وب پنهان

مقدمه

جست‌وجوی منابع اطلاعاتی در محیط کتابخانه در سال‌های اخیر و با پیدایش اینترنت و وب جهان‌گستر تغییر اساسی کرده است. در گذشته، عمل جست‌وجو عمدتاً در کتابخانه‌ها و توسط کتابداران صورت می‌گرفت. پیشینه‌های بازیابی شده نیز به صورت کتاب‌شناختی بودند. پایگاه‌های اطلاعاتی پیوسته نظیر دیالوگ، ویلسون، ابسکو و کارگزاران صفحات نوری از قبیل سیلورپلتر، دسترسی به محتوای پایگاه‌های خود را فراهم می‌کردند. امّا چند سالی است که اینترنت دورنمای بازیابی و جست‌وجوی منابع اطلاعاتی را دچار تغییر و تحول اساسی کرده است. بازیابی پیوسته دیگر در قلمرو کتابخانه‌ها نیست و کاربران از هر جایی می‌توانند عمل جست‌وجو را هر زمان که بخواهند انجام دهند. بازیابی پیوسته همچنین، دیگر در قلمرو پایگاه‌های اطلاعاتی نیست و ابزارهای جست‌وجوی شبکه‌ای از قبیل موتورهای جست‌وجو و موتورهای راهنما و ابزارهای دیگر به این بازار اضافه شده‌اند. بازیابی اطلاعات در وب جهان‌گستر توسط ابزارهای جست‌وجو، ازجمله موتورهای جست‌وجو و راهنماهای موضوعی و ابرموتورهای جست‌وجو، انجام می‌شود. هریک از ابزارهای جست‌وجو، محتوا، پوشش و واسط منحصر به فرد خود را دارند و برای انتخاب راهبرد جست‌وجوی موفق و مفید، باید از مجموعهٔ اصول و قواعد جست‌وجوی این ابزارها آگاهی داشت.

تعریف وب نامرئی یا وب پنهان

گرچه تاکنون مقالات و منابع متعددی درباره وب نامرئی یا وب پنهان منتشر شده و درباره تعریف و مفهوم آن تقریباً اتفاق‌نظر وجود دارد اما به روشنی مشخص نیست که این واژه برای اولین بار توسط چه کسی به کار رفته است. مرور نوشته‌ها نشان می‌دهد که واژه وب پنهان یا وب نامرئی برای نخستین بار در سال ۱۹۹۴ توسط دکتر ژیل السورث[۱] که از محققین جستجو و بازیابی اطلاعات در شبکه وب است در مصاحبه با «نشریه کانادایی کامپیوتر» به کار رفته است. البته برخی منابع شخص دیگری به نام ماتیو کل[۲] را مبدع این اصطلاح معرفی می‌کنند.

در ساده‌ترین شکل می‌توان این بخش پنهان از شبکه وب را به نقل از شرمن و پرایس[۳] چنین تعریف کرد:

«منظور از وب پنهان یا وب نامرئی بخشی از شبکه جهان‌گستر وب است که منابع موجود در آن خارج از حوزه جستجو و بازیابی موتورهای کاوش عمومی[۴] قرار دارد و بازیابی اطلاعات موجود در این بخش از محیط وب با استفاده و جستجوی مستقیم از این موتورها میسر نیست. امکان بازیابی منابع اطلاعاتی وب پنهان یا از نظر فنی برای موتورهای جستجو مقدور نیست و یا آن‌ها به دلیل محدودیت‌های دیگر از دسترسی و نمایه‌سازی این منابع بازمانده‌اند».

البته باید یادآور شد که گاهی واژه وب نامرئی یا وب پنهان ممکن است گمراه‌کننده به نظر برسد. کلمه نامرئی یا پنهان در این اصطلاح به معنای سرّی بودن یا غیرقابل دید بودن اطلاعات نیست. انتخاب این واژه به دلیل محبوبیت و عمومیت استفاده از موتورهای کاوش به همه بخش‌های شبکه وب دسترسی دارند.

چهار نوع اصلی ناپیدایی اطلاعات در وب پنهان

شرمن و پرایس در کتاب خود با عنوان وب نامرئی، کشف منابعی که موتورهای جستجو نمی‌توانند ببینند همه انواع ناپیدایی اطلاعات در محیط وب را به چهار نوع اصلی طبقه‌بندی کرده‌اند. این چهار نوع ناپیدایی عبارت‌اند از:

1. وب مات[۵]
2. وب خصوصی[۶]
3. وب ملکی[۷]
4. وب واقعاً پنهان[۸]

آنان بر این باورند که می‌توان طیفی از ناپیدایی اطلاعات در محیط وب برای کاربران موتورهای جستجو متصور بود که در یک طرف این طیف که از کمترین سطح ناپیدایی برخوردار است وب مات یا وب تقریباً پنهان قرار گرفته و در انتهای دیگر که کاملاً از حوزه بازیابی موتورهای کاوش خارج است وب واقعاً پنهان جای دارد. (منصوریان، ۱۳۸۲)

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

انواع منابعی که در وب پنهان قرار دارند

صفحات ابر متن ساکن[۹] که دارای پیوند (لینک) با صفحات دیگر باشند ساده‌ترین و مناسب‌ترین نوع منابع وبی برای نمایه‌سازی موتورهای کاوش محسوب می‌شوند؛ اما این نوع منابع فقط بخش کوچکی از حجم شبکه وب را تشکیل می‌دهند و منابع اطلاعاتی فراوانی در این شبکه وجود دارند که در قالب‌های دیگر هستند. بنا به دلایل فنی معمولاً نمایه‌سازی منابع غیرمتنی برای موتورهای کاوش مشکل است. از نظر ویژگی‌های منابعی که موقتاً یا به‌طور دائم خارج از حوزه بازیابی موتورهای کاوش قرار دارند می‌توان به موارد ذیل اشاره کرد:

1. جدیدترین منابع موجود در وب: امکان دستیابی به منابع روزآمد در شبکه وب، یکی از مزیت‌ها و امتیازات مهم این شبکه به شمار می‌رود؛ اما نکته‌ای که معمولاً یا از آن غفلت شده یا دست‌کم به آن کمتر اشاره می‌شود فاصله زمانی میان افزودن منابع جدید به محیط وب و امکان بازیابی آن‌ها به‌وسیله ابزارهای کاوش است. دسترسی بی‌درنگ به منابع تازه افزوده شده به وب فقط زمانی میسر است که کاربران نشانی اینترنتی این منابع را بدون واسطه موتورهای کاوش در اختیار داشته باشند.
2. منابع اطلاعاتی غیرمتنی[۱۰]: همان‌طور که گفته شد موتورهای جستجو برای نمایه‌سازی منابع ابرمتن [۱۱] طراحی شده‌اند. در حالی که بسیاری از منابع موجود در شبکه وب به قالب‌های دیگر و معمولاً غیرمتنی مثل تصویر یا منابع دیداری و شنیداری هستند و برای موتورهای کاوش امکان نمایه‌سازی بهینه این منابع به‌راحتی فراهم نیست.
3. منابع محافظت شده از طریق اسم کاربر و گذرواژه[۱۲]: چنانچه دسترسی به یک منبع اطلاعاتی نیازمند استفاده از اسم کاربر و گذرواژه باشد نرم‌افزار خزنده موتور جستجو امکان دسترسی و نمایه‌سازی آن را ندارند. این بخش از قلمرو وب پنهان را «وب شخصی» می‌خوانند.
4. پایگاه‌های اطلاعاتی متعددی در شبکه وب وجود دارند که گرچه دارای صفحه جستجو مبتنی بر وب هستند اما مندرجات آن‌ها در قالب‌های دیگر است و فقط زمانی یک صفحه ابرمتن تولید می‌شود که جستجوی مستقلی در درون پایگاه مربوطه اجرا شود. این مجموعه که به وب عمیق مشهور است، از نظر کمّی بزرگ‌ترین قسمت از قلمرو وب نامرئی را به خود اختصاص می‌دهد به‌طوری که حجم اطلاعات موجود در وب عمیق چندین برابر وب آشکار یا وب سطحی تخمین زده می‌شود.
5. منابع و صفحات وب بدون پیوند (لینک)[۱۳]: امکان برقراری پیوند مستقیم میان هر دو منبع در وب بدون در نظر گرفتن نوع یا فاصله مکانی سایت‌های آن‌ها یکی از امتیازات برجسته این شبکه است. پیوندهای موجود در صفحات وب به دو دلیل اهمیت دارد. نخست آنکه کاربران می‌توانند با دنبال کردن هریک از این پیوندها به‌راحتی بین صفحات مختلف حرکت کنند و این یکی از روش‌های رایج اطلاع‌یابی در وب است. دلیل دوم که مهم‌تر به نظر می‌رسد این است که پیوندهای مذکور تنها راه موجود برای خزنده‌های موتورهای کاوش برای یافتن صفحات جدید به شمار می‌روند این منابع مجموعه‌ای جدا مانده از مجموعه جستجوپذیر وب محسوب می‌شوند و به همین دلیل خزنده‌های موتورهای جستجو قادر به یافتن آن‌ها نیستند. در صورتی که این صفحات به موتورهای جستجو معرفی نشوند امکان شناسایی آن‌ها برای موتورهای جستجو وجود ندارد و در نتیجه هرگز در نتایج جستجوی این موتورها دیده نخواهند شد و از دسترس کاربران دور خواهند ماند.
6. عمق نمایه‌سازی موتورهای جستجو معمولاً هر وب‌سایت از بیش از یک صفحه تشکیل شده است و صفحات متعدد دیگری در لایه‌های زیرین سایت‌ها قرار دارند. منظور از لایه‌های زیرین، صفحات موجود در وب‌سایت‌ها هستند که امکان دسترسی به آن‌ها از طریق صفحه نخست وب‌سایت[۱۴] یا صفحات دیگر فراهم می‌شود. در منابع مرتبط با موتورهای جستجو اصطلاح دیگری با عنوان «عمق نمایه‌سازی»[۱۵] یا عمق عمل خزنده‌ها [۱۶] وجود دارد که بیان‌کننده تعداد صفحات یک وب‌سایت است که توسط خزنده‌های موتورهای جستجو نمایه می‌شوند.
7. حداکثر صفحات قابل مرور در نتایج بازیابی: به دلیل انبوهی فزاینده اطلاعات موجود در شبکه وب، بیشتر جستجوها در موتورهای کاوش نتایج بازیابی فراوان به همراه دارند. موتورهای جستجو معمولاً موارد بازیابی شده را در صفحات متعدد نمایش می‌دهند. تحقیقات گذشته نشان داده است که بیشتر کاربران فقط صفحه‌های اول یا دوم نتایج بازیابی را مرور می‌کنند و به‌ندرت به صفحات بعدی مراجعه می‌کنند؛ بنابراین حتی در بهترین شرایط بخشی از اطلاعات بازیابی شده که ممکن است مرتبط با نیاز کاربر باشد در لایه‌های زیرین صفحات بازیابی شده توسط موتورهای جستجو مدفون می‌ماند و به این ترتیب جزء وب پنهان به شمار می‌رود. (منصوریان، ۱۳۸۴)

اهمیت وب پنهان

به دو دلیل می‌توان گفت که وب نامرئی اهمیت دارد. نخست از نظر کمی، باید گفت که حجم اطلاعات موجود در این بخش بسیار بیشتر از سطح آشکار است. موارد زیر اهمیت وب پنهان را از نظر کمی نشان می‌دهند:

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

- بهترین موتورهای کاوش فقط قادر هستند که حدود ۱۶ درصد از اطلاعات موجود در وب را بازیابی کنند و بنابراین ۸۴ درصد آن‌ها جزو وب نامرئی به‌حساب می‌آیند.
- اندازه وب نامرئی تقریباً ۵۰۰ برابر وب مرئی است، وب نامرئی ۵۵۰ میلیون سند و وب مرئی تقریباً ۱ میلیون سند را دارا هست.

دوم اینکه از نظر کیفی، اطلاعات بخش‌های مختلف این مجموعه به‌ویژه منابع اطلاعاتی موجود در وب عمیق، معمولاً ارزشمند و مفید هستند و در بسیاری موارد پاسخگوی نیاز کاربران هستند. تقریباً بیش از نیمی از وب نامرئی را پایگاه‌های اطلاعاتی موضوعی تشکیل می‌دهند.

دلایل عدم بازیابی وب نامرئی توسط موتورهای کاوش

- دلایل فنی: بسیاری از موتورهای کاوش به دلیل محدودیت‌های نرم‌افزاری توانایی روزآمدسازی اطلاعات جدید وب را ندارند و همیشه این موتورها یک گام از سرعت روزافزون اطلاعات عقب‌تر هستند.
- دلایل بودجه‌ای: فرآیند نمایه‌سازی تمام صفات وب، هزینه‌بر خواهد بود و موتورهای کاوش نیز بنا به محدودیت بودجه ناگزیرند فقط بخشی از وب‌سایت‌ها را نمایه‌سازی کنند.
- دلایل اجتماعی و حقوقی : از آنجا که اطلاعات موجود در وب در دسترس عموم قرار می‌گیرد، بسیاری از افراد و سازمان‌ها به دلیل صرف بودجه‌های کلان در راه‌اندازی سایت‌ها و پایگاه‌های اطلاعاتی خود، حاضر نیستند این اطلاعات را به‌صورت رایگان در اختیار همه بگذارند. البته این از لحاظ اجتماعی و حقوقی حق مسلم آن‌ها است. (خسروی، ۱۳۸۳)

روش‌های اطلاع‌یابی در وب پنهان

راهکارهای مختلفی برای ارتقاء سطح دسترس‌پذیری اطلاعات در وب پنهان وجود دارد. مهم‌ترین گام در این زمینه آگاهی‌رسانی درباره وجود قلمرو تاریک وب است. 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 آن‌ها از این نکته مهم غفلت می‌کنند که روش کار موتورهای جستجو با هم متفاوت است و همچنین آنچه این موتورها جستجو می‌کنند با آنچه در زمان جستجو در وب وجود دارد متفاوت است».

اگر کاربران به این مسائل آگاهی داشته باشند می‌توانند جستجوهای دقیق‌تر و مطمئن‌تری انجام دهند.

روش دیگر، گسترش دادن حوزه جستجوست. برای مثال استفاده از ابرموتورهای کاوش[۱۷] یکی از راهکارهای پیشنهادی است. این ابرموتورها خود موتور جستجوی واقعی نیستند بلکه به کاربران این امکان را می‌دهند که کلیدواژه‌های موردنظر خود را هم‌زمان توسط چند موتور جستجو مورد کاوش قرار دهند و نتایج جستجوی همه آن‌ها را یکجا در اختیار داشته باشند. (منصوریان، ۱۳۸۴).

نتیجه‌گیری

آنچه در این مقاله آمد مروری بود بر معرفی منابع الکترونیک مخفی در وب پنهان و تبیین دلایل اصلی پنهانی اطلاعات در محیط وب. همان‌طور که گفته شد امکان بازیابی منابع موجود در این بخش از شبکه وب وجود دارد به‌شرط اینکه کاربران این شبکه ابزار جستجوی خود را به موتورهای کاوش معمولی محدود نکنند. راهکارهای متعددی برای ارتقاء سطح دسترس‌پذیری اطلاعات در محیط وب وجود دارد. از آن جمله می‌توان به دروازه‌های اطلاعاتی موضوعی اشاره کرد. در مجموع مهم‌ترین گام در این زمینه آگاهی‌رسانی در مورد وجود قلمرو تاریک وب است. متأسفانه بسیاری از کاربران از وجود این بخش از وب بی‌اطلاع هستند؛ بنابراین گرچه منابع وب پنهان خارج از دسترس موتورهای کاوش هستند اما این بدان معنا نیست که امکان دسترسی به آن‌ها وجود ندارد. بلکه برعکس این منابع کاملاً جستجو پذیر هستند.

پانویس:

. Dr.Jill Ellsworrth[۱]

. Matthew Koll[۲]

. Sherman & Price[۳]

. General-Purpose Search Engines[۴]

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.



[۵] Opaque Web .
[۶] Private Web .
[۷] Proprietary Web .
[۸] Truly Invisible Web.
[۹] Static Hyper-Text Pages(html) .
[۱۰] Non-Text Resources .
[۱۱] Hyper-Text Pages .
[۱۲] Password-Protected Pages .
[۱۳] Disconnected Webpages .
[۱۴] Homepage .
[۱۵] Indexing Depth .
[۱۶] Depth of Crawling .
[۱۷] Meta-Search Engine .

منابع:

- خسروی، عبدالرسول. «وب نامرئی» علوم اطلاع‌رسانی ۲۰ (۱ و ۲) (۱۳۸۳): ۵۸-۵۱.
- منصوریان، یزدان. «تدوین مدل اطلاع‌یابی کاربران در محیط وب پنهان: مراحل مقدماتی» مجله مطالعات ملی کتابداری و سازماندهی اطلاعات ۶۳ – ۷۵ :(۱۳۸۴)
- منصوریان، یزدان. (۱۳۸۲)«وب پنهان چیست و چه اهمیتی دارد؟»، بازیابی شده در تاریخ ۲۲/۰۹/۱۳۹۳ از سایت: irandoc.ac.ir/elect-j.htm...

نویسنده:نقی منافی

– – – – – – – –

در انتها می توانید همچنین به این ویدئو توجه کنید:

مشخصات استناددهی به این مقاله

نویسنده(ها):	نقی منافی
عنوان مقاله:	وب پنهان
عنوان مجله:	کتابدار ۲.۰ –(عنوان لاتین: i-Kitابdار 2.0)
دوره مجله(Vol):	۳
شماره مجله(Issue):	۲
سال(Year):	۱۳۹۶
شناسه دیجیتال(DOI):	
لینک کوتاه:	http://lib2mag.ir/8538

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.