**نام مقاله: اينترنت‌پنهان و منابع اطلاعاتي نهفته در اعماق نامرئي شبكة جهان‌گستر وب**

**نام نشريه: فصلنامه كتابداري و اطلاع رساني (اين نشريه در www.isc.gov.ir نمايه مي شود)**

**شماره نشريه: 25 \_ شماره اول ، جلد 7**

**پديدآور: يزدان منصوريان**

**مترجم:**

**چكيده**

**مقالة حاضر به تعريف مفهوم اينترنت پنهان يا وب نامرئي و همچنين عوامل مؤثر در ايجاد آن پرداخته است. اجزاي مختلف قلمرو تاريك اينترنت نظير وب عميق، وب مات، وب شخصي، وب ملكي و وب مرده به تفكيك تشريح‌شده‌اند و چگونگي ايجاد و اهميت هريك مورد بحث قرار گرفته است. اين مقاله ضمن اشاره به حجم و كيفيت قابل‌توجه منابع اطلاعاتي نهفته در اينترنت‌پنهان، بر اهميت مهارت‌هاي اطلاع‌يابي در محدودة خارج از حوزة جستجوي موتورهاي كاوش تأكيد مي‌كند. از آنجا كه اطلاعات نهفته در دل اينترنت پنهان معمولاً از كيفيت بالايي برخوردار است، اين بخش از اينترنت اهميت قابل توجهي دارد و نمي‌توان آن را ناديده گرفت.نقش كتابداران و متخصصان اطلاع‌رساني در آگاه‌سازي كاربران اينترنت از وجود وب نامرئي و درنتيجه ارتقاي سطح دسترس‌پذيري اطلاعات براي آنان و همچنين نقشي كه كتابداران مي‌توانند در ساخت دروازه‌هاي اطلاعاتي ايفا كنند، از ديگر مباحث مطرح‌شده در اين نوشتار است.**

**كليدواژه‌ها: اينترنت پنهان، وب پنهان، وب نامرئي، وب عميق، وب مات، بازيابي اطلاعات در وب**

**نگاهي كوتاه بر مشكلات جستجو و بازيابي در اينترنت**

**جستجو و بازيابي اطلاعات در اينترنت فرايندي پيچيده و چندبُعدي است كه مي‌توان آن را از زواياي مختلف مورد توجه و بررسي قرار داد. تاكنون نيز پژوهش‌هاي متعددي دربارة جنبه‌هاي مختلف كاوش و بازيابي اطلاعات در اين محيط صورت گرفته است. در سال‌هاي اخير اين موضوع همچنان جزء زمينه‌هاي مهم پژوهشي در حوزة بازيابي اطلاعات و مطالعات اطلاع‌يابي محسوب شده. به‌رغم تفاوت‌هاي موجود در رويكرد اين پژوهش‌ها، محور اصلي همة آن‌ها يافتن راهكارهايي عملي براي ارتقاي سطح دسترس‌پذيري اطلاعات در اين شبكه و سهولت بازيابي است. (e.g. Jansen, 2000; Jansen & Spink, 2000; Wany et.al, 2000; Palmpuist, 2000; Kim, 2001, Cothey, 20002; Hargittai, 2002; Ford et al, 2002; Ford et al, 2003)**

**موتورهاي جستجو اصلي‌ترين و معمولي‌ترين ابزار كاوش در اينترنت به شمار مي‌آيند و معمولاً اغلب كاربران از اين موتورها براي بازيابي اطلاعات اينترنتي استفاده مي‌كنند.(Gordon & Pathak, 1999) اهميت و جايگاه اين ابزارهاي كاوش به‌حدي است كه بدون آن‌ها عملاً، جستجو در اين شبكة عظيم اگر ناممكن نباشد، قطعاً بسيار دشوار خواهد بود.**

**با اين حال و با تمام ويژگي‌هاي ارزندة اين موتورها و كاربرد گستردة آن‌ها در فرايند بازيابي اطلاعات، مشكلات متعددي در پيوند با كارآيي و سودمندي آن‌ها وجود دارد. پژوهش‌هاي گذشته حاكي از آن است كه كاربران اينترنت با دشواري‌هايي در استفاده از موتورهاي جستجو مواجه هستند.(Gorden & Pathak, 1999; Chu & Rojenthal, 1996) آن‌ها گاهي براي يافتن اطلاعات مورد نياز خود ساعت‌ها وقت صرف مي‌كنند و ممكن است اين تلاش حاصلي هم درپي نداشته باشد. دلايل متعددي براي دشواري‌هاي جستجو در اينترنت وجود دارند كه تاكنون موضوع پژوهش‌هاي فراواني بوده‌اند. در يك طبقه‌بندي كلي، بخشي از اين دشواري‌ها به حجم عظيم و روزافزون اطلاعات موجود در اينترنت مربوط مي‌شود كه معمولاً نتايج بازيابي موتورهاي جستجو را با ريزش كاذب فراوان همراه مي‌سازد. بخش ديگر به ميزان مهارت‌هاي اطلاع‌يابي و تجربة كاربران در جستجو بازمي‌گردد. گاهي نيز ناكامي در جستجو ريشه در نوع نياز اطلاعاتي كاربران دارد كه ممكن است گسترده و مبهم باشد و كاربر نتواند آن را به شكل مناسبي در قالب يك راهبرد جستجوي كارآمد تعريف كند.**

**هريك از اين احتمالات خود زمينة گسترده‌اي براي پژوهش و مطالعه به‌شمار مي‌آيند و تشريح و تبيين آن‌ها نيازمند مطالعات مستقل و مفصلي است كه اين مقاله قصد پرداختن به آن‌ها را ندارد. آنچه كه محور اصلي اين نوشته را تشكيل مي‌دهد موضوع ديگري است كه تاكنون چندان مورد توجه كاربران اينترنت و پژوهشگران حوزة بازيابي اطلاعات در وب قرار نگرفته و بي‌توجهي به آن مي‌تواند به طور بالقوه در ناكامي جستجو در وب مؤثر باشد. محور اصلي اين مقاله پديده‌اي تحت عنوان اينترنت پنهان[1] يا وب نامرئي[2] مي‌باشد كه در ادامه معنا و مفهوم آن تشريح مي‌شود و به عوامل ايجادكنندة آن و اهميت آن در فرايند بازيابي اطلاعات در اينترنت اشاره مي‌گردد.**

**اينترنت پنهان چيست**

**واژة اينترنت پنهان يا وب نامرئي اشاره به يك مفهوم چند‌بُعدي و گسترده دارد، به شكلي كه ارائة تعريفي گويا و كوتاه از آن را كمي دشوار مي‌كند. به بيان ديگر، ارائة يك تعريف جامع و مانع از اينترنت پنهان مستلزم تبيين بخش‌هاي مختلف آن است. با اين حال در ساده‌ترين شكل مي‌توان اين بخش پنهان از اينترنت را به‌طور خلاصه اين چنين تعريف نمود:**

**اينترنت پنهان آن بخش از فضاي اينترنت است كه به هر دليل، خارج از حوزة جستجوي موتورهاي كاوش قرار دارد و بازيابي اطلاعات موجود در آن از طريق استفادة مستقيم از اين موتورها ميسر نيست (Pedley, 2001; Price & Sherman, 2001). افزودن قيد «به‌هردليل» در اين تعريف حاكي از آن است كه ايجاد اينترنت پنهان ناشي از وجود فقط يك عامل نيست و به عوامل متعددي مربوط مي‌شود. تأثير همين عوامل متعدد منجر به ايجاد انواع مختلف از پنهان‌شدگي اطلاعات در اينترنت گرديده كه در بخش‌هاي بعدي مقاله تشريح خواهند شد.**

**ذكر اين نكته ضروري است كه واژه‌هاي موازي ديگري نظير وب نامرئي، وب پنهان، وب عميق و وب تاريك يا وب مات[3] براي اين بخش از اينترنت در متون مختلف مشاهده مي‌شود. گرچه اين اصطلاحات گاهي به‌صورت مترادف به‌كار برده مي‌شوند، اما در حقيقت معادل يكديگر نيستند و هريك اشاره به نوع خاصي از ناپيدايي[4] در وب دارند.**

**از آنجا كه ارتباط مستقيمي بين شيوة كار موتورهاي جستجو و ايجاد اينترنت پنهان، يا به بيان دقيقتر بروز پنهاني اطلاعات در اينترنت وجود دارد، به‌نظر مي‌رسد قبل از توصيف بخش‌هاي مختلف اينترنت پنهان لازم باشد به نحوة عملكرد موتورهاي جستجو اشاره شود.**

**اجزاي تشكيل‌دهندة موتورهاي جستجو و چگونگي عملكرد آن‌ها**

**به‌رغم وجود گوناگوني و تنوع فراوان در موتورهاي جستجو، اغلب آن‌ها از ويژگي‌هاي مشتركي برخوردارند. ساختار اصلي موتور جستجو معمولاً از سه بخش عمده تشكيل شده. اين قسمت‌ها عبارت‌اند از:**

**1. نرم‌افزار خزنده يا روبات جستجوگر:[5] اين قسمت از موتور جستجو در واقع نرم‌افزاري است كه با دنبال‌كردن پيوند[6]هاي موجود در صفحات وب، صفحات جديد را شناسايي مي‌كند و اين فرايند، پيوسته تكرار مي‌شود. انتخاب نام «خزنده» براي اين نوع از نرم‌افزارها به‌دليل نوع كاري است كه انجام مي‌دهند. عملكرد آن‌ها در يافتن صفحات وب مثل خزيدن در فضاي وب است. هر موتور جستجو، نرم‌افزار خزندة مخصوص به خود را دارد و توانايي بازيابي اطلاعات در هر موتور، كاملاً وابسته به قدرت «خزندة» آن است. منظور از قدرت خزنده، سرعت عمل و دامنة فعاليت آن در دنبال‌كردن پيوندهاي موجود مي‌باشد. گستردگي و عمق دسترسي اطلاعات در هر موتور جستجو بيش از هر چيز به ويژگي‌هاي خزندة آن بستگي دارد و نماية[7] موتور جستجو، حاصل كار خزنده است. البته علاوه بر خزنده‌ها در هر موتور جستجو، بخشي براي معرفي سايت‌هاي جديد نيز وجود دارد كه كاربران مي‌توانند با ارسال نشاني[8] سايت مورد نظر خود، آن را به موتور جستجو تحويل[9] دهند.**

**2. نمايه: صفحات بازيابي‌شده در وب از طريق عملكرد خزنده، در يك پايگاه اطلاعاتي تحت عنوان «نماية موتور جستجو» (و درواقع يك فايل بزرگ) ذخيره مي‌شوند. بزرگي و روزآمدي اطلاعات موجود در نماية هر موتور جستجو وابسته به عملكرد خزندة آن است. ساختار نمايه، اندازه و حجم آن در موتورهاي جستجوي مختلف متفاوت است. به همين دليل جستجو با كليدواژه‌هاي يكسان، نتايج نسبتاً متفاوتي در موتورهاي گوناگون درپي خواهد داشت.**

**3. نرم‌افزار جستجو در نمايه:[10] سومين جزء يك موتور جستجو، نرم‌افزار بازيابي است كه به جستجو و بازيابي اطلاعات موجود در فايل نمايه براساس كليدواژه‌هاي دريافتي از طرف كاربران مي‌پردازد و نتايج بازيابي را نمايش مي‌دهد. به بيان ديگر، هنگامي كه كاربران كليدواژه‌هاي مورد نظر خود را در بخش جستجوي موتور كاوش وارد مي‌كنند، آنچه اتفاق مي‌افتد جستجوي نماية موتور بر اساس كليدواژه‌هاي مذكور است نه جستجوي فضاي اينترنت. آنچه كه كاربران مشاهده مي‌كنند بيروني‌ترين لاية موتورهاي جستجو است. وقتي آن‌ها به كمك كليدواژه‌هاي انتخابي خود به جستجو در اينترنت مي‌پردازند، درواقع آنچه را جستجو مي‌كنند پايگاه اطلاعاتي (نماية) موتور جستجو است كه پيوند آن‌ها را با ساير سايت‌هاي اينترنتي ـ سايت‌ها و صفحات وب كه قبلاً توسط خزنده شناسايي شده‌اند‌ـ برقرار مي‌كنند.**

**بسته به اين‌كه خزنده‌هاي موتور جستجو با چه تناوب[11] و چه عمقي[12] عمل كنند نتيجة جستجو توسط كاربر متفاوت خواهد بود، چرا كه به‌دليل تفاوت تناوب و عمق عملكرد خزندة موتورهاي كاوش گوناگون، حجم و اطلاعات ذخيره‌شده در نماية آن‌ها با هم متفاوت مي‌باشد و درنتيجه حاصل جستجو در هر يك از آن‌ها متفاوت خواهد بود‌ـ هرچند همواره درصدي از همپوشاني بين نتايج بازيابي وجود دارد.**

**قسمت‌هاي مختلف اينترنت پنهان**

**همانطور كه گفته شد موتورهاي جستجو براساس عملكرد نرم‌افزارهاي خزندة خود و همچنين سايت‌هايي كه به آن‌ها معرفي مي‌شوند فايل يا پايگاه اطلاعاتي خود را مي‌سازند و از اين طريق امكان جستجو در اينترنت را براي كاربران فراهم مي‌سازند. پايگاه اطلاعاتي موتورهاي جستجو مرتباً روزآمد مي‌گردند و سايت‌هاي جديد به آن‌ها اضافه مي‌شوند و مرتباً بر حجم آن‌ها افزوده مي‌گردد. بسته به فاصلة روزآمدسازي اين پايگاه‌ها، اندازة آن‌ها در موتورهاي مختلف متفاوت است. به‌دليل گستردگي و پويايي محيط وب عملاً هيچ موتور جستجويي قادر به پوشاندن تمام وب نيست و همواره بخش‌هاي مختلف يا به بيان بهتر منابع اطلاعاتي متعددي در اينترنت وجود دارند كه در اين پايگاه‌ها گنجانده نمي‌شوند. پس از بيان اين مقدمة كوتاه اكنون مي‌توان بخش‌هاي مختلف اينترنت پنهان را به‌شرح زير خلاصه كرد:**

**1. وب تاريك يا وب مات**

**اين بخش از اينترنت پنهان، بزرگترين قسمت اين فضاي گسترده را تشكيل مي‌دهد. انتخاب عنوان «وب تاريك» براي اين قسمت به خاطر اين است كه گويي منابع اطلاعاتي موجود در آن در زير سايه‌اي قرار گرفته‌اند و از دسترسي كاربران خارج شده‌اند. به بيان ديگر، اطلاعات موجود در وب تاريك مي‌تواند مورد استفادة عموم قرار گيرد و هيچ‌كس تعمدي در مخفي‌كردن آن نداشته، اما به‌دليل تأثير عواملي كه در ادامه ذكر مي‌گردد، عملاً اطلاعاتي كه مي‌توانسته توسط موتورهاي جستجو بازيابي شود از حوزه و قلمرو موتورهاي جستجو دور مانده است. اجزاي تشكيل‌دهندة وب تاريك عبارت‌اند از:**

**1ـ1. وب مرده يا نشاني‌هاي قطع‌شده:[13] محيط وب فضايي كاملاً پويا، متغير يا به بيان بهتر ناپايدار است. هر روز صفحات و منابع مختلفي به اين مجموعة عظيم افزوده مي‌شوند و منابع ديگري از آن حذف مي‌گردند. از سويي ديگر، صفحاتي در وب وجود دارند كه هيچ پيوندي با منابع ديگر برقرار نكرده‌اند و به همين دليل خزنده‌هاي موتورهاي جستجو قادر به يافتن آن‌ها نيستند. در صورتي كه اين صفحات به موتورهاي جستجو معرفي نگردند امكان شناسايي آن‌ها براي موتور جستجو وجود ندارد. طبق پژوهشي كه در سال 2000 انجام شده اين مجموعة جدامانده از محيط جستجوپذير وب حدود 20 درصد كل فضايي است كه مي‌تواند توسط موتورهاي جستجو نمايه شود.(Broder, et al. 2000) البته از آنجا كه به‌دليل گستردگي بي‌وقفة اطلاعات موجود در شبكة جهانگستر وب هرگونه برآوردي‌ـ حتي نسبي ـ از حد و مرز اين محيط خيلي زود كهنه و قديمي مي‌شود، اين رقم مي‌تواند بيشتر يا كمتر از 20 درصد باشد. آنچه كه در اينجا بيش از صحت يا نادرستي اين رقم اهميت دارد، واقعيتي است كه هم‌اكنون در وب به‌وجود آمده و حجم عظيمي از اطلاعات را براي كاربران، دسترس‌ناپذير ساخته است.**

**1ـ2. تناوب روزآمد سازي نماية موتورهاي جستجو: همانطور كه گفته ‌شد نرم‌افزارهاي خزنده مسئوليت شناسايي صفحات جديد و افزودن آن‌ها به نماية موتورهاي جستجو را برعهده دارند. توان اين نرم‌افزارها محدودتر از آن است كه بتوانند تمام محيط وب را پوشش دهند و تناوب روزآمدسازي آن‌ها كندتر از سرعت افزودن اطلاعات جديد به محيط وب است. به بيان ساده‌تر، نرم‌افزارهاي خزنده فرصت كافي براي حركت هماهنگ با صفحات جديد را ندارند و بنابراين همواره صفحات جديدي به محيط وب افزوده مي‌شوند كه هنوز مورد شناسايي اين خزنده‌ها واقع نشده‌اند و به‌همين‌دليل از حوزة جستجوي موتورها دور مي‌مانند. بنابراين، حتي به فرض آن‌كه يك موتور جستجو قادر به گسترش حوزة كاوش خود به تمام محيط وب باشد (كه البته هنوز هيچ موتوري اين توانايي يا ادعا را ندارد)، همواره فاصله‌اي بين اطلاعات واقعي موجود در وب و آنچه كه موتور مذكور جستجو مي‌كند وجود خواهد داشت. درواقع موتورهاي جستجو هيچ‌گاه نمي‌توانند پابه‌پاي سرعت افزايش اطلاعات به محيط وب حركت كنند و هميشه چندين گام عقب‌تر هستند. بنابراين اختلاف سرعت بين افزودن اطلاعات به محيط وب و شناسايي اطلاعات جديد توسط خزنده‌ها، همواره بخشي از اطلاعات موجود در وب را در تاريكي وب پنهان قرار مي‌دهد.**

**1ـ3. عمق نمايه‌سازي:[14] در متون مرتبط با موتورهاي جستجو اصطلاح ديگري تحت عنوان «عمق نمايه‌سازي» يا عمق عمل خزنده‌ها وجود دارد كه بيانگر چگونگي و عمق نمايه‌سازي صفحات وب توسط موتورهاي جستجو است. توان مالي و فناوري بسياري از موتورهاي جستجو هنوز محدودتر از آن است كه بتوانند تمام صفحات و لايه‌هاي موجود در وب‌سايت‌ها را نمايه‌سازي كنند، چرا كه اين كار فرايندي هزينه‌بر و دشوار است. بنابراين هر يك از موتورهاي جستجو بسته به ميزان توان خود و سياستي كه براي نمايه‌سازي اتخاذ كرده، تنها بخشي از وب‌سايت‌هاي شناسايي‌شده را نمايه مي‌كند. منظور از «بخشي از وب‌سايت‌ها» درواقع صفحات دروني و لايه‌هاي زيرين هر وب‌سايت است. بنابراين، همواره بخشي از اطلاعات نهفته در پايين‌ترين لايه‌هاي وب‌سايت‌ها از حوزة جستجوي كاربران دور مي‌مانند. پس اين مسئله نيز دليل ديگري است كه منجر به ايجاد بخشي از اينترنت پنهان مي‌شود.**

**1ـ4. حداكثر صفحات قابل مرور در نتايج بازيابي:[15] به‌دليل انبوهي فزاينده اطلاعات وب معمولاً اغلب جستجوها با نتايج فراوان در بازيابي همراه‌اند. موتورهاي جستجو اطلاعات بازيابي‌شده را در صفحات متعدد نمايش مي‌دهند و هر صفحه بين 10، 20 يا 30 مورد را دربرمي‌گيرد. البته كاربران مي‌توانند بسته به سليقه يا نياز خود اين تعداد را در تنظيمات موتورهاي جستجو كم يا زياد كنند.**

**پژوهش‌هاي گذشته نشان مي‌دهند كه اكثر كاربران تنها صفحات اول يا دوم نتايج بازيابي موتورهاي جستجو را مرور مي‌كنند و به هردليل به‌ندرت به صفحات بعدي مراجعه مي‌كنند (Jansen et al. 2000; Spink et al, 2001). گرچه موتورهاي جستجو براساس نظام رتبه‌بندي[16] خود سعي مي‌كنند اطلاعات مرتبط‌تر را در ابتداي فهرست بازيابي نمايش دهند اما موضوع ربط[17] يك مفهوم ساده نيست و نمي‌توان به راحتي نسبت به مربوط‌بودن يا نبودن اطلاعات مندرج در يك مدرك قضاوت كرد. بويژه آن‌كه همواره قضاوت نهايي دربارة ربط برعهدة كاربر است و سيستم نمي‌‌تواند براساس معيارهاي تعريف‌شده، اطلاعاتي را صددرصد مرتبط يا نامرتبط بداند. موضوع ربط بحث بسيار مفصل و پيچيده‌اي است كه پرداختن به جزئيات آن خارج از محدودة اين مقاله است. آنچه در حوزة بحث اين مقاله مي‌گنجد بيان اين واقعيت است كه حتي در بهترين شرايط همواره بخشي از اطلاعات بازيابي‌شده كه ممكن است مرتبط با نياز كاربر نيز باشد در لايه‌هاي زيرين صفحات بازيابي‌شده توسط موتورهاي جستجو مدفون مي‌ماند و جزء وب پنهان محسوب مي‌شود.**

**2. وب عميق**

**پس از گسترش و فراگيري شبكة جهانگستر وب به‌عنوان يكي از گسترده‌ترين و مؤثرترين محمل‌هاي اطلاع‌رساني، بسياري از پايگاه‌هاي اطلاعات الكترونيكي پيوسته[18] منابع خود را از طريق اين شبكه دسترس‌پذير ساختند. اين منابع اطلاعاتي شامل پايگاه‌هايي بودند كه قبل از پيدايش وب نيز وجود داشتند و خدمات اطلاع‌رساني ارزنده‌اي ارائه‌مي‌كردند. اكنون هزاران مورد از اين پايگاه‌هاي كوچك و بزرگ، از جمله فهرست پيوستة عمومي[19] بسياري از كتابخانه‌هاي دنيا از طريق شبكة وب دسترس‌پذير هستند. هر يك از آن‌ها نظام بازيابي اطلاعات خاص خود را دارند و از طريق صفحات جستجوي مبتني بر وب، اطلاعات و مدارك موجود در پايگاه خود را به كاربران عرضه مي‌كنند. بعضي از اين پايگاه‌ها و فهرست‌هاي پيوسته، خدمات خود را رايگان عرضه مي‌دارند و برخي ديگر در برابر ارائة خدمات، هزينة اشتراك دريافت مي‌كنند. مثلاً اطلاعات موجود در پايگاه اطلاعات علوم تربيتي «اريك»[20] به رايگان دسترس‌پذير است، ولي در مقابل، بسياري از مجله‌هاي الكترونيكي و پايگاه‌هاي اطلاعاتي (مثل «ديالوگ»[21]) هزينه‌هاي اشتراك قابل‌توجهي دريافت مي‌كنند.**

**فارغ از رايگان بودن يا نبودن خدمات اين پايگاه‌ها، همة آن‌ها در يك موضوع مشترك‌اند و آن، خارج‌بودن مندرجات آن‌ها از حوزة جستجوي موتورهاي جستجو است. هريك از اين پايگاه‌ها يك صفحة جستجوي مبتني بر وب دارند كه امكان جستجو در آن‌ها را براي كاربران فراهم مي‌كند. نتيجة جستجوها به صورت صفحات دسترس‌پذير از طريق شبكة وب به نمايش درمي‌آيد. اما هريك از اين صفحات، نتيجة جستجوي مشخص با كليدواژه‌هايي هستند كه توسط كاربران به نظام بازيابي وارد شده‌اند. بنابراين خزنده‌هاي موتورهاي جستجو توانايي نمايه‌كردن مندرجات اين پايگاه‌ها را ندارند و درواقع همواره در پشت مجموعه‌اي عظيم از اطلاعات متوقف شده‌اند و به درون پايگاه راهي ندارند. (Snow, 2000)**

**به بيان ديگر، در فضاي گستردة وب، منابع اطلاعاتي ارزشمندي در عمق اين درياي اطلاعات و درون پايگاه‌هاي اطلاعاتي وجود دارند كه از دسترس كاربران موتورهاي جستجو دور مي‌مانند. مثلاً اگر فردي در جستجوي اطلاعاتي دربارة يك موضوع تخصصي در زمينة كشاورزي باشد و جستجوي خود را تنها به موتورهاي جستجو محدود كند، درواقع بخشي عظيم از اطلاعات مرتبط را از دست داده است. بي‌ترديد يا به بيان محتاطانه‌تر، به احتمال زياد، اطلاعات مرتبط و مفيدي در زمينة مذكور در اينترنت وجود دارد اما نهفته در دل پايگاه‌هاي اطلاعاتي متعددي مي‌باشد كه دسترسي به مندرجات آن‌ها فقط با مراجعة مستقيم به پايگاه‌هاي اطلاعاتي مذكور ميسر است نه با جستجو از طريق موتورهاي جستجوي عمومي.[22]**

**كاربري كه درصدد بازيابي چنين اطلاعاتي است بايد ابتدا از طريق موتورهاي جستجو پايگاه‌هاي مرتبط با موضوع موردنظر خود را شناسايي‌كند و سپس جداگانه به جستجو در آن‌ها بپردازد. درغيراين‌صورت، با جستجو از طريق موتورهاي جستجو حوزة كاوش خود را به وب سطحي[23] محدود كرده و از دسترسي به وب عميق بازمانده است.**

**3. وب خصوصي[24] و وب ملكي[25]**

**گرچه اين بخش از اينترنت پنهان تفاوت اساسي با بخش‌هاي ديگر دارد و به نوعي خارج از بحث اصلي اين مقاله است، اما اشارة اجمالي به آن ضروري به‌نظر مي‌رسد. اشاره به وب خصوصي و وب ملكي از اين جهت ضروري است كه انواع ديگري از پنهان‌شدگي اطلاعات در محيط اينترنت نيز معرفي شود و تفاوت‌هاي موجود تشريح گردد.**

**همان‌طوركه گفته شد، تفاوتي اساسي بين اين قسمت از اينترنت پنهان با قسمتهاي ديگر وجود دارد. اين تفاوت اساسي از آنجا ناشي مي‌شود كه مخفي‌بودن اطلاعات در اين بخش، كاملاً تعمدي مي‌باشد و اطلاعات مذكور جزء دارايي‌هاي شخصي و خصوصي افراد يا سازمان‌ها محسوب مي‌شود. بي‌‌ترديد هر شخص حقيقي يا حقوقي اين حق را دارد كه اطلاعات مربوط به خود را كه خارج از مسائل عمومي است، به نحوي محافظت نمايد.**

**مثلاً سازمان‌هاي مختلف صفحات و اطلاعاتي در اينترنت دارند كه تنها به مسائل كاري آن سازمان و كاركنان آن مربوط مي‌شود و حفاظت از آن ضرورتي مسلم است. اين بخش از اينترنت پنهان از حوزة دسترسي موتورهاي جستجو خارج است و طبيعتاً هم بايد خارج باشد. منابع اطلاعاتي موجود در وب را كه مربوط به مسائل شخصي و خصوصي افراد حقيقي يا حقوقي مي‌باشد و به روش‌هاي مختلفي كه در ادامه ذكر مي‌شود از دسترس ساير كاربران خارج است، وب خصوصي يا وب شخصي مي‌نامند.**

**همچنين بعضي از منابع اطلاعاتي مثل انواع نشريه‌هاي الكترونيكي و بانك‌هاي اطلاعاتي مبتني بر وب را كه دسترسي به آن‌ها مستلزم پرداخت حق اشتراك و درواقع خريد اطلاعات است و جزء محصولات شركت‌هاي مختلف مي‌باشد، را وب ملكي مي‌نامند.**

**در اينجا فقط به انواع سازوكارهايي كه مي‌تواند يك صفحه يا منبع اطلاعاتي را از حوزة وب عمومي خارج كند و جزء وب شخصي يا وب ملكي قرار دهد اشاره مي‌گردد. اين بخش معمولاً به يكي از سه روش زير ايجاد مي‌شود:**

**3ـ1. استفاده از اسم كاربر[26] و گذرواژه[27]: صفحاتي كه دسترسي به آن‌ها مستلزم استفاده از اسم كاربر و گذرواژه باشد از حوزة جستجوي موتورهاي كاوش خارج هستند.**

**3ـ2. استفاده از فايل Robots.txt : وجود اين فايل، خزنده‌هاي موتورهاي جستجو را از نمايه‌سازي وب سايت مذكور بازمي‌دارد.**

**3ـ3. استفاده از كد noindex : افزودن كد noindex به مجموعة كدهاي يك صفحة وب، پيامي براي خزندة موتور جستجو به‌شمار مي‌رود كه آن را از نمايه‌سازي صفحة مذكور منع مي‌كند و معمولاً موتورهاي جستجو اين مسئله را رعايت مي‌كنند و از افزودن صفحات محتوي كد noindex به نماية خود خودداري مي‌كنند. بنابراين، اين كد نيز مي‌تواند صفحات معيني را از حوزة جستجوي موتورهاي جستجو خارج كند.**

**در اينجا باز هم تأكيد مي‌شود كه اين بخش خاص از اينترنت پنهان، تفاوت بنيادي با بخش‌هاي ديگر قلمرو تاريك اينترنت دارد، چرا كه اين قسمت آگاهانه از دسترسي موتورهاي جستجو خارج شده؛ در حالي‌كه بخش‌هاي ديگر اينترنت پنهان مي‌توانسته جزء قلمرو بازيابي موتورهاي جستجو باشد و عوامل ناخواسته‌اي، آن‌ها را خارج از دسترسي كاربران قرار داده است.**

**4. اينترنت واقعاً پنهان[28]**

**آنچه تاكنون دربارة اينترنت ذكر شد اشاره به بخش‌هايي از اينترنت داشت كه يا به‌دليل عوامل غيرفني ايجاد شده بود يا بنا به خواست و تمايل صاحبان منابع اطلاعات. منظور از عوامل غيرفني عواملي است كه مثلاً در وب مات به آن‌ها اشاره شد (مثل حداكثر صفحات قابل رويت در نتايج بازيابي). اما بخش‌هاي ديگري نيز در قلمرو تاريك وب وجود دارند كه در اثر مسائل فني و تكنيكي، يا به بيان ديگر در اثر ناكارآمدي ابزارهاي جستجو، از دسترسي كاربران دور مانده‌اند.**

**همان‌طور كه قبلاً گفته شد موتورهاي جستجو اغلب قادر به يافتن اطلاعات متني «اچ‌تي‌ام‌ال»[29] هستند و توانايي بازيابي فايل‌هايي از انواع ديگر نظير swf,، pppيا pdf را يا ندارند يا به‌دليل كمبود منابع مالي و فني از جستجوي فايل‌هاي غيرمتني صرف‌نظر كرده‌اند. مثلاً درحال‌حاضر موتور جستجوي معروف «گوگل»[30] قادر به بازيابي فايل‌هاي با پسوند ptt, و pdf مي‌باشد، اما اين امكان در بسياري ديگر از موتورهاي جستجو هنوز وجود ندارد و چند سال پيش نيز «گوگل» به اين امكان مجهز نبود.**

**بنابراين منابع اطلاعاتي متنوعي در وب وجود دارند كه تنها به‌دليل محدوديت تكنولوژيكي يا مالي موتورهاي جستجو، از حوزة كاوش آن‌ها و درنتيجه از دسترسي كاربران دور مانده‌اند.**

**اهميت اينترنت پنهان**

**بخش پنهان اينترنت از دو جهت اهميت دارد. نخست آن‌كه از نظر كم‍ّي اين بخش چيزي بيش از نيمة پنهان اينترنت است. به بيان ديگر، حجم اطلاعات نهفته در اينترنت پنهان خيلي بيشتر از بخش سطحي[31] يا آشكار[32] است. دوم آن‌كه از نظر كيفي نيز اين قسمت بسيار داراي اهميت است. اطلاعات موجود در بخش‌هاي مختلف اين مجموعه بويژه منابع اطلاعاتي موجود در وب‌عميق معمولاً منابع ارزشمند و مفيدي هستند و مي‌توانند در بسياري از موارد پاسخگوي نياز كاربران باشند. بنابراين، نمي‌توان بخش پنهان اينترنت را ناديده گرفت، چرا كه كاربران مي‌توانند با روش‌هاي مختلف سطح ناپيدايي در وب را در جستجوهاي خود كاهش دهند و از اين طريق، بهره‌وري كاوش خود را ارتقا بخشند.**

**يكي از زمينه‌هاي فعاليت كتابداران و متخصصان اطلاع‌رساني، آگاه‌نمودن كاربران از وجود اينترنت پنهان است تا به اين ترتيب، آنان فرايند جستجوي خود در محيط اينترنت را فقط به نتايج بازيابي موتورهاي جستجو محدود نسازند و نتايج بهتري كسب كنند.**

**اطلاع‌يابي در اينترنت پنهان**

**راهكارهاي مختلفي براي ارتقاي سطح دسترس‌پذيري اطلاعات در اينترنت پنهان وجود دارد. مهم‌ترين گام در اين زمينه، آگاهي‌رساني در مورد وجود قلمرو تاريك وب است. متأسفانه بسياري از كاربران از وجود اين بخش از اينترنت بي‌اطلاع هستند. «شرمن» و «پرايس» در بيان بخشي از مشكلات كاربران در هنگام جستجو در وب مي‌نويسند: «بسياري از كاربران بر اين باورند كه تمام اطلاعات موجود در وب در حوزة دسترسي موتورهاي جستجو قرار دارد و همة اين موتورها نيز مجموعة يكساني از وب را كاوش مي‌كنند. درضمن معمولاً از اين نكتة مهم غافل‌اند كه شيوة عمل موتورهاي جستجو با هم متفاوت است و نيز آنچه كه اين موتورها جستجو مي‌كنند با آنچه كه در زمان جستجو در وب وجود دارد متفاوت است» (Price & Sherman, 2001 (b)).**

**اگر كاربران به اين مسائل آگاهي داشته باشند، مي‌توانند جستجوهاي دقيق‌تر و مطمئن‌تري انجام دهند. تاكنون تدابير متعددي براي بهبود سطح بهره‌وري جستجو در اينترنت پيش‌بيني شده. استفاده از اين تدابير مي‌تواند به‌عنوان مكمل جستجو از طريق موتورهاي كاوش، ميزان بازيابي اطلاعات مرتبط مورد نياز را به نحو قابل‌توجهي ارتقا بخشد. از جملة اين امكانات مي‌توان به دروازه‌هاي اطلاعاتي موضوعي[33] اشاره كرد.**

**تاكنون در بسياري از رشته‌هاي علمي، دروازه‌هاي اطلاعاتي متنوعي طراحي و ساخته شده‌اند. از آنجا كه انتخاب منابع موجود در اين سايت‌ها توسط متخصصان موضوعي يا كتابداران انجام مي‌شود، معمولاً منابع برگزيده از كيفيت و اعتبار خوبي برخوردار هستند.**

**روش ديگر، استفاده از اَبرموتورهاي كاوش[34] است. اين ابرموتورها خود موتور جستجوي واقعي نيستند بلكه به كاربران اين امكان را مي‌دهند كه كليدواژه‌هاي خود را همزمان توسط چند موتور جستجو مورد كاوش قرار دهند و نتايج جستجوي تمام آن‌ها را يكجا در اختيار داشته باشند.**

**امكان ديگري كه اخيراً موردتوجه متخصصان بازيابي اطلاعات در وب قرار گرفته، استفاده از عوامل هوشمند[35] است. اين عوامل هوشمند كه توانايي جستجو، مقايسه و انتخاب منابع اطلاعاتي بر اساس نياز مطرح‌شده توسط كاربر را دارند، به‌عنوان كارگزار كاربران در اين فضاي گسترده به جستو مي‌پردازند و نتايج جستجو را بر اين اساس فيلتر مي‌كنند و به آن‌ها تحويل مي‌دهند.**

**به‌طوركلي پيش‌بيني مي‌شود با پيشرفت ابزارهاي بازيابي اطلاعات در وب و بهبود سواد اطلاعاتي كاربران، به‌تدريج از سطح ناپيدايي[36] اطلاعات در محيط وب كاسته شود.**

**نقش آموزش كتابداران در كاهش سطح پنهاني اطلاعات**

**همان‌طوركه گفته شد بخش عمده‌اي از اطلاعات مخفي‌مانده در وب پنهان را مي‌توان به راحتي براي كاربران اين شبكه دسترس‌پذير نمود. مهمترين راهكار براي نفوذ به قلمرو تاريك وب، آگاهي از وجود آن و آگاهي از روش‌هاي اطلاع‌يابي در آن است. بي‌ترديد اگر كاربران اينترنت تصوير روشني از نقاط قوت و ضعف موتورهاي جستجو و توانايي‌ها و محدوديت‌هاي آن‌ها داشته باشند، بهتر مي‌توانند از اين ابزارهاي كاوش استفاده كنند.**

**پنهاني اطلاعات در اينترنت يك موضوع با حد و مرزهاي مشخص و معلوم نيست. درواقع نمي‌توان اين شبكه را به دو بخش نهان و آشكار يا مرئي و نامرئي تقسيم نمود و محدوده و مرز مشخصي بين اين دو قسمت قائل شد. ممكن است يك منبع اطلاعاتي براي يك فرد يا در يك تجربة جستجوي مشخص، جزء قلمرو وب پنهان قرار گيرد و در همان زمان براي كاربر ديگري كاملاً دسترس‌پذير باشد.**

**بنابراين تأكيد بر اين نكته ضروري است كه نمي‌توان شبكة وب را به دو بخش پنهان و آشكار يا روشن و تاريك يا سطحي و عميق تفكيك كرد. آنچه اهميت دارد آگاه‌نمودن كاربران به اين واقعيت است كه جستجو در اينترنت فرايندي فراتر از به‌كارگيري از موتورهاي جستجو مي‌باشد. حد و مرز و پيچيدگي منابع اطلاعاتي موجود در اين شبكه، فعلاً بسي فراتر از حوزة دسترسي موتورهاي مذكور است. بازيابي بهينة اطلاعات از اينترنت نيازمند ابزارهاي متعدد و شناخت كافي از اين شبكه است. محدودكردن جستجو به حوزة كاوش موتورهاي جستجو منجر به عدم‌دسترسي به اطلاعات نهفته در اينترنت پنهان خواهد شد.**

**بنابراين، آگاهي از وجود اينترنت پنهان نخستين و مهم‌ترين گام در جهت نفوذ به آن است و اين آگاهي را كتابداران مي‌توانند به كاربران منتقل كنند.**

**نتيجه**

**وب پنهان يكي از واقعيت‌هاي موجود اين شبكة جهانگستر است. پنهاني اطلاعات در محيط وب به هر يك از دلايلي كه ذكر شد ايجاد شود، سطح دسترس‌پذيري اطلاعات را براي كاربران كاهش مي‌دهد. از آنجا كه از نظر كم‍ّي و كيفي بخش قابل‌توجهي از منابع اطلاعاتي مبتني بر وب جزء قلمرو تاريك آن قرار دارد آگاهي از وجود آن و يافتن راهكارهايي براي بازيابي اطلاعات از درون آن كاملاً ضروري است.**

**خوشبختانه قسمت عمده‌اي از اين منابع را مي‌توان براي كاربران دسترس‌پذير و جستجو‌پذير كرد. با توجه به تعاريفي كه در اين مقاله ارائه شد سهم عوامل فني و تكنيكي در پنهاني اطلاعات در محيط اينترنت به‌مراتب كمتر از عوامل غيرفني است. به بيان ديگر، نفوذ به درون منابع پنهان در وب بيش از آن‌كه به نظام‌هاي بازيابي پيچيده‌تر وابسته باشد به جنبة كاربرمدار[37] در بازيابي اطلاعات وابسته است. بنابراين حتي با همين ابزارهاي كاوش موجود نيز مي‌توان سطح پنهاني اطلاعات را به نحو قابل توجهي كاهش داد. البته از سوي ديگر بر توانايي موتورهاي جستجو نيز افزوده مي‌شود و از نمونه‌هاي بارز آن مي‌توان به موفقيت‌هاي چشمگير موتور «گوگل» اشاره كرد كه پيوسته دامنة جستجوي خود را گسترش مي‌دهد و مرتباً امكانات جستجوي جديدي به آن افزوده مي‌گردد.**

**بنابراين، به‌دليل پژوهش‌‍‌ها و پيشرفت‌هاي اخير در حوزة بازيابي اطلاعات چه در بُعد فني و چه از نظر رويكرد كاربرمدار، آيندة روشنتري در زمينة بازيابي اطلاعات در محيط وب پيش‌بيني مي‌شود.**

**با اين حال و با توجه به شرايط موجود منابع اطلاعاتي مبتني بر وب، لازم است به طرق مختلف كاربران اينترنت را از وجود اينترنت پنهان و اهميت آن آگاه ساخت و به اين ترتيب سطح دسترس‌پذيري اطلاعات را براي آنان افزايش داد. نقش كتابداران و متخصصان اطلاع‌رساني در اين زمينه كاملاً مهم و حياتي مي‌باشد. در حال حاضر صدها دروازة اطلاعاتي و راهنماي اينترنتي كه توسط كتابداران طراحي و ساخته شده‌اند نقش مهمي در بهبود بازيابي اطلاعات كاربران ايفا مي‌كنند. اميد مي‌رود كه كتابداران ايراني نيز در اين زمينه حوزة فعاليت‌هاي خود را گسترش بخشند و در آيندة نزديك شاهد موفقيت‌هاي بيشتري در اين زمينه باشيم.**

**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.**

**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.**

**--------------------------------------------------------------------------------**

**[1]. Hidden Internet**

**[2]. Invisible Web**

**[3] . Opaque Web**

**[4]. Olnvisibility**

**[5]. Crawler or Spider**

**[6]. Links**

**[7]. Index**

**[8] . URL (Universal Resource Locator)**

**[9]. Submit**

**[10]. Query Processor**

**[11]. Frequency of Crawling**

**[12]. Depth of Crawling**

**[13]. Dead Web or Disconnected URLs**

**[14]. Indexing Depth**

**[15]. Maximum Number of Viewable Results**

**[16]. Ranking Policy**

**[17]. Relevance**

**[18] . On-Line**

**[19]. OPACs**

**[20] . ERIC**

**[21] . DIALOG**

**[22]. General –purpose search engines**

**[23]. Shallow Web**

**[24]. Private Web**

**[25]. Proprietary Web**

**[26]. User name**

**[27]. Password**

**[28]. Truly Invisible Web**

**[29] . HTML**

**[30] . Google**

**[31]. Surface**

**[32]. Visible**

**[33]. Subject Gateway**

**[34]. Meta-Search Engine**

**[35]. Intelligent Agents**

**[36]. Invisibility Level**

**[37]. User-oriented**