

بررسی میزان حضور ایران در وب

علیرضا نوروزی^۱

چکیده

از اواسط دهه ۱۹۹۰ حوزه پژوهشی جدیدی به نام وب‌سنجی بر پایه روش‌های کتابسنجی و اطلاع‌سنجی به وجود آمد که کار آن، پژوهش درباره ماهیت و ویژگی‌های وب جهان‌گستر است. هدف این پژوهش، بررسی و مقایسه و رتبه‌بندی دامنه‌های فرعی ایران (.ir) است. این پژوهش نشان می‌دهد که دامنه سطح بالای ایران (.ir) بیشترین و دامنه نظامی ایران (mil.ir) کمترین میزان صفحه‌های وب را تولید کرده‌اند. همچنین قابل ذکر است که موتور کاوش گوگل، وبگاه‌های ایرانی به‌ویژه دانشگاهی را بهتر تحت پوشش قرار می‌دهد. به عبارت دیگر، گوگل نسبت به سایر موتورهای کاوش (ياهو، آلتاویستا، ام‌اس‌ان، و آلثوب)، صفحه‌های وب بیشتری از ایران را نمایه‌سازی می‌کند و موتور کاوش «ام‌اس‌ان» وبگاه‌های ایرانی را کمتر تحت پوشش قرار می‌دهد. این پژوهش همچنین نشان می‌دهد که اغلب صفحه‌های وب منتشر شده در ایران با فرمت زبان نشانه‌گذاری فرامتن منتشر شده‌اند. یافته‌ها نشان می‌دهد که فرمت‌هایی چون HTML، HTM، PHP، ASP، PDF، و DOC بیشتر مورد استفاده وبگاه‌های ایرانی قرار گرفته‌اند و موتورهای کاوش این نوع فایل‌ها را بهتر نمایه‌سازی می‌کنند. بنابراین مدیران وبگاه‌های ایرانی باید توجه بیشتری به این نوع فایل‌ها داشته باشند.

کلیدواژه‌ها

وب‌سنجی، موتورهای کاوش، نمایه‌سازی وب، وبگاه، ایران

مقدمه

به وجود آورد که برای شناخت هرچه بهتر این پدیده، مطالعه جنبه‌های مختلف آن ضروری به نظر می‌رسد. منابع گوناگون وب، نظیر چکیده و متن کامل مقاله‌های مجله‌های

اختراع وب توسط تیم برنرز-لی^۲ (۱۹۸۹) و ویژگی منحصر به فرد فرامتنی و فرایبوندی آن انقلابی شگرف در عرصه اطلاعات و اطلاع‌رسانی

۱. دانشجوی دکتری کتابداری و دانش‌شناسی، فرانسه، nouruzi@gmail.com

2. Tim Berners-Lee

الکترونیکی دسترسی آزاد^۳، متن پروانه‌های ثبت اختراع، کنفرانس‌ها، نوشته‌های پژوهشی و آموزشی، گزارش‌های دولتی و کتاب‌های الکترونیکی جزء منابع علمی ارزشمند در محیط وب هستند. در اصل، وب به منظور برآورده کردن نیازهای اطلاعاتی پژوهشگران و به اشتراک گذاشتن منابع اطلاعاتی، و دانش و تجربه‌های فردی طراحی شده است (۳).

وب با تاریخ پانزده ساله خود در میان مهم‌ترین و مؤثرترین فناوری‌های قرن بیست و یکم قرار دارد. انقلاب اطلاعات نه تنها قدرت حرکت و پیشبرد وب را فراهم می‌کند، بلکه تقاضای بی‌سابقه‌ای برای ذخیره، سازماندهی، اشاعه، و دستیابی به اطلاعات را به دنبال داشته است. اگر اطلاعات، پول رایج در نظام اقتصادی «دانش‌محور» است، وب بانکی است که اطلاعات در آن سرمایه‌گذاری می‌شود. یکی از مزایای مهم وب این است که استفاده‌کنندگان می‌توانند به منابع الکترونیکی پیوسته دستیابی پیدا کنند، یعنی چیزی که در کتابخانه‌های سنتی «کاغذمحور»^۴ وجود ندارد. این تفاوت واقعی بین وب و کتابخانه‌هاست. وب برای تمامی شهروندان مجازی^۵ (کاربران) جالب است. کاربران وب می‌توانند از منابع اطلاعاتی وب به منظور آموزش و پژوهش، تفریح و سرگرمی خلاق، عدالت اجتماعی، آزادی مردمی و دموکراسی، توسعه اقتصادی و تجارتي، سوادآموزی، آموزش مادام‌العمر، و توسعه فرهنگی استفاده کنند. نتیجه این استفاده بهبود وضعیت

اجتماعی، فرهنگی، و اقتصادی افراد و جامعه است. بنابراین، وب باید نیازهای اطلاعاتی کاربران را شناسایی کرده و برآورده سازد و زمینه‌های توسعه خدمات را فراهم کند (۶).

در واقع، وب مجموعه‌ای است بسیار پیچیده از انواع منابع اطلاعاتی که به وسیله افراد متفاوت تولید می‌شود و توسط کاربران مختلف مورد جست‌وجو قرار می‌گیرد (۱): ۶۴-۸۶). با توجه به اینکه وب برای خدمت به جامعه به وجود آمده است، بنابراین باید علایق و نیازهای جامعه را برآورده سازد.

در ایران وبگاه‌ها و وبنوشت (وبلاگ)‌ها به یکی از ابزارهای مهم ارتباطی تبدیل شده‌اند که جهت برآوردن اهداف مختلف علمی و پژوهشی، اخبار ملی و بین‌المللی، گسترش زبان و ادبیات فارسی، و تفریح و سرگرمی به کار گرفته می‌شوند. وبگاه‌های ایرانی می‌توانند حضور فعال سازمان‌های ایرانی، گروه‌های پژوهشی، دانشگاه‌ها و مؤسسه‌های علمی، گروه‌های آموزشی، کتابخانه‌ها، مراکز تجاری و افراد را آشکار سازند و موجب ترقی و پیشرفت آنها شوند و در نهایت روند دسترسی به اطلاعات و پیشرفت اقتصادی و فرهنگی کشور تسهیل کنند.

بررسی و ارزیابی میزان «حضور کشورهای خاورمیانه در محیط وب و ضریب تأثیرگذاری آنها» با استفاده از موتور کاوش یاهو نشان می‌دهد که از لحاظ میزان تولید صفحه‌های وب، ایران بعد از ترکیه و اسرائیل در رتبه

3. Open Access (OA)

4. Paper-based

5. Cybercitizens

6. Weblog

سوم قرار دارد (۷). لذا پژوهشی جامع به منظور بررسی میزان حضور دامنه‌های مختلف ایران در عرصه وب ضروری به نظر می‌رسد. البته پژوهش‌های دیگری در مورد «ضریب تأثیرگذاری وبگاه‌های دانشگاهی ایران در محیط وب» انجام شده است (۲: ۱۰۵-۱۱۹؛ ۸).

تعریف مسئله

اهمیت مسئله به این دلیل است که امروزه وب در بسیاری از کشورهای دنیا به عنوان ابزاری ارتباطی و اطلاعاتی به منظورهای چندگانه آموزشی، پژوهشی، تجاری، و تبلیغاتی به کار گرفته می‌شود. ایران نیز همگام با جامعه جهانی امکان دسترسی به اینترنت را فراهم کرده است و بسیاری از سازمان‌ها اقدام به راه‌اندازی وبگاه‌های ویژه خود کرده‌اند. امروزه وب یکی از منابع اطلاعاتی علمی و دانشگاهی در ایران است. بنابراین، باید به بررسی جنبه‌های مختلف این پدیده پرداخت و جایگاه ایران را در بین کشورهای مختلف نشان داد. با گذشت بیش از یک دهه از عمر وب در ایران، بررسی عملکرد جامعه ایرانی از نقطه‌نظرهای مختلف به منظور اتخاذ راهبردی مناسب برای سیاست‌گذاری و تصمیم‌گیری در مورد میزان حضور ایران در عرصه وب در سطح ملی و بین‌المللی و نیز بررسی میزان تولید صفحه‌های وب در دامنه‌های مختلف ایران منطقی به نظر

می‌رسد.

یکی از ابزارهایی که می‌تواند برای بررسی و مقایسه میزان حضور یک کشور در محیط وب به کار گرفته شود، موتورهای کاوش عمومی هستند. این موتورها به‌طور مرتب (ماهانه و هفتگی) وبگاه‌های گوناگون را نمایه‌سازی می‌کنند. قابل ذکر است که «موتورهای کاوش، فقط صفحه‌های وب عمومی را که از طریق دنبال کردن پیوندها قابل نمایه‌سازی باشند در پایگاه خود گردآوری می‌کنند» (۵). چنانچه یک صفحه وب از سایر صفحه‌های وب هیچ پیوندی دریافت نکرده باشد از طریق موتورهای کاوش هرگز بازیابی نخواهد شد و اگر وبگاهی از طریق نمایه یا نقشه خود به صفحه‌های درونی هیچ «خود-پیوندی»^۹ نداده باشد، صفحه‌های درونی آن نیز توسط ربات‌های موتورهای کاوش شناسایی و نمایه‌سازی نخواهند شد و این یکی از عوامل مهم در وب پنهان است (۲: ۱۰۵-۱۱۹).

هدف پژوهش

هدف از پژوهش حاضر مقایسه و ارزیابی دامنه‌های مختلف ایران مانند ac.ir، id.ir، gov.ir و sch.ir، org.ir، net.ir، co.ir و شناسایی میزان حضور ایران در عرصه وب جهان‌گستر است. قابل ذکر است که شناسایی وضعیت وب در ایران از لحاظ کمی با استفاده از موتورهای کاوش قابل ارزیابی است.

7. Domain
8. Site map

9. Self-link

توسط موتورهای کاوش را نشان می‌دهد. در این جدول، در رتبه‌بندی دامنه‌ها، گوگل در اولویت قرار داده شده است.

| دامنه | MSN | AlltheWeb | AltaVista | Yahoo | Google |
|---------|-------------|---------------|---------------|---------------|-------------|
| ir/ | ۶۷۰,۵۳۲ | ۴,۷۴۰,۰۰۰ | ۵,۰۴۰,۰۰۰ | ۵,۰۶۰,۰۰۰ | ۶,۹۱۰,۰۰۰ |
| ac.ir/ | ۲۵۴,۸۳۳ | ۵۶۳,۰۰۰ | ۵۸۵,۰۰۰ | ۵۸۹,۰۰۰ | ۱,۱۸۰,۰۰۰ |
| org.ir/ | ۱۶,۱۷۸ | ۵۴,۳۰۰ | ۶۳,۴۰۰ | ۶۴,۰۰۰ | ۱۳۳,۰۰۰ |
| gov.ir/ | ۴۳,۴۷۵ | ۵۴,۰۰۰ | ۵۹,۵۰۰ | ۶۰,۴۰۰ | ۷۰,۳۰۰ |
| co.ir/ | ۲۳,۲۹۲ | ۳۵,۳۰۰ | ۳۸,۱۰۰ | ۳۸,۳۰۰ | ۴۸,۱۰۰ |
| sch.ir/ | ۱,۱۲۷ | ۴,۹۰۰ | ۵,۱۷۰ | ۵,۲۸۰ | ۱۲,۳۰۰ |
| id.ir/ | ۱۵۷ | ۱,۲۵۰ | ۱,۷۴۰ | ۱,۶۸۰ | ۹۰۲ |
| net.ir/ | ۱,۶۰۵ | ۸۹۱ | ۹۶۴ | ۱,۰۴۰ | ۵۵۱ |
| mil.ir/ | * | * | * | * | * |
| مجموع | ۱۰۱۱۱۹۹ | ۵۴۵۳۶۴۱ | ۵۷۹۳۸۷۴ | ۵۸۱۹۷۰۰ | ۸۳۵۵۱۵۳ |
| دستور | site:ac.ir/ | domain:ac.ir/ | domain:ac.ir/ | domain:ac.ir/ | site:ac.ir/ |

جدول ۱. صفحه‌های وب نمایه‌سازی شده از دامنه‌های ایران توسط موتورها (۲۷ مهر ۱۳۸۵)

پرسش‌های اساسی پژوهش

نتایج پژوهش حاضر با توجه به هدف و بیان مسئله باید بتواند به پرسش‌های زیر پاسخ دهد:

۱. کدام موتور کاوش وبگاه‌های ایرانی را بهتر پوشش می‌دهد؟
۲. کدام دامنه ایران بیشترین صفحه‌های وب را منتشر کرده است؟
۳. صفحه‌های نمایه‌سازی شده از وبگاه‌های ایرانی به کدام فرمت بیشتر منتشر شده‌اند؟
۴. پراکندگی دامنه‌های مختلف از نظر برخورداری از انواع فایل‌ها چگونه است؟

روش‌شناسی پژوهش

کلید و ابزار اساسی پژوهش‌های وب‌سنجی موتورهای کاوش گسترده‌ای همچون گوگل و یاهو هستند که اجازه می‌دهند تعداد کل صفحه‌های موجود در یک دامنه خاص و یا وبگاه اندازه‌گیری شود؛ هر چند که موتورهای کاوش کاربر را در مشاهده کل نتایج بازیابی شده محدود می‌کنند. قابل ذکر است که تعداد کل نتایج، معمولاً همان عددی در نظر گرفته می‌شود که این موتورها به عنوان تعداد تقریبی نتایج نشان می‌دهند. در این پژوهش برای گردآوری داده‌ها، از موتورهای کاوش گوگل، یاهو، ایم‌اس‌ان، آلتوب^{۱۰} و آلتاویستا^{۱۱} استفاده شده است، زیرا این موتورها بیشتر مورد استفاده کاربران قرار می‌گیرند. جدول ۱ میزان صفحه‌های وب نمایه‌سازی شده از هر دامنه خاص

10. AlltheWeb.com

11. Alta Vista.com

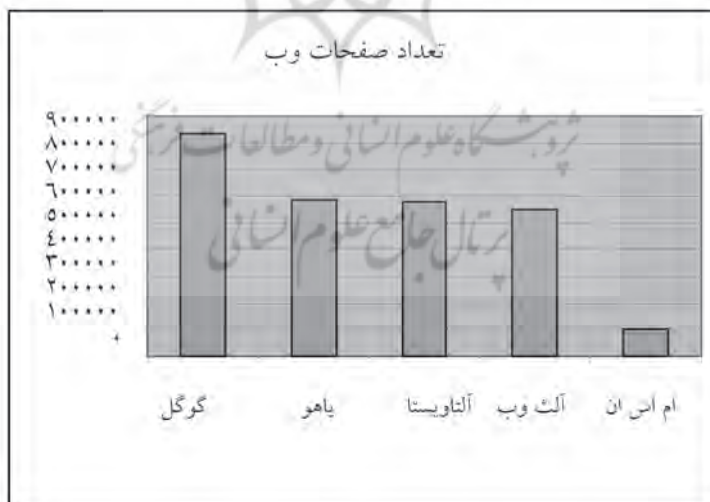
در موتور کاوش گوگل و ایماسان نیز از دستور زیر استفاده شده است:

site:ir/ -site:ac.ir/ -site:gov.ir/ -site:org.ir/ -site:co.ir/ -site:net.ir/ -site:sch.ir/ -site:id.ir/

با توجه به اینکه گوگل جامع‌ترین و پُراستفاده‌ترین موتور کاوش است و امکان جست‌وجوی پیشرفته را فراهم می‌کند، در نتیجه در مطالعه حاضر برای گردآوری داده‌ها بیشتر روی گوگل تأکید شده است. همان‌گونه که پیش‌تر بیان شد یکی از اهداف این پژوهش، بررسی نوع فایل‌هایی است که توسط وبگاه‌های ایرانی منتشر شده است. جدول ۲ فراوانی انواع فایل‌ها را به تفکیک دامنه‌های خاص در گوگل نشان می‌دهد. البته تأکید روی فایل‌هایی است که عمومیت بیشتری دارند. در صورتی که از نوع فایل

در این پژوهش برای به‌دست آوردن تعداد کل صفحه‌های وب نمایه‌سازی شده از یک دامنه خاص در موتور کاوش گوگل ایماسان از دستور «site:» و در موتور کاوش یاهو و سایر موتورهای از دستور «domain:» استفاده شده است. در جدول ۱، قسمت دستور جست‌وجو «ac.ir» به عنوان مثال به‌کار رفته و بیانگر دامنه فرعی (یا دامنه سطح دوم)^{۱۲} است. با توجه به اینکه بسیاری از وبگاه‌های ایرانی با «.ir» به ثبت رسیده‌اند، در نتیجه برای محاسبه تعداد صفحه‌های وب منتشر شده در دامنه سطح بالای^{۱۳} ایران «.ir» در موتورهای کاوش یاهو، آلت‌وب، و آلتاویستا از دستور زیر استفاده شده است:

domain:ir/ NOT (domain:ac.ir/ OR domain:gov.ir/ OR domain:org.ir/ OR domain:co.ir/ OR domain:net.ir/ OR domain:sch.ir/ OR domain:id.ir/)



نمودار ۱. مقایسه موتورهای کاوش

جدول ۲. نوع فایل‌های نمایه‌سازی شده از دامنه‌های ایران توسط گوگل (۲۷ مهر ۱۳۸۵)

| مجموع | ir/ | Id.ir/ | sch.ir/ | net.ir/ | gov.ir/ | org.ir/ | Co.ir/ | Ac.ir/ | نوع فایل |
|---------|---------|--------|---------|---------|---------|---------|--------|--------|----------|
| ۳۱۸۲۵۸۰ | ۳۰۱۰۰۰۰ | ۰ | ۲۲۵ | ۸۵ | ۱۰۸۰۰ | ۴۲۰۰ | ۳۲۵۰ | ۱۵۴۰۰۰ | HTM |
| ۳۶۵۶۲۴ | ۳۵۳۰۰۰ | ۱۹ | ۵۸ | ۱۲ | ۲۱۰۰ | ۹۵۵ | ۶۸۰ | ۸۸۰۰ | HTML |
| ۲۷۶۸۱۸ | ۲۷۱۰۰۰۰ | ۲۳ | ۱۰۳ | ۱۴ | ۸۳۴ | ۲۶۲ | ۸۴۲ | ۳۷۴۰ | ASPX |
| ۲۵۶۸۲۸ | ۲۵۳۰۰۰۰ | ۷ | ۳۰ | ۲ | ۶۳۱ | ۱۰۴ | ۴۱۴ | ۲۶۴۰ | PHP |
| ۱۷۹۳۱۶ | ۱۴۱۰۰۰۰ | ۰ | ۱۴۶۰ | ۱۶ | ۱۶۷۰ | ۳۳۸۰ | ۱۱۹۰ | ۳۰۵۰۰۰ | PDF |
| ۱۵۲۸۹۹ | ۱۴۲۰۰۰۰ | ۰ | ۱۴۹ | ۲۴ | ۴۲۰۰ | ۲۰۶۰ | ۸۲۶ | ۳۶۴۰ | ASP |
| ۲۴۹۴۰ | ۲۳۱۰۰۰ | ۰ | ۱۵ | ۱ | ۱۱۷ | ۸۰ | ۱۳۷ | ۱۴۹۰ | SWF |
| ۲۴۵۸۴ | ۱۹۵۰۰۰ | ۰ | ۱۲ | ۰ | ۹۱۴ | ۴۳۰ | ۱۳۸ | ۳۵۹۰ | DOC |
| ۱۰۷۶۴ | ۱۰۶۰۰۰ | ۰ | ۰ | ۳۳ | ۰ | ۱۳۳ | ۵ | ۳ | SHTML |
| ۲۹۵۴ | ۲۲۹ | ۰ | ۱ | ۰ | ۱۰۰ | ۱۱۵ | ۳۹ | ۲۴۷۰ | PPT |
| ۱۷۱۱ | ۹۹۹ | ۰ | ۲۶ | ۰ | ۲ | ۶۴ | ۲ | ۶۱۶ | JSP |
| ۱۰۹۴ | ۱۱۳ | ۰ | ۰ | ۰ | ۸۷ | ۶۴۸ | ۲۶ | ۲۲۰ | XLS |
| ۷۶۲ | ۵۶۵ | ۰ | ۰ | ۰ | ۶۳ | ۲ | ۲ | ۱۳۰ | XML |
| ۶۴۰ | ۲۱۱ | ۰ | ۰ | ۰ | ۴ | ۱ | ۰ | ۲۲۴ | TXT |
| ۴۴۴ | ۱۵ | ۰ | ۰ | ۰ | ۰ | ۱ | ۰ | ۴۲۸ | PS |
| ۲۰۱ | ۱۵۳ | ۰ | ۰ | ۰ | ۰ | ۲ | ۱۹ | ۲۷ | CFM |
| ۱۵۹ | ۴۸ | ۰ | ۳ | ۰ | ۱۳ | ۷ | ۲ | ۸۶ | PPS |
| ۱۲۳ | ۵۳ | ۰ | ۰ | ۰ | ۱۱ | ۵ | ۴ | ۵۰ | MHT |
| ۱۱۵ | ۹۸ | ۰ | ۰ | ۲ | ۰ | ۰ | ۲ | ۱۳ | CGI |
| ۷۸ | ۰ | ۰ | ۰ | ۰ | ۰ | ۲۷ | ۰ | ۵۱ | PHTML |
| ۵۸ | ۹ | ۰ | ۰ | ۰ | ۰ | ۱ | ۰ | ۴۸ | RTF |
| ۴۱ | ۲ | ۰ | ۰ | ۰ | ۰ | ۳ | ۰ | ۳۶ | PL |
| ۳۳ | ۰ | ۰ | ۰ | ۰ | ۰ | ۲۷ | ۰ | ۶ | DAT |
| ۱۴ | ۴ | ۰ | ۰ | ۰ | ۰ | ۳ | ۰ | ۷ | VBS |
| ۱۲ | ۲ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۱۰ | SHTML |
| ۱۲ | ۷ | ۰ | ۰ | ۰ | ۵ | ۰ | ۰ | ۰ | XSLT |
| ۶ | ۶ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | PHP3 |
| ۵ | ۴ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۱ | SQL |
| ۲ | ۲ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | XHTML |
| ۱ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۱ | EPS |

site:ac.ir/ filetype:html

تعداد کل فایل‌های نمایه‌سازی شده از هر دامنه خاص در موتور کاوش گوگل از دستور «site:» و «filetype:» (نوع فایل) استفاده شده است (۹). به عنوان مثال:

مشخصی در دامنه‌های ایرانی استفاده نشده باشد، در جدول زیر نام آن نخواهد آمد. برای آشنایی با انواع فایل‌ها به مقاله دانشنامه ویکی‌پدیا^{۱۴} (۲۰۰۶) رجوع شود. در این قسمت از پژوهش، برای به دست آوردن

14. Wikipedia

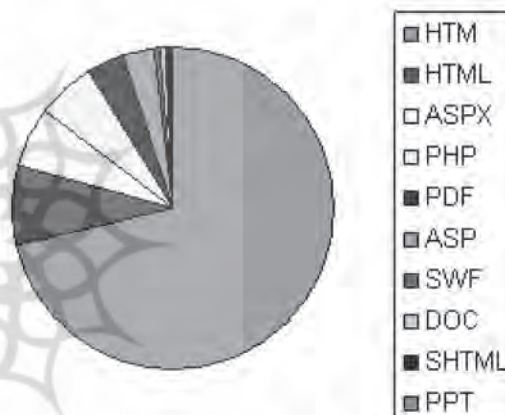
با توجه به اینکه بسیاری از وبگاه‌های ایرانی با «.ir» به ثبت رسیده‌اند، در نتیجه برای کشف نوع فایل‌های منتشر شده در دامنه سطح بالای ایران «.ir» در موتور کاوش گوگل از دستور زیر استفاده شده است. به عنوان مثال:

site:ir/ filetype:html -site:ac.ir/
-site:gov.ir/ -site:org.ir/ -site:co.ir/ -
site:net.ir/ -site:sch.ir/ -site:id.ir/

صفحه‌های وب به زبان نشانه‌گذاری فرامتن بوده‌اند، به این دلیل که نوع فایل آنها مشخص نبوده است و یا به این فرمت منتشر شده‌اند (۴: ۹۹۳). قابل توجه است که برخی از انواع فایل‌ها هرگز در وبگاه‌های ایرانی استفاده نشده‌اند. از این‌رو، در جدول ۲ فایل‌هایی که در وبگاه‌های ایرانی به کار رفته و توسط موتور کاوش گوگل نمایه‌سازی شده‌اند، نشان داده شده است.

یافته‌ها

در جدول ۱ مشاهده می‌شود که تعداد کل صفحه‌های وب نمایه‌سازی شده از دامنه سطح بالای ایران «.ir» توسط موتورهای کاوش بیشتر از مجموع کل صفحه‌های نمایه‌سازی شده در دامنه‌های فرعی است. دلیل این امر ممکن است این باشد که بسیاری از وبگاه‌های ایرانی با «.ir» به ثبت رسیده‌اند و این دامنه فعال‌تر از سایر دامنه‌هاست. به عنوان مثال، خبرگزاری‌ها، روزنامه‌ها، مجله‌های الکترونیکی، و برخی سازمان‌های دیگر از پسوند دامنه فرعی خاصی استفاده نمی‌کنند.



نمودار ۲. نوع فایل‌های پراستفاده در دامنه‌های ایران

همان‌گونه که در جدول ۲ و نمودار ۲ مشاهده می‌شود، اغلب صفحه‌های وب ایرانی با فرمت HTML و HTM ساخته شده‌اند. به عبارت دیگر، زبان نشانه‌گذاری فرامتن (HTML) فرمت محبوب مدیران وبگاه‌های ایرانی و فرمت محبوب موتور کاوش گوگل در نمایه‌سازی است. بری^{۱۶} (۱۹۹۵) نشان می‌دهد که بیش از ۸۰ درصد

بحث

از آنجا که وب ارگانیک پویا و در حال رشد است، در نتیجه امکان کشف دقیق تعداد صفحه‌های وب موجود در یک دامنه و یا یک کشور خاص، به دلیل ماهیت پویا، متنوع، و توزیعی وب و همچنین نواقص و محدودیت‌های موتورهای کاوش تقریباً

غیرممکن است.

بی‌گمان تسهیل دسترسی به فناوری‌های اطلاعاتی و ارتباطی (ITC) به‌ویژه اینترنت در دانشگاه‌ها، سازمان‌های دولتی و خصوصی، کتابخانه‌های عمومی، روستایی، آموزشی، و دانشگاهی موجب افزایش حضور ایران در عرصه اینترنت و بالا رفتن میزان تولید صفحه‌های وب فارسی زبان می‌شود که برای گسترش زبان فارسی و فرهنگ ایرانی نیز ضروری است.

برخی از مشکلات وضعیت موجود وب در ایران عبارتند از:

۱. به‌طور کلی صفحه‌های وب به‌دست مدیران وبگاه‌ها ساخته می‌شود. به عنوان نمونه در بسیاری از دانشگاه‌ها، دانشجویان و استادان هنوز هم صفحه‌های وب شخصی حتی در حد معرفی فعالیت‌های علمی و پژوهشی خود ندارند و وبگاه‌ها به‌دست مدیران بخش اطلاع‌رسانی دانشگاه‌ها اداره می‌شود. این یکی از دلایلی است که بسیاری از دانشجویان و استادان به وبگاه‌ها و وبلاگ‌های رایگان روی می‌آورند که با توجه به امکانات موجود موتورهای کاوش قابل شناسایی نیستند. جدول ۳ نشان می‌دهد که تعداد صفحه‌های وب نمایه‌سازی شده از بلاگفا توسط گوگل به‌طور تقریبی برابر تعداد

کل صفحه‌های وب دانشگاه‌ها و مؤسسه‌های دانشگاهی دولتی و آزاد است.

۲. بسیاری از وبگاه ایرانی از دامنه «.ir» استفاده نمی‌کنند، بلکه از دامنه‌های مانند edu و net com org استفاده می‌کنند که با استفاده از امکانات موجود موتورهای کاوش قابل شناسایی نیستند که این یکی از محدودیت‌های این پژوهش است.

۳. مدیران وبگاه‌ها به اندازه کافی با روش‌های نمایه‌سازی وب و طراحی صفحه‌های وب آشنا نیستند. پژوهش‌های پیشین نشان می‌دهند که بسیاری از صفحه‌های وب ایرانی فاقد اَبَر-برچسب‌های عنوان، کلیدواژه، توصیفگر، زبان، و محل انتشار هستند. در حالی که این اَبَر-برچسب‌ها در رتبه‌بندی صفحه‌های وب و در روند بازیابی اطلاعات توسط جست‌وجوگران اطلاعات در محیط وب اهمیت ویژه‌ای دارند. جدول ۴ میزان صفحه‌های وب ایرانی بدون عنوان در موتور کاوش گوگل را نشان می‌دهد. در مجموع ۴۳,۳۷۴ صفحه وب ایرانی فاقد عنوان هستند. هنگام جست‌وجو برای بازیابی مدارک فاقد عنوان در دامنه سطح بالای ایران یعنی ردیف نخست، سایر دامنه‌ها با استفاده از منطق بولی حذف شده‌اند.

جدول ۳. مقایسه دامنه دانشگاهی با بلاگفا

| تعداد صفحه | دستور جست‌وجو | |
|------------|--|----------------|
| ۱۲۱۰۰۰۰ | site:blogfa.com OR site:www.blogfa.com | بلاگفا |
| ۱۱۸۰۰۰۰ | site:ac.ir/ | دامنه دانشگاهی |

جدول ۴. صفحه‌های وب ایرانی منتشر شده

بدون عنوان

| دامنه | تعداد صفحه‌ها |
|----------------------------------|---------------|
| allintitle:Untitled site:ir/ | ۳۴۴۰۰ |
| allintitle:Untitled site:ac.ir/ | ۸۱۷۰ |
| allintitle:Untitled site:co.ir/ | ۲۷۲ |
| allintitle:Untitled site:org.ir/ | ۲۲۱ |
| allintitle:Untitled site:gov.ir/ | ۲۱۳ |
| allintitle:Untitled site:sch.ir/ | ۹۷ |
| allintitle:Untitled site:net.ir/ | ۱ |
| مجموع | ۴۳۳۷۴ |

سازمان‌های نظامی ایران از دامنه سطح بالای .ir باشد. به عنوان نمونه، وبگاه پلیس ایران با دامنه سطح بالای ایران (Police.ir) به ثبت رسیده است.

موتور کاوش گوگل، وبگاه‌های ایرانی، به ویژه دانشگاهی را بهتر تحت پوشش قرار می‌دهد. به گفته‌ای دیگر، گوگل به نسبت سایر موتورهای کاوش، صفحه‌های وب بیشتری از ایران را نمایه‌سازی می‌کند و موتور کاوش «ام‌اس‌ان» وبگاه‌های ایرانی را کمتر تحت پوشش قرار می‌دهد. لذا مدیران وبگاه‌ها باید معیارهای لازم برای نمایه‌سازی وبگاه‌ها توسط گوگل را فراهم کنند. تعداد رکوردهای بازیابی شده از سه موتور کاوش یاهو، آلتاویستا، و آلتوب تقریباً برابر است. یکی از دلایل آن این است که فناوری نمایه‌سازی موتور کاوش آلتاویستا در سال ۲۰۰۲ به وسیله یاهو خریداری شد و اکنون آلتاویستا به وسیله یاهو حمایت می‌شود.

این پژوهش نشان می‌دهد که اغلب صفحه‌های وب منتشر شده در ایران با فرمت‌هایی چون HTML HTM PHP و ASP PDF و DOC منتشر شده‌اند. در کل تعداد صفحه‌های وب منتشر و نمایه‌سازی شده از وبگاه‌های ایرانی پایین است و لازم است که دلایل پایین بودن حجم تولید صفحه‌های وب در ایران و عدم استفاده از وب به عنوان یک ابزار ارتباطی و اطلاعاتی در مقیاس وسیع در جامعه مورد بررسی قرار گیرد.

همان‌گونه که مشاهده می‌شود تعداد زیادی از صفحه‌های وب ایرانی در دامنه‌های مختلف بدون عنوان منتشر شده‌اند. در مورد «عنوان صفحه وب و اهمیت آن در رتبه‌بندی صفحه‌های وب توسط موتورهای کاوش» پژوهش دیگری در دست انجام است که در این مقال نمی‌گنجد.

نتیجه‌گیری

در پژوهش حاضر دامنه سطح بالای ایران و دامنه دانشگاهی به ترتیب بالاترین میزان صفحه‌های وب را در مقایسه با سایر دامنه‌های ایرانی دارا هستند. نتایج این پژوهش نشان می‌دهد که در مجموع، حجم دامنه سطح بالای ایران بالاتر از سایر دامنه‌های ایران است. قابل ذکر است که در دامنه نظامی ایران (mil.ir) کمترین میزان صفحه‌های وب منتشر شده است. شاید این امر ناشی از پایین بودن میزان فعالیت دامنه نظامی ایران در محیط اینترنت به دلایل امنیتی یا فنی و استفاده

Computer Networks and ISDN Systems, Vol. 28, No. 7-11 (1995): 993. [on-line]. Available: <http://www.cs.ubc.ca/local/reading/proceedings/www5/www134/overview.htm>. [19 Oct.2006].

5. Lawrence, S; Giles, C.L. "Accessibility of information on the Web". *Nature*, No.400 (1999): 107-109.

6. Noruzi, A. "Application of Ranganathan's laws to the Web". *Webology*, Vol.1, No.2 (2004). [on-line]. Available: <http://www.webology.ir/2004/v1n2/a8.html>.

7. Ibid. "Web impact factors for Iranian Universities". *Webology*, Vol.2, No.1 (2005). [on-line]. Available: <http://www.webology.ir/2005/v2n1/a11.html>.

8. Ibid. "Web presence and impact factors for Middle-Eastern Countries". *Online Magazine*, Vol.30, No.2 (2006) 22-28.

9. Wikipedia. "List of file formats". *Wikipedia, the free encyclopedia*. 2006. [on-line]. Available: http://en.wikipedia.org/wiki/List_of_file_formats. [19 Oct. 2006].

تاریخ دریافت: ۱۳۸۴/۷/۲۴
بازنگری شده در مهرماه ۱۳۸۵



پیشنهادها

پیشنهاد می‌شود که پژوهش دیگری در مورد دلایل بالا بودن میزان صفحه‌های وینوشت‌ها نسبت به تعداد صفحه‌های وبگاه‌ها در ایران انجام شود. پژوهش‌های دیگری برای آشنایی بیشتر با ساختار و مشکلات زبان فارسی در محیط وب و نحوه‌ی نمایه‌سازی و رتبه‌بندی صفحه‌های وب ایرانی توسط موتورهای کاوش ضروری به نظر می‌رسد.

منابع

۱. بجورنبورن، لنارت؛ اینگورسن، پیتر. «چشم‌اندازهایی بر وب‌سنجی». ترجمه‌ی علیرضا نوروزی و زهرا بیگدلی. *اطلاع‌رسانی*، دوره‌ی نوزدهم، ۱ و ۲ (پاییز و زمستان ۱۳۸۲): ۶۴-۸۶.
۲. نوروزی، علیرضا. «ضریب تأثیرگذاری وب و سنجش آن در برخی سایت‌های دانشگاهی ایران». *مطالعات تربیتی و روانشناسی*، ویژه‌نامه کتابداری و اطلاع‌رسانی، دوره‌ی پنجم، ۲ (۱۳۸۳): ۱۰۵-۱۱۹.
3. Berners-Lee, T. "Information management: a proposal." (24Sep.1989). [on-line]. Available: <http://www.w3.org/History/1989/proposal.html>. [19Oct. 2006].
4. Bray, T. "Measuring the Web".