

مهارت در جستجوی اطلاعات فارسی از اینترنت [۱]

محمد صابر راثی ساریانقلی [۲]

چکیده:

خط فارسی دارای مشکلات مختلفی می باشد که در جستجو و بازیابی اطلاعات مسائل و مشکلات فراوانی را فراوانی را قرار می دهد. به خصوص با رشد سریع انتشارات الکترونیکی بر روی وب در شکل های مختلف پایگاه های اطلاعاتی، وبلاگ و ... و اینکه هیچ قاعده مشخص و ثابتی برای رسم الخط فارسی وجود ندارد باعث شده است که جستجوگران مطالب فارسی با مشکلات فراوانی روبرو بشوند. این مقاله سعی دارد تا با اشاره به موارد مختلفی که می تواند در جستجو و بازیابی اطلاعات سرعت و دقت و جامعیت و مانعیت جستجو را بالا ببرد موجب افزایش مهارت کاربران اینترنت فارسی بشود.

کلید واژه ها: اینترنت، خط فارسی، جستجو و بازیابی اطلاعات.

مقدمه

اینترنت به عنوان يك محمل اطلاعاتي عظیم، منابع اطلاعاتي را در مقیاسي وسیع در دسترس مخاطبان بالقوه قرار داده است. اغلب سهولت دسترسی به منابع اطلاعاتي اعم از متن و سایر رسانه ها عمده ترین مزیت اینترنت محسوب می شود. اما این توانایی که هرکس ناشر آثار خود باشد عواقب ناخواسته ای را نیز در پی خواهد داشت و آشکارترین معضل، آن است که انبوهی از منابع بسیار متنوع و غیر قابل مدیریت را فراهم می آورد. افزایش سریع منابع اینترنتی نیازمند يك سازمان دهنی مفید و موثر است. هرچند در حال حاضر راهنماهایی برای منابع اینترنتی تهیه شده است که براساس فایل های مقلوب ساخته شده توسط موتورهای جستجو و با استفاده از قابلیت های مختلف این موتورها از جمله: استفاده از عملگرهای بولی، جستجوی دقیق عبارت، محدود کردن يك جستجو به بخش خاصی از رکورد (مانند عنوان، آدرس)، کوتاه سازی کلمات، جستجوی نزدیک یابی واژه ها، ایجاد محدودیت زمانی و منطقه ای و زبانی، و ... به جستجوی اطلاعات کمک می کند، اما باید تاکید کرد که در امر بازیابی اطلاعات از اینترنت بدون نمایه سازی نظام یافته نمی توان انتظار بازیابی مفید و موثر را داشت. هرچند بیش تر اطلاعات موجود بر روی اینترنت به زبان انگلیسی است، ولی حجم اطلاعات به زبان فارسی نیز با سرعت در حال افزایش است و کاربران به دلایل مختلفی علاقه زیادی به اطلاعات فارسی نشان می دهند و از آنجائی که زبان غالب در اینترنت انگلیسی است جستجو به زبان های غیر انگلیسی از جمله فارسی، مسایل و مشکلات مختلفی را جدای از مشکلات عمومی اینترنت دارد.

خط فارسی

اشکال و نقصی که در همه خطوط جهان است دو علت دارد که یکی در اصل خط است و دیگری بر اثر تغییر و تحول زبان ایجاد می شود. دقت فراوان در ثبت همه دقایق تلفظ اغلب موجب دشواری شیوه خط است و این دقت زمانی ضرورت می یابد که زبانی توسعه بسیار بیابد و در کشورهای دیگری که به آن زبان سخن نمی گویند رایج شود. به عنوان مثال در خط عربی نقطه و علامت های حرکات وقتی به وجود آمد که زبان عربی نزد ملت های غیر عرب معمول شد، در خط یونانی نیز نشانه های آهنگ و تکیه [۳] پس از رواج آن زبان در مصر ایجاد شد تا کسانی که زبان مادری شان یونانی نبود و با تلفظ آن مانوس نبودند بتوانند کلمات و عبارات یونانی را هر چه درست تر ادا کنند. با این حال هیچ خطی هر قدر دقیق و شماره علامت آن فراوان باشد، ممکن نیست که کاملاً نشانه شیوه تلفظ باشد. و با کمک علامت متعدد علم حروف نیز تا کسی چگونگی تلفظ زبانی را نشنود نمی تواند عبارت و کلمات آنرا مانند اهل آن زبان ادا کند.

اما نقصی که بر اثر تحول زبان و به تدریج در خط حاصل می‌شود، مشکلی است که همه ملت‌ها با آن رو به رو هستند. بعضی از حروف و اصوات زبان در طی زمان تغییر می‌پذیرند و این تغییر در گفتار حاصل می‌شود، اما خط همیشه صورت کهن تلفظ را حفظ می‌کند، و از اینجا میان "گفتار" و "نوشتار" اختلاف روی می‌دهد. دیگر آن که هر زبانی ناگزیر لغاتی از زبان‌های دیگر به عاریت می‌گیرد و اگر علائم خط در این دو زبان یکی باشد کلمه خارجی به همان املاي اصلي در نوشتن به کار می‌رود که اغلب با املاي کلمه مشابه در زبان ثانوي تفاوت دارد و از اینجا برای اصوات واحد علائم خطي متعدد پدید می‌آید. در خط فارسي نمونه همه این موارد را می‌توان یافت. چون خط عربي برای نوشتن فارسي به کار رفت کلماتي که از آن زبان اخذ شده بود به همان صورت اصلي نوشته شد. حال آنکه به یقین در هیچ دوره‌اي حروف خاص عربي را فارسي زبان‌ها درست مثل اصل تلفظ نکرده‌اند. در زبان‌های دیگر نیز این گونه موارد نمونه‌های متعدد دارد. شاید دو زبان انگلیسی و فرانسه بیش از همه زبان‌های جهان دچار اختلاف تلفظ و خط باشند. به طور کلی نقائص و معایبی که در خطوط معمول جهان است را می‌توان به طریق زیر طبقه‌بندی کرد:

۱. شکل واحدی اصوات مختلف را بیان می‌کند. چنانکه در فارسي حرف "ي" را گاهی برای حرف لین بکار می‌بریم (یک) و گاهی برای حرف مد (بی) و گاهی به جای الف (عیسی) و گاهی برای نشان دادن مصوت مرکب (ری). و یا حرف «و» در کلمات (سوار، سود، تو)

۲. اصوات واحد به صورت‌های مختلف نوشته می‌شود. در فارسي حرف "س" سه صورت (س - ص - ث) و حرف "ز" چهار صورت (ز - ذ - ض - ظ) دارد؛ در زبان فرانسه حروفی که "سن" خوانده می‌شود پنج رسم الخط دارد که اگر صورت‌های جمع را نیز به حساب بیاوریم ده شکل می‌شود از این قرار (saint, ceint, sein, seing, sain)

۳. بسیاری از حروف نوشته می‌شود ولی خوانده نمی‌شود. یعنی علاماتی بی‌فایده در نوشتن به کار می‌رود در فارسي نوشتن "واو معدوله" و "هـاء غیر ملفوظ" از این قبیل است. در انگلیسی نمونه این مورد بسیار است مانند high که دو حرف آخر آن به کلی از تلفظ ساقط است. و یا "K" در کلمه "Know".

۴. اصواتی هستند که تلفظ می‌شود اما در خط نشانه‌ای برای آن‌ها نیست. در فارسي سه مصوت کوتاه (اَ ، اِ ، اُ) از این قبیل است هم چنین الف در کلمات اسحق و الله که در کتابت نمی‌آید. [۴]

زبان و خط فارسي نیز مشکلات خاصی را دارا می‌باشد و نظام نوشتاری فارسي برای ثبت دقیق گفتار، نارسائی دارد و قواعد نگارش آن مدون نیست، از این رو فاصله میان گفتار و نوشتار در فارسي قابل توجه است. بیشترین مشکلات نیز به جهت نبود يك رسم الخط واحد که عموم اساتید و اهل فن روی آن اجماع کرده باشند به وجود آمده است. به طوری که در حال حاضر جدای از چندین شیوه‌نامه رسمي همچون "شیوه‌نامه سمت، نشر دانشگاهی، فرهنگستان، آموزش و پرورش" به تعداد افراد جامعه، رسم الخط و شیوه نگارش زبان وجود دارد، هر ناشری برای خود به قاعده‌ای دلخواه عمل می‌کند که این تعددها موجب پریشانی و پراکندگی شده و با يك دیگر تفاوت‌هایی دارند. از دیگر دلایل می‌توان به عاریتی بودن خط فارسي و چاره‌اندیشی برای حرکات و عدم تطابق واج‌ها با حروف اشاره کرد. متصل و منفصل‌نویسی نیز یکی دیگر از حوزه‌های مورد اختلاف است از دیگر مشکلات: گوناگونی معادل‌های علمی، انواع مختلف ضبط اسامي خارجی، سرهم‌نویسی، جدانویسی، بی‌فاصله‌نویسی، انواع جمع‌ها، صورت‌های مختلف نوشتاری، آوانویسی اسامي عناصر و ترکیبات شیمیایی، سرواژه‌ها و کوتاه‌نوشت‌ها می‌باشد.

به طور کلی نقص‌هایی که برای زبان فارسي شمرده‌اند به شرح زیر می‌توان عنوان کرد:

۱. سه مصوت کوتاه یعنی حرکات زیر و زیر و پیش (اَ ، اِ ، اُ) را از نوشتن ساقط می‌کنیم. و این باعث می‌شود به جای این که از خط و نوشتار پی به معنی ببریم بایستی از معنی کلمه و جایگاه آن در جمله آنرا درست بخوانیم مانند کلمات (کَرَم، کَرَم، کَرَم، کَرَم، کَرَم) و (مَلِك، مَلِك، مَلِك، مَلِك، مَلِك) و یا سه کلمه (حَکَم، حُکَم، جَکَم) و نیز نوشتن مصوت‌های کوتاه در داخل متن باعث می‌شود که برای تلفظ صحیح اجباراً لاتین کلمات به صورت پانویس متن آورده شود که همین امر باعث اتلاف وقت و انرژی می‌شود. که البته همین لاتین‌نویسی هم قاعده خاصی ندارد و هر ناشر و نویسنده‌ای سلیقه خاص خودش را برای آوانویسی حروف فارسي به لاتین دارد. که به عنوان نمونه برای نشان دادن حرکت فتحه و الف و آ هیچ‌گونه هماهنگی در کتاب‌ها و خصوصاً فرهنگ‌های مختلف دیده نمی‌شود. "هر چند برخی معتقدند همین نوشتن حرکات مزیتی است و موجب تندنویسی می‌شود" [۵].

۲. برای يك حرف چند علامت مختلف داریم مانند علامت‌های (س، ص، ث) که هر سه در فارسي یکسان خوانده می‌شوند و هم چنین (ذ، ز، ض، ظ) و نیز (ت، ط). البته این امر در زبان انگلیسی هم وجود دارد چنان که «ف» ممکن است به شکل‌های «F. GH. PH. V» باشد.

۳. يك علامت را برای دلالت بر چند حرف مختلف استعمال می‌کنیم مانند "و" که پنج مورد نوشتن دارد یکی برای بیان ضمه در کلمات "خوش" و "تو". دیگر بیان مصوت ممدود یا "واو ماقبل مضموم" مانند "شور" و "او". سوم بیان حرف صامت "واو" در

کلماتي چون "آواز" و "والی" و "عفو". چهارم بیان حرف مصوت مرکبی که در کلمات "نو" و "جوشن" و مانند آنهاست. پنجم حرفی که در زبان کنونی خوانده نمی‌شود مانند "واو معدوله" در کلمات "خواهر" و "خواستن" و "واو" در کلمه "عمرو" [۶].
 ۴. حرف‌هایی هم هست که در کلمات خاصی از نوشتن حذف می‌شود مانند "الف" در کلمات "اسحق" و "اسمعیل" و "الله".
 ۵. نقطه‌هایی متعدد در بالا و پائین حرف که هم سبب دشواری و هم موجب اشتباه در خواندن می‌شود. اهمیت بیش از حد نقطه در خط فارسی هنگام تشخیص نوری کاراکترها [۷] تولید اشکال اساسی می‌کند. به عنوان مثال در نظر بگیرید که تفاوت «ر» و «ز» و «د» و «ذ» و «پ» و «ت» فقط در نقطه است و چون نقطه جزء بسیار کوچکی است در این امر مشکلات زیادی را فرا روی متخصصین قرار می‌دهد. و یا کلمات زیر را در نظر بگیرید که با يك یا چند نقطه عوض می‌شوند (بر، پر، تر، پز، بز، تیز).

۶. يك عیب دیگر هم که برای خط فارسی ذکر کرده‌اند این است که از راست به چپ نوشته می‌شود. و برای این مورد دلایل مختلفی ذکر شده است از جمله عدم هماهنگی و ایجاد مشکل در نوشتن متون ریاضی و شیمی و نت‌های موسیقی و دستورات شطرنج و این که خط تصویری یعنی علائم گرافیکی که در کل جهان استفاده می‌شود مانند علائم راهنمایی و رانندگی تماماً از چپ خوانده می‌شوند.

۷. پیوسته‌نویسی و جدانویسی کلمات مرکب که در اکثر موارد به صورت سلیقه‌ای عمل می‌شود مانند تنوع استفاده از «می» چسبان و غیر چسبان و یا تنوع نحوه به کار بردن «علامت‌های جمع «ها، ان، جات» ، هم، هیچ، که، (ضمایر شخصی متصل مان، تان، شان)، شناسی، را، چه، چون، تر، ترین، بی (پیشوند نفی)، به، ای (نشانه ندا)، آن و این» در کلمات به صورت پیوسته و یا جدا گانه: (آنچه ، آن چه)؛ (همچنانکه، همچنان که) ؛ (جناب‌عالی، جناب‌عالی)؛ (هیچکس، هیچکس)؛ (میتواند، می‌تواند)؛ (آن ها، آنها) در این مورد کلماتی که پیشوند و یا پسوند دارند نیز در شکل‌های مختلف نوشته می‌شوند. برخی از کلمات در دو شکل متصل‌نویسی و منفصل‌نویسی به دو شکل مختلف ظاهر می‌شوند، مانند «علاقمند و علاقه‌مند؛ اندیشمند و اندیشه‌مند». مصدرها و فعل‌های مرکب و اسم‌های مشتق از آنها نیز به دو صورت متصل و منفصل نوشته می‌شوند مانند «نگه‌داشتن و نگهداشتن». در جستجوی مطالب از اینترنت این مورد تولید اشکال می‌کند چنانکه جستجوی «هیچکس» نتایج متفاوتی را با جستجوی «هیچکس» می‌آورد و یا جستجوی «کتاب‌شناسی» و «کتابشناسی» در موتور جستجوی گوگل نتایج متفاوتی را ارائه می‌کند. این گونه کلمات با این که در خواندن متن اشکال کمی به وجود می‌آورند و هر آشنای به زبان فارسی به راحتی می‌تواند آن را بخواند اما در فن‌آوری امروزه و تجزیه و تحلیل کلمات به کمک رایانه اشکال اساسی تولید می‌کند و شاید اگر قاعده‌ای جامع و مانع برای آن وضع گردد، بتوان گفت بزرگ‌ترین مشکل خط فارسی حل شده است. منظور این که، برای مثال خواندن سه کلمه «بی‌حوصلگی، بی‌حوصلگی، بی‌حوصله‌گی» مشکلی ایجاد نمی‌کند. اما در محیط الکترونیکی و شبکه اینترنت برای بازیابی این کلمه بایستی برای تمام اشکال این کلمه، جستجو را انجام دهیم، البته اگر آگاهی از تمام اشکال نوشتاری آن داشته باشیم. آ.

۸. سی و دو حرف الفبای فارسی همراه با چهار علامت مد، همزه، تنوین، تشدید به ۱۳۰ شکل مختلف ظاهر می‌شوند و تفاوت این اشکال در اتوماسیون خط فارسی تولید اشکال می‌کند. «تنوع و تعدد نویسگان، یادگیری زبان و خط فارسی را برای آموزگار و آموزنده دشوار و برای نوآموز توان‌فرسا می‌سازد. تعداد زیاد نویسگان در رابطه با اتوماسیون زبان توسط رایانه مشکلاتی در خصوص تعداد و ترتیب قرار گرفتن نویسگان در جداول کد ایجاد می‌نماید و طراحان کد در جای دادن این تعداد نویسه در جداول با مساله کمبود جا رو به رو هستند. هر چند که مشکل جا با کد ۱۶ بیتی حل شده است اما مسایل دیگری همچنان باقی می‌مانند که احتیاج به برطرف شدن دارند» [۸]

۹. نوشتن ك و گ (ك گ گ گ گ گ) در اشکال مختلف نیز باعث سردرگمی و عدم جستجوی صحیح می‌شود.

۱۰. در اغلب اوقات يك فاصله اضافی معنی متفاوتی و یا متضادی را می‌دهد (مثل مادر، ما در).

۱۱. سه کرسی مختلف برای حرف‌های مختلف الفبا باعث می‌شود که در مقایسه با اکثر زبان‌ها تعداد سطرهای هر صفحه به مراتب بیش‌تر گردد چون برخی حروف روی خط کرسی قرار می‌گیرند و برخی پائین خط کرسی و برخی بالای خط کرسی مثل (ا ب م)

۱۲. از آنجائیکه حروف در نوشتن غالباً به صورت چسبیده و پیوسته نوشته می‌شوند و این امر تشخیص حرف به حرف نوشته به وسیله رایانه را، دچار مشکل می‌کند.

۱۳. در او. سی. آر. فارسی هم چنین اعداد نیز مشکل ساز هستند چنانچه صفر در فارسی يك نقطه كوچك است که می‌تواند رایانه را به اشتباه بیاندازد و نیز اعداد ۱ و ۲ و ۳ بسیار شبیه هم هستند و تفاوت‌شان در يك دندان كوچك است.

۱۴. تنوع املائی یا تنوع در رسم الخط بعضی از کلمات که همه شکل‌های آن نیز درست است مانند (اتاق و اطاق) و یا (امپراتور و امپراطور). و کلماتی که فقط يك شکل آنها صحیح می‌باشد ولی شکل ناصحیح آن نیز زیاد استفاده می‌شود مانند «ذغال و زغال؛ خوشنود و خشنود». البته این جدای از تنوع در مفهوم کلمات است که در دیگر زبان‌ها نیز وجود دارد، یعنی

برای بعضی از مفاهیم ممکن است کلمات متنوعی استفاده بشود. مانند کامپیوتر و رایانه.

۱۵. بکار بردن همزه در صورت‌های مختلف مانند (مساله، مسئله) ؛ (مسئول، مسؤول)

۱۶. استفاده از «ا» و «آ» به جای یکدیگر مانند (فرایند و فرآیند).

۱۷. شکل‌های مختلف ضبط نام‌های بیگانه در فارسی: ورود واژه‌های بیگانه معمولاً از راه ورود پدیده‌های فرهنگی نو در عرصه‌های مختلف فنی، علمی، اجتماعی، سیاسی و هنری و و یا از طریق افراد دو زبانه انجام می‌گیرد که به قرض‌گیری زبان معروف است و کم و بیش در تمام زبان‌ها وجود دارد. واژه‌های بیگانه اغلب برای پر کردن خلاء واژه‌های علمی و یا ارتباطی سودمند هستند، اما وجود آن‌ها مسائلی از قبیل چگونگی ضبط آن‌ها در زبان قرض‌گیرنده را به وجود می‌آورد. برای ضبط واژه‌های قرضی به سبب اختلاف فاحش نشانه‌های الفبایی فارسی با نشانه‌های الفبایی خارجی مشکلات جدی وجود دارد. از جمله این که الفبایی فارسی آوانگار نیست و به همین جهت در ضبط دقیق تلفظ واژه‌های زبان فارسی نیز ناتوان است و این ناتوانی در ضبط واژه‌های بیگانه به مراتب بیشتر است و این که در مورد برگردان اسامی خارجی به خط فارسی قاعده خاصی وجود ندارد و هر کس بنا بر سلیقه و ذوق خود این کار را انجام می‌دهد که در نتیجه یک کلمه واحد به صورت‌های مختلف نوشته می‌شود. برای مثال (اتومبیل و اتوموبیل)؛ (کلسیم، کلسیوم، کالسیوم) و یا اسم Franklin به صورت (فرانکلین، فرانکلن، فرنکلین، فرنکلن) ضبط شده است. خانم صدیق بهزادی این مشکلات را به سه دسته تقسیم کرده است: " ۱- نام‌هایی که در برگردان آن‌ها همخوانی ایجاد مشکل می‌کنند. ۲- نام‌هایی که در برگردان آن‌ها واژه‌های ساده مشکلاتی را به وجود می‌آورند. ۳- و سوم نام‌هایی که در برگردان آن‌ها مشکل اصلی مربوط به واژه‌های مرکب است [۹].

۱۸. استفاده یا عدم استفاده از «ی» در کلمات مختوم به «الف» مانند (موسی و موسا).

۱۹. استفاده یا عدم استفاده از «ء» برای کلمات مختوم به های بیان حرکت در حالت مضاف مانند (خانه مسکونی و خانه مسکونی و یا خانه‌ی مسکونی).

۲۰. استفاده یا عدم استفاده از اعراب برای کلمات.

۲۱. انواع مختلف جمع برای یک واژه مفرد: به عنوان مثال جمع بستن یک واژه با علایم جمع فارسی و علایم جمع عربی و نیز جمع بستن بی قاعده (جمع مکسر)، استفاده از جمع جمع، مانند (معلم، معلمین، معلمان، معلم‌ها).

۲۲. تنوین‌های زبان عربی نیز از جمله دشواری‌های رعایت اصل همخوانی نوشتاری و گفتاری هستند.

۲۳. در نگارش یاء وحدت یا نکره در آخر کلماتی که به هاء مختفی یا غیر ملفوظ ختم می‌شوند سه نوع املاء دیده می‌شود. (خانه‌ای، خانه‌یی، خانه).

۲۴. کلمه‌های عربی در شکل‌های گوناگون در زبان فارسی نوشته می‌شوند. (مبدأ، مبداء)؛ (ابتدا، ابتداء)؛ (نسبتاً، نسبته، نسبتاً) و

۲۵. ناتوانی خط فارسی در نشان دادن تلفظ واژه‌های ایران باستان و میانه و گویش‌ها و لهجه‌های ایرانی و واژه‌های بیگانه حتی با نشانه‌ها.

۲۶. وجود دندانه‌های متعدد در کلمات خواندن کلمات و به خصوص در اوستا، آر. فارسی ایجاد اشکال می‌کند مانند کلمات: نشستن و استشهاد.

۲۷. حروف فارسی غالباً مشابه‌اند و با اندکی غفلت به جای هم نوشته می‌شوند و مطلب را به کلی دگرگون می‌کنند مانند (در، رد، ور).

زبان و خط فارسی در اینترنت :

حجم اطلاعات به زبان فارسی در روی اینترنت در اشکال مختلف آن به سرعت رشد کرده است. در حال حاضر توسعه وبلاگ‌های فارسی و سایت‌های علمی و تبلیغاتی و دانشگاهی به زبان فارسی باعث شده است که جایگاه زبان فارسی تا حد زبان اول ارتباطات اینترنتی نزد ایرانیان و فارسی‌زبانان در سراسر جهان ارتقا یابد. شاید بتوان گفت که اولین مرجع وبلاگ‌نویسی فارسی با انتشار راهنمای ساخت وبلاگ فارسی آغاز شده است. بدون شک دومین موج نیز با شروع به کار سایت پرشین بلاگ که امکان راه‌اندازی وبلاگ برای کاربران فارسی زبان را با سهولت بیش‌تری فراهم می‌کند آغاز شده است. اما پیامد قابل توجه دیگری که رشد وبلاگ‌نویسی در ایران داشته است پیدایش سایت‌های اینترنتی فارسی زبانی است که صاحبان وبلاگ‌ها ایجاد کرده‌اند و این خود موج جدیدی از گسترش کاربرد اینترنت در جامعه ایران به حساب می‌آید. اکنون روی آوردن برخی از روزنامه‌نگاران، پژوهش‌گران، دانشجویان و به وب فارسی و استفاده از منابع خبری و علمی و آن موجب تقویت نقش رسانه‌ای وب فارسی شده است.

پدیده دیگری که باعث گسترش زبان و خط فارسی در اینترنت شده است ایجاد کتابخانه‌های دیجیتال فارسی در شبکه

جهاني است، با اين كه از شكل گيري كتابخانه هاي فارسي در شبكه جهاني مدت زيادي نمي گذرد با اين حال به سرعت در حال رشد و گسترش است. شماری از اين كتابخانه ها در پايفاه هاي اينترنتي شكل گرفته اند و بسياري وبلاگ هايي هستند كه براي اين كار راه اندازي شده اند. از ويژگي هاي اين كتابخانه ها اين است كه هيچ يك جنبه تجاري ندارند و نيز به جز عده معدودي اكثر كتابخانه ها كوشيده اند جانب بي طرفي را رعايت كرده و از اعمال سليقه شخصي پرهيز كنند. آنچه در بسياري از كتابخانه هاي مجازي فارسي در دسترس است تنها شامل كتاب نيست بلكه نوشته هايي اعم از داستان ، مقاله ، تك نگاهت و نيز در ميان مجموعه ها ديده مي شود. هم چنين است آثاري كه احتمالاً هيچ گاه چاپ كاغذي ندارند و البته وجود كتاب هايي كه مدت ها است ناپايد و مجال انتشار دوباره نيافته اند و يا آثاري كه امروز به دلایلي بازچاپ آنها مقدور نيست از جاذبه هاي كتابخانه هاي مجازي اند. در اينجا شماری از اين كتابخانه ها ذكر مي شوند: پايفاه اينترنتي كتاب هاي رايفان فارسي، پايفاه اينترنتي باني تك، كتابخانه مجازي داستان هاي فارسي، آواي آزاد، پايفاه اينترنتي خوابگرد، كتابخانه دوات، پايفاه اينترنتي سخن، وبلاگ كتابخانه هرمس، پايفاه اينترنتي گفتمان، پايفاه تاريخ و فرهنگ ايران زمين، پايفاه مركز جهاني اطلاع رساني آل البيت، كتابخانه پايفاه اينترنتي حوزه، پايفاه اينترنتي امام علي (ع)، پايفاه اينترنتي كتابخانه ديجيتال و

كه لازم به ذكر است غلبه با كتاب هاي دو حوزه ادبيات و دين است. [۱۰]

كاربران به دلایل مختلفي از قبيل " دسترسي آسان و ارزان به حجم عظيم اطلاعات ، عدم نياز اطلاعات يافته شده از اينترنت به تايپ مجدد ، دسترسي سريع و اطلاعات جديد، صرفه جويي در وقت و مهم ترين دليل، عدم تسلط اكثر کاربران به زبان انگليسي " كه زبان غالب بر اينترنت است" به دنبال اطلاعات فارسي از اينترنت هستند. گسترش زبان و انبوهي از نوشتارها ايجاب مي كند كه خط ضابطه داشته باشد و از سوي ديگر پيشرفت فن آوري و پيدايش اينترنت خواستار ضابطه و قانونمندي است. اطلاع رساني كه جنبه بين المللي پيدا كرده است بدون دستور خطي سامان يافته و نظام مند ميسر نيست و دست كم دشواري ها مي آفزيند. در حال حاضر وبلاگ هاي فارسي مقام دوم يا سوم را در جهان دارا مي باشد. به نظر دكتور آشوري " اگر زبان فارسي به همين صورت بي دقت در اينترنت به كار رود در سطح زباني براي تفنن باقي خواهد ماند و كم تر حرفي جدي به اين زبان زده خواهد شد. آينده زبان فارسي در اينترنت بستگي به اين دارد كه نويسندگان فارسي تا چه حد كار خود را جدي بگيرند و اين زبان را بازسازي كنند كه از لحاظ قدرت بيان و دقت مفاهيم و استواري ساختار دستوري به زبان انگليسي نزديك شود". [۱۱]

نيود استاندارد ثابت رسم الخط فارسي موجب اين شده است كه به تعداد صفحات وب فارسي سيك و سياق نگارش به كار رفته باشد لكن مي توان چنين ارزيابي نمود كه اكثر وب هاي فارسي در برخي خصوصيات مشترك مي باشند از جمله اين كه نگارش برخي از آنها زبان غير رسمي و محاوره اي مي باشد و به خصوص در متون علمي اغلب واژه هاي بيگانه به دفعات استفاده مي شود. رسم الخط مورد استفاده نيز متفاوت و سليقه اي است و برخي از آنها غلط هاي تايپي و نگارشي فراواني دارند و اين خصوصيات، اغلب به جهت محدوديت هاي محيط الكترونيكي و عدم تطابق رسم الخط فارسي با آن مي باشد كه نمايه سازي و سپس جستجو به اين زبان را با دشواري هايي رو به رو مي سازد.

با توجه به اين نكته كه اطلاعات ارزشمند فراواني در اينترنت وجود دارد و اينترنت با شتابي فراوان به يك منبع اطلاعاتي ممتاز تبديل شده است. موتورهاي جستجو به عنوان يكي از اساسي ترين دروازه هاي ورود به منابع اينترنتي داراي ضعف هايي هستند. كه مي توان به اين موارد اشاره كرد:

- در يك مجموعه از يافته هاي بازيايي شده مدخل هاي تكراري فراواني ملاحظه مي شود.
- نتايج غير قابل پيش بيني هستند.
- نتايج چه بسا گمراه كننده باشند: ممكن است جستجويي در يك موتور كاوش نتيجه اي نداشته، ولي در موتور ديگر داراي يافته هاي فراوان باشد.
- موتورهاي كاوش محتويات پايفاه هاي اطلاعاتي خودشان را نشان نمي دهند و از معيارهايي كه براي گنجاندن يك مدرك در فايل هايشان دارند حتي شرحي ارائه نمي كنند.
- مهار واژگاني وجود ندارد و قواعد نقطه گذاري و بزرگ نويسي نيز استاندارد نيست.
- بدون بررسي عملي هر عنصر، اغلب نمي توان ميزان ربط و رابطه ها را تحليل كرد. يعني اطلاعات كافي در مدخل نمايه نيست تا فرد بتواند دست به انتخاب بزند. [۱۲]
- عدم توان موتورهاي جستجو در تمايز ميان مداركي كه توسط فرد الف نوشته شده و مداركي كه در باره فرد الف نوشته شده است.

- منابع قابل توجهي در شبكه وب وجود دارند كه توسط موتورهاي جستجو نمايه نمي شوند. به اين بخش از وب اصطلاحاً وب نامرئي مي گويند. "وب نامرئي بخش بزرگي از وب است كه موتورهاي جستجو آنها را نمايه نمي كنند يا نمي توانند نمايه كنند و عبارتند از: سايت هاي داراي رمز عبور، فايل هاي پي. دي. اف از متون آرشيو شده، ابزارهاي تعاملتي نظير ماشين

حساب‌ها و برخی از واژه‌نامه‌ها و همچنین بعضی از پایگاه‌های اطلاعاتی، منابع محافظت شده از طریق اسم کاربر و گذرواژه، منابع و صفحات وب بدون پیوند و صفحات افزون بر حداکثر تعداد صفحات قابل مرور [۱۳].

جستجوی اطلاعات در اینترنت به دو روش می‌تواند صورت گیرد یکی استفاده از جملات زبان محاوره‌ای است و دیگری بکارگیری کلمات کلیدی. در روش استفاده از جملات زبان محاوره‌ای که اغلب به کاربران تازه‌کار پیشنهاد می‌گردد، مورد سوال خود را در قالب یک جمله سوالی مطرح می‌سازند. یکی از عیب‌های بزرگ این روش تعداد نتایج جستجوی زیادی است که بازگردانده می‌شود. به همین دلیل این روش توسط کاربران حرفه‌ای و حتی توسط همه، کمتر استفاده می‌شود. اما چنانچه از این روش استفاده بشود بایستی سعی در انتخاب بهترین نوع جمله بشود و توصیه می‌شود در انتخاب یک یک کلمات لحظه‌ای درنگ نموده و با ظرافت خاصی جمله نهایی را مطرح نمود.

یکی از کاراترین و مقتدرترین روش‌های جستجوی اطلاعات در دنیای وب استفاده از واژه‌هایی است که اصطلاحاً کلمات کلیدی نامیده می‌شوند. اغلب کاربران حرفه‌ای و جستجوگران ورزیده دنیای اینترنت می‌توانند با طرح بهترین کلمات کلیدی و بکار بستن قوانین ترکیب آن‌ها با هم برای نیازهای اطلاعاتی خود پاسخی در خور بیابند. در این روش توصیه‌های زیر برای انتخاب کلمات کلیدی و نیز جستجوی دقیق و مفید پیشنهاد می‌شود:

- ۱- حتی‌المقدور سعی شود کلمات کلیدی از میان اصطلاحات منحصر به فرد و اسامی خاص انتخاب بشود.
- ۲- حتی‌المقدور از آوردن کلمات عمومی که عناوین بسیاری را در زیر مجموعه خود شامل می‌شوند جداً خودداری کنید.
- ۳- همیشه اسم شخص یا نام شی یا هر چیز دیگری را که مد نظر دارید بطور کامل وارد کنید.
- ۴- دقت کنید که اگر موتور جستجو میان حروف بزرگ و کوچک تفاوتی می‌گذارد، این مسئله را در طرح کلمات کلیدی خود مد نظر داشته باشید.

۵- در نظر داشته باشید اگر نتیجه جستجو صفر بود به احتمال زیاد می‌تواند از یک اشتباه تایپی باشد.

۶- اگر املای صحیح و کامل کلمه‌ای را نمی‌دانید از کارکتر جانشین که اغلب * و یا ؟ است استفاده کنید.

۷- اگر یک کلمه کلیدی را برای طرح دقیق و تمام و کمال یک مورد جستجو کفایت نمی‌کند از تکنیک‌های جستجوی عبارتی، استفاده از اپراتورهای جبر بولین (AND, OR, NOT) استفاده کنید. جستجوی عبارتی یکی از مهم‌ترین و قدرتمندترین امکانات جستجو در اغلب موتورهای جستجو می‌باشد و می‌توان یک عبارت یا جمله مشخص را به همان ترتیبی که کلمات وارد شده‌اند مورد جستجو قرار داد. برای این روش جستجو عبارت مورد نظر را داخل گیومه "" بگذارید

۸- استفاده از عملگر and : AND به مفهوم "و" برای محدود کردن دامنه جستجو از طریق ترکیب کلیدواژه‌های مختلف به کار می‌رود و برای ترکیب کلیدهای جستجو زمانی که برای شما مهم است که دو یا چند کلمه کلیدی حتماً وجود داشته باشد و علامت آن در پایگاه‌های مختلف به صورت استفاده از عبارت and ، استفاده از + ، انتخاب عبارت all the word از منو، انتخاب عبارت (match on all words) بوسیله کلیک کردن بر روی دکمه‌های رادیویی می‌باشد.

۹- استفاده از عملگر OR: اپراتور OR به مفهوم "یا" و برخلاف عملگر AND باعث گسترش دامنه جستجو و بازیابی اطلاعات بیشتر شده برای ترکیب کلیدواژه‌های جستجو زمانی که انتظار دارید تنها یک، دو یا چند کلمه کلیدی حضور داشته باشند و علامت آن استفاده از عبارت or، نحوه اجرای ساده و معمولی آن، انتخاب عبارت any of the words از منو، انتخاب عبارت (match on any words) با کلیک بر روی دکمه‌های رادیویی می‌باشد. یکی از کاربردهای مهم این عملگر پوشش مفاهیم یا اصطلاحات مترادف، مرتبط، یا با املاهای متفاوت می‌باشد.

۱۰- استفاده از عملگر NOT : اپراتور Not به مفهوم "نه" و یا به جز که در این صورت تمامی جواب‌های بازگشتی که حاوی عبارت یا کلمه کلیدی هستند حذف خواهند گردید و برای اجرای آن تنها کافیست که not را قبل از عبارت یا کلمه کلیدی مورد نظران با یک فاصله بیاورید.

۱۱- استفاده از کوتاه‌سازی [۱۴] کلید واژه‌ها: این تکنیک به ما امکان می‌دهد که با وارد کردن بخشی از یک کلیدواژه بتوانیم مشتقات مختلف آن را نیز در فرآیند جستجو بازیابی کنیم. اکثر موتورهای جستجو این تکنیک را با استفاده از علامت ستاره (*) ارائه می‌دهند. یکی از مشکلات استفاده از این تکنیک این است که باعث بازیابی اطلاعات غیرمرتبط و ناخواسته زیادی می‌شود.

۱۲- استفاده از عملگر نزدیک‌یابی [۱۵]: در بسیاری از موارد استفاده از عملگر and باعث بازیابی اطلاعاتی شود که برای ما مفید نمی‌باشد، به این دلیل که این عملگر کلیدواژه‌ها را در هر کجای متن که باشند بازیابی می‌کند. در این موارد استفاده از تکنیک نزدیک‌یابی می‌تواند از ریزش کاذب اطلاعات و یا بازیابی اطلاعات غیر مرتبط جلوگیری نماید. همه موتورهای جستجو قابلیت استفاده از این تکنیک را ندارند ولی به عنوان مثال در موتور جستجوی آلتاویستا می‌توان با استفاده از عملگر NEAR از این تکنیک استفاده نمود.

۱۳- جستجوی ترکیبی با استفاده از پرانتز: این تکنیک یکی از مهم‌ترین تکنیک‌های جستجو می‌باشد که به وسیله آن

می‌توان تا حدود زیادی از بازیابی موارد غیر مرتبط در محیط وب جلوگیری کرد. در این روش می‌توان از همه عملگرهای جستجو که در بالا گفته شده یک‌جا استفاده کرد و آن‌ها را با هم‌دیگر ترکیب نمود.

۱۴ - جستجوی کلیدواژه در عنوان صفحات وب: این تکنیک با این پیش فرض که عنوان یک صفحه وب تا حدود زیادی نمایان‌گر محتوای اطلاعات موجود در آن است به جستجوی واژه‌های کلیدی در عنوان سایت‌ها می‌پردازد. علامت آن در موتورهای جستجو متفاوت است ولی اغلب موتورهای جستجو از طریق فهرست انتخابی و یا گزینه‌های دیگر این امکان را فراهم می‌آورند.

۱۵ - جستجوی حوزه سایت‌ها: با توجه به این که به صورت قراردادی هر کشوری حوزه خاصی در محیط وب دارد، قابلیت جستجوی حوزه سایت‌ها به ما این امکان را می‌دهد که فرایند جستجو را به حوزه خاصی نظیر سایت‌های وب ایران (ir) و یا سایت‌های وب سازمان‌های غیر انتفاعی (org) محدود کنیم. دستورات استفاده از این تکنیک در موتورهای جستجو مختلف می‌باشد.

۱۶ - محدود کردن جستجو به زبان‌های مختلف؛ باعث می‌شود نتایج جستجو به زبان‌های دیگر آورده نشود و انتخاب مطلب مورد نظر آسان‌تر است.

۱۷ - محدود کردن جستجو به تاریخ انتشار منابع در وب: تاریخ انتشار یا به اصطلاح روزآمدی مطلب به خصوص در منابع علمی اصل مهمی است و این‌گونه محدودیت باعث می‌شود بنا به نیاز کاربر جدیدترین و یا قدیمی‌ترین منبع بازیابی بشود.

۱۸ - جستجوی رسانه‌های مختلف: موسیقی، عکس، ویدئو؛ زمانی که فقط نوع خاصی از رسانه مورد نیاز است به عنوان مثال زمانی که به عکس یک شخصیت نیاز داریم، جستجو در میان عکس‌ها باعث می‌شود نتیجه جستجو شامل اطلاعات دیگری در مورد آن شخصیت نباشد.

۱۹ - جستجوی صفحات با فرمت‌های مختلف: PDF, Word, MP3, MPEG, زمانی که فرمت خاصی مورد نظر است می‌توان از این تکنیک استفاده کرد. به عنوان مثال اگر مایل باشیم منبع بازیابی شده در فرمت PDF باشد، این تکنیک می‌تواند مفید باشد.

۲۰ - آگاهی از پیش‌فرض‌های جستجو در موتور جستجو: با توجه به این که هر موتور جستجو برای ترکیب واژه‌ها یک پیش‌فرض دارد و اگر از هیچ گونه عملگری استفاده نشود، کلیدواژه‌ها را به صورت پیش‌فرض با یکی از عملگرهای جبر بولی ترکیب می‌کند؛ آگاهی از این پیش‌فرض موتورهای جستجوی مختلف مهارت ما را در جستجو بالا می‌برد.

۲۱ - وب نامرئی: وب نامرئی به دو دلیل کمی و کیفی اهمیت دارد کمی از این نظر که موتورهای جستجو فقط قادر هستند حدود ۱۶ درصد از اطلاعات موجود در اینترنت را بازیابی کنند و اندازه وب نامرئی تقریباً ۵۰۰ برابر وب مرئی است و کیفی از این نظر که منابع اطلاعاتی موجود در وب عمیق معمولاً ارزشمند و مفید هستند و در بسیاری از موارد پاسخ‌گویی نیاز کاربران می‌باشند. آشنایی با ابزارهایی که برای شناسایی منابع وب نامرئی به وجود آمده‌اند و کاربران را به سایت‌های مناسب راهنمایی می‌کنند، باعث دسترسی به این بخش عظیم از اطلاعات مفید و ارزشمند می‌شود. مثل سایت Invisibleweb که فهرستی از منابع نامرئی را و سایت Completeplaset که فهرستی از تقریباً ۴۰۰۰۰ پایگاه اطلاعاتی وب نامرئی را ارائه می‌دهد، [۱۶]

راهبرد جستجو در اینترنت

جستجو عبارت از جستجو در منابعی مشخص با استفاده از کلیدواژه‌ها و عبارت‌های خاص در حوزه‌های موضوعی ویژه است. طراحی نظام‌مند مراحل انجام یک جستجو را راهبرد جستجو می‌گویند به نظر پائو «راهبرد جستجو عبارت است از فرایندی که از طریق آن فایلی مورد جستجو قرار می‌گیرد تا مدارک متناسب با نیاز کاربر شناسایی شود. این مدارک بر اساس مجموعه‌ای از معیارهایی که شخص متقاضی مطرح می‌کند بازیابی می‌شود» [۱۷] هر فرایند جستجو می‌تواند به مراحل ارائه درخواست دقیق، انتخاب منابع اطلاعاتی مناسب، آماده کردن جستجو و اجرای جستجو تقسیم شود. بر خلاف منابع نمایه‌سازی شده در پایگاه‌های اطلاعاتی کتاب‌شناختی، مدارک در اینترنت از طریق واژگان کنترل شده قابل بازیابی نیستند. بنابراین جستجوگر برای بازیابی باید بر فنون خاص اینترنت متکی باشد. نخست آگاهی از ابزارهای مختلف جستجو در اینترنت و در ادامه انتخاب یکی از این ابزار برای جستجوی اطلاعات مورد نیاز می‌باشد. هزاران موتور جستجو، صدها ابرموتور جستجو و راهنماهای موضوعی وب و پایگاه‌های تخصصی وجود دارد و انتخاب درست ابزار جستجو در ابتدای کار جستجو می‌تواند یک جستجوی موفق را باعث گردد. در زیر چند معیار برای انتخاب ابزار جستجو آورده می‌شود:

- اگر در جستجوی اطلاعات خاصی باشید بهتر است از موتورهای جستجو استفاده کنید.
- اگر در جستجوی یک واژه مبهم یا منحصر به فرد هستید از ابرموتورهای جستجو استفاده نمایید.
- اگر در جستجوی اطلاعات عمومی روی موضوعات عام هستید از راهنماهای موضوعی وب استفاده کنید.

- اگر در حال جستجوی اطلاعات علمی هستید از کتابخانه‌های مجازی استفاده کنید.

- اگر در جستجوی آخرین اطلاعات یا برای تغییر پویای فهرست مطالب، آخرین خبرها، راهنماهای دفتر تلفن، دسترسی به زمان پروازهای هوایی و غیره هستید از پایگاه‌های تخصصی استفاده کنید.[۱۸]

برای جستجوی اطلاعات از اینترنت چهار شیوه وجود دارد شیوه نخست دسترسی به اطلاعات از طریق نشانی پایگاه اطلاعاتی مورد نظر بر روی اینترنت (URL) است، که در این صورت نشانی پایگاه اطلاعاتی در سطر نشانی برنامه مرورگر وب تایپ می‌شود و برنامه مرورگر وب مراجعه کننده را به وب سایت آن نشانی هدایت خواهد کرد. اما اگر فقط یک حرف یا علائم نقطه‌گذاری از قلم بیفتد، برنامه مرورگر نخواهد توانست آن پایگاه را باز نماید. روش دوم دنبال کردن لینک‌های موجود در صفحات وب است که کاربران را از صفحه‌ای به صفحه دیگر هدایت می‌کند. این سهولت دسترسی به منابع در وب از امتیازات بزرگ آن است و برای کاربران امکان مرور سریع و آسان در منابع مختلف را فراهم می‌کند. روش سوم بازیابی گزینشی اطلاعات است که در آن در واقع به جای آنکه کاربران شخصا در جستجوی اطلاعات مورد نظر باشند، موضوعات مورد نیاز خود را به سیستم‌های بازیابی گزینشی می‌سپارند و سپس در طول زمان، اطلاعات دریافتی جدید توسط سیستم برای آنها به طور خودکار ارسال خواهد شد.

چهارمین روش که در واقع معمول‌ترین و متداول‌ترین راه بازیابی اطلاعات در وب است استفاده از موتورهای جستجو است. هنگام جستجو باید دقت کرد که موتور جستجو به طور معمول هوشمند نیست و معمولاً به دنبال کلیه کلیدواژه‌هایی که شما به دستگاه داده‌اید بدون توجه به معنای آنها می‌گردد.

نکات کلیدی جستجو به زبان فارسی

برای جستجوی مطالب فارسی طبق گفته‌های پیشین چنانچه آدرس سایت به خصوصی که در زمینه موضوعی مورد نظر ما فعالیت می‌کند را داشته باشیم؛ می‌توان مستقیماً به آن سایت رفته و از مطالب آن استفاده نمود. به عنوان مثال سایت تخصصی برنامه‌نویس مطالب مفیدی در زمینه رایانه و علوم وابسته، به ما ارائه می‌دهد و یا سایت عمران در زمینه موضوعی عمران فعالیت می‌نماید و نیز سایت‌های انجمن ریاضی در زمینه ریاضی، سایت انجمن فیزیک ایران در زمینه فیزیک، سایت انجمن روان‌شناسی ایران در زمینه روان‌شناسی و علوم تربیتی فعالیت می‌نمایند، مرکز اطلاعات و مدارک علمی ایران با دارا بودن پایگاه‌های اطلاعاتی مختلف مخصوصاً پایگاه پایان‌نامه‌ها می‌تواند مورد استفاده متخصصین تمام رشته‌ها گردد. ولی چنانچه امکان استفاده از این سایت‌ها نباشد و یا آدرس این سایت‌ها را نداشته باشیم بایستی مطلب مورد نظر خود را بوسیله یکی از موتورهای جستجو پیدا بکنیم.

انتخاب موتور جستجو عامل مهمی در فرایند جستجو است. در حال حاضر ابزارهای کاوش مختلفی در ایران ظهور پیدا کرده‌اند. لیکن ابزارهای جستجویی که امکان جستجوی اطلاعات به زبان فارسی را در اختیار قرار می‌دهند، محدودند. از طرف دیگر، امکانات و قابلیت‌های آنها برای بازیابی موثر و مناسب اطلاعات متغیر هستند. برخی از ابزارهای کاوش با امکانات جستجوی فارسی عبارتند از: ان.پی. ایران‌NPiran، ایران‌هوIranhoo، ایران‌مه‌IranMehre، پارسیک Parseek، گوگل Google.

در بین ابزارهای کاوش فوق، تنها موتور کاوش گوگل دارای برنامه روبات به منظور شناسایی و نمایه‌سازی صفحات یا سایت‌های وب به زبان فارسی و نمایه‌سازی خودکار می‌باشد و قادر است صفحات فارسی را در قالب یونی‌کد شناسایی و در پایگاه خود نمایه کند و سایت پارسیک نیز از پایگاه گوگل برای جستجو و بازیابی اطلاعات استفاده می‌کند. به تعبیر دیگر، چهار ابزار کاوش دیگر توسط نمایه‌سازی انسانی اداره می‌شوند و از این لحاظ راهنمای موضوعی تلقی می‌شوند و انسان، فرآیند شناسایی، بررسی و نمایه‌سازی سایت‌ها یا صفحات وب را بر عهده دارد.[۱۹]

معمولاً به جهت دامنه وسیع موضوعی و نیز صفحه به زبان فارسی گوگل اکثر کاربران از این موتور جستجو استفاده می‌نمایند. برای جستجوی بهتر توجه به نکات زیر ضروری به نظر می‌رسد:

- با ترکیب چند واژه کلیدی مهم خیلی سریع می‌توانیم مطلب مورد نظر خود را بدست بیاوریم.
- دقت در انتخاب کلید واژه‌ها به طوری‌که واژه‌های انتخابی بطور دقیق نماینده نیاز اطلاعاتی ما باشند کمک خواهد کرد تا از نتایج جستجوی گسترده‌ای که در اکثر موارد بار اطلاعاتی مفید ندارند دوری گزینیم.
- استفاده از تکنیک جستجوی عبارتی که در آن عبارت جستجوی مورد نظر خود را داخل گیومه " " می‌گذاریم و به این ترتیب به موتور جستجو می‌گوییم که مطلب مورد نظر ما بایستی عین این عبارت باشد، نیز در محدود کردن نتایج جستجو کمک فراوان می‌کند.
- استفاده از انواع محدودگرهای زبانی، زمانی، مکانی، شکلی، و موضوعی و ... در جستجوی پیشرفته گوگل به ما در رسیدن سریعتر به مطلب مورد نظر کمک فراوانی می‌کند.

- به علت این که منابع و اطلاعات موجود در اینترنت بوسیله افراد مختلف و بدون کنترل در شیوه‌های رسم‌الخط و بدون ویرایش صاحب‌نظران منتشر می‌گردد آشنایی با گونه‌های مختلف نوشتاری و املاهای مختلف یک واژه و یا یک مفهوم در زبان فارسی به ما کمک می‌کند که با جستجوی گونه‌های مختلف نوشتاری یک واژه یا یک مفهوم و استفاده از واژه‌های مترادف و متشابه و شکل‌های دیگر نوشتاری آن واژه و نیز استفاده از انواع شکل‌های جمع و مفرد یک واژه جامعیت جستجوی خود را بالا ببریم. به عنوان مثال برای جستجوی مطلبی در زمینه بتن بایستی آنرا به دو صورت «بتون» و «بتن» جستجو نمائیم تا به تمام مطالبی که در زمینه بتن می‌باشد دسترسی داشته باشیم و یا به عنوان مثال دوم برای جستجوی مطلبی در باره «آبگرمکن» برای دستیابی به همه اطلاعات موجود بایستی آن را به چهار شکل زیر بنویسیم «آب گرم کن، آب گرمکن، آبگرم کن، آبگرمکن» واضح است که هر کدام از این کلمات نتایج متفاوتی را در موتور جستجو بدست می‌دهد. «استاد، اساتید، استادان، استاده‌ها» «آمریکا، امریکا» «تیدروژن، هیدروژن» «آنلاین، پیوسته، درون خطی» از مثال‌های دیگری هستند که جستجو به تمام این شکل‌ها جامعیت جستجوی ما را زیاد می‌کند و به ما در از دست ندادن مطالب مفید کمک می‌کند.
- با استفاده از عملگرهای بولی، دقت جستجو را بالا برده و نتایج جستجوی کم و مفیدی را بدست می‌آوریم.
- مترادفات: با استفاده از شکل‌های مختلف مترادفات موجود برای یک مفهوم و هم چنین شبه مترادفات و یا حتی گاهی کلمات متضاد مثل بی‌سواد و سوادآموزی در جستجو می‌توان جامعیت جستجو را بالا برد.
- اسامی مشهور و اسامی علمی: آگاهی از شکل‌های مختلف اسامی علمی و مشهور عامیانه و اسامی تجاری یک پدیده و یا وسیله و ... و استفاده از آن‌ها می‌تواند جامعیت جستجو را بالا ببرد.
- با توجه به این که در اکثر وب‌ها از واژه خارجی یک کلمه به همان صورت و با همان الفبا استفاده می‌شود استفاده از شکل خارجی این لغات و واژه‌ها نیز می‌تواند جامعیت جستجوی ما را بالاتر ببرد.
- با توجه به این که در برخی از سایت‌ها و نیز وبلاگ‌ها روش خاصی برای رفع مشکلات فارسی پیشنهاد کرده‌اند و واضح است که خودشان نیز از آن رسم‌الخط استفاده می‌کنند، آگاهی از این شکل‌های مختلف و جستجو به این شکل‌ها می‌تواند باعث جامعیت جستجو گردد. از این موارد می‌توان به حذف واو معدوله در برخی سایت‌ها و وبلاگ‌ها اشاره کرد که به عنوان مثال «خواهر» را به صورت «خاهر» می‌نویسند و یا حذف تنوین در برخی منابع که به عنوان مثال «عملاً» را به صورت «عملن» می‌نویسند.
- در نهایت این که «در تشکیل صفحات وب فارسی، جای یک استاندارد حاکم بر عملکرد تالیف نویسندگان وب، خالی است. استاندارد که انتخاب بعضی کلمات دارای چندین رسم‌الخط و حتی انتخاب بعضی کلمات که بر مفاهیم متنوعی دلالت دارند را منحصر به فرد نماید و مولفان را از طرفی ترغیب به انتخاب گونه زبانی مناسب، برای تضمین کیفیت ارتباط و انتقال مؤثر پیام و از طرف دیگر موظف به حفظ سلامت زبان و رعایت استانداردهای آن به عنوان یک وظیفه رسانه‌ای نماید. ایجاد و گسترش چنین استانداردی به عهده "فرهنگستان زبان و ادب فارسی" و با هماهنگی انجمن‌ها و شوراهای علمی یا صنفی انفورماتیک در ایران است. تعویق در تنظیم این استاندارد، با توجه به رشد روز افزون وب‌های فارسی زبان، هزینه‌های جبران ناپذیری در بر خواهد داشت.» [۲۰]

[۱] بر گرفته از: محمد صابر راثی ساریانقلی. " بررسی مشکلات جستجو و بازیابی اطلاعات به زبان فارسی از اینترنت با مطالعه موردی بر روی کاربران مرکز اینترنت دانشگاه آزاد اسلامی واحد شبستر ". پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشگاه آزاد اسلامی، واحد تهران شمال، ۱۳۸۴

[۲] کارشناس ارشد کتابداری و اطلاع‌رسانی دانشگاه آزاد اسلامی واحد شبستر

[3] accents

۴ پرویز نائل خانلری. زبان‌شناسی و زبان فارسی. (تهران: توس، ۱۳۷۳). ص. ۲۵۶

[۵] مجتبی مینوی. مینوی بر گستره ادبیات فارسی، به کوشش ماه منیر مینوی. (تهران: توس، ۱۳۸۰)، ص. ۵۱۰

[۶] واو معدوله واوی است که در این زمان عموماً نوشته می‌شود ولی خوانده نمی‌شود، مانند خواهش. اما در زمان قدیم آن را با کیفیت خاصی تلفظ می‌کرده‌اند و چون در هنگام تلفظ ضمه به فتحه عدول می‌کرده‌اند، آن را واو معدوله نامیده‌اند. هنوز در برخی از لهجه‌ها تلفظ آن به صورت قدیم مانده است. پیش از واو معدوله همیشه حرف «خ» و پس از آن یکی از حروف «د،ر،ز،س،ش،ن،و،ه،ی» آمده است.

[۷] OCR= Optical Character Reader فرایندی که در طی آن یک وسیله الکترونیکی کاراکترهای چاپ شده بر روی کاغذ را آزمایش می‌کند و شکل آن‌ها را با بررسی الگوهای تیره و روشن تعیین می‌کند. پس از تعیین اشکال توسط اسکنر یا وسیله مورد استفاده برای خواندن، روش‌های تشخیص نوری کاراکترها برای تبدیل اشکال به متون کامپیوتری مورد استفاده قرار

می‌گیرند. (فرهنگ تشریحی اصطلاحات کامپیوتری میکروسافت، مترجم فرهاد قلی‌زاده نوری، [تهران: کانون نشر علوم، ۱۳۷۹]، ص. ۴۴۵)

[۸] محمداصادق محقق زاده، کاظم زارعیان، "ارائه راه حل برای برخی مسائل اتوماسیون و نگارش فارسی" فصلنامه اطلاع‌رسانی، (دوره ۱۹، شماره ۳ و ۴) ص.

[۹] ماندانا صدیق بهزادی، "ناهماهنگی ضبط نام‌های بیگانه در فارسی"، فرهنگ (کتاب سیزدهم، زمستان ۱۳۷۱) ص. ۱۰۳-۱۱۶

[۱۰] مجید رهبانی، "قند پارسی در شبکه جهانی: کتاب‌های دیجیتال و کتابخانه‌های مجازی فارسی در اینترنت"، کتاب، (۱۸۳) ص.

[۱۱] "تابوی اصلاح خط"، جام جم، (۴ بهمن ۱۳۸۳)

[۱۲] براندا پاریس سیبلی، "فهرست‌نویسی منابع اینترنت: سازماندهی وب در کتابخانه‌های محلی و غیر آن" ترجمه محسن حاجی زین العابدینی، فصلنامه اطلاع‌رسانی (دوره ۱۶، شماره ۳ و ۴) ص. ۱

[۱۳] عبدالرسول خسروی، "وب نامرئی" فصلنامه اطلاع‌رسانی (دوره ۲۰، شماره ۱ و ۲) ص. ۵۳

[14] Truncation

[15] Proximity search

[۱۶] عبدالرسول خسروی، "وب نامرئی" فصلنامه اطلاع‌رسانی (دوره ۲۰، شماره ۱ و ۲) ص. ۵۴

[۱۷] میرندا لی پائو، مفاهیم بازیابی اطلاعات، ترجمه اسدالله آزاد و رحمت الله فتاحی، (مشهد: دانشگاه فردوسی، ۱۳۷۸) ص. ۳۱۴

[۱۸] دانیل بازاک، "جستجوی وب بطور کارآمدتر: رهنمودها، فنون و راهبردها"، مترجمین مریم اسدی، اکرم اسدی، مجله الکترونیکی مرکز اطلاعات و مدارک علمی ایران، شماره چهارم دوره دوم

[۱۹] کیوان کوشا، "معیارهای ارزیابی ابزارهای کاوش اینترنت: مطالعه مقایسه‌ای بر روی ابزارهای کاوش وب با واسط جستجوی فارسی"، مجله الکترونیکی کتابدار.

[۲۰] محسن صدیقی، کامران زمانی‌فر، "روشی برای رفع چالشهای محتوا کاوی وب‌های فارسی زبان" مجله الکترونیکی مرکز اطلاعات و مدارک علمی

منابع و مأخذ

۱- آزادی، قاسم، "اینترنت: سازمان‌دهی و جستجو"، ابرار اقتصادی، ۱۴.

۲- ابوالقاسمی، محسن، تاریخ زبان فارسی، تهران: سمت، ۱۳۸۰.

۳- احمدی فصیح، صدیقه، "آشنایی با شبکه جهانی وب"، فصلنامه اطلاع‌رسانی، دوره ۱۸، شماره ۱ و ۲.

۴- ادیب سلطانی، میر شمس‌الدین، راهنمای آماده ساختن کتاب: برای مولفان، مترجمان، ویراستاران، رسانه‌گران، کتابداران...، تهران، علمی و فرهنگی، ۱۳۸۱.

۵- ----- درآمدی بر چگونگی شیوه‌ی خط فارسی، تهران: امیرکبیر، ۱۳۷۸.

۶- اشرف‌زاده، بهرام، "زبان فارسی در وبلاگ‌های فارسی".

<http://www.persianfarsi.com/articles/zabaneweblog.htm>

۷- بازاک، دانیل، "جستجوی وب بطور کارآمدتر: رهنمودها، فنون و راهبردها"، مترجمین مریم اسدی، اکرم اسدی،

http://www.irandoc.ac.ir/data/E_J/vol2/Search_Web.htm

۸- باقری، مه‌ری، تاریخ زبان فارسی، تهران: قطره، ۱۳۷۸.

۹- بهار، محمدتقی، سبک‌شناسی یا تاریخ تطور نثر فارسی، تهران: امیرکبیر، ۱۳۷۰.

۱۰- پائو، میراندا لی، مفاهیم بازیابی اطلاعات، ترجمه اسدالله آزاد و رحمت الله فتاحی، مشهد: دانشگاه فردوسی، ۱۳۷۸.

۱۱- "تابوی اصلاح خط"، جام جم، ۴ بهمن ۱۳۸۳.

۱۲- خانلری، پرویز، زبان‌شناسی و زبان فارسی، تهران: توس، ۱۳۷۳.

۱۳- خسروی، عبدالرسول، "وب نامرئی"، فصلنامه اطلاع‌رسانی، دوره ۲۰، شماره ۱ و ۲.

۱۴- خلخالی، نازیلا، بررسی علمی شیوه‌ی خط فارسی، تهران: ققنوس، ۱۳۷۵.

۱۵- خوانساری، حیران، "تکامل وب و مقایسه ابزارهای جستجو در اینترنت"، فصلنامه اطلاع‌رسانی، دوره ۱۶، شماره ۳ و ۴.

- ۱۶- داورپناه، محمد رضا. جستجوی اطلاعات علمی و پژوهشی در منابع چاپی و الکترونیکی؛ شامل حوزه های علوم، فنی مهندسی ... تهران: دبیزش، ۱۳۸۱.
- ۱۷- دراگولانسکو، نیکلای جورج. "ارزیابی کیفی وب سایت ها: ابزارها و معیارها". ترجمه غلام حیدری.
http://www.irandoc.ac.ir/Data/E_J/vol4/haidari.htm
- ۱۸- دشتی، افشین. "بررسی سه پیشنهاد در شیوه نگارش خط فارسی حتما مثلن، خاهر". روزنامه شرق ۱۰ تیر ۱۳۸۳.
- ۱۹- "دگرگونیهای زبان و خط فارسی در محیطهای رایانه ای: گفتگو با دکتر عاصی". پیام ارتباطات، ۳۵.
- ۲۰- "ده نکته برای جستجوی سریع در گوگل". دانشمند، ۴۸۷.
- ۲۱- رقابی، فرنوش؛ شریفی، شهرزاد. "نگاهی به اینترنت و نقش آن در دستیابی به منابع رایگان پژوهش".
http://www.irandoc.ac.ir/data/E_J/vol3/sharfi_reghabi_2.htm
- ۲۲- رهبانی، مجید. "قند پارسی در شبکه جهانی: کتابهای دیجیتال و کتابخانه های مجازی فارسی در اینترنت". جهان کتاب، ۱۸۳.
- ۲۳- رئیس، محمد رضا. "OCR : آموزش الفبای فارسی به رایانه".
<http://iranwsis.org/Default.asp?C=IRNW&R=&I=93>
- ۲۴- سیلی، برندا پاریس. "فهرست نویسی منابع اینترنت: سازمان دهی وب در کتابخانه های محلی و غیر آن". ترجمه محسن حاجی زین العابدینی . فصلنامه اطلاع رسانی، دوره ۱۶، شماره ۳ و ۴.
- ۲۵- صدیق بهزادی، ماندانا، "ناهماهنگی های ضبط نام های بیگانه در فارسی". فرهنگ. کتاب سیزدهم، زمستان ۱۳۷۱.
- ۲۶- صدیقی، محسن، زمانی فر، کامران. "روشی برای رفع چالشهای محتوای کای وب های فارسی زبان".
http://www.irandoc.ac.ir/Data/E_J/vol4/shahidi.htm
- ۲۷- طباطبایی، علاءالدین. "در دشواری های رایانه ای زبان فارسی". نشر دانش، ۱۰۳.
- ۲۸- عاصی، مصطفی. "نقش رایانه در ایجاد استانداردهای زبانی". فرهنگ. سال چهاردهم، شماره های اول - دوم، بهار - تابستان ۱۳۸۰.
- ۲۹- عزیز محمدی، فاطمه، "بررسی برخی فرآیندهای رایج قرض گیری در زبان فارسی". فصلنامه اطلاع رسانی، دوره ۱۸، شماره ۳ و ۴.
- ۳۰- فتاحی، رحمت الله. "چالش های سازمان دهی منابع دانش در آغاز قرن بیست و یکم با نگاهی بر دانش فهرست نویسی در ایران". فصلنامه کتاب، ۴۸.
- ۳۱- فرهنگ تشریحی اصطلاحات کامپیوتری میکروسافت. مترجم فرهاد قلی زاده نوری. تهران: کانون نشر علوم، ۱۳۷۹.
- ۳۲- قاسمی، علی حسین، اطلاع یابی در اینترنت. تهران: چاپار، ۱۳۸۰.
- ۳۳- کوشا، کیوان، ابزارهای کاوش اینترنت : اصول، مهارتها و امکانات جستجو در وب . تهران: نشر کتابدار، ۱۳۸۱.
- ۳۴- ----- "معیارهای ارزیابی ابزارهای کاوش اینترنت: مطالعه مقایسه ای بر روی ابزارهای کاوش وب با واسط جستجوی فارسی".
- <http://www.ketabdar.org/magazine/detailarticle.asp?number=25>
- ۳۵- کوک، آلیسون. راهنمای یافتن اطلاعات با کیفیت در اینترنت، راهبردهای گزینش و ارزیابی. ترجمه مهدی خادمیان. مشهد: کتابخانه رایانه ای، ۱۳۸۲.
- ۳۶- گزنی، علی. "جست و جوی اطلاعات و ساز و کارهای بهینه سازی آن". فصلنامه کتاب، ۴۵.
- ۳۷- گلاسبرنر، آلفرد. کلید طلایی جستجو در اینترنت. ترجمه رضا مجری، لیلا ملکان، عبدالله تبار. تهران: انتشارات خلیج فارس، ۱۳۸۲.
- ۳۸- محقق زاده، محمدصادق ؛ زارعیان، کاظم. "ارائه راه حل برای برخی مسائل اتوماسیون و نگارش فارسی" فصلنامه اطلاع رسانی شماره ۳ و ۴ دوره ۱۹.
- ۳۹- محمدی فرد، داود؛ آباقری، محمد. کامپیوتر برای کتابداران و اطلاع رسانی. تهران: چاپار، ۱۳۸۳.
- ۴۰- مختاری نبی، ابراهیم. "سازمان دهی منابع اینترنت: چالشها و ضرورتها".
- http://www.irandoc.ac.ir/data/E_J/vol1/organaizing.htm
- ۴۱- مرتضائی، لیلا . "مسائل زبان و خط فارسی در ذخیره و بازیابی اطلاعات". فصلنامه اطلاع رسانی. دوره ۱۷. شماره ۱ و ۲ .
- ۴۲- منصوریان، یزدان "عوامل موثر بر جستجو و بازیابی اطلاعات در شبکه جهانگستر وب".
- <http://www.ketabdar.org/magazine/detailarticle.asp?number=23>

۴۳- مینوی، مجتبی. مینوی بر گستره ادبیات فارسی. به کوشش ماه منیر مینوی. تهران: توس، ۱۳۸۰.

۴۴- "نگاهی به مشکلات خط فارسی در ارتباط با فناوری اطلاعات".

<http://www.itna.ir/archives/report/001948.htm>

۴۵- نوتس، گری. راهبردها و شیوه‌های جستجو در اینترنت. ترجمه سیمین نیازی. فصلنامه کتاب،

۴۶- نوری، علی‌رضا. "جستجو در اینترنت: آشنایی با موتور جستجوی گوگل". فصلنامه اطلاع‌رسانی، دوره ۱۶، شماره ۳ و ۴