

مطالعه مدل‌های اطلاع‌یابی کاربران در محیط وب پنهان: مراحل مقدماتی

تدوین مدل اطلاع‌یابی کاربران در محیط وب پنهان: مراحل مقدماتی

یزدان منصوریان^۱

چکیده

مقاله حاضر گزارشی است از مراحل اولیه پژوهشی درباره جنبه‌های کاربرمدار جستجو و بازیابی اطلاعات در شبکه جهانگستر وب، که طی آن بر تأثیر وب پنهان یا وب نامرئی بر اطلاع‌یابی در وب که موضوع اصلی پژوهش است توجه شده است. مرور مدل‌های موجود اطلاع‌یابی در محیط وب حاکی از آن است که به رغم اهمیت وب پنهان، تأثیر این پدیده در مدل‌های مذکور لحاظ نشده است و تدوین مدلی در این زمینه که مراحل اطلاع‌یابی در وب پنهان را نیز در نظر گرفته باشد اهمیت زیادی دارد. به همین منظور ابتدا انواع ناپیدایی اطلاعات در وب و انواع منابعی که به طور موقت یا دائم خارج از حوزه کاوش متورهای جستجو قرار دارند در این مقاله تشریح و تبیین شده است و در ادامه و پس از بیان مفاهیم پایه درباره وب پنهان، گزارش یافته‌های تحقیق مقدماتی در این زمینه بخش‌های دیگر مقاله را به خود اختصاص می‌دهد. در این مقاله دلایل انتخاب اصطلاح "سطح پنهانی اطلاعات" که از گویایی بیشتری نسبت به اصطلاح "وب پنهان" برخوردار است تشریح شده است. نتایج تحقیق مقدماتی نشان داد که می‌توان برای هر کاربر و در هر مرحله از فرایند جستجو در وب، سطحی پنهانی از اطلاعات قابل شد که میزان آن به عوامل مختلف از جمله مهارت‌های اطلاع‌یابی کاربر، موضوع جستجو و همچنین توانمندی ابزار کاوش بستگی دارد.

کلیدواژه‌ها

وب نامرئی، وب پنهان، وب عمیق، جستجو و بازیابی اطلاعات در محیط وب ، مدل‌های اطلاع‌یابی در وب

مقدمة

طی سه دهه گذشته فرایند اطلاع‌یابی کاربران در منابع الکترونیکی پیوسته^۲ موضوع تحقیقات فراوانی در حوزه ذخیره و بازیابی اطلاعات بوده است. پس از اواسط دهه ۱۹۹۰ و با فرآگیر شدن دسترسی افراد به منابع شبکه وب،^۳ فصل تازه‌ای در این مطالعات گشوده شد و بسیاری از محققان حوزه بازیابی اطلاعات به مطالعات مبتنی بر وب علاقه‌مند شدند (۲۴۶-۲۳۵:۹). از طرف دیگر متخصصان علوم رایانه نیز کوشش‌های مستمری برای طراحی ابزارهای نوین بازیابی اطلاعات در وب آغاز کردند و حاصل این تلاش‌ها، موتورهای جستجوی گوناگون برای تسهیل اطلاع‌یابی در این شبکه جهان گستر بود. در حال حاضر تعداد قابل توجهی از موتورهای جستجو در وب دسترسی‌پذیر هستند البته فقط محدودی از آنها از جمله گوگل، یاهو و آلتاویستا و موارد مشابه توسعه پیشید کاربران استفاده می‌شود.

در کنار توسعه روزافرون موتورهای جستجو، کتابداران و متخصصان اطلاع‌رسانی نیز به تدوین راهنمایها و دروازه‌های موضوعی^۳ متعددی پرداختند که شیوه طبقه‌بندی اطلاعات در آنها معمولاً براساس مبانی نظری و دسته‌بندی موضوعی علوم کتابداری و اطلاع‌رسانی استوار است. در حال حاضر همین موتورهای جستجو و راهنمایها از ابزارهای اصلی جستجو و بازیابی عموم کاربران به شمار می‌روند.

به موازات گسترش این ابزارهای کاوش، پژوهش‌هایی درباره جنبه‌های مختلف آنها شروع شد. از موضوعاتی که طی سال‌های اخیر توجه محققان را جلب کرده، دامنه جستجو و میزان سودمندی و کارایی آنهاست (۱۴۱-۱۸۰). از آنجا که این موتورهای کاوش از معمول ترین ابزارهای کاوش در محیط وب محسوب می‌شوند، انتظار می‌رود نتایج جستجوی آنها جامعیت و مانعیت^۴ رضایت‌بخشی داشته باشند، اما نتایج تحقیقات قبلی حاکی از آن است که به رغم اهمیت شایان توجه این ابزارها در بازیابی اطلاعات در محیط وب، هنوز مشکلات و محدودیت‌های فراوانی در کارآمدی آنها وجود دارد.

بدیهی است که پرداختن به همه جنبه‌های مربوط به امتیازات موتورهای کاوش و مشکلات و محدودیت‌های آنها در یک مقاله میسر و ممکن نیست. از میان موضوعات مربوط به این حوزه به موضوع "وب پنهان" یا "وب نامرئی"^۵ کمتر توجه شده است.

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 اتفاق نظر وجود دارد اما به روشنی مشخص نیست که این واژه برای اولین بار توسط چه کسی به کار رفته است. مرور نوشته‌های نشان می‌دهد که واژه وب پنهان یا وب نامرئی برای تختستین بار در سال ۱۹۹۴ توسط دکتر ژیل سورث^۶ که از محققین جستجو و بازیابی اطلاعات در شبکه وب است در مصاحبه با شریه کانادایی کامپیوٹر^۷ به کار رفته است (۴). البته برخی منابع شخص دیگری به نام ماتیو کل^۸ را مبدع این اصطلاح معرفی می‌کنند.

"وب پنهان یا وب نامرئی" به یک مفهوم چند بعدی و گسترده اطلاق می شود به شکلی که ارائه تعریفی گویا و کوتاه از آن را کمی دشوار می کند. با این حال در ساده ترین شکل می توان این بخش پنهان از شبکه وب را به نقل از شرمن و پرایس^۹ 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^{۱۰} 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 میانده‌اند» (۱۶: ۵۷-۵۸).

البته باید یادور شد که گاهی واژه وب نامرئی یا وب پنهان ممکن است گمراه کننده 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، و این تصور عمومی است که گویا موتورهای کاوش به همه بخش‌های شبکه وب دسترسی دارند. به بیان دیگر این بخش از محیط وب برای کاربرانی پنهان خواهد ماند که جستجوی خود در این شبکه را به استفاده از موتورهای کاوش محدود می‌کنند. بنابراین، منابع وب پنهان قابل بازیابی و دسترسی سنت اما نه به سهولت اطلاعاتی و ب پنهان مستلزم استفاده از ابزارهای کاوشی غیر از موتورهای جستجوی معمولی، نظیر "درگاهها"^{۱۱} و

2. On-line Resources

3.Directories & Subject Gateways

4. Recall & Precision

5. The Invisible Web or Hidden Web

6.Dr. Jill Ellsworth

7. Canadian Computer Magazine

8. Matthew Koll

9.Sherman &Price

10.General-Purpose Search Engines

- 10.General-Purpose
- 11.Subject Portals

دروازه‌های اطلاعاتی موضوعی و پایگاه‌های اطلاعاتی مبتنی بر وب است.

کمتر اشاره‌می شود فاصله زمانی میان افزودن منابع جدید به محیط وب و امكان بازیابی آنها به وسیله 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. اما اگر درصد یافتن این منابع به کمک ابزارهای بازیابی باشیم همواره یک فاصله زمانی میان افزودن یک منبع جدید به محیط وب و امكان بازیابی آن از طریق ابزارهای کاوش وجود دارد. فرایند شناسایی و نمایه‌سازی منابع جدید توسط موتورهای جستجو به دو شکل میسر است. احتمال نخست شناسایی منبع جدید توسط نرم‌افزارهایی موسوم به خزنه‌یا عنکبوت^{۱۷} است که بخش مهمی از هر موتور جستجو را تشکیل می‌دهند. این نرم‌افزارها پیوندهای موجود روی صفحات وب را دنبال می‌کنند و به این ترتیب صفحات جدید را شناسایی می‌نمایند. این فرایند را اصطلاحاً خزینه^{۱۸} در وب می‌خوانند.

خزینه در وب مرتبأً تکرار می‌شود، صفحات جدید شناسایی و صفحات قبلی بازیابی می‌شوند و در صورت تغییر در آنها، تغییرات به نمایه موتور جستجو منتقل شده و در واقع این نمایه‌ها مرتبأً روزآمد می‌شوند. این خزنه‌ها هرقدر هم که سریع باشند نمی‌توانند با سرعت افزوده شدن منابع جدید و سرعت تغییرات اعمال شده در منابع قدیمی هماهنگی داشته باشند. بنابراین همواره منابعی در وب وجود دارند که هنوز نمایه نشده‌اند و جزو وب پنهان محسوب می‌شوند. به بیان دیگر تناوب روزآمدسازی نمایه‌های موتورهای جستجو^{۱۹} که ممکن است هفتگی یا حتی ماهانه باشد منجر به بروز نوعی ناپیدایی در وب می‌شود. این بخش از محیط وب را وب مات یا وب تاریک و گاهی با عنوان وب تقریباً پنهان^{۲۰} معرفی می‌کنند (۱۲). این بخش از وب پنهان که بخش قابل توجه‌ای را به خود اختصاص می‌دهد دچار نوعی ناپیدایی موقعیت است و امكان پیوستن آن به بخش آشکار یا مرئی وب محتمل‌تر از بخش‌های دیگر وب ناممی‌ست.

روش دیگر شناسایی یک منبع جدید توسط موتورهای جستجو یا معرفی آن به موتورهای جستجوست.^{۲۱} این روش نیز به زمان نیاز دارد و همواره یک فاصله زمانی میان زمان معرفی سایت جدید و ظهور آن در نتایج جستجو وجود دارد و در طول این فاصله زمانی گرچه منبع موردنظر در شبکه وب موجود است اما امكان بازیابی آن برای کاربران موتورهای

چهار نوع اصلی ناپیدایی اطلاعات در وب پنهان: شرمن و پرایس در کتاب خود با عنوان وب ناممی‌کشf منابعی که موتورهای جستجو نمی‌توانند بینند همه انواع ناپیدایی اطلاعات در محیط وب را به چهار نوع اصلی طبقه‌بندی کرده‌اند. این چهار نوع ناپیدایی عبارتند از: ۱. وب مات^{۱۲} ۲. وب خصوصی^{۱۳} ۳. وب ملکی^{۱۴} ۴. وب واقعاً پنهان^{۱۵}. 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 (۱۶).

در این مقاله ضمن آنکه به معنا و مفهوم این چهار گونه ناپیدایی اطلاعات در وب اشاره می‌شود در یک تقسیم‌بندی مشابه درباره انواع منابعی که جزء این بخش‌های وب پنهان هستند نیز توضیح داده می‌شود و چهار نوع ناپیدایی موردنظر شرمن و پرایس نیز تشریح خواهد شد.

انواع منابعی که در وب پنهان قرار دارند
صفحات ابرمتن ساکن^{۱۶} که دارای پیوند (لينک) 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 به طور دائم خارج از حوزه بازیابی موتورهای کاوش قرار دارند می‌توان به موارد ذیل اشاره کرد:

۱. جدیدترین منابع موجود در وب امکان دستیابی به منابع روزآمد در شبکه وب، یکی از مزیت‌ها و امتیازات مهم این شبکه به شمار می‌رود که البته در صحت این ادعا تردیدی نیست و با ظهور وب امكان دسترسی سریع به بسیاری از منابع روزآمد فراهم شده است. اما نکته‌ای که معمولاً یا از آن غفلت شده یا دست‌کم به آن

12.Opaque Web

16.Static Hyper-Text Pages (html)

19.Frequency of Crawling

13.Private Web

17.Crawler/Spider/Robot

20.Almost Invisible Web

14.Proprietary Web

18.Web Crawling

21.Manual Submitting

15.Truly Invisible Web

تکست است. زیرا این فایل خزنده‌های موتورهای جستجو را از نمایه‌سازی سایت مذکور بازمی‌دارد. سومین روش ممکن استفاده از کد nofollow یا noindex است به طوری که افزودن یکی از این کدها به مجموعه کدهای یک صفحه وب پیامی برای خزنده موتور جستجو به شمار می‌رود که آن را از نمایه‌سازی صفحه مذکور منع خواهد کرد. در اینجا باز هم تأکید می‌شود که این بخش خاص از وب پنهان تفاوت بنیادی با بخش‌های دیگر قلمرو تاریک و ب دارد چرا که این قسمت آگاهانه از دسترسی موتورهای جستجو خارج شده است و باید که خارج از این حوزه هم باشد و جزء موضوع این مقاله محسوب نمی‌شود. با این حال چون وب خصوصی بخشی از وب پنهان است اشاره اجمالی به آن مفید به نظر می‌رسد.

علاوه بر وب خصوصی بعضی از منابع اطلاعاتی دیگر مثل انواع نشریات الکترونیکی^{۲۸} و بانک‌های اطلاعاتی مبتنی بر وب^{۲۹} که دسترسی به آنها مستلزم پرداخت حق اشتراک و در واقع خرید اطلاعات است و جزء محصولات شرکت‌های مختلف هستند نیز از حوزه دسترسی موتورهای کاوشن خارجند. این بخش از وب نامرئی را "وب ملکی"^{۳۰} می‌نامند. از نظر فنی معمولاً میان منابع وب ملکی و وب عمیق شباهت وجود دارد، اما از آنجا که در این بخش از وب پنهان منابع مالی تولیدکنندگان اطلاعات منجر به خارج کردن این منابع از حوزه دسترسی موتورهای جستجو شده است، این گروه در دسته‌های جداگانه طبقه‌بندی می‌شوند. منابع وب ملکی اغلب از اعتبار و ارزش علمی بالایی برخوردار هستند. بسیاری از اندیشه‌ها و کتابخانه‌های دنیا سالانه مبالغ قابل توجهی برای دریافت حق اشتراک و برخورداری از امکان دسترسی به این منابع اطلاعاتی می‌پردازند.

۴. پایگاه‌های اطلاعاتی با صفحات جستجو مبتنی بر وب

پایگاه‌های اطلاعاتی متعددی در شبکه وب وجود دارند که گرچه دارای صفحه جستجو مبتنی بر وب^{۳۱} هستند اما متدرجات آنها در قالب‌های دیگر است و فقط زمانی یک صفحه ابرمنتن تولید می‌شود که جستجوی مستقلی در درون پایگاه مربوطه اجرا شود. پس از گسترش شبکه وب بسیاری از پایگاه‌ها اطلاعات الکترونیکی خود را از طریق این شبکه دسترسی‌پذیر ساختند. این مجموعه که به وب عمیق مشهور است، از نظر کمی بزرگ‌ترین قسمت از قلمرو وب نامرئی را

کاوشن میسر نیست و به همین دلیل جزء وب پنهان به شمار می‌رود.

۲. منابع اطلاعاتی غیرمنتی^{۲۲}

همان‌طور که گفته شد موتورهای جستجو برای نمایه‌سازی منابع ابرمنتن^{۲۳} طراحی شده‌اند. در حالی که بسیاری از منابع موجود در شبکه وب به قالب‌های دیگر و معمولاً غیرمنتی مثل تصویر یا منابع دیداری و شنیداری هستند و برای موتورهای کاوشن امکان نمایه‌سازی بهینه این منابع به راحتی فراهم نیست و برای این منظور آنها همچنان به یک منبع منتی برای توصیف منابع مذکور نیازمند هستند. به این ترتیب منابع فراوانی در وب وجود دارند که برای موتورهای جستجو به سادگی قابل نمایه‌سازی نیستند. شرمن و پرایس این بخش از دسترسی از وب نامرئی را "وب واقعاً پنهان"^{۲۴} خوانده‌اند و معتقد هستند که احتمال دور ماندن منابع اطلاعاتی این بخش از دسترسی کاربران بیشتر است (۱۵).

۳. منابع محافظت شده از طریق اسم کاربر و گذرواژه^{۲۵}

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 خصوصی" یا "وب شخصی" می‌خوانند. البته تفاوتی اساسی میان این قسمت از وب پنهان با قسمت‌های دیگر وجود دارد چرا که مخفی بودن اطلاعات در این بخش کاملاً تعمدی است و اطلاعات مذکور جزء دارایی‌های شخصی و خصوصی افراد یا سازمان‌ها به شمار می‌رود. کاملاً بدیهی است که هر شخص حقیقی یا حقوقی این حق را دارد که از اطلاعات خصوصی خود به نحوی محافظت نماید. برای مثال سازمان‌های مختلف صفحات و اطلاعاتی در شبکه وب دارند که فقط مربوط به مسائل کاری آن سازمان و کارکنان آن می‌شود و حفاظت از آن مسلم است. این بخش از وب پنهان از حوزه دسترسی موتورهای جستجو خارج است و به طبع باید هم خارج باشد. این بخش از وب پنهان معمولاً به یکی از سه روش زیر ایجاد می‌شود.

نخستین شیوه حفاظت صفحات موردنظر استفاده از اسم کاربر و گذرواژه^{۲۶} است. همان‌طور که گفته شد منابعی که دسترسی به آنها مستلزم استفاده از اسم کاربر و گذرواژه باشد از دسترس نرم‌افزارهای خزنده موتورهای جستجو دور می‌مانند. دومین روش، استفاده از فایلی با عنوان روبوت

۶. عمق نمایه سازی موتورهای جستجو معمولاً هر وب سایت از بیش از یک صفحه تشکیل شده است و صفحات متعدد دیگری در لایه های زیرین سایت ها قرار دارند. منظور از لایه های زیرین، صفحات موجود در وب سایت ها هستند که امکان دسترسی به آنها از طریق صفحه نخست وب سایت^{۳۳} یا صفحات دیگر فراهم می شود. بسته به نوع سایت تعداد این صفحات ممکن است خیلی محدود یا برعکس بسیار گسترده باشد و گاه به ده ها و یا صد ها برسد. در منابع مرتبط با موتورهای جستجو اصطلاح دیگری با عنوان "عمق نمایه سازی"^{۳۴} یا عمق عمل خزنه ها^{۳۵} وجود دارد که بیان کننده تعداد صفحات یک وب سایت است که توسط خزنه های موتورهای جستجو نمایه می شوند. این کار فرایندی هزینه بر است و توان مالی و فناوری بسیاری از موتورهای جستجو به آنها اجازه نمی دهد که بتوانند همه صفحات و لایه های موجود در همه وب سایت هایی را که می باند نمایه سازی کنند. بنابراین هر یک از موتورهای جستجو بسته به توان مالی خود سیاستی مشخص برای نمایه سازی اتخاذ کرده اند و فقط تعداد معینی از لایه های بالایی وب سایت های شناسایی شده را نمایه می کنند. بنابراین، همواره بخشی از اطلاعات نهفته در پایین ترین لایه های وب سایت ها، از حوزه جستجوی کاربران دور می مانند و این مسئله خود موجب تشكیل بخشی از وب پنهان می شود.

۷. حداقل صفحات قابل مرور در نتایج بازیابی^{۳۶} به دلیل این بحثی 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 می دهند. تحقیقات گذشته نشان داده است که بیشتر کاربران فقط صفحه های اول یا دوم نتایج بازیابی را مورد می کنند و به ندرت به صفحات بعدی مراجعه می کنند (۲۰۷:۸؛ ۲۲۶:۱۸؛ ۲۲۷-۲۰۷:۲۲۴). گرچه این موتورهای جستجو تلاش می کنند براساس نظام رتبه بندی^{۳۷} موارد مرتبطتر را در ابتدای فهرست بازیابی نمایش دهند، اما موضوع ربط یک مفهوم ساده نیست و نظام بازیابی نمی تواند همواره براساس معیارهای از قبل تعریف شده یک مدرک را با اطمینان مرتبط یا غیر مرتبط تشخیص دهد و قضاوت نهایی در این باره بر عهده کاربر است. بنابراین حتی در بهترین شرایط بخشی از اطلاعات بازیابی شده که ممکن است مرتبط با نیاز کاربر

به خود اختصاص می دهد به طوری که حجم اطلاعات موجود در وب عمیق چندین برابر وب آشکار یا وب سطحی تخمین زده می شود. برگمن در مطالعه پژوهشی خود حجم این بخش از وب نامرئی را بین ۴۰۰ تا ۵۵۰ برابر حجم محیط وب مرئی تخمین زده است (۱).

هر چند به دلیل گستردگی شگفت آور محیط وب و تغییرات بدون وقفه این محیط می توان به هر رقم و عدد در این مورد با تردید نگریست و هر رقم به زودی کهنه می شود و از طرفی این رقم می تواند خیلی تقریبی به نظر برسد اما دو نکته مهم در اینجا قابل توجه است: نخست آنکه به دلیل محسوب نمودن پایگاه های اطلاعاتی مبتنی بر وب در این مطالعه این رقم چندان دور از ذهن و اغراق آمیز به نظر نمی رسد و دوم آنکه حتی اگر این رقم را مبالغه آمیز فرض کنیم (که البته مبالغه ای در کار نیست و روش پژوهش برگمن گواهی این ادعای است) ارقام خیلی کمتر از این نیز دلالت بر حجم قبل توجهی از اطلاعات دارد که نمی تواند به سادگی نادیده گرفته شود.

۵. منابع و صفحات وب بدون پیوند (لينک)^{۳۲}

امکان برقراری پیوند مستقیم میان هر دو منبع در وب بدون در نظر گرفتن نوع یا فاصله مکانی سایت های آنها یکی از امتیازات بر جسته این شبکه است. پیوندهای موجود در صفحات وب به دو دلیل اهمیت دارد. نخست آنکه کاربران می توانند با دنبال کردن هر یک از این پیوندها به راحتی بین صفحات مختلف حرکت کنند و این یکی از روش های رایج اطلاع یابی در وب است. دلیل دوم که مهم تر به نظر می رسد این است که پیوندهای مذکور تنها راه موجود برای خزنه های موتورهای کاوش برای یافتن صفحات جدید به شمار می روند. هر چند وجود صفحات وب بدون پیوند با صفحات دیگر غیر معمول به نظر می رسد اما هنوز صفحات فراوانی در این شبکه وجود دارند که هیچ پیوندی با هیچ منبع یا صفحه دیگری برقرار نکرده اند و منابع دیگر نیز پیوندی با آنها ایجاد نکرده اند. این منابع مجموعه ای جدا مانده از مجموعه جستجو پذیر وب محسوب می شوند و به همین دلیل خزنه های 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.

32. Disconnected WebPages

33. Homepage

34. Indexing Depth

35. Depth of Crawling

36. Maximum Number of Viewable Results

37. Ranking Policy

جستجوهای دقیق تر و مطمئن تری انجام دهند. تاکنون تدابیر متعددی برای بهبود سطح بهرهوری جستجو در شبکه وب پیش یینی شده است. استفاده از این تدابیر می‌تواند مکمل جستجو از طریق موتورهای کاوش باشد و میزان بازیابی اطلاعات مرتبط را به نحو قابل توجهی ارتقاء بخشد. از جمله این امکانات می‌توان به دروازه‌های اطلاعاتی موضوعی اشاره کرد. از آنجا که انتخاب منابع موجود در این سایت‌ها توسط متخصصان موضوعی یا کتابداران انجام می‌شود معمولاً منابع برگزیده از کیفیت و اعتبار قابل قبولی برخوردار هستند چرا که با د. نظر گفت: معاهده، مدون، انتخاب شده‌اند.

روش دیگر، گسترش دادن حوزه جستجوست. برای مثال استفاده از ابرموتورهای کاوش^{۳۹} 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. امکان دیگری که اخیراً مورد توجه متخصصان بازیابی اطلاعات در وب قرار گرفته، استفاده از عوامل هوشمند^{۴۰} است.

این عوامل هوشمند که توانایی جستجو، مقایسه و انتخاب منابع اطلاعاتی را براساس نیاز مطرح شده توسط کاربر دارند، به عنوان کارگزار کاربران در این فضای گسترده به جستجو می پردازند و نتایج جستجو را بر اساس نیاز مطرح شده توسط کاربران، فیلتر کرده و به آنها تحویل می دهند. به طور کلی پیش بینی می شود با پیشرفت ابزارهای بازیابی اطلاعات در وب و بهبود سواد اطلاعاتی کاربران به تدریج از سطح ناپیدایی اطلاعات د محيط وب کاسته شود.

همان طور که گفته شد بخش عمده‌ای از اطلاعات مخفی مانده در وب پنهان را می‌توان به راحتی برای کاربران این شبکه دسترسی پذیر نمود. مهم‌ترین راهکار عملی در شرایط موجود برای نفوذ به قلمرو تاریک وب آگاهی از وجود آن و آگاهی از داشتگی های اطلاعاتی، در آن است.

متخصصان رفتارهای اطلاع‌یابی طی سال‌های اخیر تلاش کرده‌اند فرایند جستجوی کاربران در وب را از دیدگاه‌های مختلف بررسی کنند. مرور پژوهش‌های مذکور نشان می‌دهد که پدیده وب پنهان به مثابه یکی از عوامل مؤثر در اطلاع‌یابی در، وب مورد توجه قرار نگرفته است (۱۴:۸).

باشد در لایه‌های زیرین صفحات بازیابی شده توسط موتورهای جستجو مدفون می‌ماند و به این ترتیب جزء وب پنهان به شمار می‌رود.

اہمیت و ب پنہاں

پس از بیان مفهوم و تعریف وب پنهان ممکن است این پرسش مطرح شود که اصولاً چه ضرورتی برای جستجو و دسترسی به منابع موجود در وب پنهان وجود دارد؟ و با توجه به حجم روزافزون منابع موجود در شبکه وب شاید پنهانی بخشنده از آن خلی جدی در دسترس پذیری منابع ایجاد نکند. اساساً اهمیت منابع موجود در بخش پنهان وب از دو جهت قابل بررسی است. نخست آنکه از نظر کمی این بخش به مراتب حجمی بیش از نیمه آشکار شبکه وب دارد. به بیان دیگر، حجم اطلاعات نهفته در وب پنهان خیلی بیشتر از بخش سطحی یا آشکار ^{۳۸} آن است. دوم آنکه از نظر کیفی نیز این قسمت بسیار حائز اهمیت است. اطلاعات مندرج بخش‌های مختلف این مجموعه به ویژه 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 وب را نادیده گرفت، چرا که کاربران می‌توانند با روش‌های مختلف سطح ناپیدایی اطلاعات در وب را در جستجوهای خود کاوش دهند و از این طریق سطح بهره‌وری خود را بهبود بخشنده.

روش‌های اطلاع‌یابی در وب‌پنهان

راهکارهای مختلفی برای ارتقاء سطح دسترس پذیری اطلاعات در وب پنهان وجود دارد. مهم‌ترین گام در این زمینه آگاهی رسانی درباره وجود قلمرو تاریک وب است. متأسفانه بسیاری از کاربران از وجود این بخش از وب 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 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 مازان جستجو در وب وجود دارد متفاوت است»(۱۶).

اگر کاربران به این مسائل آگاهی داشته باشند می‌توانند

بیان مسئله

این حوزه توجه نشان دادند و از آن تاریخ تاکنون رشد فزاینده‌ای در تعداد و تنوع رویکردهای پژوهشی مطالعات مذکور مشاهده می‌شود (۱۸۵-۱۶۷:۶). به طور کلی دو رویکرد اصلی در این مطالعات قابل تفکیک و طبقه‌بندی سنت به طوری که این تحقیقات عمدتاً از رویکرد نظام مدار^{۴۱} برخوردارند یا رویکرد کاربر مدار^{۴۲} را برگزیده‌اند. مطالعات نظام مدار در جستجوی راهکارهایی برای ارتقاء سودمندی و بهبود عملکرد نظامهای بازیابی اطلاعات در وب از جمله موتورهای جستجو هستند. در مقابل پژوهش‌های کاربر مدار به مطالعه رفتارهای اطلاع‌یابی کاربران در محیط وب و چگونگی تعامل آنها با این محیط می‌پردازند. تأثیر عوامل گوناگون از جمله مهارت‌های یارانه‌ای افراد، تجربه جستجوی آنان در وب و عوامل دیگر بر بازیابی اطلاعات از مهم‌ترین موضوعات این تحقیقات به شمار می‌روند (۲۳۳-۲۵۵:۱۰). نکته قابل توجه در مطالعات بازیابی اطلاعات در وب این است که به رغم اهمیت وب پنهان در فرایند جستجو، این موضوع در مطالعات قبلی مورد توجه قرار نگرفته است. منابعی که تاکنون به موضوع وب پنهان و پنهانی اطلاعات در محیط وب پرداخته‌اند بیشتر جنبه توصیفی دارند و فقط به معرفی آن و مرور عواملی که در ایجاد آن مؤثر هستند محدود می‌شوند. مبحث وب پنهان در برخی کتاب‌های این حوزه به طور مختصر و گذرآ مطرح شده است. با این حال دو کتاب به طور اختصاصی به موضوع وب پنهان پرداخته‌اند.

کتاب نخست تحت عنوان وب نامئی: کشف منابع اطلاع‌یابی که موتورهای جستجو نمی‌توانند بیانند توسط کریس شرمن و گری پرایس در سال ۲۰۰۱ منتشر شده و جامع ترین منبع در زمینه وب پنهان محسوب می‌شود و تاکنون بیش از هر منبع دیگر مورد استناد قرار گرفته است (۱۶). دومین کتاب توسط پاول پدلی در همین سال منتشر شده است و منبع مختصر و مفیدی برای آشنایی با مقدمات این موضوع به شمار می‌آید.

مرور نوشتارها

پژوهش درباره جنبه‌های مختلف وب پنهان، بخشی از مطالعات در زمینه‌ای گسترده‌تر است که از آن به عنوان "حوزه تحقیقات جستجو و بازیابی اطلاعات در محیط وب" یاد می‌شود. مرور مقالات سال‌های اخیر در این حوزه حاکی از آن است که محققان بازیابی اطلاعات از سال ۱۹۹۵ به بعد به

پرسش‌های پژوهش

۱. آیا کاربران از وجود وب پنهان آگاه هستند و وجود آن را در تجارب جستجوی قبلی خود تجربه کرده‌اند؟
۲. احساس و نظر کلی آنان از جستجو در شبکه وب چگونه است؟
۳. آنان معمولاً از چه ابزارهای کاوشی برای اطلاع‌یابی در

41. System-Oriented or Technical-Oriented

42. User-Oriented or User-Centred

به همین منظور هفت نفر از جامعه پژوهش به عنوان نمونه در این مصاحبه شرکت کردند. هر مصاحبه بین ۴۵ دقیقه تا یک ساعت به طول انجامید و جریان مصاحبه‌ها از ابتدای آنها توسط ضبط صوت دیجیتالی خبیط شد. در مرحله بعد همه مصاحبه‌ها به طور کامل پیاده‌سازی و تایپ شد. گرچه پیاده‌سازی مصاحبه‌ها کاری طولانی بود اما از آنگاه که پژوهش حاضر از نوع کیفی است و پژوهشگر به همه جزئیات نیاز داشت، این کار ضروری به نظر می‌رسید. پیاده‌سازی هر ساعت مصاحبه بین ۷ تا ۸ ساعت زمان نیاز داشت و تحلیل داده‌ها براساس متن به دست آمده از پیاده‌سازی انجام شد. استفاده از ضبط دیجیتالی این امکان را برای پژوهشگر فراهم آورد که جریان کامل هر مصاحبه را به صورت یک فایل صوتی به ایانه منتقل کند.

خلاصه‌ای از نتایج پژوهش مقدماتی

نتایج تحقیق حاکی از آن بود که گرچه هیچ یک از کاربران قبل‌آوازه و پنهان را نشنیده بوده و تعریف روشی از آن در ذهن نداشتند اما همه آنان وجود آن را به نوعی در جستجوهای قبلی تجربه کرده و هر کدام راهبردهای خاص خود را برای اطلاع‌یابی از این بخش از وب داشتند. به بیان دیگر هر یک از آنها براساس تجارت قبلی جستجو در وب، توصیه دوستان یا استادان راهنمایی، کمک کتابداران دانشگاه و یا روش‌های دیگر از منابعی که براساس تعریف ارائه شده در این مقاله جزء وب پنهان محسوب می‌شود به نحوی استفاده می‌کردند بدون آنکه این منابع را نامرئی فرض کنند. فرایندی که طی آن مصاحبه‌شوندگان تجربه جستجوی خود را تشریح می‌کردند نشان می‌داد که عوامل متعددی آنان را از بخش سطحی وب به بخش عمیق رهمنمون ساخته است بی‌آنکه آنها آگاهانه

تجربه جستجوی آنان در شبکه و بین هشت تا ده سال بود و به طور متوسط هر روز در وب به جستجو می پرداختند و به همین دلیل جزء کاربران کاملاً با تجربه طبقه بندی می شوند. همه آنها به جز یک نفر همواره از موتور جستجوی گوگل^{۴۵} برای کاوشن در وب استفاده می کنند و جالب است که هر یک از آنان قبل از معرفی و شهرت گوگل (از سال ۱۹۹۹ به بعد) از سایر موتورها مثل یاهو^{۴۶} یا آلتاویستا^{۴۷} استفاده می کردند و پس از آشنایی با گوگل دیگر هیچ موتور جستجوی دیگری را به آن ترجیح ندادند. مهم ترین دلایل

و ب استفاده می کنند و چه دلایلی برای انتخاب این ابزارها
دارند؟

۴. کاربران معمولاً جستجو در وب را چگونه آغاز می‌کنند و پس از طی چه مراحلی و در چه مقطعی جستجوی خود را خاتمه یافته‌اند (اقرئی می‌کنند؟)

۵. بدون در نظر گرفتن اینکه کاربران وب قبل‌اصل‌لاح و ب پنهان را شنیده یا نشنیده‌اند، براساس تجارب قبلی آنان از جستجو در وب، تعریف آنان از این پدیده چیست و برای آن چه درجه‌ای از اهمیت در بازیابی اطلاعات در وب قائل هستند؟

جامعه پژوهش

کل دانشجویان دوره دکترا (۳۹ نفر) و پژوهشگران (۱۷ نفر) گروه اطلاع رسانی^{۴۳} دانشگاه شفیلد انگلستان مجموعاً ۶۵ نفر، جامعه پژوهش این مطالعه مقدماتی را تشکیل می‌دادند. با انتخاب این جامعه پژوهشی، پژوهشگر قصد داشت که این مطالعه مقدماتی را در جامعه‌ای انجام دهد که انتظار می‌رود از بیشترین تجربه و دانش برای بازیابی اطلاعات در وب برخوردار باشند. به بیان دیگر پژوهش مقدماتی در فضایی انجام شود که افراد جامعه پژوهشی از بیشترین دانش تخصصی برای بازیابی اطلاعات در وب بهره‌مند هستند و مشخصات افراد نمونه نیز این فرض اولیه را تأیید می‌کرد. همه آنان از سال‌ها تجربه در جستجو در منابع الکترونیکی برخوردار بودند و اغلب آنان جستجو در منابع الکترونیکی پیوسته را حتی قبل از ظهور شیکه وب و از طریق خدمات اینترنتی پیش از وب آغاز کرده بودند.

روش و ابزار گردآوری داده‌ها

صاحبہ نیمہ ساختار یافتہ با پرسش‌های باز، روش اصلی گردآوری داده‌ها در این پژوهش مقدماتی بود که پرسش‌های صاحبہ ضمیمه مقاله است. دلیل انتخاب این شیوه گردآوری داده‌ها به رویکرد پژوهش بازمی‌گردد. از آنجا که پژوهش حاضر از نوع کاملاً کیفیست، به آن دسته از ابزارهای گردآوری داده‌ها نیاز است که امکان فراهم آوردن داده‌های کیفی و عمیق را مهیا سازند. برای مثال پرسشنامه ابزار مناسب برای این پژوهش محسوب نمی‌شود چرا که در این تحقیق پژوهشگر بیش از آنکه به حجم زیادی از داده‌های سطحی از یک نمونه بزرگ نیاز داشته باشد به داده‌هایی عمیق از یک نمونه کوچک نیازمند است.

یکی از آنان وب پنهان را سخت‌افزارها و نرم‌افزارهای تلقی می‌کرد که از دید کاربران پنهان هستند ولی در واقع امکان فنی جستجو در وب توسط این امکانات فراهم می‌شود. یکی دیگر از آنان بر این باور بود که شاید منظور از وب پنهان این است که افراد می‌توانند نام و نشان خود را در وب مخفی نگه دارند!

نتایج تحقیق مقدماتی نشان داد که موضوع پنهان بودن برخی منابع اطلاعاتی در شبکه وب، یک مبحث با حد و مرزهای تفکیک شده و مشخصی نیست. در واقع نمی‌توان و نباید این شبکه را به دو بخش نهان و آشکار یا مرئی و نامرئی تقسیم کرد و محدوده و مرز مشخصی بین این دو قسمت قائل شد. ممکن است یک منبع اطلاعاتی برای یک فرد یا در یک تجربه جستجوی جزء قلمرو وب پنهان قرار گیرد و در همان زمان برای کاربر دیگری کاملاً دسترس پذیر باشد. به بیان دیگر اصطلاح "سطح پنهانی اطلاعات" در وب از گویایی بیشتری نسبت به "وب پنهان" برخوردار است و یکی از نتایج این پژوهش تا این مرحله ارائه تعریفی تازه با عنوان "سطح پنهانی اطلاعات در وب" و به شرح زیر است: برای هر کاربر و در هر مرحله از فرایند جستجو در محیط وب، سطحی از پنهانی اطلاعات وجود دارد که میزان آن بسته به عوامل مختلف از جمله مهارت‌های اطلاع‌یابی و تجربه کاربر در جستجو، موضوع جستجو، انگیزه و هدف کاربر از جستجو و میزان اهمیت موضوع جستجو برای او و انتخاب ابزار جستجو، متغیر است. بعضی از عوامل مذکور معمولاً سطح پنهانی اطلاعات در وب را کاهش و برخی دیگر افزایش می‌دهند. مثلاً افزایش تجربه کاربر در جستجو در وب معمولاً سطح ناپیدایی اطلاعات در وب را برای او کاهش و گستردنگی موضوع جستجو این سطح را افزایش می‌دهد. انتخاب ابزار کاوش مناسب با موضوع جستجو نیز می‌تواند در کاهش سطح ناپیدایی اطلاعات در وب مؤثر باشد.

مراحل بعدی پژوهش

آنچه در این مقاله آمد گزارش مختصری از مراحل مقدماتی پژوهش درباره رفتارهای اطلاع‌یابی کاربران در شبکه وب و بهویژه چگونگی نگرش و اطلاع‌یابی آنها از منابع وب پنهان بود. در حال حاضر مراحل بعدی پژوهش در یک جامعه پژوهش به مراتب بزرگ‌تر و با پرسش‌های ساختاریافته در حال انجام است و امید آن می‌رود که در پایان تحقیق مدلی

محبوبیت گوگل سادگی صفحه جستجوی آن و نظام رتبه‌بندی کارآمدی که دارد عنوان شد و آنها از نتایج جستجوی این موتور در اغلب موارد نسبتاً راضی بودند. هرچند گاهی از ناتوانی آن در رفع نیازهای اطلاعاتی خود شکایت داشتند اما در مجموع استفاده از گوگل را به هر موتور جستجوی دیگر ترجیح می‌دادند.

تجارب جستجوی آنها در گستره‌ای قرار داشت که از جستجوهای ساده برای مقاصد شخصی تا جستجوی پیچیده برای یافتن اطلاعات مربوط به کار یا تحصیل را در بر می‌گرفت. توصیف تجارب موفق جستجو در وب حاکی از آن بود که آنان معمولاً جستجوهای یک مرحله‌ای از طریق موتورهای کاوش را که منجر به یافتن پاسخ سریع می‌شود بسیار موفق ارزیابی می‌کنند و تجاربی که پس از جستجو در موتور جستجو به جستجویی مجزا در منبع اطلاعاتی دیگری نیازمند بوده و صرف وقت بیشتری را می‌طلبیده، جستجوی نسبتاً موفق و جستجوهایی را که نهایتاً بی‌نتیجه مانده بود کاملاً ناموفق توصیف می‌کردند. به رغم آنکه همه مصاحبه‌شوندگان از مهارت بالایی در بازیابی اطلاعات در وب برخوردار بودند، اما همه آنها دست‌کم یک جستجوی ناموفق در وب را اخیراً تجربه کرده بودند. تشریح مراحلی که در این جستجوی ناموفق طی شده بود نشان می‌داد که چنانچه آنان از وجود وب پنهان آگاهی داشتند ممکن بود راهبرد جستجوی دیگری انتخاب می‌کردند که احتمال موفقیت آن بیشتر بود.

همه مصاحبه‌شوندگان تجارب قبلی جستجوهای خود در وب را براساس هدف و موضوع هر جستجو طبقه‌بندی می‌کردند. برای مثال جستجو در شبکه وب برای مسائل روزمره و شخصی در مقابل جستجوهای معمولاً جدی تر برای مسائل کاری و تحصیلی. توصیف آنان از تجارب جستجو در وب نشان داد که آنان بسته به نوع و هدف موضوع جستجو معمولاً راهبردهای جستجوی گوناگونی برمی‌گزینند با این حال چنانچه سایت مشخصی را از قبل در نظر نداشته باشند معمولاً موتور جستجوی گوگل نقطه شروع فرایند اطلاع‌یابی آنها در وب محسوب می‌شود. نکته جالب توجه دیگر در این پژوهش مقدماتی تأیید این فرض اولیه بود که واژه وب پنهان از نظر لغوی از گویایی لازم برخوردار نیست به طوری که هر یک از مصاحبه‌شوندگان تعریف متفاوتی از وب پنهان در ذهن داشتند. از آن جمله

از رفتارهای اطلاع‌یابی کاربران در وب پنهان تدوین و ترسیم شود. این مدل می‌تواند در کنار مدل‌های موجود از اطلاع‌یابی در محیط وب خلاً ناشی از در نظر نگرفتن وب پنهان در مدل‌های قبلی را پر کند و تصویری روشن‌تر از چگونگی تعامل کاربران با شبکه وب ارائه دهد.

وب پنهان در ذهن داشتند که البته هیچ کدام از این پیش‌فرضها با تعریف وب پنهان همخوانی نداشت و فقط بیانگر نخستین برداشت آنان از معنای تحت‌اللفظی این اصطلاح بود.

گرچه چگونگی تعامل کاربران با منابع شبکه وب در مدل‌های اطلاع‌یابی کاربران در وب مورد پژوهش قرار گرفته‌اند اما به جایگاه و اهمیت بخش پنهان وب توجه کافی نشده است و امید آن می‌رود که پژوهش حاضر در مراحل بعدی خود موفق به شناسایی عناصر اصلی در این زمینه و ترسیم مدل اطلاع‌یابی در وب پنهان شود. ارائه مفهوم سطح ناپیدایی اطلاعات در وب، نخستین گام در تدوین این مدل محسوب می‌شود.

پیوست ۱. متن پرسش‌های مصاحبه

توضیح: این مصاحبه مجموعاً شامل ۲۵ پرسش است که این پرسش‌ها از پنج پرسش اصلی تشکیل شده و هر کدام در برگیرنده چند پرسش فرعی است. پرسش‌های فرعی فقط زمانی از مصاحبه شوندگان پرسیده شده‌اند که پاسخ آنان به پرسش‌های اصلی در برگیرنده موارد مذکور در این پرسش‌ها نبود.

۱. آیا ممکن است نظر کلی خود را درباره جنبه‌های مختلف جستجو و بازیابی اطلاعات از شبکه وب بیان دارید؟ ارزیابی کلی شما از مزایا و محدودیت‌های فرایند جستجو در این شبکه چیست؟

۱-۱. چه مدت است که شما برای یافتن نیازهای اطلاعاتی خود به شبکه وب مراجعه می‌کنید؟

۱-۲. معمولاً هر چند وقت یک بار به جستجو در وب می‌پردازید؟

۱-۳. معمولاً در وب در جستجوی چه نوع اطلاعاتی هستید؟

۱-۴. آیا معمولاً پاسخ پرسش‌های خود را به راحتی و به سرعت در این شبکه می‌یابید؟

۱-۵. نتایج جستجوی ابزارهای کاوش در وب چقدر برای شما رضایت‌بخش است؟

۱-۶. آیا از موتورهای جستجو برای بازیابی اطلاعات استفاده می‌کنید؟ کدام موتور را ترجیح می‌دهید؟ و چه ویژگی یا ویژگی‌هایی از موتور جستجوی مورد نظر برای شما مهم بوده است؟

۱-۷. آیا به نظر شما موتورهای جستجو امکان دسترسی به

بحث و نتیجه

منابع موجود در وب پنهان از اهمیت قابل توجهی در پاسخگویی به نیازهای اطلاعاتی کاربران وب برخوردار است اما به دلیل آنکه بازیابی اطلاعات از این بخش شبکه وب به سهولت اطلاع‌یابی از وب سطحی نیست، معمولاً احتمال دور ماندن آن از نظر کاربران بیشتر است. آنچه در این مقاله آمد تأکیدی بود بر اهمیت این قسمت از محیط وب و به ویژه بر جنبه‌های کاربرمدار آن. گرچه واژه "پنهان" یا "نامرئی" در این اصطلاح دلالت بر نوعی دسترسی ناپذیری این منابع دارد، اما در واقع منابع اطلاعاتی وب پنهان برای کاربران دسترسی پذیر است در صورتی که آنها جستجوی خود در وب را فقط به نتایج موتورهای کاوش محدود نکنند.

نتایج تحقیق مقدماتی نشان داد که هیچ یک از مصاحبه‌شوندگان در این تحقیق قبل از اصطلاح وب پنهان را نشنیده بودند و تعریف روشی از آن در ذهن نداشتند. با این حال همه آنان وجود آن را به نوعی در جستجوهای قبلی تجربه کرده و هر یک راهبردهای خاص خود را برای اطلاع‌یابی از این بخش از وب داشتند. به طور کلی تجارب جستجوی آنها در گسترده‌ای قرار داشت که از جستجوهای ساده برای مقاصد شخصی تا جستجوی پیچیده برای یافتن اطلاعات مربوط به کار یا تحصیل را شامل می‌شد. توصیف تجارب موفق جستجو در وب حاکی از آن بود که آنان معمولاً جستجوهای یک مرحله‌ای و ساده از طریق موتورهای جستجو را که منجر به پاسخ سریع و سهل الوصول برای نیازهای اطلاعاتی آنان می‌شود، کاملاً موفق ارزیابی می‌کنند و تجاربی که در پی کاوش در موتور جستجو به جستجویی مجزا در منبع اطلاعاتی دیگری نیازمند باشد و به صرف وقت بیشتری نیاز داشته باشد جستجوی نسبتاً موفق؛ و جستجوهایی را که نهایتاً می‌نتیجه مانده بود کاملاً ناموفق بر می‌شمرند. یکی دیگر از نکات مهم این است که اصطلاح وب پنهان از نظر لغوی از گویایی لازم برخوردار نیست به طوری که هر یک از مصاحبه‌شوندگان تعریف متفاوتی از

شما مؤثر است؟

۴. به نظر شما واژه "وب نامرئی" یا "وب پنهان" می‌تواند به چه معنا باشد؟ آیا قبلاً این اصطلاح را شنیده بودید؟ و اگر تاکنون با این واژه برخورد نکردید این واژه دلالت به چه مفهومی ممکن است داشته باشد؟

۵-۱. آیا پنهانی بخشی از منابع اطلاعاتی وب شما را نسبت به افزایش احتمال عدم بازیابی اطلاعات مورد نیازتان نگران می‌سازد؟

۵. آیا مطلب خاصی در مورد جستجو در وب به نظر شما می‌رسد که تمایل داشته باشید در اینجا به آن اشاره داشته باشید؟ اگر مطلبی هست لطفاً بفرمایید.

منابع

1.Bergman, M.K. "The Deep Web: Surfacing Hidden Value". *The Journal of Electronic Publishing*, Vol. 7, No. 1 (2001). [on-line]. Available: <http://www.press.umich.edu/jep/07-01/bergman.html>. [25 Jan. 2004].

2.Broder, A. ... [et al]. "Graph structure in the Web". In Proceedings of WWW9 Conference, 2000. [on-line]. Available: <http://www9.org/w9cdrom/160/160.html>. [2 Jan. 2004].

3.Chu, H.; Rosenthal, M. "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, PP.127-135.

4.Garcia, F. "Business and Marketing on the Internet". 1996. [on-line]. Available: <http://web.archive.org/web/19980113075450/http://www.tcp.ca/Jan96/BusandMark.html>. [10 Dec. 2003].

5.Gordon, M.; Pathak, P. "Finding information on the World Wide Web: the retrieval effectiveness of search engines". *Information Processing and*

تمام منابع موجود در وب را فراهم می‌کنند؟

۸-۱. معمولاً نتایج بازیابی موتورهای جستجو را چه اندازه مربوط با نیاز خود ارزیابی می‌کنید؟

۲. آیا ممکن است یکی از موارد جستجوی موفق را که اخیراً در وب انجام داده‌اید تشریح کنید؟ منظور از جستجوی موفق در اینجا هر جستجویی است که در آن شما موفق به بازیابی اطلاعات مورد نیاز خود شدید، اطلاعاتی که برای شما مهم بوده و یا منجر به اتخاذ تصمیمی شده است.

۱-۲. آیا به خاطر می‌آورید که کدام موتور جستجو یا هر ابزار بازیابی اطلاعات دیگر برای شما در این جستجوی موفق مفید و مؤثر بود؟

۲-۲. شما معمولاً جستجوی خود را در وب از کجا و چگونه آغاز می‌کنید؟

۳-۲. آیا به خاطر دارید که معمولاً کدام سایت نقطه شروع جستجوی شما در وب است؟

۴-۲. چرا شما این سایت را برای نقطه شروع انتخاب کرده‌اید؟

۵-۲. آیا ویژگی یا ویژگی‌هایی در این سایت وجود دارد که شما را به این انتخاب تشویق نموده است؟

۳. آیا ممکن است یکی از موارد جستجوی ناموفق را که اخیراً در وب تجربه کرده‌اید تشریح کنید؟ منظور از جستجوی ناموفق جستجویی است که در آن شما موفق به بازیابی اطلاعات مورد نیاز خود نشدید.

۱-۳. به نظر شما در تجربه جستجوی ناموفقی که مطرح نمودید چه چیز مانع دستیابی شما به اطلاعات موردنیاز شما شده بود؟ چرا شما موفق به بازیابی آن نشدید؟

۲-۳. آیا شما تاکنون تصور کرده‌اید که به احتمال زیاد اطلاعات بیشتری در شبکه وب درباره موضوع موردنیاز شما وجود داشته باشد اما به هر دلیل دسترسی به آن برای شما میسر نشده است؟

۳-۳. وقتی در جستجوی مطلبی در شبکه وب هستید در چه زمانی جستجوی خود را متوقف می‌کنید؟

۴-۳. در چه زمانی جستجوی خود را در وب ارزیابی می‌کنید؟

۵-۳. آیا شما همواره به یک شیوه مشخص در شبکه وب جستجو می‌کنید؟ یا برای موضوعات مختلف شیوه‌های جستجوی منفاوتی بر می‌گزینید؟

۶-۳. آیا موضوع مورد جستجو بر انتخاب شیوه جستجوی

information review, Vol. 19, No. 1 (2002): 23-31.

14.Ross, N. ; Wolfram, D. "End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine". *Journal of the American Society for Information Science*, Vol. 51, No. 10 (2000): 949-958.

15.Sherman, C. ; Price, G. "Exploring the Invisible Web". *Online*, Vol. 25, No. 4 (2001): 32-35.

16.Ibid. "The Invisible Web: Uncovering Information Sources Search Engines Can't See". *CyberAge Book*, (2001): 56-58.

17.Smith, C.B. "Getting to Know the Invisible Web". *Library Journal*, Vol. 126, No. 11 (2001): 16-19.

18.Spink, A. ... [et al]. "Searching the Web: the public and their queries". *Journal of the American Society for Information Science and Technology*, Vol. 52, No. 3 (2001): 226-234.

19.Wang, P. ; Hawk, W.B.; Tenopir, C. "Users' interaction with World Wide Web resources: an exploratory study using a holistic approach". *Information Processing and Management*, Vol. 36, No. 2 (2000): 229-251.

20.Wildemuth, M.B. "Effective Methods for studying information seeking and use". *Journal of the American Society for Information Science and Technology*, Vol. 53, No. 14 (2002): 1218-1222

تاریخ دریافت: ۱۳۸۴/۱۲/۲۴

Management, Vol. 35, No. 2 (1999): 141-180.

6.Hsieh-Yee, I. "Research on Web search behaviour". *Library & Information Science Research*, Vol. 23, No. 2 (2001): 167-185.

7.Jansen, B.J.; Spink, A. "The Excite research project: A study of searching characteristics by Web users". *Bulletin of the American Society for Information Science*, Vol. 27, No. 1 (2000): 15-17.

8.Jansen, B.J.; Spink, A. ; Saracevic, T. "Real life, real users and real needs: A study and analysis of users queries on the Web". *Information Processing and Management*, Vol. 36, No. 2 (2000): 207-227.

9.Jansen, B.J.; Pooch, U. "A review of Web searching studies and a framework for future research". *Journal of the American Society for Information Science and Technology*, Vol. 52, No. 3 (2001): 235-246.

10.Kim, K. "Information-seeking on the Web: Effects of user and task variables". *Library & Information Science Research*, Vol. 23, No. 3 (2001): 233-255.

11.Palmquist, R.A.; Kim, K. "Cognitive style and online database search experience as predictors of Web search performance". *Journal of the American Society for Information Science*, Vol. 51, No. 6 (2000): 558-566.

12.Redley,P. "The Invisible Web: searching the hidden parts of the Internet". Aslib.

13.Pedley, P. "Why you can't afford to ignore the Invisible Web". *Business*