابع پنهان در وب بیش از آن که به نظام‌های بازیابی پیچیده‌تر وابسته باشد به جنبه کاربرمدار³⁸ در بازیابی اطلاعات وابسته است. بنابراین حتی با همین ابزارهای کاوش موجود نیز می‌توان سطح پنهانی اطلاعات را به نحو قابل توجهی کاهش داد. البته از سوی دیگر بر توانایی موتورهای جستجو نیز افزوده می‌شود و از نمونه‌های بارز آن می‌توان به موفقیت‌های چشمگیر موتور «گوگل» اشاره کرد که پیوسته دامنه جستجوی خود را گسترش می‌دهد و مرتباً امکانات جستجوی جدیدی به آن افزوده می‌گردد.

بنابراین، به دلیل پژوهش‌ها و پیشرفت‌های اخیر در حوزه بازیابی اطلاعات چه در بُعد فنی و چه از نظر رویکرد کاربرمدار، آینده روشنتری در زمینه بازیابی اطلاعات در محیط وب پیش‌بینی می‌شود. با این حال و با توجه به شرایط موجود منابع اطلاعاتی مبتنی بر وب، لازم است به طرق مختلف کاربران اینترنت را از وجود اینترنت پنهان و اهمیت آن آگاه ساخت و به این ترتیب سطح دسترس‌پذیری اطلاعات را برای آنان افزایش داد. نقش کتابداران و متخصصان اطلاع‌رسانی در این زمینه کاملاً مهم و حیاتی می‌باشد. در حال حاضر صدها دروازه اطلاعاتی و راهنمای اینترنتی که توسط کتابداران طراحی و ساخته شده‌اند نقش مهمی در بهبود بازیابی اطلاعات کاربران ایفا می‌کنند. امید می‌رود که کتابداران ایرانی نیز در این زمینه حوزه فعالیت‌های خود را گسترش بخشند و در آینده نزدیک شاهد موفقیت‌های بیشتری در این زمینه باشیم.

منابع

Broder, A., Kumar, R. Maghoul, F. Raghavan, P. Rajagopalan, S. Stata, R. Tomkins, A. Wiener, J. (2000). "Graph structure in the Web," in **Proceedings of WWW9 Conference, 2000**. [Online], <http://www9.org/w9cdrom/160/160.html>, [10 Sep 2003].

Chu, H., & Rosenthal, M. (1996). "Search engines for the World Wide Web: A comparative study and evaluation methodology." **Proceedings of the 59th Annual Meeting of the American Society for Information Science**, 33, 127-135.

Cothey, V. (2002). "A longitudinal study of World Wide Web users' information-searching behaviour." *Journal of the American Society for Information Science and Technology*, 53 (2), 67-78.

³⁸. User-oriented

Ford, N., Miller, D., & Moss, N. (2002). "Web search strategies and retrieval effectiveness: An empirical study". *Journal of Documentation*, 58(1), 30-48.

Ford, N., Miller, D., & Moss, N. (2003). "Web search strategies and approaches to studying". *Journal of the American Society for Information Science and Technology*, 54 (6), 473-489.

Gordon, M., & Pathak, P. (1999). "Finding information on the World Wide Web: the retrieval effectiveness of search engines". *Information Processing and Management*, 35 (2), 141-180.

Hargittai, E. (2002). "Beyond logs and surveys: In-depth measures of people's Web use skills". *Journal of the American Society for Information Science and Technology*, 53(14), 1239-1244.

Jansen, B. J., & Spink, A. (2000). "The Excite research project: A study of searching characteristics by Web users". *Bulletin of the American Society for Information Science*, 27(1), 15-17.

Jansen, B.J., Spink, A. & Saracevic, T. (2000). "Real life, real users and real needs: A study and analysis of users' queries on the Web". *Information Processing and Management*, 36 (2), 207-227.

Kim, K. (2001). "Information-seeking on the Web: Effects of user and task variables". *Library & Information Science Research*, 23 (3), 233-255.

Palmquist, R.A. & Kim, K. (2000). "Cognitive style and online database search experience as predictors of Web search performance". *Journal of the American Society for Information Science*, 51 (6), 558-566.

Pedley, P. (2001). **The Invisible Web: searching the hidden parts of the Internet**. Aslib.

Pedley, P. (2002). "Why you can't afford to ignore the Invisible Web". **Business information review**, 19 (1), 23-31.

Price, G. & Sherman, C.(2001a)" Exploring the Invisible Web". *Online*, 25 (4), 32-35.

Price, G. & Sherman, C.(2001b). **The Invisible Web: Uncovering Information Sources Search Engines Can't See**. CyberAge Book.

Smith, C. B. (2001). "Getting to Know the Invisible Web". **Library Journal**, 126 (11), 16-19.

Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001). "Searching the Web: the public and their queries". *Journal of the American Society for Information Science and Technology*, 52 (3), 226-234.

Snow, B. (2000). "The Internet's Hidden Content and How to Find It". **Online**, 24 (3), 61-66.

Wang, P. , Hawk, W.B. & Tenopir, C. (2000). "Users' interaction with World Wide Web resources: an exploratory study using a holistic approach". *Information Processing and Management*, 36 (2), 229-251.