

خط فارسی و مهارت در جستجوی اطلاعات فارسی از اینترنت

برای استادان و طلاب جامعه المصطفی اصفهان

چکیده:

خط فارسی دارای مشکلات مختلفی می-باشد که در جستجو و بازیابی اطلاعات مسائل و مشکلات فراوانی را فراروی کاربران اینترنت قرار می-دهد. به خصوص با رشد سریع انتشارات الکترونیکی بر روی وب در شکل‌های 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 یک محمل اطلاعاتی عظیم، منابع اطلاعاتی را در مقیاسی وسیع در دسترس مخاطبان بالقوه قرار داده است. اغلب سهولت دسترسی به منابع اطلاعاتی اعم از متن و سایر رسانه‌ها عمده‌ترین مزیت اینترنت محسوب می‌شود. اما این توانایی که هرکس ناشر آثار خود باشد عواقب ناخواسته‌ای 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 یک سازمان-دهی مفید و موثر است. هرچند در حال حاضر راهنماهایی برای منابع اینترنتی تهیه شده است که براساس فایل‌های مقلوب ساخته شده توسط موتورهای جستجو و با استفاده از قابلیت‌های مختلف این موتورها از جمله : استفاده از عملگرهای بولی، جستجوی دقیق عبارت، محدود کردن یک جستجو به بخش خاصی از رکورد (مانند عنوان ، آدرس) ، کوتاه‌سازی کلمات، جستجوی نزدیک‌یابی واژه‌ها، ایجاد محدودیت زمانی و منطقه‌ای و زبانی، و به جستجوی اطلاعات کمک می-کند، اما باید تاکید کرد که در امر بازیابی اطلاعات از اینترنت بدون نمایه‌سازی نظام یافته نمی‌توان انتظار بازیابی مفید و موثر را داشت. هرچند بیش‌تر 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 می‌دهند و از آنجائی-که زبان غالب در اینترنت انگلیسی است جستجو به زبان‌های 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 می‌شود. دقت فراوان در ثبت همه دقایق تلفظ اغلب موجب دشواری شیوه خط است و این دقت زمانی ضرورت می‌یابد که زبانی توسعه بسیار بیابد و در کشورهای دیگری که به آن زبان سخن نمی‌گویند رایج شود. به عنوان مثال در خط عربی نقطه و علامت‌های حرکات وقتی به وجود آمد که زبان عربی نزد ملت‌های غیر عرب معمول شد، در خط یونانی نیز نشانه‌های آهنگ و تکیه [۳] پس از رواج آن زبان در مصر ایجاد شد تا کسانی که زبان مادری‌شان یونانی نبود و با تلفظ آن مانوس نبودند بتوانند کلمات و عبارات یونانی را هر چه درست‌تر ادا کنند. با این حال هیچ خطی هر قدر دقیق و شماره‌علامات آن فراوان باشد، ممکن نیست که کاملاً نشانه شیوه تلفظ باشد. و با کمک علامات متعدد علم حروف نیز تا کسی چگونگی تلفظ زبانی را نشنود نمی‌تواند 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 املائی اصلی در نوشتن به کار می‌رود که اغلب با املائی کلمه مشابه در زبان ثانوی تفاوت دارد و از اینجا برای اصوات واحد علائم خطی متعدد پدید می‌آید. در خط فارسی نمونه همه این موارد را می‌توان یافت. چون خط عربی برای نوشتن فارسی به کار رفت کلماتی که از آن زبان اخذ شده بود به همان صورت اصلی نوشته شد. حال آنکه به یقین در هیچ دوره‌ای حروف خاص عربی را فارسی 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 فارسی حرف "ی" 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 فارسی 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 فارسی نوشتن "واو معدوله" و "ه‌اء 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 فارسی سه مصوت کوتاه (َ ، ِ ، ُ) از این قبیل است هم چنین الف در کلمات اسحق و الله که در کتابت نمی‌آید. [۴]

زبان و خط فارسی نیز مشکلات خاصی را دارا می‌باشد و نظام نوشتاری فارسی برای ثبت دقیق گفتار، نارسائی دارد و قواعد نگارش آن مدون نیست، از این رو فاصله میان گفتار و نوشتار در فارسی قابل توجه است. بیش‌ترین مشکلات نیز به جهت نبود یک 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 رسمی 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 یک-دیگر تفاوت‌هایی دارند. از دیگر دلایل می‌توان به عاریتی بودن خط فارسی 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 اسامی خارجی، سرهم‌نویسی، جدانویسی، بی‌فاصله‌نویسی، انواع جمع‌ها، صورت‌های مختلف نوشتاری، آوانویسی اسامی عناصر و ترکیبات شیمیایی، سرواژه‌ها و کوتاه‌نوشت‌ها می‌باشد.

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

۱. سه مصوت کوتاه یعنی حرکات زیر و زیر و پیش (َ ، ِ ، ُ) 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 آن-را درست بخوانیم مانند کلمات (کَرَم، کَرَم، کَرَم، کَرَم) و (مَلک، مَلک، مَلک، مَلک) و یا سه کلمه (حُکَم، حُکَم، حُکَم) 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 فارسی 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 یک حرف چند علامت مختلف داریم مانند علامت‌های (س، ص، ث) که هر سه در فارسی 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. یک 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 کلماتی 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 یا تفاوت «د» و «ذ» و یا تفاوت «ب» «ت» «پ» «ث» 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 یک یا چند نقطه عوض می‌شوند (بر، بر، پر، پر، تر، پز، بز، تز).

۶. یک 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، راه، 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)؛ (آن‌ها، آنها) در این مورد کلماتی که پیشوند و یا پسوند دارند نیز در شکل‌های مختلف نوشته می‌شوند. برخی از کلمات در دو شکل متصل نویسی و منفصل نویسی 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، بی‌حوصلگی» مشکلی ایجاد نمی‌کند. اما در محیط الکترونیکی و شبکه اینترنت برای بازبازی 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 هم-چنان باقی می‌مانند که احتیاج به برطرف شدن دارند» [۸]

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

۱۰. در اغلب اوقات یک 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 یک نقطه کوچک است که می‌تواند رایانه را به اشتباه بیاندازد و نیز اعداد ۱ و ۲ و ۳ بسیار شبیه هم هستند و تفاوت‌شان در یک دندان کوچک 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 یک 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، مسوول)

۱۶. استفاده از «ا» و «آ» به جای یک-دیگر 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 الفبای فارسی با نشانه‌های الفبای خارجی مشکلات جدی وجود دارد. از جمله این که الفبای 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 هم‌خوان‌ها 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، خانهٔ).

۲۴. کلمه‌های عربی در شکل‌های گوناگون در زبان فارسی نوشته می‌شوند. (مبدأ، مبداء)؛ (ابتداء، 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 استشهاده.

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 دیجیتالی فارسی در شبکه جهانی است، با این که از شکل‌گیری کتابخانه‌های فارسی در شبکه جهانی مدت زیادی نمی‌گذرد با این حال به سرعت در حال رشد و گسترش است. شماری از این کتابخانه‌ها در پایگاه‌های اینترنتی شکل گرفته‌اند و بسیاری وبلاگ‌هایی هستند که برای این کار راه‌اندازی شده‌اند. از ویژگی‌های این کتابخانه‌ها این است که هیچ یک جنبه تجاری ندارند و نیز به جز عده معدودی اکثر کتابخانه‌ها کوشیده‌اند جانب بی‌طرفی را رعایت کرده و از اعمال سلیقه شخصی پرهیز کنند. آنچه در بسیاری از کتابخانه‌های مجازی فارسی در دسترس است تنها شامل کتاب نیست بلکه نوشته‌هایی اعم از داستان، مقاله، تک‌نگاشت و ... نیز در میان مجموعه‌ها دیده می‌شود. هم‌چنین است آثاری که احتمالاً هیچ‌گاه چاپ کاغذی ندارند و البته وجود کتاب‌هایی که مدت‌هاست نایاب‌اند و مجال انتشار دوباره نیافته‌اند و یا آثاری که امروز به دلایلی بازچاپ آن‌ها مقدور نیست از جاذبه‌های کتابخانه‌های مجازی‌اند. در اینجا شماری از این کتابخانه‌ها ذکر می‌شوند: پایگاه اینترنتی کتاب‌های رایگان فارسی، پایگاه اینترنتی بانی تک، کتابخانه مجازی داستان‌های فارسی، آوای آزاد، پایگاه اینترنتی خوابگرد، کتابخانه دوات، پایگاه اینترنتی سخن، وبلاگ کتابخانه هرمس،

پایگاه اینترنتی گفتمان، پایگاه تاریخ و فرهنگ ایران زمین، پایگاه مرکز جهانی اطلاع‌رسانی آل‌البیت، کتابخانه پایگاه اینترنتی حوزه، پایگاه اینترنتی امام علی (ع)، پایگاه اینترنتی کتابخانه دیجیتال و ... که لازم به ذکر است غلبه با کتاب‌های دو حوزه ادبیات و دین است. [۱۰]

کاربران به دلایل مختلفی از قبیل "دسترسی آسان و ارزان به حجم عظیم اطلاعات، عدم نیاز اطلاعات یافته شده از اینترنت به تایپ مجدد، دسترسی سریع و اطلاعات جدید، صرفه‌جویی در وقت و مهم‌ترین دلیل، عدم تسلط اکثر کاربران به زبان انگلیسی" که زبان غالب بر اینترنت است" به دنبال اطلاعات فارسی از اینترنت هستند. گسترش زبان و انبوهی از نوشتارها ایجاب می‌کند که خط ضابطه داشته باشد و از سوی دیگر پیشرفت فن‌آوری و پیدایش اینترنت خواستار ضابطه و قانونمندی است. اطلاع‌رسانی که جنبه بین‌المللی پیدا کرده است بدون دستور خطی سامان یافته و نظام‌مند میسر نیست و دست‌کم دشواری‌ها می‌آفریند. در حال حاضر وبلاگ‌های فارسی مقام دوم یا سوم را در جهان دارا می‌باشد. به نظر دکتر آشوری "اگر زبان فارسی به همین صورت بی‌دقت در اینترنت به کار رود در سطح زبانی برای تفنن باقی خواهد ماند و کم‌تر حرفی جدی به این زبان زده خواهد شد. آینده زبان فارسی در اینترنت بستگی به این دارد که نویسندگان فارسی تا چه حد کار خود را جدی بگیرند و این زبان را بازسازی کنند که از لحاظ قدرت بیان و دقت مفاهیم و استواری ساختار دستوری به زبان انگلیسی نزدیک شود". [۱۱]

نبود استاندارد ثابت رسم الخط فارسی موجب این شده است که به تعداد صفحات وب فارسی سبک و سیاق نگارش به کار رفته باشد لکن می‌توان چنین ارزیابی نمود که اکثر وب‌های فارسی در برخی خصوصیات مشترک می‌باشند از جمله این که نگارش برخی از آن‌ها زبان غیر رسمی و محاوره‌ای می‌باشد و به خصوص در متون علمی اغلب واژه‌های بیگانه به دفعات استفاده می‌شود. رسم الخط مورد استفاده نیز متفاوت و سلیقه‌ای است و برخی از آن‌ها غلط‌های تایپی و نگارشی فراوانی دارند و این خصوصیات، اغلب به جهت محدودیت‌های محیط الکترونیکی و عدم تطابق رسم الخط فارسی با آن می‌باشد که نمایه‌سازی و سپس جستجو به این زبان را با دشواری‌هایی رو به رو می‌سازد.

با توجه به این نکته که اطلاعات ارزشمند فراوانی در اینترنت وجود دارد و اینترنت با شتابی فراوان به یک منبع اطلاعاتی ممتاز تبدیل شده است. موتورهای جستجو به عنوان یکی از اساسی‌ترین دروازه‌های ورود به منابع اینترنتی دارای ضعف‌هایی هستند. که می‌توان به این موارد اشاره کرد:

- در یک مجموعه از یافته‌های بازیابی شده مدخل‌های تکراری فراوانی ملاحظه می‌شود.
- نتایج غیر قابل پیش‌بینی هستند.
- نتایج چه بسا گمراه‌کننده باشند: ممکن است جستجویی در یک موتور کاوش نتیجه‌ای نداشته، ولی در موتور دیگر دارای یافته‌های فراوان باشد.
- موتورهای کاوش محتویات پایگاه‌های اطلاعاتی خودشان را نشان نمی‌دهند و از معیارهایی که برای گنجاندن یک مدرک در فایل‌هایشان دارند حتی شرحی ارائه نمی‌کنند.

- مهار واژگانی وجود ندارد و قواعد نقطه-گذاری و بزرگ-نویسی نیز استاندارد نیست.

- بدون بررسی عملی هر عنصر، اغلب نمی-توان میزان ربط و رابطه-ها را تحلیل کرد. یعنی اطلاعات کافی در مدخل نمایه نیست تا فرد بتواند دست به انتخاب بزند.[۱۲]

- عدم توان موتورهای جستجو در تمایز میان مدارکی که توسط فرد الف نوشته شده و مدارکی که در باره فرد الف نوشته شده است.

- منابع قابل توجهی در شبکه وب وجود دارند که توسط موتورهای جستجو نمایه نمی-شوند. به این بخش از وب اصطلاحاً وب نامرئی می-گویند. "وب نامرئی بخش بزرگی از وب است که موتورهای جستجو آن-ها را نمایه نمی-کنند یا نمی-توانند نمایه کنند و عبارتند از: سایت-های دارای رمز عبور، فایل-های پی.دی.اف از متون آرشیو شده، ابزارهای تعاملی نظیر ماشین حساب-ها و برخی از واژه-نامه-ها و همچنین بعضی از پایگاه-های 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 می-تواند 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، کم-تر 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 روش-های 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 نظرتان 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 هیچ‌گونه 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 پاسخ‌گوی 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 نمائید.
- 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 نقطه گذاری 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 کلمه 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 روان-شناسی ایران در زمینه روان-شناسی 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 چنان-چه 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. ایران NPiran ، ایران هو Iranhoo ، ایران مهر IranMehre ، پارسیک Parseek ، گوگل Google.

در بین ابزارهای کاوش فوق ، تنها موتور کاوش گوگل دارای برنامه روبات به منظور شناسایی و نمایه-سازی صفحات یا سایت-های وب به زبان فارسی و نمایه-سازی خودکار می-باشد و قادر است صفحات فارسی را در قالب یونی کد شناسایی و در پایگاه خود نمایه کند و سایت پارسیک نیز از پایگاه گوگل برای جستجو و بازیابی اطلاعات استفاده می-کند. 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 طوری که 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 رسم الخط 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 بیاوریم.
- 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. استانداری 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 به‌عنوان یک وظیفه رسانه‌ای نماید. ایجاد و گسترش چنین استانداری 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 ساربانقلی. " بررسی مشکلات جستجو و بازیابی اطلاعات به زبان فارسی از اینترنت با مطالعه موردی بر روی کاربران مرکز اینترنت دانشگاه آزاد اسلامی واحد شبستر ". پایان-نامه کارشناسی ارشد کتابداری و اطلاع-رسانی، دانشگاه آزاد اسلامی، واحد تهران شمال، ۱۳۸۴
- [۲] کارشناس ارشد کتابداری و اطلاع-رسانی دانشگاه آزاد اسلامی واحد شبستر
- [۳] accents
- ۴ پرویز ناتل 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 می شود ولی خوانده نمی شود، مانند خواهش. اما در زمان قدیم آن را با کیفیت خاصی تلفظ می کرده اند و چون در هنگام تلفظ ضمه به فتحه عدول می کرده اند، آن را واو معدوله نامیده اند. هنوز در برخی از لهجه ها 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 می کند و شکل آن ها را با بررسی الگوهای تیره و روشن تعیین می کند. پس از تعیین اشکال توسط اسکنر یا وسیله مورد استفاده برای خواندن، روش های تشخیص نوری کاراکترها برای تبدیل اشکال به متون کامپیوتری مورد استفاده قرار می گیرند. (فرهنگ تشریحی اصطلاحات کامپیوتری میکروسافت. مترجم فرهاد قلی-زاده نوری. [تهران: کانون نشر علوم، ۱۳۷۹]، ص. ۴۴۵)
- [۸] محمدصادق محقق زاده، کاظم زارعیان. "ارائه راه حل برای برخی مسائل اتوماسیون و نگارش فارسی" فصلنامه اطلاع رسانی. (دوره ۱۹، شماره ۳ و ۴) ص.
- [۹] ماندانا صدیق بهزادی. "ناهماهنگی ضبط نام های بیگانه در فارسی". فرهنگ (کتاب سیزدهم، زمستان ۱۳۷۱) ص. ۱۰۳-۱۱۶
- [۱۰] مجید رهبانی. "قند پارسی در شبکه جهانی: کتاب های دیجیتالی و کتابخانه های مجازی فارسی در اینترنت. "جهان کتاب، (۱۸۳) ص.
- [۱۱] "تابوی اصلاح خط". جام جم. (۴ بهمن ۱۳۸۳)
- [۱۲] براندا پاریس سیلی. "فهرست-نویسی منابع اینترنت: سازماندهی وب در کتابخانه های محلی و غیر آن" ترجمه محسن حاجی زین العابدینی. فصلنامه اطلاع-رسانی (دوره ۱۶، شماره ۳ و ۴) ص. ۱
- [۱۳] عبدالرسول خسروی. "وب نامرئی" فصلنامه اطلاع-رسانی (دوره ۲۰، شماره ۱ و ۲) ص. ۵۳
- [۱۴] Truncation
- [۱۵] Proximity search
- [۱۶] عبدالرسول خسروی. "وب نامرئی" فصلنامه اطلاع-رسانی (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 زمانی-فر. "روشی برای رفع چالشهای محتوا کاوی وب‌های فارسی زبان" مجله الکترونیکی مرکز اطلاعات و مدارک علمی

منابع و مأخذ

۱- آزادی، قاسم. "اینترنت: سازمان-دهی 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 شمس الدین. 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 تاریخ‌تطور 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. زبان شناسی 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، جیران. "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 سایتها: ابزارها و معیارها". ترجمه غلام حیدری.

http://www.irandoc.ac.r/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 رایانه‌ای: 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 کتابخانه‌های مجازی فارسی در اینترنت". جهان کتاب، ۱۸۳.

۲۳- رئیسی، محمد رضا. "OCR: آموزش الفبای فارسی به رایانه".

<http://iranwsis.org/Default.asp?C=IRNW&R=&I=93>

۲۴- سیبلی، برندا پاریس. "فهرست‌نویسی منابع اینترنت: سازمان-دهی وب در کتابخانه‌های محلی و غیر آن". ترجمه محسن حاجی زین‌العابدینی. فصلنامه اطلاع‌رسانی، دوره ۱۶، شماره ۳ و ۴.

۲۵- صدیق بهزادی، ماندانا، "ناهماهنگی‌های ضبط نام‌های بیگانه در فارسی". فرهنگ. کتاب سیزدهم، زمستان ۱۳۷۱.

۲۶- صدیقی، محسن، زمانی‌فر، کامران. "روشی برای رفع چالش‌های محتوای کاوی وب‌های فارسی زبان".

http://www.irandoc.ac.ir/Data/E_J/vol4/shahidi.htm

۲۷- طباطبایی، علاءالدین. "در دشواری‌های رایانه‌ای زبان فارسی". نشر دانش، ۱۰۳.

۲۸- عاصی، مصطفی. "نقش رایانه در ایجاد استانداردهای زبانی". فرهنگ. سال چهاردهم، شماره‌های اول - دوم، بهار - تابستان ۱۳۸۰.

۲۹- عزیز محمدی، فاطمه، "بررسی برخی فرآیندهای رایج قرض‌گیری در زبان فارسی". فصلنامه اطلاع‌رسانی، دوره ۱۸، شماره ۳ و ۴.

۳۰- فتاحی، رحمت‌الله. "چالش‌های سازمان‌دهی منابع دانش در آغاز قرن بیست و یکم با نگاهی بر دانش فهرست‌نویسی در ایران". فصلنامه کتاب. ۴۸.

۳۱- فرهنگ تشریحی اصطلاحات کامپیوتری میکروسافت. مترجم فرهاد قلی‌زاده نوری. تهران: کانون نشر علوم، ۱۳۷۹.

۳۲- قاسمی، علی حسین، اطلاع‌یابی 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 مقایسه‌ای 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 رایانه‌ای، ۱۳۸۲.

۳۶- گزنی، علی. "جست و جوی اطلاعات و ساز و کارهای بهینه‌سازی 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 اطلاع‌رسانی شماره ۳ و ۴ دوره ۱۹.

۳۹- محمدی-فرد، داود؛ آباقری، محمد. کامپیوتر برای کتابداران و اطلاع‌رسانان. 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 اطلاع‌رسانی. 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- نوروزی، علی‌رضا. "جستجو در اینترنت: آشنایی با موتور جستجوی گوگل". فصلنامه اطلاع‌رسانی، دوره ۱۶، شماره ۳ و ۴.