نما

مجله الکترونیکی مرکز اطلاعات و مدارک علمی ایران

شماره اول دوره پنجم

تاریخ انشار این مقاله: ۵ شهریور ۱۳۸٤

مهارت در جستجوي اطلاعات فارسي از اينترنت[۱]

محمد صابر راثي ساربانقلي[۲]

چكيده:

خط فارسي داراي مشكلات مختلفي مي‌باشد كه در جستجو و بازيابي اطلاعات مسائل و مشكلات فراواني را فراروي كابران اينترنت قرار مي‌دهد. به خصوص با رشد سريع انتشارات الكترونيكي بر روي وب در شكلهاي مختلف پايگاه‌هاي اطلاعاتي، وبلاگ و ... و اينكه هيچ قاعده مشخص و ثابتي براي رسم‌الخط فارسي وجود ندارد باعث شده است كه جستجوگران مطالب فارسي با مشكلات فراواني روبرو بشوند. اين مقاله سعي دارد تا با اشاره به موارد مختلفي كه مي‌تواند در جستجو و بازيابي اطلاعات سرعت و دقت و جامعيت و مانعيت جستجو را بالابرد موجب افزايش مهارت كاربران اينترنت فارسي بشود.

كليد واژه ها: اينترنت، خط فارسي، جستجو و بازيابي اطلاعات.

مقدمه

اينترنت به عنوان يك محمل اطلاعاتي عظيم، منابع اطلاعاتي را در مقياسي وسيع در دسترس مخاطبان بالقوه قرار داده است. اغلب سهولت دسترسي به منابع اطلاعاتي اعم از متن و ساير رسانه‌ها عمده‌ترين مزيت اينترنت محسوب مي‌شود. اما اين توانايي كه هركس ناشر آثار خود باشد عواقب ناخواسته‌اي را نيز در پي خواهد داشت و آشكارترين معضل، آن است كه انبوهي از منابع بسيار متنوع و غير قابل مديريت را فراهم مي‌آورد. افزايش سريع منابع اينترنتي نيازمند يك سازمان‌دهي مفيد و موثر است. هرچند در حال حاضر راهنماهايي براي منابع اينترنتي تهيه شده است كه براساس فايل‌هاي مقلوب ساخته شدهِ توسط موتورهاي جستجو و با استفاده از قابليت‌هاي مختلف اين موتورها از جمله : استفاده از عملگرهاي بولي، جستجوي دقيق عبارت، محدود كردن يك جستجو به بخش خاصي از ركورد (مانند عنوان ، آدرس) ، كوتاه‌سازي كلمات، جستجوي نزديك‌يابي واژه‌ها، ايجاد محدوديت زماني و منطقه‌اي و زباني، و .... به جستجوي اطلاعات كمك مي‌كند، اما بايد تاكيد كرد كه در امر بازيابي اطلاعات از اينترنت بدون نمايه‌سازي نظام يافته نمي‌توان انتظار بازيابي مفيد و موثر را داشت. هرچند بيش‌تر اطلاعات موجود بر روي اينترنت به زبان انگليسي است، ولي حجم اطلاعات به زبان فارسي نيز با سرعت در حال افزايش است و كاربران به دلايل مختلفي علاقه زيادي به اطلاعات فارسي نشان مي‌دهند و از آنجائي‌كه زبان غالب در اينترنت انگليسي است جستجو به زبان‌هاي غير انگليسي از جمله فارسي، مسايل و مشكلات مختلفي را جداي از مشكلات عمومي اينترنت دارد.

خط فارسي

اشكال و نقصي كه در همه خطوط جهان است دو علت دارد كه يكي در اصل خط است و ديگري بر اثر تغيير و تحول زبان ايجاد مي‌شود. دقت فراوان در ثبت همه دقايق تلفظ اغلب موجب دشواري شيوه خط است و اين دقت زماني ضرورت مي‌يابد كه زباني توسعه بسيار بيابد و در كشورهاي ديگري كه به آن زبان سخن نمي‌گويند رايج شود. به عنوان مثال در خط عربي نقطه و علامت‌هاي حركات وقتي به وجود آمد كه زبان عربي نزد ملت‌هاي غير عرب معمول شد، در خط يوناني نيز نشانه‌هاي آهنگ و تكيه[۳] پس از رواج آن زبان در مصر ايجاد شد تا كساني كه زبان مادري‌شان يوناني نبود و با تلفظ آن مانوس نبودند بتوانند كلمات و عبارات يوناني را هر چه درست‌تر ادا كنند. با اين حال هيچ خطي هر قدر دقيق و شماره علامات آن فراوان باشد، ممكن نيست كه كاملاً نشانه شيوه تلفظ باشد. و با كمك علامات متعدد علم حروف نيز تا كسي چگونگي تلفظ زباني را نشنود نمي‌تواند عبارت و كلمات آنرا مانند اهل آن زبان ادا كند.

اما نقصي كه بر اثر تحول زبان و به تدريج در خط حاصل مي‌شود، مشكلي است كه همه ملت‌ها با آن رو به رو هستند. بعضي از حروف و اصوات زبان در طي زمان تغيير مي‌پذيرند و اين تغيير در گفتار حاصل مي‌شود، اما خط هميشه صورت كهن تلفظ را حفظ مي‌كند، و از اينجا ميان "گفتار" و "نوشتار" اختلاف روي مي‌دهد. ديگر آن كه هر زباني ناگزير لغاتي از زبان‌هاي ديگر به عاريت مي‌گيرد و اگر علائم خط در اين دو زبان يكي باشد كلمه خارجي به همان املاي اصلي در نوشتن به كار مي‌رودكه اغلب با املاي كلمه مشابه در زبان ثانوي تفاوت دارد و از اينجا براي اصوات واحد علائم خطي متعدد پديد مي‌آيد. در خط فارسي نمونه همه اين موارد را مي‌توان يافت. چون خط عربي براي نوشتن فارسي به كار رفت كلماتي كه از آن زبان اخذ شده بود به همان صورت اصلي نوشته شد. حال آنكه به يقين در هيچ دوره‌اي حروف خاص عربي را فارسي زبان‌ها درست مثل اصل تلفظ نكرده‌اند. در زبان‌هاي ديگر نيز اين گونه موارد نمونه‌هاي متعدد دارد. شايد دو زبان انگليسي و فرانسه بيش از همه زبان‌هاي جهان دچار اختلاف تلفظ و خط باشند. به طور كلي نقائص و معايبي كه در خطوط معمول جهان است را مي‌توان به طريق زير طبقه‌بندي كرد:

۱. شكل واحدي اصوات مختلف را بيان مي كند. چنانكه در فارسي حرف "ي" را گاهي براي حرف لين بكار مي‌بريم (يك) و گاهي براي حرف مد (بي) و گاهي به جاي الف (عيسي) و گاهي براي نشان دادن مصوت مركب (ري). و يا حرف «و» در كلمات (سوار، سود، تو)

۲. اصوات واحد به صورت‌هاي مختلف نوشته مي شود. در فارسي حرف "س" سه صورت (س – ص – ث) و حرف "ز" چهار صورت (ز – ذ- ض- ظ) دارد؛ در زبان فرانسه حروفي كه "سن" خوانده مي‌شود پنج رسم الخط دارد كه اگر صورت‌هاي جمع را نيز به حساب بياوريم ده شكل مي‌شود از اين قرار (saint, ceint, sein, seing, sain)

۳. بسياري از حروف نوشته مي‌شود ولي خوانده نمي‌شود. يعني علاماتي بي‌فايده در نوشتن به كار مي‌رود در فارسي نوشتن "واو معدوله" و "هاء غير ملفوظ" از اين قبيل است. در انگليسي نمونه اين مورد بسيار است مانند high كه دو حرف آخر آن به كلي از تلفظ ساقط است. و يا "K" در كلمه "Know" .

۴. اصواتي هستند كه تلفظ مي‌شود اما در خط نشانه‌اي براي آن‌ها نيست. در فارسي سه مصوت كوتاه (ـَ ـُ) از اين قبيل است هم چنين الف در كلمات اسحق و الله كه در كتابت نمي‌آيد. [٤]

زبان و خط فارسي نيز مشكلات خاصي را دارا مي‌باشد و نظام نوشتاري فارسي براي ثبت دقيق گفتار، نارسائي دارد و قواعد نگارش آن مدون نيست، از اين رو فاصله ميان گفتار و نوشتار در فارسي قابل توجه است. بيش‌ترين مشكلات نيز به جهت نبود يك رسم الخط واحد كه عموم اساتيد و اهل فن روي آن اجماع كرده باشند به وجود آمده است. به طوري كه در حال حاضر جداي از چندين شيوه‌نامه رسمي همچون" شيوه‌نامه سمت، نشر دانشگاهي، فرهنگستان، آموزش و پرورش" به تعداد افراد جامعه، رسم الخط و شيوه نگارش زبان وجود دارد، هر ناشري براي خود به قاعده‌اي دلخواه عمل مي‌كندكه اين تعددها موجب پريشاني و پراكندگي شده و با يك‌ديگر تفاوت‌هايي دارند. از ديگر دلايل مي‌توان به عاريتي بودن خط فارسي و چاره‌انديشي براي حركات و عدم تطابق واج‌ها با حروف اشاره كرد. متصل و منفصل‌نويسي نيز يكي ديگر از حوزه‌هاي مورد اختلاف است از ديگر مشكلات: گوناگوني معادل‌هاي علمي، انواع مختلف ضبط اسامي خارجي، سرهم‌نويسي، جدانويسي، بي‌فاصله‌نويسي، انواع جمع‌ها، صورت‌هاي مختلف نوشتاري، آوانويسي اسامي عناصر و تركيبات شيميايي، سرواژه‌ها و كوته‌نوشت‌ها مي‌باشد.

به طور كلي نقص‌هايي كه براي زبان فارسي شمرده‌اند به شرح زير مي‌توان عنوان كرد:

۱.سه مصوت كوتاه يعني حركات زير و زبر و پيش (ـَ ـُ) را از نوشتن ساقط مي‌كنيم. و اين باعث مي‌شود به جاي اين كه از خط و نوشتار پي به معني ببريم بايستي از معني كلمه و جايگاه آن در جمله آن‌را درست بخوانيم مانند كلمات (كَرَم، كَرَم، كِرم، كُرُم، كِرْم) و (مَلَك، مَلِك، مُلك، مِلك) و يا سه كلمه (حَكَم، حُكم، حِكَم) و نيز ننوشتن مصوت‌هاي كوتاه در داخل متن باعث مي‌شود كه براي تلفظ صحيح اجباراً لاتين كلمات به صورت پانويس متن آورده شود كه همين امر باعث اتلاف وقت و انرژي مي‌شود. كه البته همين لاتين‌نويسي هم قاعده خاصي ندارد و هر ناشر و نويسنده‌اي سليقه خاص خودش را براي آوانويسي حروف فارسي به لاتين دارد. كه به عنوان نمونه براي نشان دادن حركت فتحه و الف و آ هيچ‌گونه هماهنگي در كتاب‌ها و خصوصا فرهنگ‌هاي مختلف ديده نمي‌شود. "هر چند برخي معتقدند همين ننوشتن حركات مزيتي است و موجب تندنويسي مي‌شود"[٥].

۲.براي يك حرف چند علامت مختلف داريم مانند علامت‌هاي (س،ص، ث) كه هر سه در فارسي يكسان خوانده مي‌شوند و هم چنين (ذ، ز، ض، ظ) و نيز (ت، ط). البته اين امر در زبان انگليسي هم وجود دارد چنان كه «ف» ممكن است به شكل‌هاي « F. GH. PH. V» باشد.

۳.يك علامت را براي دلالت بر چند حرف مختلف استعمال مي‌كنيم مانند "و" كه پنج مورد نوشتن دارد يكي براي بيان ضمه در كلمات "خوش" و "تو". ديگر بيان مصوت ممدود يا "واو ماقبل مضموم" مانند "شور" و "او". سوم بيان حرف صامت "واو" در

كلماتي چون "آواز" و "والي" و "عفو" . چهارم بيان حرف مصوت مركبي كه در كلمات "نو" و "جوشن" و مانند آن‌هاست. پنجم حرفي كه در زبان كنوني خوانده نمي‌شود مانند "واو معدوله" در كلمات "خواهر" و "خواستن" و "واو" در كلمه "عمرو"[٦]
۴.حرف‌هايي هم هست كه در كلمات خاصي از نوشتن حذف مي‌شود مانند "الف" در كلمات "اسحق" و "اسمعيل" و "اللّه"
۵.نقطه‌هايي متعدد در بالا و پائين حرف كه هم سبب دشواري و هم موجب اشتباه در خواندن مي‌شود. اهميت بيش از حد نقطه درخط فارسي هنگام تشخيص نوري كاراكترها[٧] توليد اشكال اساسي مي‌كند. به عنوان مثال در نظر بگيريد كه تفاوت ‹ر› و ‹ز› و يا تفاوت ‹د› و ‹ذ› و يا تفاوت ‹ب› ‹ت› ‹ پ› ‹ث› فقط در نقطه است و چون نقطه جزء بسيار كوچكي است در اين امر مشكلات زيادي را فرا روي متخصصين قرار مي‌دهد. و يا كلمات زير را در نظر بگيريد كه با يك يا چند نقطه عوض مي‌شوند (بُر، بَر، پُر، پَر، تَر، پُز، پَز، بُز، تِز).
۶. يك عيب ديگر هم كه براي خط فارسي ذكر كرده‌اند اين است كه از راست به چپ نوشته مي‌شود. و براي اين مورد دلايل مختلفي ذكر شده است از جمله عدم هماهنگي و ايجاد مشكل در نوشتن متون رياضي و شيمي و نت‌هاي موسيقي و دستورات شطرنج و اين كه خط تصويري يعني علائم گرافيكي كه در كل جهان استفاده مي‌شود مانند علائم راهنمائي و رانندگي تماماً از چپ خوانده مي‌شوند.
۷.پيوسته‌نويسي و جدانويسي كلمات مركب كه در اكثر موارد به صورت سليقه‌اي عمل مي‌شود مانند تنوع استفاده از ‹مي› چسبان و غير چسبان و يا تنوع نحوه به كار بردن «علامت‌هاي جمع ‹ها، ان، جات› ، هم، هيچ، كه، (ضماير شخصي متصل مان، تان، شان)، شناسي، را، چه، چون، تر، ترين، بي (پيشوند نفي)، به، اي (نشانه ندا)، آن و اين» در كلمات به صورت پيوسته و يا جدا گانه: (آنچه ، آن چه)؛ (همچنانكه، همچنان‌كه) ؛ (جنابعالي، جناب‌عالي)؛ (هيچكس، هيچ‌كس)؛ (ميتواند، مي‌تواند)؛ (آن ها، آنها) در اين مورد كلماتي كه پيشوند و يا پسوند دارند نيز در شكل‌هاي مختلف نوشته مي‌شوند. برخي از كلمات در دو شكل متصل‌نويسي و منفصل‌نويسي به دو شكل مختلف ظاهر مي‌شوند، مانند «علاقمند و علاقه‌مند؛ انديشمند و انديشه‌مند». مصدرها و فعل‌هاي مركب و اسم‌هاي مشتق از آنها نيز به دو صورت متصل و منفصل نوشته مي‌شوند مانند «نگه‌داشتن و نگهداشتن». در جستجوي مطالب از اينترنت اين مورد توليد اشكال مي‌كند چنانكه جستجوي «هيچ‌كس» نتايج متفاوتي را با جستجوي «هيچكس» مي‌آورد و يا جستجوي «كتاب‌شناسي» و «كتابشناسي» در موتور جستجوي گوگل نتايج متفاوتي را ارائه مي‌كند. اين گونه كلمات با اين كه در خواندن متن اشكال كمي به وجود مي‌آورند و هر آشناي به زبان فارسي به راحتي مي‌تواند آن را بخواند اما در فن‌آوري امروزه و تجزيه و تحليل كلمات به كمك رايانه اشكال اساسي توليد مي‌كند و شايد اگر قاعده‌اي جامع و مانع براي آن وضع گردد، بتوان گفت بزرگ‌ترين مشكل خط فارسي حل شده است. منظور اين كه، براي مثال خواندن سه كلمه «بي‌حوصلگي، بيحوصلگي، بي‌حوصله‌گي» مشكلي ايجاد نمي‌كند. اما در محيط الكترونيكي و شبكه اينترنت براي بازيابي اين كلمه بايستي براي تمام اشكال اين كلمه، جستجو را انجام دهيم، البته اگر آگاهي از تمام اشكال نوشتاري آن داشته باشيم.آآ
۸.سي و دو حرف الفباي فارسي همراه با چهار علامت مد، همزه، تنوين، تشديد به ۱۳۰ شكل مختلف ظاهر مي‌شوند و تفاوت اين اشكال در اتوماسيون خط فارسي توليد اشكال مي‌كند. « تنوع و تعدد نويسگان، يادگيري زبان و خط فارسي را براي آموزگار و آموزنده دشوار و براي نوآموز توان‌فرسا مي‌سازد. تعداد زياد نويسگان در رابطه با اتوماسيون زبان توسط رايانه مشكلاتي در خصوص تعداد و ترتيب قرار گرفتن نويسگان در جداول كد ايجاد مي‌نمايد و طراحان كد در جاي دادن اين تعداد نويسه در جداول با مساله كمبود جا رو به رو هستند. هر چند كه مشكل جا با كد ١٦ بيتي حل شده است اما مسايل ديگري همچنان باقي مي‌مانند كه احتياج به برطرف شدن دارند»[٨]
۹. نوشتن ك و گ (كـ گـ ك گ گ ك) در اشكال مختلف نيز باعث سردرگمي و عدم جستجوي صحيح مي‌شود.
۱۰. در اغلب اوقات يك فاصله اضافي معني متفاوتي و يا متضادي را مي‌دهد (مثل مادر ، ما در).
۱۱. سه كرسي مختلف براي حرف‌هاي مختلف الفبا باعث مي شود كه در مقايسه با اكثر زبان‌ها تعداد سطرهاي هر صفحه به مراتب بيش‌تر گردد چون برخي حروف روي خط كرسي قرار مي‌گيرند و برخي پائين خط كرسي و برخي بالاي خط كرسي مثل (ا ب م)
۱۲.از آنجائيكه حروف در نوشتن غالباً به صورت چسبيده و پيوسته نوشته مي‌شوند و اين امر تشخيص حرف به حرف نوشته به وسيله رايانه را، دچار مشكل مي‌كند.
۱۳. در او. سي. آر. فارسي هم چنين اعداد نيز مشكل ساز هستند چنانچه صفر در فارسي يك نقطه كوچك است كه مي‌تواند رايانه را به اشتباه بياندازد و نيز اعداد ۱ و ۲ و ۳ بسيار شبيه هم هستند و تفاوت‌شان در يك دندانه كوچك است.
۱۴. تنوع املائي يا تنوع در رسم الخط بعضي از كلمات كه همه شكل‌هاي آن نيز درست است مانند(اتاق و اطاق) و يا (امپراتور و امپراطور). و كلماتي كه فقط يك شكل آنها صحيح مي‌باشد ولي شكل ناصحيح آن نيز زياد استفاده مي‌شود مانند «ذغال و زغال؛ خوشنود و خشنود». البته اين جداي از تنوع در مفهوم كلمات است كه در ديگر زبان‌ها نيز وجود دارد، يعني

براي بعضي از مفاهيم ممكن است كلمات متنوعي استفاده بشود. مانند كامپيوتر و رايانه.
۱۵. بكار بردن همزه درصورت‌هاي مختلف مانند (مساله، مسئله) ؛ (مسئول، مسوول)
۱۶. استفاده از ‹ا› و ‹آ› به جاي يك‌ديگر مانند (فرايند و فرآيند).
۱۷. شكل‌هاي مختلف ضبط نام‌هاي بيگانه در فارسي: ورود واژه‌هاي بيگانه معمولا از راه ورود پديده‌هاي فرهنگي نو در عرصه‌هاي مختلف فني ، علمي، اجتماعي، سياسي و هنري و .... و يا از طريق افراد دو زبانه انجام مي‌گيرد كه به قرض‌گيري زبان معروف است و كم و بيش در تمام زبان‌ها وجود دارد. واژه‌هاي بيگانه اغلب براي پر كردن خلاء واژه‌هاي علمي و يا ارتباطي سودمند هستند، اما وجود آن‌ها مسائلي از قبيل چگونگي ضبط آن‌ها در زبان قرض‌گيرنده را به وجود مي‌آورد. براي ضبط واژه‌هاي قرضي به سبب اختلاف فاحش نشانه‌هاي الفباي فارسي با نشانه‌هاي الفباي خارجي مشكلات جدي وجود دارد. از جمله اين كه الفباي فارسي آوانگار نيست و به همين جهت در ضبط دقيق تلفظ واژه‌هاي زبان فارسي نيز ناتوان است و اين ناتواني در ضبط واژه‌هاي بيگانه به مراتب بيش‌تر است و اين كه در مورد برگردان اسامي خارجي به خط فارسي قاعده خاصي وجود ندارد و هر كس بنا بر سليقه و ذوق خود اين كار را انجام مي‌دهد كه در نتيجه يك كلمه واحد به صورت‌هاي مختلف نوشته مي‌شود. براي مثال (اتومبيل و اتوموبيل)؛ (كلسيم، كلسيوم، كالسيوم) و يا اسم Franklin به صورت (فرانكلين، فرانكلن، فرنكلين، فرنكلن) ضبط شده است. خانم صديق بهزادي اين مشكلات را به سه دسته تقسيم كرده است: " ۱- نام‌هايي كه در برگردان آن‌ها هم‌خوان‌ها ايجاد مشكل مي‌كنند. ۲ – نام‌هايي كه در برگردان آن‌ها واكه‌هاي ساده مشكلاتي را به وجود مي‌آورند . ۳ – و سوم نام‌هايي كه در برگردان آن‌ها مشكل اصلي مربوط به واژه‌هاي مركب است [۹].
۱۸.استفاده يا عدم استفاده از ‹ي› در كلمات مختوم به ‹الف › مانند (موسي و موسا).
۱۹.استفاده يا عدم استفاده از ‹ء› براي كلمات مختوم به هاي بيان حركت در حالت مضاف مانند (خانه مسكوني و خانهء مسكوني و يا خانه‌ي مسكوني).
۲۰.استفاده يا عدم استفاده از اعراب براي كلمات.
۲۱.انواع مختلف جمع براي يك واژه مفرد: به عنوان مثال جمع بستن يك واژه با علايم جمع فارسي وعلايم جمع عربي و نيز جمع بستن بي قاعده (جمع مكسر)، استفاده ازجمع جمع، مانند (معلم، معلمين، معلمان، معلم‌ها).
۲۲.تنوين‌هاي زبان عربي نيز از جمله دشواري‌هاي رعايت اصل هم‌خواني نوشتاري و گفتاري هستند.
۲۳.در نگارش ياء وحدت يا نكره در آخر كلماتي كه به هاء مختفي يا غير ملفوظ ختم مي‌شوند سه نوع املاء ديده مي‌شود. (خانه‌اي، خانه‌يي، خانهٔ).
۲۴.كلمه‌هاي عربي در شكل‌هاي گوناگون در زبان فارسي نوشته مي‌شوند. (مبدا، مبداء)؛ (ابتدا، ابتداء)؛ (نسبتاً، نسبته، نسبتا) و ....
۲۵.ناتواني خط فارسي در نشان دادن تلفظ واژه‌هاي ايران باستان و ميانه و گويش‌ها و لهجه‌هاي ايراني و واژه‌هاي بيگانه حتي با نشانه‌ها.
۲۶.وجود دندانه‌هاي متعدد درکلمات خواندن کلمات و به خصوص در او.سي.آر. فارسي ايجاد اشكال مي‌كند مانند كلمات: نشستن و استشهاد.
۲۷. حروف فارسي غالباً مشابه‌اند و با اندكي غفلت به جاي هم نوشته مي‌شوند و مطلب را به كلي دگرگون مي‌كنند مانند (در، رد، ور).

زبان و خط فارسي در اينترنت :

حجم اطلاعات به زبان فارسي در روي اينترنت در اشكال مختلف آن به سرعت رشد كرده است. در حال حاضر توسعه وبلاگ‌هاي فارسي و سايت‌هاي علمي و تبليغاتي و دانشگاهي به زبان فارسي باعث شده است كه جايگاه زبان فارسي تا حد زبان اول ارتباطات اينترنتي نزد ايرانيان و فارسي‌زبانان در سراسر جهان ارتقا يابد. شايد بتوان گفت كه اولين مرجع وبلاگ‌نويسي فارسي با انتشار راهنماي ساخت وبلاگ فارسي آغازشده است. بدون شك دومين موج نيز با شروع به كار سايت پرشين بلاگ كه امكان راه‌اندازي وبلاگ براي كاربران فارسي زبان را با سهولت بيش‌تري فراهم مي‌كند آغاز شده است. اما پيامد قابل توجه ديگري كه رشد وبلاگ‌نويسي در ايران داشته است پيدايش سايت‌هاي اينترنتي فارسي زباني است كه صاحبان وبلاگ‌ها ايجاد كرده‌اند و اين خود موج جديدي از گسترش كاربرد اينترنت در جامعه ايران به حساب مي‌آيد. اكنون روي آوردن برخي از روزنامه‌نگاران ، پژوهش‌گران، دانشجويان و .... به وب فارسي و استفاده از منابع خبري و علمي و .... آن موجب تقويت نقش رسانه‌اي وب فارسي شده است.
پديده ديگري كه باعث گسترش زبان و خط فارسي در اينترنت شده است ايجاد كتابخانه‌هاي ديجيتالي فارسي در شبكه

جهاني است، با اين كه از شكل‌گيري كتابخانه‌هاي فارسي در شبكه جهاني مدت زيادي نمي‌گذرد با اين حال به سرعت در حال رشد و گسترش است. شماري از اين كتابخانه‌ها در پايگاه‌هاي اينترنتي شكل گرفته‌اند و بسياري وبلاگ‌هايي هستند كه براي اين كار راه‌اندازي شده‌اند. از ويژگي‌هاي اين كتابخانه‌ها اين است كه هيچ يك جنبه تجاري ندارند و نيز به جز عده معدودي اكثر كتابخانه‌ها كوشيده‌اند جانب بي‌طرفي را رعايت كرده و از اعمال سليقه شخصي پرهيز كنند. آنچه در بسياري از كتابخانه‌هاي مجازي فارسي در دسترس است تنها شامل كتاب نيست بلكه نوشته‌هايي اعم از داستان ، مقاله، تك نگاشت و .... نيز در ميان مجموعه‌ها ديده مي‌شود. هم چنين است آثاري كه احتمالاً هيچ گاه چاپ كاغذي ندارند و البته وجود كتاب‌هايي كه مدت‌هاست نايابند و مجال انتشار دوباره نيافته‌اند و يا آثاري كه امروز به دلايلي بازچاپ آن‌ها مقدور نيست از جاذبه‌هاي كتابخانه‌هاي مجازي‌اند. در اينجا شماري از اين كتابخانه‌ها ذكر مي‌شوند: پايگاه اينترنتي كتاب‌هاي رايگان فارسي، پايگاه اينترنتي باني تك، كتابخانه مجازي داستان‌هاي فارسي، آواي آزاد، پايگاه اينترنتي خوابگرد، كتابخانه دوات، پايگاه اينترنتي سخن، وبلاگ كتابخانه هرمس، پايگاه اينترنتي گفتمان، پايگاه تاريخ و فرهنگ ايران زمين،پايگاه مركز جهاني اطلاع‌رساني آل البيت، كتابخانه پايگاه اينترنتي حوزه، پايگاه اينترنتي امام علي (ع)، پايگاه اينترنتي كتابخانه ديجيتال و .... كه لازم به ذكر است غلبه با كتاب‌هاي دو حوزه ادبيات و دين است.[۱۰]

كاربران به دلايل 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، صرفه‌جويي در وقت و مهم‌ترين دليل، عدم تسلط اكثر كاربران به زبان انگليسي "كه زبان غالب بر اينترنت است" به دنبال اطلاعات فارسي از اينترنت هستند. گسترش زبان و انبوهي از نوشتارها ايجاب مي‌كند كه خط ضابطه داشته باشد و از سوي ديگر پيشرفت فن‌آوري و پيدايش اينترنت خواستار ضابطه و قانونمندي است. اطلاع‌رساني كه جنبه بين‌المللي پيدا كرده است بدون دستور خطي سامان يافته و نظام‌مند ميسر نيست و دست‌كم دشواري‌ها مي‌آفريند. در حال حاضر وبلاگ‌هاي فارسي مقام دوم يا سوم را در جهان دارا مي‌باشد. به نظر دكتر آشوري" اگر زبان فارسي به همين صورت بي‌دقت در اينترنت به كار رود در سطح زباني براي تفنن باقي خواهد ماند و كم‌تر 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]

نبود استاندارد ثابت رسم الخط فارسي موجب اين شده است كه به تعداد صفحات وب فارسي سبك و سياق نگارش به كار رفته باشد لكن مي‌توان چنين ارزيابي نمود كه اكثر وب‌هاي فارسي در برخي خصوصيات مشترك مي‌باشند از جمله اين كه نگارش برخي از آن‌ها زبان غير رسمي و محاوره‌اي مي‌باشد و به خصوص در متون علمي اغلب واژه‌هاي بيگانه به دفعات استفاده مي‌شود. رسم‌الخط مورد استفاده نيز متفاوت و سليقه‌اي است و برخي از آن‌ها غلط‌هاي تايپي و نگارشي فراواني دارند و اين خصوصيات، اغلب به جهت محدوديت‌هاي محيط الكترونيكي و عدم تطابق رسم الخط فارسي با آن مي‌باشد كه نمايه‌سازي و سپس جستجو به اين زبان را با دشواري‌هايي رو به رو مي‌سازد.

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 اساسي‌ترين دروازه‌هاي ورود به منابع اينترنتي داراي ضعف‌هايي هستند. كه مي‌توان به اين موارد اشاره كرد:

- در يك مجموعه از يافته‌هاي بازيابي شده مدخل‌هاي تكراري فراواني ملاحظه مي‌شود.
- نتايج غير قابل پيش‌بيني هستند.

- نتايج چه بسا گمراه كننده باشند: ممكن است جستجويي در يك موتور كاوش نتيجه‌اي نداشته، ولي در موتور ديگر داراي يافته‌هاي فراوان باشد.

- موتورهاي كاوش محتويات پايگاه‌هاي اطلاعاتي خودشان را نشان نمي‌دهند و از معيارهايي كه براي گنجاندن يك مدرك در فايل‌هايشان دارند حتي شرحي ارائه نمي‌كنند.

- مهار واژگاني وجود ندارد و قواعد نقطه‌گذاري و بزرگ‌نويسي نيز استاندارد نيست.

- بدون بررسي عملي هر عنصر، اغلب نمي‌توان ميزان ربط و رابطه‌ها 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 نمي‌شوند. به اين بخش از وب اصطلاحاً وب نامرئي مي‌گويند. "وب نامرئي بخش بزرگي از وب است كه موتورهاي جستجو آن‌ها را نمايه نمي‌كنند يا نمي‌توانند نمايه كنند و عبارتند از: سايت‌هاي داراي رمز عبور، فايل‌هاي پي. دي. اف از متون آرشيو شده، ابزارهاي تعاملي نظير ماشين

حسـاب‌ها و برخي از واژه‌نامه‌ها و همچنين بعضي از پايگاه‌هاي اطلاعاتي، منابع محافظت شده از طريق اسم كاربر و گذرواژه، منابع و صفحات وب بدون پيوند و صفحات افزون بر حداكثر تعداد صفحات قابل مرور"[۱۳].

جسـتجوي اطلاعـات در اينتـرنت بـه دو روش مـي‌توانـد صـورت گيـرد يـكي اسـتفاده از جمـلات زبـان محـاوره‌اي است و ديگري بكـارگيري كلمات كليدي. در روش استفاده از جملات زبان محاوره‌اي كه اغلب به كاربران تازه‌كار پيشـنهاد مي‌گردد، مورد سوال خـود را در قـالب يـك جملـه سـوالي مطـرح مي‌سازند. يكي از عيب‌هاي بزرگ اين روش تعداد نتايج جستجوي زيادي است كه بـازگردانده مـي‌شـود. به همين دليل اين روش توسط كاربران حرفه‌اي و حتي توسط همه، كمتر استفاده مي‌شود. اما چنانچه از ايـن روش اسـتفاده بشـود بايسـتي سـعي در انتخـاب بهتـرين نـوع جملـه بشـود و توصيه مي‌شود در انتخاب يك يك كلمات لحظه‌اي درنگ نموده و با ظرافت خاصي جمله نهايي را مطرح نمود.

يـكي از كـاراترين و مقتدرتـرين روش‌هـاي جسـتجوي اطلاعـات در دنيـاي وب اسـتفاده از واژه‌هـايي اسـت كـه اصطلاحاً كلمات كليـدي نـاميده مـي‌شوند. اغلب كاربران حرفه‌اي و جستجوگران ورزيده دنياي اينترنت مي‌توانند با طرح بهترين كلمات كليدي و بكـار بسـتن قـوانين تـركيب آن‌هـا بـا هـم براي نيازهاي اطلاعاتي خود پاسخي در خور بيابند. در اين روش توصيه‌هاي زير براي انتخاب كلمات كليدي و نيزجستجوي دقيق و مفيد پيشنهاد مي‌شود:

۱- حتي‌المقدور سعي شود كلمات كليدي از ميان اصطلاحات منحصر به فرد و اسامي خاص انتخاب بشود.

۲- حتي‌المقدور از آوردن كلمات عمومي كه عناوين بسياري را در زير مجموعه خود شامل مي‌شوند جداً خودداري كنيد.

۳ – هميشه اسم شخص يا نام شي يا هر چيز ديگري را كه مد نظر داريد بطور كامل وارد كنيد.

۴ – دقت كنيد كه اگر موتور جستجو ميان حروف بزرگ و كوچك تفاوتي مي‌گذارد، اين مسئله را در طرح كلمات كليدي خود مد نظر داشته باشيد.

۵ – در نظر داشته باشيد اگر نتيجه جستجو صفر بود به احتمال زياد مي‌تواند از يك اشتباه تايپي باشد.

۶ – اگر املاي صحيح و كامل كلمه‌اي را نمي‌دانيد از كاركتر جانشين كه اغلب * و يا ؟ است استفاده كنيد.

۷ – اگـر يـك كلمه كليدي را براي طرح دقيق و تمام و كمال يك مورد جستجو كفايت نمي‌كند از تكنيك‌هاي جستجوي عبارتي، اسـتفاده از اپراتورهـاي جبر بولين (AND, OR, NOT) استفاده كنيد. جستجوي عبارتي يكي از مهم‌ترين و قدرتمندترين امكانات جسـتجو در اغـلب موتورهـاي جسـتجو مـي‌باشـد و مـي‌تـوان يـك عبـارت يـا جمله مشخص را به همان ترتيبي كه كلمات وارد شده‌اند مورد جستجو قرار داد. براي اين روش جستجو عبارت مورد نظر را داخل گيومه"" بگذاريد

۸ - اسـتفاده از عملگـر and : AND بـه مفهوم "و" براي محدود كردن دامنه جستجو از طريق تركيب كليدواژه‌هاي مختلف به كـار مـي‌رود و بـراي تـركيب كليدهـاي جسـتجو زمـاني‌كـه براي شما مهم است كه دو يا چند كلمه كليدي حتماً وجود داشته باشـد و عـلامت آن در پايگاه‌هاي مختلف به صورت استفاده از عبارت and ،استفاده از + ، انتخاب عبارت all the word از منو، انتخاب عبارت match on all words (and) بوسيله كليك كردن بر روي دكمه‌هاي راديويي مي‌باشد.

۹ – اسـتفاده از عملگـر OR: اپراتور OR به مفهوم "يا" و برخلاف عملگر AND باعث گسترش دامنه جستجو و بازيابي اطلاعات بيـش‌تـر شـده براي تركيب كليدواژه‌هاي جستجو زماني كه انتظار داريد تنها يك، دو يا چند كلمه كليدي حضور داشته باشند و عـلامت آن اسـتفاده از عبـارت or، نحـوه اجـراي سـاده و معمـولي آن، انتخـاب عبارت any of the words از منو، انتخاب عبارت match on any words (or) بـا كليك بر روي دكمه هاي راديويي مي‌باشد. يكي از كاربردهاي مهم اين عملگر پوشش مفاهيم يا اصطلاحات مترادف، مرتبط، يا با املاهاي متفاوت مي‌باشد.

۱۰ – اسـتفاده از عملگـر NOT : اپراتور Not به مفهوم "نه" و يا به جز كه در اين صورت تمامي جواب‌هاي بازگشتي كه حاوي عبارت يا كلمه كليدي هستند حذف خواهند گرديد و براي اجراي آن تنها كافيست كه not را قبل از عبارت يا كلمه كليدي مورد نظرتان با يك فاصله بياوريد.

۱۱ – اسـتفاده از كوتـاه‌سـازي [۱٤]كليد واژه‌ها: اين تكنيك به ما امكان مي‌دهد كه با وارد كردن بخشي از يك كليدواژه بتوانيم مشـتقات مختلـف آن را نيـز در فـرآيند جسـتجو بازيـابي كنيم. اكثر موتورهاي جستجو اين تكنيك را با استفاده از علامت ستاره (*) ارائـه مـي‌دهنـد. يـكي از مشكلات استفاده از اين تكنيك اين است كه باعث بازيابي اطلاعات غيرمرتبط و ناخواسته زيادي مي‌شـود.

۱۲ – اسـتفاده از عملگـر نزديك‌يابي[۱۵]: در بسياري از موارد استفاده از عملگر and باعث بازيابي اطلاعاتي شود كه براي ما مفيـد نمي‌باشد، به اين دليل كه اين عملگر كليدواژه‌ها را در هر كجاي متن كه باشند بازيابي مي‌كند. در اين موارد استفاده از تكنيـك نـزديك‌يـابي مـي‌توانـد از ريـزش كاذب اطلاعات و يا بازيابي اطلاعات غير مرتبط جلوگيري نمايد. همه موتورهاي جستجو قـابليت اسـتفاده از اين تكنيك را ندارند ولي به عنوان مثال در موتور جستجوي آلتاويستا مي‌توان با استفاده از عملگر NEAR از اين تكنيك استفاده نمود.

۱۳ – جسـتجوي تـركيبي بـا اسـتفاده از پرانتـز: ايـن تكنيـك يـكي از مهـم‌تـرين تكنيـك‌هـاي جستجو مي‌باشد كه به وسيله آن

مي‌توان تا حدود زيادي از بازيابي موارد غير مرتبط در محيط وب جلوگيري كرد. در اين روش مي‌توان از همه عملگرهاي جستجو كه در بالا گفته شده يك‌جا استفاده كرد و آن‌ها را با همديگر تركيب نمود.
۱۴ – جستجوي كليدواژه در عنوان صفحات وب: اين تكنيك با اين پيش فرض كه عنوان يك صفحه وب تا حدود زيادي نمايانگر محتواي اطلاعات موجود در آن است به جستجوي واژه‌هاي كليدي در عنوان سايت‌ها مي‌پردازد. علامت آن در موتورهاي جستجو متفاوت است ولي اغلب موتورهاي جستجو از طريق فهرست انتخابي و يا گزينه‌هاي ديگر اين امكان را فراهم مي‌آورند.
۱۵ – جستجوي حوزه سايت‌ها: با توجه به اين كه به صورت قراردادي هر كشوري حوزه خاصي در محيط وب دارد، قابليت جستجوي حوزه سايت‌ها به ما اين امكان را مي‌دهد كه فرايند جستجو را به حوزه خاصي نظير سايت‌هاي وب ايران (ir) و يا سايت‌هاي وب سازمان‌هاي غير انتفاعي (org) محدود كنيم. دستورات استفاده از اين تكنيك در موتورهاي جستجو مختلف مي‌باشد.
۱۶ – محدود كردن جستجو به زبان‌هاي مختلف؛ باعث مي‌شود نتايج جستجو به زبان‌هاي ديگر آورده نشود و انتخاب مطلب مورد نظر آسان‌تر است.
۱۷ – محدود كردن جستجو به تاريخ انتشار منابع در وب: تاريخ انتشار يا به اصطلاح روزآمدي مطلب به خصوص در منابع علمي اصل مهمي است و اين‌گونه محدوديت باعث مي‌شود بنا به نياز كاربر جديدترين و يا قديمي‌ترين منبع بازيابي بشود.
۱۸ – جستجوي رسانه‌هاي مختلف: موسيقي، عكس، ويدئو : زماني كه فقط نوع خاصي از رسانه مورد نياز است به عنوان مثال زماني كه به عكس يك شخصيت نياز داريم، جستجو در ميان عكس‌ها باعث مي‌شود نتيجه جستجو شامل اطلاعات ديگري در مورد آن شخصيت نباشد.
۱۹ – جستجوي صفحات با فرمت‌هاي مختلف: PDF, Word, MP3, MPEG, ...: زماني كه فرمت خاصي مورد نظر است مي‌توان از اين تكنيك استفاده كرد . به عنوان مثال اگر مايل باشيم منبع بازيابي شده در فرمت PDF باشد، اين تكنيك مي‌تواند مفيد باشد.
۲۰ – آگاهي از پيش‌فرض‌هاي جستجو در موتور جستجو: با توجه به اين كه هر موتور جستجو براي تركيب واژه‌ها يك پيش‌فرض دارد و اگر از هيچ گونه عملگري استفاده نشود، كليدواژه‌ها را به صورت پيش‌فرض با يكي از عملگرهاي جبر بولي تركيب مي‌كند؛ آگاهي از اين پيش‌فرض موتورهاي جستجوي مختلف مهارت ما را در جستجو بالا مي‌برد.
۲۱ – وب نامرئي: وب نامرئي به دو دليل كمّي و كيفي اهميت دارد كمّي از اين نظر كه موتورهاي جستجو فقط قادر هستند حدود ۱۶ درصد از اطلاعات موجود دراينترنت را بازيابي كنند و اندازه وب نامرئي تقريبا ۵۰۰ برابر وب مرئي است و كيفي از اين نظر كه منابع اطلاعاتي موجود در وب عميق معمولا ارزشمند و مفيد هستند و در بسياري از موارد پاسخ‌گوي نياز كاربران مي‌باشند. آشنايي با ابزارهايي كه براي شناسايي منابع وب نامرئي به وجود آمده‌اند و كاربران را به سايت‌هاي مناسب راهنمايي مي‌كنند، باعث دسترسي به اين بخش عظيم از اطلاعات مفيد و ارزشمند مي‌شود. مثل سايت Invisibleweb كه فهرستي از منابع نامرئي را و سايت Completeplaset كه فهرستي از تقريبا ۴۰۰۰۰ پايگاه اطلاعاتي وب نامرئي را ارائه مي‌دهد.[۱۶]

راهبرد جستجو در اينترنت

جستجو عبارت از جستجو در منابعي مشخص با استفاده از كليدواژه‌ها و عبارت‌هاي خاص در حوزه‌هاي موضوعي ويژه است. طراحي نظام‌مند مراحل انجام يك جستجو را راهبرد جستجو مي‌گويند به نظر پائو «راهبرد جستجو عبارت است از فرايندي كه از طريق آن فايلي مورد جستجو قرار مي‌گيرد تا مدارك متناسب با نياز كاربر شناسايي شود. اين مدارك بر اساس مجموعه‌اي از معيارهايي كه شخص متقاضي مطرح مي‌كند بازيابي مي‌شود»[۱۷] هر فرايند جستجو مي‌تواند به مراحل ارائه درخواست دقيق، انتخاب منابع اطلاعاتي مناسب، آماده كردن جستجو و اجراي جستجو تقسيم شود. بر خلاف منابع نمايه‌سازي شده در پايگاه‌هاي اطلاعاتي كتاب‌شناختي؛ مدارك در اينترنت از طريق واژگان كنترل شده قابل بازيابي نيستند. بنابراين جستجوگر براي بازيابي بايد بر فنون خاص اينترنت متكي باشد. نخست آگاهي از ابزارهاي مختلف جستجو در اينترنت و در ادامه انتخاب يكي از اين ابزار براي جستجوي اطلاعات مورد نياز مي‌باشد. هزاران موتور جستجو، صدها ابرموتور جستجو و راهنماهاي موضوعي وب و پايگاه‌هاي تخصصي وجود دارد و انتخاب درست ابزار جستجو در ابتداي كار جستجو مي‌تواند يك جستجوي موفق را باعث گردد. در زير چند معيار براي انتخاب ابزار جستجو آورده مي‌شود:
- اگر در جستجوي اطلاعات خاصي باشيد بهتر است از موتورهاي جستجو استفاده كنيد.
- اگر در جستجوي يك واژه مبهم يا منحصر به فرد هستيد از ابرموتورهاي جستجو استفاده نمائيد.
- اگر در جستجوي اطلاعات عمومي روي موضوعات عام هستيد از راهنماهاي موضوعي وب استفاده كنيد.

- اگر در حال جستجوي اطلاعات علمي هستيد از كتابخانه‌هاي مجازي استفاده كنيد.
- اگـر در جسـتجوي آخـرين اطلاعـات يـا براي تغيير پوياي فهرست مطالب، آخرين خبرها، راهنماهاي دفتر تلفن، دسترسـي به زمان پروازهاي هوايي و غيره هستيد از پايگاه‌هاي تخصصي استفاده كنيد.[۱۸]

بـراي جسـتجوي اطلاعـات از اينتـرنت چهـار شـيوه وجـود دارد شـيوه نخسـت دسترسـي بـه اطلاعات از طريق نشاني پايگاه اطلاعـاتي مورد نظر بر روي اينترنت (URL) است، كه در اين صورت نشاني پايگاه اطلاعاتي در سطر نشاني برنامه مرورگر وب تـايپ مـي‌شود و برنامه مرورگر وب مراجعه كننده را به وب سايت آن نشاني هدايت خواهد كرد. اما اگر فقط يك حرف يا علائم نقطـه‌گـذاري از قلـم بيفتـد، برنامـه مرورگـر نخواهـد توانسـت آن پايگـاه را بـاز نمايد. روش دوم دنبال كردن لينک‌های موجود در صـفحات وب است كه كاربران را از صفحه‌ای به صفحه ديگر هدايت می‌كند. اين سـهولت دسترسـی به منابع در وب از امتيازات بـزرگ آن اسـت و بـرای كـاربران امكـان مـرور سـريع و آسـان در منـابع مختلـف را فـراهم مـی‌كند. روش سوم بازيابی گزينشـی اطلاعـات اسـت كه در آن در واقع به جای آنكه كاربران شخصا در جستجوی اطلاعات مورد نظر باشـند، موضوعات مورد نياز خود را بـه سيسـتم‌هـای بازيـابی گزينشـی می‌سپارند و سپس در طول زمان، اطلاعات دريافتی جديد توسط سيستم برای آنها به طور خودكار ارسال خواهد شد.

چهـارمين روش كـه در واقـع معمـول‌تـرين و متداول‌ترين راه بازيابی اطلاعات در وب است استفاده از موتورهاي جستجو است. هنگام جستجو بايد دقت كرد كه موتور جستجو به طور معمول هوش‌مند نيست و معمولاً به دنبال كليه كليدواژه‌هايي كه شما به دستگاه داده‌ايد بدون توجه به معناي آن‌ها مي‌گردد.

نكات كليدي جستجو به زبان فارسي

براي جستجوي مطالب فارسي طبق گفته‌هاي پيشين چنانچه آدرس سايت به خصوصي كه در زمينه موضوعي مورد نظر ما فعـاليت مـي‌كنـد را داشـته باشـيم؛ مـي‌تـوان مسـتقيماً به آن سايت رفته و از مطالب آن استفاده نمود. به عنوان مثال سايت تخصصـي برنامـه‌نويس مطالب مفيدي در زمينه رايانه و علوم وابسته، به ما ارائه مي‌دهد و يا سايت عمران در زمينه موضوعي عمـران فعـاليت مـي‌نمـايد و نيـز سـايت‌هاي انجمن رياضي در زمينه رياضي، سايت انجمن فيزيك ايران در زمينه فيزيك، سايت انجـمن روان‌شنـاسـي ايران در زمينه روان‌شناسي و علوم تربيتي فعاليت مي‌نمايند، مركز اطلاعات و مدارك علمي ايران با دارا بـودن پايگـاه‌هـاي اطلاعـاتي مختلـف مخصوصـا پايگـاه پايان‌نامه‌ها مي‌تواند مورد استفاده متخصصين تمام رشته‌ها گردد. ولي چنـان‌چـه امكـان اسـتفاده از ايـن سـايت‌هـا نباشـد و يا آدرس اين سايت‌ها را نداشته باشيم بايستي مطلب مورد نظر خود را بوسيله يكي از موتورهاي جستجو پيدا بكنيم.

انتخـاب موتـور جستجو عامل مهمي در فرايند جستجو است. در حال حاضر ابزارهاي كاوش مختلفي در ايران ظهور پيدا كرده‌انـد. ليـكن ابزارهـاي جسـتجويي كـه امكـان جسـتجوي اطلاعـات به زبان فارسـي را در اختيار قرار مي‌دهند، محدودند. از طرف ديگـر، امكانـات و قـابليت‌هـاي آن‌هـا بـراي بازيـابي موثـر و مناسـب اطلاعات متغير هستند. برخي از ابزارهاي كاوش با امكانات جسـتجوي فارسـي عبـارتند از: ان.پـي. ايـران NPiran ، ايـران‌هـو Iranhoo ، ايـران‌مهـر IranMehre ، پارسـيك Parseek ، گوگل Google.

در بيـن ابزارهاي كاوش فوق ، تنها موتور كاوش گوگل داراي برنامه روبات به منظور شناسايي و نمايه‌سازي صفحات يا سايت‌هـاي وب بـه زبـان فارسـي و نمايـه‌سـازي خودكـار مـي‌باشد و قادر است صفحات فارسـي را در قالب يونيكد شناسايي و در پايگاه خود نمايه كند و سايت پارسيك نيز از پايگاه گوگل براي جستجو و بازيابي اطلاعات استفاده مي‌كند. به تعبير ديگر، چهار ابـزار كـاوش ديگر توسط نمايه‌سازي انساني اداره مي‌شوند و از اين لحاظ راهنماي موضوعي تلقي مي‌شوند و انسان، فرآيند شناسايي، بررسـي و نمايه‌سازي سايت‌ها يا صفحات وب را بر عهده دارد.[۱۹]

معمـولاً بـه جـهت دامنـه وسـيع موضـوعي و نيـز صـفحه بـه زبـان فارسـي گوگـل اكثـر كـاربران از ايـن موتـور جستجو استفاده مي‌نمايند. براي جستجوي بهتر توجه به نكات زير ضروري به نظر مي‌رسد:

- با تركيب چند واژه كليدي مهم خيلي سريع مي‌توانيم مطلب مورد نظر خود را بدست بياوريم.
- دقـت در انتخـاب كليد واژه‌ها به طوري‌كه واژه‌هاي انتخابي بطور دقيق نماينده نياز اطلاعاتي ما باشند كمك خواهد كرد تا از نتايج جستجوي گسترده‌اي كه در اكثر موارد بار اطلاعاتي مفيد ندارند دوري گزينيم.
- اسـتفاده از تكنيـك جستجوي عبارتي كه در آن عبارت جستجوي مورد نظر خود را داخل گيومه" " مي‌گذاريم و به اين ترتيب بـه موتـور جسـتجو مـي‌گوييـم كـه مطـلب مـورد نظر ما بايستي عين اين عبارت باشد، نيز در محدود كردن نتايج جستجو كمك فراوان مي‌كند.
- اسـتفاده از انـواع محدودگرهـاي زبـاني، زمـاني، مكـاني، شـكلي، و موضـوعي و ... در جسـتجوي پيشرفته گوگل به ما در رسيدن سريعتر به مطلب مورد نظر كمك فراواني مي‌كند.

- به علت اين كه منابع و اطلاعات موجود در اينترنت بوسيله افراد مختلف و بدون كنترل در شيوه‌هاي رسم‌الخط و بدون ويرايش صاحب‌نظران منتشر مي‌گردد آشنايي با گونه‌هاي مختلف نوشتاري و املاهاي مختلف يك واژه و يا يك مفهوم در زبان فارسي به ما كمك مي‌كند كه با جستجوي گونه‌هاي مختلف نوشتاري يك واژه يا يك مفهوم و استفاده از واژه‌هاي مترادف و متشابه و شكل‌هاي ديگر نوشتاري آن واژه و نيز استفاده از انواع شكل‌هاي جمع و مفرد يك واژه جامعيت جستجوي خود را بالا ببريم . به عنوان مثال براي جستجوي مطلبي در زمينه بتن بايستي آنرا به دو صورت «بتون» و « بتن» جستجو نمائيم تا به تمام مطالبي كه در زمينه بتن مي‌باشد دسترسي داشته باشيم و يا به عنوان مثال دوم براي جستجوي مطلبي در باره «آبگرمكن» براي دستيابي به همه اطلاعات موجود بايستي آن را به چهار شكل زير بنويسيم «آب گرم كن، آب گرمكن، آبگرم كن، آبگرمكن» واضح است كه هر كدام از اين كلمات نتايج متفاوتي را در موتور جستجو بدست مي‌دهد. «استاد، اساتيد، استادان، استادها» «آمريكا، امريكا» «ئيدروژن، هيدروژن» «آنلاين، پيوسته، درون خطي» از مثال‌هاي ديگري هستند كه جستجو به تمام اين شكل‌ها جامعيت جستجوي ما را زياد مي‌كند و به ما در از دست ندادن مطالب مفيد كمك مي‌كند .
- با استفاده از عملگرهاي بولي، دقت جستجو را بالا برده و نتايج جستجوي كم و مفيدي را بدست بياوريم.
- مترادفات: با استفاده از شكل‌هاي مختلف مترادفات موجود براي يك مفهوم و هم چنين شبه مترادفات و يا حتي گاهي كلمات متضاد مثل بي‌سوادي و سوادآموزي در جستجو مي‌توان جامعيت جستجو را بالا برد.
- اسامي مشهور و اسامي علمي: آگاهي از شكل‌هاي مختلف اسامي علمي و مشهور عاميانه و اسامي تجاري يك پديده و يا وسيله و ... و استفاده از آن‌ها مي‌تواند جامعيت جستجو را بالا ببرد.
- با توجه به اين كه در اكثر وب‌ها از واژه خارجي يك كلمه به همان صورت و با همان الفبا استفاده مي‌شود استفاده از شكل خارجي اين لغات و واژه‌ها نيز مي‌تواند جامعيت جستجوي ما را بالاتر ببرد.
- با توجه به اين كه در برخي از سايت‌ها و نيز وبلاگ‌ها روش خاصي براي رفع مشكلات فارسي پيشنهاد كرده‌اند و واضح است كه خودشان نيز از آن رسم‌الخط استفاده مي‌كنند، آگاهي از اين شكل‌هاي مختلف و جستجو به اين شكل‌ها مي‌تواند باعث جامعيت جستجو گردد. از اين موارد مي‌توان به حذف واو معدوله در برخي سايت‌ها و وبلاگ‌ها اشاره كرد كه به عنوان مثال «خواهر» را به صورت «خاهر» مي‌نويسند و يا حذف تنوين در برخي منابع كه به عنوان مثال «عملاً» را به صورت «عملن» مي‌نويسند.

در نهايت اين كه «در تشكيل صفحات وب فارسي ، جاي يك استاندارد حاكم بر عملكرد تاليف نويسندگان وب، خالي است. استانداردي كه انتخاب بعضي كلمات داراي چندين رسم‌الخط و حتي انتخاب بعضي كلمات كه بر مفاهيم متنوعي دلالت دارند را منحصر به فرد نمايد و مولفان را از طرفي ترغيب به انتخاب گونه زباني مناسب، براي تضمين كيفيت ارتباط و انتقال مؤثر پيام و از طرف ديگر موظف به حفظ سلامت زبان و رعايت استانداردهاي آن به‌عنوان يك وظيفه رسانه‌اي نمايد. ايجاد و گسترش چنين استانداردي به عهده "فرهنگستان زبان و ادب فارسي" و با هماهنگي انجمن‌ها و شوراهاي علمي يا صنفي انفورماتيك در ايران است. تعويق در تنظيم اين استاندارد ، با توجه به رشد روز افزون وب‌هاي فارسي زبان ، هزينه‌هاي جبران ناپذيري در بر خواهد داشت.»[۲۰]

---

[۱] بر گرفته از: محمد صابر راثي ساربانقلي. " بررسي مشكلات جستجو و بازيابي اطلاعات به زبان فارسي از اينترنت با مطالعه موردي بر روي كاربران مركز اينترنت دانشگاه آزاد اسلامي واحد شبستر" . پايان‌نامه كارشناسي ارشد كتابداري و اطلاع‌رساني، دانشگاه آزاد اسلامي، واحد تهران شمال، ۱۳۸٤

[۲] كارشناس ارشد كتابداري و اطلاع‌رساني دانشگاه آزاد اسلامي واحد شبستر

[3] accents

۴ پرويز ناتل خانلري. زبان‌شناسي و زبان فارسي. (تهران: توس،۱۳۷۳). ص. ۲۵٦

[۵] مجتبي مينوي. مينوي بر گستره ادبيات فارسي، به كوشش ماه منير مينوي. (تهران: توس، ۱۳۸۰)، ص. ۵۱۰

[۶] واو معدوله واوي است كه در اين زمان عموماً نوشته مي‌شود ولي خوانده نمي‌شود، مانند خواهش. اما در زمان قديم آن را با كيفيت خاصي تلفظ مي‌كرده‌اند و چون در هنگام تلفظ ضمه به فتحه عدول مي‌كرده‌اند، آن را واو معدوله ناميده‌اند. هنوز در برخي از لهجه‌ها تلفظ آن به صورت قديم مانده است. پيش از واو معدوله هميشه حرف «خ» و پس از آن يكي از حروف «د.ر.ز.س.ش.ن.و.هـ.ي» آمده است.

[۷] OCR= Optical Character Reader فرايندي كه در طي آن يك وسيله الكترونيكي كاراكترهاي چاپ شده بر روي كاغذ را آزمايش مي‌كند و شكل آن‌ها را با بررسي الگوهاي تيره و روشن تعيين مي‌كند. پس از تعيين آشكال توسط اسكنر يا وسيله مورد استفاده براي خواندن، روش‌هاي تشخيص نوري كاراكترها براي تبديل اشكال به متون كامپيوتري مورد استفاده قرار

مي‌گيرند. (فرهنگ تشريحي اصطلاحات كامپيوتري ميكروسافت. مترجم فرهاد قلي‌زاده نوري. [تهران: كانون نشر علوم، ١٣٧٩]، ص. ٤٤٥
[٨] محمدصادق محقق زاده، كاظم زارعيان. "ارائه راه حل براي برخي مسائل اتوماسيون و نگارش فارسي" فصلنامه اطلاع‌رساني. (دوره ١٩، شماره‌ء ٣ و ٤) ص.

[٩] ماندانا صديق بهزادي. "ناهماهنگي ضبط نام‌هاي بيگانه در فارسي" . فرهنگ (كتاب سيزدهم، زمستان ١٣٧١) ص.١٠٣-١١٦
[١٠] مجيد رهباني. "قند پارسي در شبكه جهاني: كتاب‌هاي ديجيتال و كتابخانه‌هاي مجازي فارسي در اينترنت. "جهان كتاب، (١٨٣) ص.
[١١] "تابوي اصلاح خط". جام جم. (٤ بهمن ١٣٨٣)
[١٢] براندا پاريس سيبلي. " فهرست‌نويسي منابع اينترنت: سازماندهي وب در كتابخانه‌هاي محلي و غير آن" ترجمه محسن حاجي زين العابديني .فصلنامه اطلاع رساني (دوره ١٦، شماره ٣ و ٤) ص. ١
[١٣] عبدالرسول خسروي. " وب نامرئي" فصلنامه اطلاع‌رساني (دوره ٢٠، شماره ١ و ٢) ص. ٥٣
[14] Truncation
[15] Proximity search
[١٦] عبدالرسول خسروي. " وب نامرئي" فصلنامه اطلاع‌رساني (دوره ٢٠، شماره ١ و ٢) ص. ٥٤
[١٧] ميرندا لي پائو. مفاهيم بازيابي اطلاعات. ترجمه اسدالله آزاد و رحمت الله فتاحي. (مشهد: دانشگاه فردوسي. ١٣٧٨) ص. ٣١٤
[١٨] دانيل بازاك. "جستجوي وب بطور كارآمدتر: رهنمودها، فنون و راهبردها". مترجمين مريم اسدي، اكرم اسدي. مجله الكترونيكي مركز اطلاعات و مدارك علمي ايران. شماره چهارم دوره دوم
[١٩] كيوان كوشا. "معيارهاي ارزيابي ابزارهاي كاوش اينترنت: مطالعه مقايسه‌اي‌بر روي ابزارهاي كاوش وب با واسط جستجوي فارسي" . مجله الكترونيكي كتابدار.

[٢٠] محسن صديقي، كامران زماني‌فر. " روشي براي رفع چالشهاي محتوا كاوي وب‌هاي فارسي زبان"مجله الكترونيكي مركز اطلاعات و مدارك علمي

منابع و مآخذ

١- آزادي، قاسم." اينترنت: سازمان‌دهي و جستجو" . ابرار اقتصادي. ١٤.
٢- ابوالقاسمي، محسن. تاريخ زبان فارسي. تهران: سمت، ١٣٨٠.
٣- احمدي فصيح، صديقه، "آشنايي با شبكه جهاني وب". فصلنامه اطلاع‌رساني، دوره ١٨، شماره ١ و ٢.
٤- اديب سلطاني، مير شمس الدين. راهنماي آماده ساختن كتاب: براي مولفان، مترجمان، ويراستاران، رسانه‌گران، كتابداران ....تهران، علمي و فرهنگي، ١٣٨١.
٥- ---------- درآمدي بر چگونگي شيوه‌ي خط فارسي. تهران: اميركبير، ١٣٧٨.
٦- اشرف‌زاده ، بهرام. "زبان فارسي در وبلاگ‌هاي فارسي".
http://www.persianfarsi.com/articles/zabaneweblog.htm
٧- بازاك، دانيل، "جستجوي وب بطور كارآمدتر: رهنمودها، فنون و راهبردها" . مترجمين مريم اسدي، اكرم اسدي،
http://www.irandoc.ac.ir/data/E_J/vol2/Search_Web.htm
٨- باقري، مهري. تاريخ زبان فارسي. تهران: قطره، ١٣٧٨.
٩- بهار، محمدتقي. سبك شناسي يا تاريخ تطور نثر فارسي. تهران: اميركبير، ١٣٧٠.
١٠- پائو، ميراندا لي. مفاهيم بازيابي اطلاعات. ترجمه اسدالله آزاد و رحمت الله فتاحي. مشهد: دانشگاه فردوسي، ١٣٧٨.
١١- "تابوي اصلاح خط". جام جم، ٤ بهمن ١٣٨٣.
١٢- خانلري، پرويز. زبان شناسي و زبان فارسي. تهران: توس، ١٣٧٣.
١٣- خسروي، عبدالرسول. "وب نامرئي". فصلنامه اطلاع‌رساني، دوره ٢٠، شماره ١ و ٢.
١٤- خلخالي، نازيلا، بررسي علمي شيوه‌ي خط فارسي. تهران: ققنوس، ١٣٧٥.
١٥- خوانساري، جيران. "تكامل وب و مقايسه ابزارهاي جستجو در اينترنت". فصلنامه اطلاع‌رساني، دوره ١٦، شماره ٣ و ٤.

۱۶- داورپناه، محمد رضا. جستجوي اطلاعات علمي و پژوهشي در منابع چاپي و الكترونيكي؛ شامل حوزه هاي علوم، فني مهندسي ... .تهران: دبيزش، ۱۳۸۱.
۱۷- دراگولانسكو، نيكلاي جورج. "ارزيابي كيفي وب سايت‌ها: ابزارها و معيارها". ترجمه غلام حيدري.
http://www.irandoc.ac.r/Data/E_J/vol4/haidari.htm
۱۸- دشتي، افشين."بررسي سه پيشنهاد درشيوه نگارش خط فارسي حتا مثلن، خاهر". روزنامه شرق ۱۰ تير ۱۳۸۳.
۱۹- "دگرگونيهاي زبان و خط فارسي در محيطهاي رايانه‌اي: گفتگو با دكتر عاصي". پيام ارتباطات، ۳۵.
۲۰- "ده نكته براي جستجوي سريع درگوگل". دانشمند، ٤٨٧.
۲۱- رقابي، فرنوش؛ شريفي، 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- رهباني، مجيد. "قند پارسي در شبكه جهاني: كتاب‌هاي ديجيتال و كتابخانه‌هاي مجازي فارسي در اينترنت". جهان كتاب، ۱۸۳.
۲۳- رئيسي، محمد رضا." OCR : آموزش الفباي فارسي به رايانه".
http://iranwsis.org/Default.asp?C=IRNW&R=&I=93
۲۴- سيبلي، برندا پاريس." فهرست‌نويسي منابع اينترنت: سازمان‌دهي وب در كتابخانه‌هاي محلي و غير آن". ترجمه محسن حاجي زين‌العابديني . فصلنامه اطلاع‌رساني، دوره ١٦، شماره ٣ و ٤.
۲۵- صديق بهزادي، ماندانا، "ناهماهنگي‌هاي ضبط نامهاي بيگانه در فارسي". فرهنگ. كتاب سيزدهم، زمستان ۱۳۷۱.
۲۶- صديقي، محسن، زماني‌فر، كامران. "روشي براي رفع چالشهاي محتواي كاوي وب‌هاي فارسي زبان".
http://www.irandoc.ac.ir/Data/E_J/vol4/shahidi.htm
۲۷- طباطبايي، علاءالدين. "در دشواري‌هاي رايانه‌اي زبان فارسي". نشر دانش، ۱۰۳.
۲۸- عاصي، مصطفي. "نقش رايانه در ايجاد استانداردهاي زباني". فرهنگ. سال چهاردهم، شماره‌هاي اول – دوم،بهار – تابستان ۱۳۸۰.
۲۹- عزيز محمدي، فاطمه، "بررسي برخي فرآيندهاي رايج قرض‌گيري در زبان فارسي". فصلنامه اطلاع‌رساني، دوره ۱۸، شماره ۳ و ٤.
۳۰- فتاحي، رحمت الله. "چالش هاي سازمان‌دهي منابع دانش در آغاز قرن بيست و يكم با نگاهي بر دانش فهرست‌نويسي در ايران" . فصلنامه كتاب. ٤٨.
۳۱- فرهنگ تشريحي اصطلاحات كامپيوتري ميكروسافت. مترجم فرهاد قلي‌زاده نوري. تهران: كانون نشر علوم، ۱۳۷۹.
۳۲- قاسمي، علي حسين، اطلاع يابي در اينترنت. تهران: چاپار، ۱۳۸۰.
۳۳- كوشا، كيوان، ابزارهاي كاوش اينترنت : اصول، مهارتها و امكانات جستجو در وب . تهران: نشر كتابدار، ۱۳۸۱.
۳۴- ---------- "معيارهاي ارزيابي ابزارهاي كاوش اينترنت: مطالعه مقايسه‌اي‌بر روي ابزارهاي كاوش وب با واسط جستجوي فارسي ".
http://www.ketabdar.org/magazine/detailarticle.asp?number=25
۳۵- كوك، آليسون. راهنماي يافتن اطلاعات با كيفيت در اينترنت ،راهبردهاي گزينش و ارزيابي. ترجمه مهدي خادميان. مشهد: كتابخانه رايانه‌اي، ۱۳۸۲.
۳۶- گزني، علي. "جست و جوي اطلاعات و ساز و كارهاي بهينه‌سازي آن". فصلنامه كتاب، ٤٥.
۳۷- گلاسبرنر، آلفرد. كليد طلائي جستجو در اينترنت. ترجمه رضا مجري، ليلا ملكان، عبدالله تباره . تهران: انتشارات خليج فارس، ۱۳۸۲.
۳۸- محقق زاده، محمدصادق ؛ زارعيان، كاظم. " ارائه راه حل براي برخي مسائل اتوماسيون و نگارش فارسي" فصلنامه اطلاع‌رساني شماره ۳و ٤ دوره ۱۹.
۳۹- محمدي‌فرد،داود؛ آباقري، محمد. كامپيوتر براي كتابداران و اطلاع‌رسانان. تهران: چاپار، ۱۳۸۳.
۴۰- مختاري نبي، ابراهيم." سازمان‌دهي منابع اينترنت: چالشها و ضرورتها ".
http://www.irandoc.ac.ir/data/E_J/vol1/organaizing.htm
۴۱- مرتضائي، ليلا . "مسائل زبان و خط فارسي در ذخيره و بازيابي اطلاعات". فصلنامه اطلاع‌رساني. دوره ۱۷. شماره ۱ و ۲ .
۴۲- منصوريان، يزدان"عوامل موثر بر جستجو و بازيابي اطلاعات در شبكه جهانگستر وب" .
http://www.ketabdar.org/magazine/detailarticle.asp?number=23

۴۳- مينوي، مجتبي. مينوي بر گستره ادبيات فارسي. به كوشش ماه منير مينوي. تهران: توس، ۱۳۸۰.
۴۴- "نگاهي به مشكلات خط فارسي در ارتباط با فناوري اطلاعات".
http://www.itna.ir/archives/report/001948.htm
۴۵- نوتس، گري. راهبردها و شيوه‌هاي جستجو در اينترنت. ترجمه سيمين نيازي. فصلنامه كتاب،
۴۶- نوروزي، علي‌رضا. "جستجو در اينترنت:آشنايي با موتور جستجوي گوگل". فصلنامه اطلاع‌رساني، دوره ١٦، شماره ٣ و ٤
.