

وب پنهان چیست و چه اهمیتی دارد؟

نگارش: یزدان منصوریان

دانشجو دکترا کتابداری و اطلاع رسانی دانشگاه شفیلد انگلستان

بهمن 1382

v.mansourian@sheffield.ac.uk

Department of Information Studies, University of Sheffield
Regent Court, 211 Portobello Street, Sheffield, S1 4DP, England, UK

کلیدواژه‌ها: شبکه جهانگستر وب / وب نامرئی / وب پنهان / وب عمیق / بازیابی اطلاعات در محیط وب / محدودیتهای موتورهای کاوش

مقدمه:

هر چه بر حجم و تنوع منابع موجود در شبکه وب افزوده می گردد، بحث جستجو و اطلاع یابی¹ در محیط وب ابعاد و جنبه های گسترده تر و پیچیده تر به خود می گیرد. گاهی یافتن پاسخ یک پرسش در این شبکه به شکلی ساده و سریع انجام می شود و تنها به تایپ کلید واژه ای مناسب در یک موتور جستجوی آشنا مثل گوگل نیاز است. اما همیشه اطلاع یابی در این شبکه به این سهولت نیست و معمولاً به تدبیر و تامل بیشتری نیاز دارد. با این حال گاهی نیز با اتخاذ تمام تدابیر ممکن باز هم جستجو در وب و بویژه از طریق موتورهای کاوش عمومی نتیجه ای رضایت بخش در پی ندارد. عدم توفیق در بازیابی اطلاعات مورد نیاز از شبکه وب ممکن است به دلیل عدم وجود منابع مورد نیاز به شکل الکترونیکی باشد و یا به دلیل عدم تناسب استراتژی اتخاذ شده برای جستجو و نوع نیاز اطلاعاتی باشد. اما چنانچه منبعی مورد نظر در محیط وب موجود باشد اما موتورهای جستجو قادر به فراهم آوری امکان بازیابی آن نباشند اطلاعات مورد نظر در سایه وب پنهان/وب نامرئی مخفی مانده است.

در این مقاله تلاش می شود معنا و مفهوم وب پنهان و سطح پنهانی اطلاعات در محیط وب به طور اجمالی مورد بررسی قرار گیرد و به چند پرسش اساسی در این زمینه از جمله اینکه وب پنهان چه نوع منابع اطلاعاتی را در بر می گیرد و در چه مواقعی به اطلاع یابی از این بخش وب نیاز است پاسخ داده شود.

از آنجا که موضوع وب پنهان و بویژه جنبه های کاربر مدار آن موضوع رساله دکتری نگارنده است امید آن می رود که با پیشرفت کار پژوهش در این مقطع مقالات مفصلتری در این زمینه تهیه و منتشر شود.

¹ . Information Seeking & Retrieval

تعریف وب نامرئی یا وب پنهان چیست و مبدع این اصطلاح کیست؟

مرور متون تخصصی موجود نشان می دهد که به رغم آنکه تاکنون مقالات متعددی به زبان انگلیسی درباره وب نامرئی منتشر شده و تقریباً اتفاق نظری جامع در مورد تعریف و مفهوم آن وجود دارد اما هنوز به روشنی معلوم نیست که چه کسی این واژه را برای اولین بار به کار برده است. مرور نوشتارها نشان می دهد که به احتمال قریب به یقین گویا واژه وب پنهان یا وب نامرئی² برای نخستین بار در سال 1994 توسط دکتر ژیل السورث³ ابداع شده است (برگمن، 2001). البته معدودی از منابع موجود در این زمینه نیز شخص دیگری به نام ماتیو کل⁴ را به عنوان اولین مبدع این اصطلاح معرفی می کنند (شرمن، 2003). بیشتر منابع موجود بجای معرفی نخستین مبدع این اصطلاح بیشتر به تبیین معنی و مفهوم آن می پردازند و در مواردی نیز که اشاره شده است نام دکتر ژیل السورث بیشتر به چشم می خورد. به هر حال شواهد موجود حاکی از آن است که مبدع این اصطلاح باید یکی از این دو نفر باشد.

"وب پنهان یا وب نامرئی" اشاره به يك مفهوم چند بعدي و گسترده دارد به شكلي كه ارائه تعريفی 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 (2001:56) به طور خلاصه این چنین تعریف نمود:

وب پنهان آن بخش از فضاي شبکه جهانگستر وب است که عمدتاً شامل منابع اطلاعاتی غیر متنی و پویا⁵ است که به هر دلیل به طور موقت یا دائم خارج از حوزه جستجو و بازیابی موتورهای کاوش قرار دارند و بازیابی اطلاعات موجود در آن از طریق استفاده مستقیم از این موتورها میسر نمی باشد. امکان بازیابی منابع پنهان در وب نامرئی یا برای موتورهای کاوش از نظر فنی میسر نیست و یا محدودیهای مالی مانع از نمایه سازی این منابع شده است. (شرمن و پرایس، 2001 b: 56)

آیا منابع اطلاعاتی وب پنهان/وب نامرئی واقعاً پنهان و نامرئی هستند؟

گاهی واژه وب نامرئی یا وب پنهان ممکن است کمی گمراه کننده به نظر برسد. کلمه نامرئی یا پنهان در این واژه به این معنا نیست که اطلاعات موجود در این بخش از وب به معنای کلمه نامرئی یا غیر قابل رویت هستند. انتخاب این واژه از این واقعیت سرچشمه می گیرد که به دلیل محبوبیت و عمومیت استفاده از موتورهای کاوش این ابزار بازیابی اطلاعات در وب بیشترین استفاده را برای کاربران داشته و این تصور عمومی کم و بیش وجود دارد که موتورهای کاوش به تمام بخشهای شبکه وب دسترسی دارند و این در حالی است که با توضیحاتی که در ادامه می آید محدودیتهای این موتورها روشنتر خواهد شد. به بیان دیگر این بخش از وب بیشتر برای کاربرانی پنهان خواهد ماند که فرایند جستجوی خود را عمدتاً تنها به موتورهای کاوش محدود می کنند. بنابراین اطلاعات موجود در وب پنهان قابل بازیابی و دسترسی است اما نه به سهولت منابع موجود در وب سطحی. به طوری که دسترسی به منابع وب پنهان مستلزم استفاده از ابزارهای کاوشی غیر از موتورهای جستجو معمولی است مثل استفاده از ورودگاهها⁶ دروازه های اطلاعاتی موضوعی⁷.

وب پنهان چه نوع منابع اطلاعاتی را در بر می گیرد؟

مقاله حاضر هفت گروه اصلی از این منابع را به شرح زیر خلاصه و طبقه بندی می کند:

1. منابع اطلاعاتی تازه افزوده شده به محیط وب:

هر منبع اطلاعاتی که به محیط وب افزوده می شود تنها در صورتی توسط موتورهای جستجو قابل بازیابی خواهد بود که قبلاً توسط موتورهای جستجو به نحوی شناسایی و نمایه سازی شده باشد. فرایند شناسایی و نمایه سازی منابع جدید به دو صورت ممکن است. احتمال نخست شناسایی منبع جدید توسط نرم

² . The Invisible Web

³ . Dr. Jill Ellsworth

⁴ . Matthew Koll

⁵ . Dynamic Pages

⁶ . Portals

⁷ . Subject Gateways

افزارهایی موسوم به خزنده یا عنکبوت⁸ است که بخشی مهم از موتورهای جستجو را تشکیل می دهند. این نرم افزارها پیوندهای⁹ موجود روی صفحات وب را دنبال کرده و به این ترتیب صفحات جدید را مرتباً شناسایی می کنند. این فرایند را اصطلاحاً خزیدن¹⁰ در وب می خوانند که بی وقفه تکرار می شود و صفحات جدید شناسایی شده و صفحات قبلی بازبینی می گردند و در صورت تغییر در آنها تغییرات داده شده به نمایه موتور جستجو منتقل شده و به بیان بهتر این نمایه ها روزآمد خواهند شد. این خزنده ها هر چند هم که سریع باشند نمی توانند با سرعت افزوده شدن منابع جدید و سرعت تغییرات اعمال شده در منابع قدیمی هماهنگی لازم را داشته باشند. بنابراین همواره منابعی در وب وجود دارند که هنوز نمایه نشده اند و جزء وب پنهان محسوب می شوند. به بیان دیگر تناوب روزآمدسازی نمایه های موتورهای جستجو¹¹ که ممکن است هفتگی یا حتی ماهانه باشد منجر به بروز نوعی ناپایداری در وب می شود. این بخش از محیط وب را وب مات یا وب تاریک و گاهی تحت عنوان وب تقریباً پنهان¹² نیز معرفی می کنند (پدلی، 2001). این بخش از وب پنهان که بخش قابل توجه ای را نیز به خود اختصاص می دهد دچار نوعی ناپایداری موقتی است و امکان پیوستن آن به بخش آشکار یا مرئی وب محتمل تر از بخشهای دیگر وب نامرئی است. اطلاعات موجود در وب تاریک/وب مات می تواند مورد استفاده کاربران قرار گیرد و هیچ کس تعمداً در مخفی کردن آن نداشته است اما از قلمرو موتورهای جستجو دور مانده است. از آنجا که امکان دسترسی به اطلاعات روزآمد در وب همواره یکی از امتیازات این شبکه محسوب می شود مخفی ماندن بخشی از اطلاعات تازه افزوده شده یا تازه تغییر یافته می تواند یک بحث جدی به شمار آید.

دومین روش شناسایی یک منبع جدید توسط موتورهای جستجو این است که طراح سایت یا مالک منابع مورد نظر به صورت دستی سایت خود را به موتورهای جستجو معرفی کند¹³ که این روش نیز نیازمند زمان است و همواره یک فاصله زمانی بین زمان معرفی شدن سایت جدید و ظهور آن در نتایج جستجو وجود دارد.

2. منابع اطلاعاتی در فرمتهای غیر متنی¹⁴:

موتورهای جستجو معمولاً برای نمایه سازی منابع متنی و بویژه منابع ابرمتن¹⁵ طراحی شده اند. این در حالی است که بسیاری از منابع موجود در شبکه وب به فرمتهای دیگر و معمولاً غیر متنی مثل تصویر یا منابع دیداری و شنیداری هستند و برای موتورهای کاوش امکان نمایه سازی بهینه این منابع به راحتی فراهم نیست و برای این منظور آنها همچنان به یک منبع متنی برای توصیف منابع مذکور نیازمند هستند. به این ترتیب منابع فراوانی در وب وجود دارند که برای موتورهای جستجو به سادگی قابل نمایه سازی نیستند. مطالعات و پروژه های فراوانی در مراکز تحقیقات بازیابی اطلاعات دنیا در دست اجراست که روشهایی موثرتر برای نمایه سازی این قبیل منابع توسعه یابد. اما به رغم تمام تلاشهای صورت گرفته هنوز راهکارهایی عملی برای حل این معضل در محیط وب ارائه نشده است. گرچه نتایج خوبی در سطوح آزمایشگاهی و محدود گزارش می شود اما این نتایج کاربرد عام در شبکه وب نیافته و همچنان اطلاعات غیر متنی فراوانی در وب نمایه نشده باقی مانده است و جزء وب پنهان محسوب می شود. شرمن و پرایس (2001) این بخش از وب نامرئی را اصطلاحاً "وب واقعا پنهان"¹⁶ خوانده اند و معتقد هستند که احتمال دور ماندن منابع اطلاعاتی این بخش از دسترس کاربران بیشتر از سایر منابع وب پنهان است.

3. صفحات وبی که تنها در پاسخ یک پرسش¹⁷ از نظام بازیابی تولید می شوند:

همانطور که ذکر شده مناسبترین فرمت منابع برای نمایه سازی توسط موتورهای جستجو فرمت متنی و ابر متنی ساکن¹⁸ است. پایگاههای اطلاعاتی فراوانی در وب وجود دارند که گرچه دارای صفحه جستجو

⁸ . Crawler/Spider/Robot

⁹ . Links

¹⁰ . Web Crawling

¹¹ . Frequency of Crawling

¹² . Almost Invisible Web

¹³ . Manual Submitting

¹⁴ . Non-Text Resources

¹⁵ . Hyper-Text Pages

¹⁶ . The Truly Invisible Web

¹⁷ . Query

¹⁸ . Static Hyper-Text

مبتنی بر وب هستند اما مندرجات آنها به فرمتهای دیگر ذخیره شده است و تنها زمانی یک صفحه اینترنت تولید می شود که جستجوی مستقلی در درون پایگاه مربوطه اجرا شود. پس از گسترش شبکه وب به عنوان یکی از گستردهترین محملهای اطلاع رسانی، بسیاری از پایگاههای اطلاعات الکترونیکی پیوسته منابع خود را از طریق این شبکه دسترس پذیر ساختند. اما بدیهی است که امکان نمایه سازی مندرجات این پایگاهها برای موتورهای کاوش وجود ندارد و به این ترتیب خارج از حوزه موتورهای کاوش قرار دارند و جزئی از وب پنهان محسوب می شوند که به "وب عمیق" موسوم است.

به نظر می رسد که از نظر کمی این بخش از قلمرو وب نامرئی باید بزرگترین قسمت آن را به خود اختصاص دهد بطوری که حجم اطلاعات موجود در وب عمیق چندین برابر وب آشکار یا وب سطحی تخمین زده می شود. برگمن (2001) در مطالعه پژوهشی خود حجم این بخش از وب نامرئی را بین 400 تا 550 برابر حجم محیط وب مرئی تخمین زده است. هرچند به دلیل گستردگی شگفت آور محیط وب و تغییرات مستمر و بدون وقفه این محیط می توان به هر رقم و عدد در این مورد به دیده تردید نگریست و هر رقم به زودی کهنه خواهد شد 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 پایگاههای 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 متدولوژی پژوهش برگمن (2001) گواهی بر این ادعاست) ارقام خیلی کمتر از این نیز دلالت بر حجم بسیار قابل توجهی از اطلاعات دارد که بسیار در خور توجه است.

4. منابع محافظت شده توسط اسم کاربر و گذرواژه¹⁹:

چنانچه دسترسی به یک منبع اطلاعاتی نیازمند استفاده از اسم کاربر و گذرواژه باشد نرم افزار خزنده موتور جستجو امکان دسترسی و نمایه سازی آن را ندارد. این بخش از وب پنهان را "وب خصوصی"²⁰ یا "وب شخصی" می خوانند. البته تفاوتی اساسی بین این قسمت از وب پنهان با قسمتهای دیگر وجود دارد چرا که مخفی بودن اطلاعات در این بخش کاملاً عمدی می باشد و اطلاعات مذکور جزء داراییهای افراد یا سازمانها محسوب می شود. این بخش از وب پنهان معمولاً به یکی از سه روش زیر ایجاد می شود.

نخستین شیوه حفاظت صفحات مورد نظر با کمک اسم کاربر و گذرواژه²¹ است. همانطور که گفته شد منابعی که دسترسی به آنها مستلزم استفاده از اسم کاربر و گذرواژه باشد از دسترس نرم افزارهای خزنده موتورهای جستجو خارج می مانند. دومین روش استفاده از فایل تحت عنوان Robots.txt است چرا که این فایل خزندههای موتورهای جستجو را از نمایه سازی سایت مذکور باز می دارد. سومین روش ممکن استفاده از کد noindex است به طوری که افزودن این کد به مجموعه کدهای يك صفحه وب پیامی برای خزنده موتور جستجو به شمار می رود که آن را از نمایه سازی صفحه مذکور منع خواهد کرد.

علاوه بر وب خصوصی بعضی از منابع اطلاعاتی دیگر مثل انواع نشریه های الکترونیکی²² و بانکهای اطلاعاتی مبتنی بر وب²³ که دسترسی به آنها مستلزم پرداخت حق اشتراک و در واقع خرید اطلاعات است و جزء محصولات شرکت های مختلف می باشد نیز خارج از حوزه دسترسی موتورهای کاوش هستند. این بخش از وب نامرئی را "وب ملکی"²⁴ می نامند. از نظر فنی معمولاً بین منابع وب ملکی و وب عمیق مشابهت وجود دارد اما از آنجا که در این بخش از وب پنهان منافع مالی تولید کنندگان اطلاعات منجر به خارج نمودن این منابع از حوزه دسترسی موتورهای جستجو شده این گروه در دسته ای جداگانه طبقه بندی می شوند. منابع وب ملکی اغلب از اعتبار و ارزش علمی بالایی برخوردار هستند. بسیاری از دانشگاهها و کتابخانه های دنیا سالانه مبالغ قابل توجهی برای دریافت حق اشتراک و امکان دسترسی به این منابع می پردازند.

¹⁹ . Password-Protected Pages

²⁰ . Private Web

²¹ . User name & Password

²² . E-Journals

²³ . Web-Based Databanks

²⁴ . Proprietary Web

5. صفحات وب بدون پیوند با منابع دیگر²⁵:

به رغم سهولت برقراری پیوند (لینک) بین هر دو نقطه در محیط وب هنوز صفحات فراوانی در این شبکه وجود دارند که هیچ پیوندی با هیچ منبع یا صفحه دیگری برقرار نکرده‌اند و درست 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 و طبیعتاً هرگز در نتایج جستجوی این موتورها ظاهر نخواهند شد.

6. لایه های زیرین وب سایتها که در عمقی بیشتر از عمق نمایه سازی موتورها قرار دارند:

معمولاً هر وب سایت از بیش از صفحه تشکیل شده است و صفحات متعدد دیگری در لایه های دیگر قرار دارند. بسته به نوع سایت تعداد صفحات لایه های زیرین ممکن است خیلی محدود یا بالعکس خیلی گسترده باشد و گاه به دهها و یا حتی صفحه برسد. در منابع مرتبط با موتورهای جستجو اصطلاح دیگری تحت عنوان «عمق نمایه سازی²⁶» یا عمق عمل خزنده‌ها²⁷ وجود دارد که بیانگر تعداد صفحات یک وب سایت است که توسط خزنده 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 تمام صفحات و لایه‌های موجود در همه وب سایت‌هایی را که می یابند نمایه سازی کنند. بنابراین هر یک از موتورهای جستجو بسته به میزان توان مالی خود سیاستی مشخص برای نمایه سازی اتخاذ کرده است و تنها تعداد معینی از لایه های بالایی از وبسایت‌های شناسایی شده را نمایه می‌کند. بنابراین، همواره بخشی از اطلاعات نهفته در پایین ترین لایه‌های وب سایتها از حوزه جستجوی کاربران دور می‌مانند و بنابراین این مسئله نیز دلیل دیگری است که منجر به ایجاد بخشی از وب پنهان می‌شود.

7. حداکثر صفحات قابل مرور در نتایج بازیابی²⁸

به دلیل انبوهی فراینده اطلاعات موجود در شبکه وب اغلب جستجوها در موتورهای کاوش با نتایج بازیابی فراوان همراه می‌باشد. موتورهای جستجو معمولاً موارد بازیابی شده را در صفحات متعدد نمایش می‌دهند و هر صفحه بین 10، 20 یا 30 مورد را دربرمی‌گیرد. بر اساس تحقیقات اکثر کاربران تنها صفحه اول یا دوم نتایج بازیابی را مرور نموده و به ندرت به صفحات بعدی مراجعه می‌کنند (جانسن و دیگران، 2000؛ اسپینک و دیگران، 2001). گرچه این موتورهای جستجو تلاش می کنند به کمک نظام رتبه‌بندی²⁹ موارد مرتبطتر را در ابتدای فهرست بازیابی نمایش دهند اما موضوع ربط³⁰ یک مفهوم ساده نیست و نظام بازیابی نمی‌تواند همواره براساس معیارهای از قبل تعریف شده یک مدرک را با اطمینان مرتبط یا غیرمرتبط تشخیص دهد و قضاوت نهایی در این مورد برعهده کاربر است. بنابراین حتی در بهترین شرایط نیز معمولاً بخشی از اطلاعات بازیابی شده که ممکن است مرتبط با نیاز کاربران نیز باشند در صفحات بعدی بازیابی شده توسط موتورهای جستجو مدفون مانده و به این ترتیب جزء وب پنهان محسوب می‌شود.

چرا وب پنهان مهم است؟

شاید این پرسش در ذهن شما شکل گرفته باشد که اساساً چه ضرورتی برای دسترسی به منابع وب پنهان وجود دارد؟ البته این سؤال کاملاً طبیعی است چرا که با توجه به حجم روز افزون منابع موجود در شبکه وب شاید پنهانی بخشی از آن چندان خالی در دسترس پذیری منابع ایجاد نکند. اما در واقع بخش پنهان وب از دو جهت بسیار اهمیت دارد. نخست آنکه از نظر کمی این بخش به مراتب چیزی بیش از نیمه پنهان وب است. به بیان دیگر، حجم اطلاعات نهفته در وب پنهان خیلی بیشتر از بخش سطحی³¹ یا آشکار³² آن است (برگمن،

²⁵ . Disconnected WebPages

²⁶ . Indexing Depth

²⁷ . Depth of Crawling

²⁸ . Maximum Number of Viewable Results

²⁹ . Ranking Policy

³⁰ . Relevance

³¹ . Surface

³² . Visible

2001 و پدلی، 2001). دوم آنکه از نظر کیفی نیز این قسمت بسیار حائز اهمیت می باشد. اطلاعاتی که در بخشهای مختلف این مجموعه به ویژه منابع اطلاعاتی موجود در وب عمیق و وب ملکی معمولاً منابع ارزشمند و مفیدی هستند. این امر بویژه در مورد منابع دانشگاهی و علمی اهمیت بیشتری دارد چرا که بیشتر منابع علمی معتبر دنیا در وب عمیق و وب ملکی یافت می شوند نه در نتایج بازیابی موتورهای کاوش.

بحث و نتیجه:

آنچه در این مقاله آمد مروری بود بر معرفی منابع الکترونیکی 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 به شرط آنکه کاربران این شبکه ابزار جستجو خود را به موتورهای کاوش معمولی محدود نکنند. راهکارهای متعددی برای ارتقاء سطح دسترسی پذیری اطلاعات در محیط وب وجود دارد. از آن جمله می توان به دروازه های اطلاعاتی موضوعی³³ اشاره کرد. در بسیاری از رشته های علمی دروازه های اطلاعاتی متوعی ساخته شده است. از آنجا که انتخاب منابع موجود در این سایتها توسط متخصصان موضوعی یا کتابداران انجام می شود معمولاً منابع برگزیده از کیفیت و اعتبار قابل قبولی برخوردار هستند چرا که با در نظر گرفتن معیارهای مدونی انتخاب شده اند. در مجموع مهمترین گام در این زمینه آگاهی رسانی در مورد وجود قلمرو تاریک وب است. متأسفانه بسیاری از کاربران از وجود این بخش از وب بی اطلاع هستند. شرمین و پرایس (2001b) در بیان بخشی از مشکلات کاربران در هنگام جستجو در وب می نویسند: «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 می کنند. ضمناً آنها معمولاً از این نکته مهم غفلت می کنند که شیوه عمل موتورهای جستجو با هم متفاوت است و همچنین آنچه که این موتورها جستجو می کنند با آنچه که در زمان جستجو در وب وجود دارد متفاوت است». بنابراین، گرچه منابع وب پنهان خارج از دسترس موتورهای کاوش هستند اما این بدان معنا نیست که امکان دسترسی به آنها وجود ندارد. بلکه بالعکس این منابع کاملاً جستجو پذیر می باشند. همانطور که گفته شد دروازه های اطلاعاتی متعددی در شبکه وب وجود دارد که به شکلی ساختار یافته و طبقه بندی شده به معرفی منابع وب پنهان می پردازند. از جمله معروفترین آنها می توان به سایت "راهنمای وب نامرئی"³⁴ <http://www.invisible-web.net> اشاره نمود که در برگیرنده صدها منبع ارزشمند در این زمینه است.

منابع برای مطالعه بیشتر:

وب لاگی تحت عنوان "وب لاگ وب نامرئی"³⁵ از مرداد 1382 توسط نگارنده این مقاله راه اندازی شده که اختصاصاً به موضوع وب نامرئی و مباحث مربوط به آن می پردازد. این وب لاگ به نشانی اینترنتی زیر دسترس پذیر می باشد:

<http://ciquest.shef.ac.uk/invisible/>

علاقه مندان به مطالعه بیشتر در خصوص موضوع وب نامرئی می توانند به این وب لاگ مراجعه کنند. این وب لاگ به طور متوسط هفته ای دو تا سه نوبت روزآمد می شود و حاوی مطالب متنوع و پیوندها (لینکهای) فراوان به سایتهای مرتبط به این موضوع می باشد.

مقاله ای تحت عنوان "نگاهی به وب نامرئی: مروری بر پژوهشها" در همایش وب سایت کتابخانه ها: طراحی و ارزیابی در مورخه 24 دی 1382 توسط نگارنده این مقاله ارائه شده است که متن کامل آن که در برگیرنده جزییات بیشتری در مورد وب پنهان است ان شاءالله به زودی در مجموعه مقالات همایش مذکور به چاپ خواهد رسید و در دسترس علاقه مندان قرار خواهد گرفت. ضمناً چنانچه خوانندگان محترم این مقاله علاقه مند به در اختیار داشتن منابع بیشتری در مورد جنبه های مختلف وب پنهان هستند لطفاً با نشانی نگارنده این مقاله که در صفحه نخست آمده مکاتبه نمایند.

³³ . Subject Gateway

³⁴ . The Invisible Web Directory

³⁵ . The Invisible Web Weblog

Reference:

- Bergman, M. K. (2001). "The Deep Web: Surfacing Hidden Value". The Journal of Electronic Publishing [On-line], 7 (1) [Accessed 25 January, 2004]
<http://www.press.umich.edu/jep/07-01/bergman.html>
- 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> [15 July 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.
- Dahn, M. (2000). "Spotlight on the Invisible Web". Online, **24** (4), 57-62.
- Garcia, F. (1996). "Business and Marketing on the Internet". [Online], <http://web.archive.org/web/19980113075450/http://www.tcp.ca/Jan96/BusandMark.html> [Accessed 20 November, 2003]
- 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.
- 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.
- 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.
- Sherman, C. & Price, G. (2001a) "Exploring the Invisible Web". Online, **25** (4), 32-35.
- Sherman, C. & Price, G. (2001b). "The Invisible Web: Uncovering Information Sources Search Engines Can't See. CyberAge Book.
- Sherman, C. (2003). "Invisible Web". [Online]
http://my.simmons.edu/gslis/alumni/Sherman_2003_Presentation.pdf [Accessed 2 February 2004]
- Smith, C. B. (2001). "Getting to Know the Invisible Web". Library Journal, **126** (11), 16-19.

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

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