**نام مقاله: بررسي ميزان اثر بخشي عناصر ابرداده اي بر رتبه بندي صفحات وب توسط موتورهاي كاوش عمومي**

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

**شماره نشريه: 38 \_ شماره دوم، جلد 10**

**پديدآور: عاطفه شريف**

**چكيده**

**پژوهش حاضر با هدف تعيين ميزان اثربخشي عناصر ابرداده‌اي بر رتبه‌بندي صفحات وب توسط سه موتور كاوش عمومي «گوگل»، «ياهو»، و «ام‌اس‌ان» انجام پذيرفت. اثربخشي پنج عنصر ابرداده‌اي نشانه عنوان زبان «اچ‌.تي.‌ام.‌ال»، ابرنشانه‌هاي كليدواژه‌ها و توصيف زبان «اچ.تي.ام.ال»، ابرداده‌هاي عنوان و موضوع از قالب ابرداده‌اي «دابلين كور» كه بر بازنمون موضوعي صفحات وب متمركزند، با روش تجربي آزموده شد. 84 صفحة وب در گروههاي گواه و آزمون به عنوان جامعة مورد مطالعه تحت دامنة فرعي [http://metadata.irandoc.ac.ir](http://metadata.irandoc.ac.ir/) منتشر شد. از ميان موتورهاي كاوش انتخابي، دو موتور كاوش ياهو و گوگل، صفحات را نمايه‌سازي كردند. كليدواژه‌هاي منحصر به فرد و تكرار شونده كه در مرحله طراحي در قالب عناصر ابرداده‌اي به صفحات گروه آزمون افزوده شده بود، بررسي شد و ميزان معني‌داري تفاوت رتبه صفحات دو گروه با استفاده از آزمون غيرپارامتري «يو.من.وايتني» محاسبه گرديد. نتايج، نشانگر اثربخشي دو عنصر «عنوان» و «توصيف زبان اچ.تي.ام.ال» در هر دو موتور كاوش مورد مطالعه است. كليدواژه‌هاي منحصر به فرد در ابرنشانه كليدواژه‌هاي «اچ.تي.ام.ال» و ابرداده موضوع دابلين كور به بهبود رتبه صفحات در ياهو انجاميد. در حالي كه تكرار كليدواژه‌ها در بخش سرآيند صفحه وب، باعث بهبود رتبه در موتور كاوش ياهو شد، گوگل عملكردي معكوس داشت.**

**كليدواژه‌ها: ابرداده، ابرنشانه‌هاي اچ.تي.ام.ال، دابلين كور، بازيابي اطلاعات، رتبه‌بندي نتايج كاوش، موتورهاي كاوش عمومي، وب، قابليت دسترسي صفحات وب.**

**مقدمه**

همزمان با تولد وب، جهان شاهد تغييرات شگرفي در حوزة ذخيره و بازيابي اطلاعات بوده است. هر روز بر تعداد صفحات وب افزوده مي گردد و حجم وسيعي از اطلاعات، در بستر آن، به صورت ساختار نيافته [[2]](http://192.168.0.110/editor/main.htm" \l "_ftn2" \o ") (يالتاقيان[[3]](http://192.168.0.110/editor/main.htm" \l "_ftn3" \o ")، 2002) و فارغ از كنترل محتوايي و كتاب‌شناختي منتشر مي‌شود (زانگ و جاستريم [[4]](http://192.168.0.110/editor/main.htm" \l "_ftn4" \o ")، 2004؛ گوتليب و اليوپولوس[[5]](http://192.168.0.110/editor/main.htm" \l "_ftn5" \o ") ، 2003). در چنين وضعيتي، مسئله اساسي، چگونگي كنترل و مديريت بدنة ساختارنيافته و رشد سريع اين بدنه است (اسدي و جمالي مهمويي،[[6]](http://192.168.0.110/editor/main.htm" \l "_ftn6" \o ") 2004). تاكنون ابزارهاي كاوش[[7]](http://192.168.0.110/editor/main.htm" \l "_ftn7" \o ") از قبيل «موتورهاي كاوش»[[8]](http://192.168.0.110/editor/main.htm" \l "_ftn8" \o ")«ابرموتورهاي كاوش»،[[9]](http://192.168.0.110/editor/main.htm" \l "_ftn9" \o ")«راهنماهاي موضوعي»[[10]](http://192.168.0.110/editor/main.htm" \l "_ftn10" \o ") و «نرم‌افزارهاي كاوش»[[11]](http://192.168.0.110/editor/main.htm" \l "_ftn11" \o ") محيط مجازي وب را تا اندازه اي تحت كنترل و مديريت خويش درآورده‌اند.

كاربران از ميان ابزارهاي رايج كاوش، موتورهاي كاوش را به عنوان نقطه آغازين ورود به اينترنت تلقي مي كنند (اسپينك و ديگران[[12]](http://192.168.0.110/editor/main.htm" \l "_ftn12" \o ")، 2001 نقل در دوال و واگان [[13]](http://192.168.0.110/editor/main.htm" \l "_ftn13" \o ")، 2004؛ بار- ايلان[[14]](http://192.168.0.110/editor/main.htm" \l "_ftn14" \o ") ، 2005؛ زانگ و ديميتروف[[15]](http://192.168.0.110/editor/main.htm" \l "_ftn15" \o ")، 2004)، بيش از 95% ترافيك كاوش در اينترنت به موتورهاي كاوش مربوط است و 80% كاربران، اطلاعات مورد نياز خود را از طريق موتورهاي كاوش به دست مي آورند (هاتلي[[16]](http://192.168.0.110/editor/main.htm" \l "_ftn16" \o ") ، 2002 نقل در زانگ و ديميتروف،a 2005 (. يافتن اطلاعات موضوعي ويژه در وب دشواريهايي دارد و هر روز بر حجم اين دشواريها افزوده مي‌گردد (دروت[[17]](http://192.168.0.110/editor/main.htm" \l "_ftn17" \o ")، 2000، ص209). تعداد نتايج بازيابي شدة موتورهاي كاوش، اغلب چنان فراوان است كه كاربر عملاً جز مرور چند صفحه نخست نتايج، از ساير صفحات منصرف مي‌شود (جانسن، اسپينك، و ساراسويك،[[18]](http://192.168.0.110/editor/main.htm" \l "_ftn18" \o ") 2000؛ فدايي عراقي[[19]](http://192.168.0.110/editor/main.htm" \l "_ftn19" \o ")، 2005، ص13؛ يالتاقيان، 2002) و به ناچار به رتبه‌بندي[[20]](http://192.168.0.110/editor/main.htm" \l "_ftn20" \o ") ارائه شدة موتورهاي كاوش اعتماد مي‌كند (بارـ ايلان، 2005).در اين وضعيت، چنانچه صفحه اي مرتبط، در رتبه‌هاي اول جاي نگيرد، ممكن است از ديد كاوشگر پنهان بماند (زانگ و جاستريم ، 2005، ص92؛ گوتليب و اليوپولوس، 2003).

از سويي، يكي از اولين دغدغه‌هاي ناشران وب‌سايتها دستيابي به رتبه‌هاي برتر در ميان وب‌سايتهاي مشابه و هم موضوع است. بدين منظور همواره سياهه‌اي از عناوين،[[21]](http://192.168.0.110/editor/main.htm" \l "_ftn21" \o ")كليدواژه‌ها[[22]](http://192.168.0.110/editor/main.htm" \l "_ftn22" \o ") و توصيفهايي[[23]](http://192.168.0.110/editor/main.htm" \l "_ftn23" \o ") كه احتمال كسب رتبه‌هاي برتر را دارند، تهيه و در طراحي صفحات لحاظ مي‌شود (ريچاردسون،[[24]](http://192.168.0.110/editor/main.htm" \l "_ftn24" \o ") 2003 نقل در زانگ و ديميتروف، a2005). طراحي نرم‌افزارهايي چون «تحليلگر چگالي كليدواژه»،[[25]](http://192.168.0.110/editor/main.htm" \l "_ftn25" \o ")«ورد تركر»،[[26]](http://192.168.0.110/editor/main.htm" \l "_ftn26" \o ")«وب پزيشن گولد»،[[27]](http://192.168.0.110/editor/main.htm" \l "_ftn27" \o ") و شكل‌گيري و گسترش وب‌سايتهايي كه خدمات توصيه‌اي و مشاوره‌اي بهينه‌سازي صفحات[[28]](http://192.168.0.110/editor/main.htm" \l "_ftn28" \o ") را به منظور كسب رتبه‌هاي برتر در موتورهاي كاوش ارائه مي دهند، تأييدي بر حساسيت و توجه به اين مسئله است.

نتايج تحقيقات، تفاوتهاي قابل ملاحظه‌اي را ميان الگوريتمهاي رتبه‌بندي موتورهاي كاوش عمومي نشان مي‌دهد (بار ـ ايلان، 2005). اطلاعات مربوط به الگوريتمهاي رتبه‌بندي موتورهاي كاوش به صورت طبقه بندي شده[[29]](http://192.168.0.110/editor/main.htm" \l "_ftn29" \o ") و به عنوان اسرار تجاري [[30]](http://192.168.0.110/editor/main.htm" \l "_ftn30" \o ") محافظت مي‌شود. حفظ حالت رقابتي و جلوگيري از سوء استفادة طراحان وب‌سايتها از اين اطلاعات، از جمله دلايل حفاظت هستند (بارـ ايلان، 2005،ص1512). با وجود ابهامهاي موجود در زمينه الگوريتم، رتبه‌بندي موتورهاي كاوش، پژوهش پيرامون چگونگي رتبه‌بندي نتايج متوقف نشده است و تلاشهايي در زمينه كشف عوامل اثرگذار و تعيين ميزان اثرگذاري آن عوامل، انجام پذيرفته است (ترنر و برك بيل[[31]](http://192.168.0.110/editor/main.htm" \l "_ftn31" \o ")، 1998؛ زانگ و ديميتروف، 2004؛ a2005، b2005، صفري[[32]](http://192.168.0.110/editor/main.htm" \l "_ftn32" \o ")،2005؛ محمد[[33]](http://192.168.0.110/editor/main.htm" \l "_ftn33" \o ")، 2006؛ هنشاو           والاسكاس[[34]](http://192.168.0.110/editor/main.htm" \l "_ftn34" \o ") ،2001).

**محدودة بررسي**

صفحات وب متشكل از سه جزء اند: 1. معنا[[35]](http://192.168.0.110/editor/main.htm" \l "_ftn35" \o ") يا محتوا[[36]](http://192.168.0.110/editor/main.htm" \l "_ftn36" \o ") 2. بستر نحوي[[37]](http://192.168.0.110/editor/main.htm" \l "_ftn37" \o ") يا ساختار[[38]](http://192.168.0.110/editor/main.htm" \l "_ftn38" \o ") و 3. پيوندهاي فرامتني[[39]](http://192.168.0.110/editor/main.htm" \l "_ftn39" \o "). محتوا در بستر نحوي زبانهاي نشانه گذاري[[40]](http://192.168.0.110/editor/main.htm" \l "_ftn40" \o ") كه قالب يا ساختار ارائه محتوا را فراهم مي‌آورند، جاي مي‌گيرد و ارتباط ميان اجزاي اطلاعاتي از طريق پيوندهاي فرامتني حاصل مي‌شود. هر يك از اين عناصر - محتوا، ساختار و پيوندهاي فرامتني ـ ويژگيهاي خاصي دارند كه بستر ارزيابي صفحات را فراهم مي آورد و به طور بالقوه در بهبود كيفي رتبه‌بندي حاصل از كاوش اثر گذار است (كوودو-تررو[[41]](http://192.168.0.110/editor/main.htm" \l "_ftn41" \o ")، 2004).

گوگل بيش از هزار عامل را در نظام رتبه‌بندي نتايج خود مدنظر دارد،اما به دليل ماهيت تجاري و حفظ يكپارچگي نتايج كاوش، از ذكر جزئيات بيشتر خودداري مي‌كند (گوگل[[42]](http://192.168.0.110/editor/main.htm" \l "_ftn42" \o ")، 2004). توافقي بر سر مؤثرتر بودن يك عامل نسبت به ساير عوامل وجود ندارد (فيشكين[[43]](http://192.168.0.110/editor/main.htm" \l "_ftn43" \o ")، 2005) و هر يك از موتورهاي كاوش، الگوريتم رتبه‌بندي خاصي را دنبال مي‌كنند (هنشا و والاسكاس، 2001، ص92). با وجود اين، ساختار ابرداده‌اي،[[44]](http://192.168.0.110/editor/main.htm" \l "_ftn44" \o ") محتواي صفحه[[45]](http://192.168.0.110/editor/main.htm" \l "_ftn45" \o ") و(عوامل داخلي)، و وضعيت ارجاعات فرامتني[[46]](http://192.168.0.110/editor/main.htm" \l "_ftn46" \o ") وـ (عوامل خارجي) - از جمله عوامل مؤثر بر رتبه‌بندي ذكر شده است (زانگ و جاستريم، 2005).

در اين پژوهش، از ميان عوامل مؤثر شناخته شده بر رتبه‌بندي نتايج كاوش، تمركز بر عناصر ابرداده‌اي[[47]](http://192.168.0.110/editor/main.htm" \l "_ftn47" \o ") است و از ميان عناصر ابرداده‌اي، پنج عنصر به عنوان معيار ربط فني[[48]](http://192.168.0.110/editor/main.htm" \l "_ftn48" \o ") انتخاب شده و ميزان اثرگذاري هر يك (به صورت منحصر به فرد و در صورت تكرار) بر رتبة صفحات بازيابي شده توسط موتورهاي كاوش عمومي، بررسي و تجزيه و تحليل شده است. عناصر ابرداده‌اي مورد بررسي در اين پژوهش عبارتند از: نشانه عنوان[[49]](http://192.168.0.110/editor/main.htm" \l "_ftn49" \o ") از زبان «اچ.تي.ام.ال»، ابرنشانه‌هاي كليدواژه‌ها[[50]](http://192.168.0.110/editor/main.htm" \l "_ftn50" \o ") و توصيف[[51]](http://192.168.0.110/editor/main.htm" \l "_ftn51" \o ") از زبان «اچ.تي.ام.ال»، دو ابردادةعنوان[[52]](http://192.168.0.110/editor/main.htm" \l "_ftn52" \o ")  و موضوع[[53]](http://192.168.0.110/editor/main.htm" \l "_ftn53" \o ") از قالب ابرداده‌اي دابلين كور.

اگر چه نشانه عنوان زبان «اچ.تي.ام.ال»، عنصر ابرداده‌اي محسوب نمي‌شود؛ اما به واسطه اهميت ويژه اش (ساليوان[[54]](http://192.168.0.110/editor/main.htm" \l "_ftn54" \o ")، 2002؛ نوروزي[[55]](http://192.168.0.110/editor/main.htm" \l "_ftn55" \o ")، 2005) در ميان ساير نشانه‌هاي «اچ.تي.ام.ال»، در كنار چهار عنصر ابرداده‌اي ديگر بررسي شده است.

پنج عنصر منتخب، تنها تعدادي از عناصر ابرداده‌اي موجود در قالب ابرداده‌اي دابلين كور و زبان «اچ.تي.ام.ال»هستند. اين عناصر بر بازنمون موضوعي مدرك متمركزند و در برخي منابع، كاربرد آنها توصيه شده است (زانگ و ديمتروف، 2004 ؛ ساليوان، 2002؛ لي – اسملتزر[[56]](http://192.168.0.110/editor/main.htm" \l "_ftn56" \o ")، 2000، ص206؛ والكي، فرير[[57]](http://192.168.0.110/editor/main.htm" \l "_ftn57" \o ")، 2001، ص272).

**پرسشهاي اساسي**

پژوهش حاضر فاقد فرضيه است و پرسشهاي اساسي آن بدين قرار است:

1.   حضور كليدواژة مورد كاوش در هر يك از عناصر مورد بررسي (نشانه عنوان، ابرنشانه‌هاي كليدواژه‌ها و توصيف زبان نشانه گذاري فرامتن، ابرداده‌هاي عنوان و موضوع قالب ابرداده‌اي دابلين كور) تا چه ميزان بر رتبه‌بندي صفحات در سه موتور كاوش مورد آزمون اثرگذار است؟

2.   تكرار كليدواژة مورد كاوش در دو تا پنج عنصر ياد شده در بخش سرآيند[[58]](http://192.168.0.110/editor/main.htm" \l "_ftn58" \o ") صفحه و صرف‌نظر از نوع عنصر، تا چه ميزان بر بهبود رتبة صفحه اثرگذار است؟

**روش‌شناسي**

پژوهشهايي كه به تعيين ميزان اثربخشي عناصر ابرداده‌اي بر رتبه‌بندي صفحات وب در موتورهاي كاوش پرداخته‌اند، با تفاوتهايي اندك، از روش تجربي بهره برده اند.در اين پژوهش نيز با آگاهي از وجود ساير عواملي كه بر رتبه‌بندي صفحات اثرگذارند، جهت اعمال متغيرهاي مستقل، كنترل متغيرهاي دخيل، و مشاهدة تغييرات در متغير وابسته، از روش تجربي استفاده كرده‌ايم.

جامعةمورد پژوهش، نشريه اي الكترونيكي با 84 صفحة وب است كهپژوهشگر آن را طراحي كرده است. اين نشريه با روندي كه در ادامه مي‌آيد، طراحي شد و سپس در معرض نمايه سازي موتورهاي كاوش قرار گرفت.

**1. مراحل طراحي صفحات**

در مرحله نخست، تعداد چهارده مقاله در موضوعات وب معنايي[[59]](http://192.168.0.110/editor/main.htm" \l "_ftn59" \o ")، هستي‌شناختي[[60]](http://192.168.0.110/editor/main.htm" \l "_ftn60" \o ")، وب‌سنجي[[61]](http://192.168.0.110/editor/main.htm" \l "_ftn61" \o ") و ابرداده‌هاي دابلين كور از ميان مقالات منتشر شده در نشريات رايگان حوزة كتابداري و اطلاع‌رساني كه در فهرست راهنماي نشريات پيوسته رايگان[[62]](http://192.168.0.110/editor/main.htm" \l "_ftn62" \o ") معرفي شده است، انتخاب گرديد.

كليدواژه‌هاي نمايه‌اي اين مقالات به شيوة كنترل نشده (زبان طبيعي) و از بستر عنوان، كليدواژه‌ها (درصورت وجود) چكيده، و در برخي موارد متن مقاله برگزيده شد. پژوهش بر دو دسته كليدواژه متمركز است: الف) كليدواژه‌هاي منحصر به فردو ب) كليدواژه‌هاي تكرارشونده. كليدواژه‌هاي منحصر به فرد آن دسته از كليدواژه‌هايي هستند كه منحصراً به يك عنصر ابرداده‌اي اختصاص يافته‌اند؛ با ديگر كليدواژه‌هاي مربوط به ساير عناصر، همپوشاني و شباهت ندارند و پاسخگويي به نخستين پرسش را ممكن مي‌سازند. كليدواژه‌هاي تكرارشونده، كليدواژه‌هايي هستند كه در تمام عناصر ابرداده‌اي به يك شيوه و ترتيب، قبل از كليدواژه‌هاي منحصر به فرد و به منظور سنجش اثربخشي تكرار كليدواژه‌ها در دو تا پنج عنصر ابرداده‌اي (پرسش دوم) افزوده شده‌اند.

به طور نمونه، كليدواژه‌هاي منحصر به فرد و تكرار شوندة مقاله هشتم با عنوان «A Metadata Registry for the Semantic Web» در جدول 1 آمده است.

**جدول1. نمونه‌اي از كليدواژه‌هاي منحصر به فرد و تكرار شونده**

|  |  |  |
| --- | --- | --- |
| Keywords | | Element |
| **Repeated** | **Unique** |
| Semantic web, Metadata Registry | --- | E1[[63]](http://192.168.0.110/editor/main.htm" \l "_ftn63" \o ") |
| Semantic web, Metadata Registry | Metadata schemas, DCMI vocabulary | E2 |
| Semantic web, Metadata Registry | Automated sharing of metadata, Dublin core metadata Initiative, DCMI Registry | E3 |
| Semantic web, Metadata Registry | --- | E4 |
| Semantic web, Metadata Registry | Registry applications | E5 |

پس از تعيين كليدواژه‌هاي مربوط به هر يك از عناصر مورد بررسي در مرحلة بعد، با توجه به پرسشهاي اساسي پژوهش، گروههاي گواه و آزمون تعيين گرديد. پاسخگويي به پرسشها با وجود شش گروه صفحه امكان‌پذير است. گروه‌بندي صفحات در جدول 2 آمده است. به منظور حفظ هم ترازي در دو گروه گواه و آزمون، تعداد صفحات گروه گواه با تعداد صفحات گروه آزمون، برابر است و محتواي صفحه آزمون در هر مقاله جز در مورد عناصر ابرداده‌اي افزوده شده، مشابه صفحه گروه گواه خود است. بدين ترتيب، چهارده مقالة منتخب در شش شمارة نشريه تكرار شده اند (84= 6\*14).

**جدول2. گروه‌بندي صفحات**

|  |  |  |
| --- | --- | --- |
| **گروه** | **حاوي كليدواژة منحصر به فرد در** | **حاوي تكرار كليدواژه‌ها در** |
| اول | ---- | ---- |
| دوم | ---- | نشانة عنوان «اچ.تي.ام.ال» (عنصر 1) |
| سوم | ابرنشانة كليدواژه‌هاي «اچ.تي.ام.ال» (عنصر 2) | نشانة عنوان و ابرنشانة كليدواژه‌هاي «اچ.تي.ام.ال» (عناصر 1، و 2) |
| چهارم | ابرنشانة توصيف «اچ.تي.ام.ال»    (عنصر 3) | نشانة عنوان، ابرنشانه‌هاي كليدواژه‌ها و توصيف «اچ.تي.ام.ال» (عناصر 1، 2، و3) |
| پنجم | ---- | نشانة عنوان، ابرنشانه‌هاي كليدواژه‌ها و توصيف «اچ.تي.ام.ال»، ابرداده عنوان دابلين‌كور  (عناصر 1، 2، 3، و 4) |
| ششم | ابردادة موضوع قالب ابرداده‌اي دابلين كور (عنصر 5) | نشانة عنوان، ابرنشانه‌هاي كليدواژه‌ها و توصيف «اچ.تي.ام.ال»، ابرداده‌هاي عنوان و موضوع دابلين‌كور (عناصر 1، 2، 3، 4، و 5) |

سپس كليدواژه‌هاي منتخب در بستر نحوي «اچ.تي.ام.ال» نشانه گذاري و در كدمنبع[[64]](http://192.168.0.110/editor/main.htm" \l "_ftn64" \o ") مقالات گنجانده شد. به طور نمونه، كدگذاري مقالة هشتم چنين است: در اين مثال، كليدواژه‌هاي تكرار شونده به صورت توپر نمايش داده شده است. گروه نخست، فاقد كليدواژه است.

**در گروه دوم**

**در گروه سوم**

**در گروه چهارم**

**در گروه پنجم**

**در گروه ششم**

**2. انتشار صفحات طراحي شده**

بدين ترتيب، 84 صفحه ايستا[[65]](http://192.168.0.110/editor/main.htm" \l "_ftn65" \o ") به كمك واژه‌پرداز Wordاز مجموعه مايكروسافت آفيس 2003،[[66]](http://192.168.0.110/editor/main.htm" \l "_ftn66" \o ") طراحي و از24 خرداد تا 8 شهريور1385[[67]](http://192.168.0.110/editor/main.htm" \l "_ftn67" \o ") در دامنه فرعي «ابرداده» از سايت «پژوهشگاه اطلاعات و مدارك علمي ايران»[[68]](http://192.168.0.110/editor/main.htm" \l "_ftn68" \o ") منتشر شد و در معرض نمايه سازي سه موتور كاوش گوگل[[69]](http://192.168.0.110/editor/main.htm" \l "_ftn69" \o ")، ياهو[[70]](http://192.168.0.110/editor/main.htm" \l "_ftn70" \o ")، ام اس ان[[71]](http://192.168.0.110/editor/main.htm" \l "_ftn71" \o ") – سه موتور كاوش پراستفاده به استناد «الكسا»،[[72]](http://192.168.0.110/editor/main.htm" \l "_ftn72" \o ") «سرچ اينجين واچ»،[[73]](http://192.168.0.110/editor/main.htm" \l "_ftn73" \o ")«سرچ اينجين گايد»[[74]](http://192.168.0.110/editor/main.htm" \l "_ftn74" \o ") و سايتهاي مشابه[[75]](http://192.168.0.110/editor/main.htm" \l "_ftn75" \o ") ـ قرار گرفت. در دهمين روز انتشار صفحات، موتور كاوش گوگل، تمامي صفحات منتشر شده را نمايه‌سازي نمود؛ اين در حالي است كه ياهو روندي كندتر داشت و پس از گذشت بازه زماني مشخص شده، 76 صفحه مربوط به مقالات نشريه نمايه شد. اما از ميان سه موتور كاوش انتخابي، «ام.اس.ان» به دليل عدم نمايه سازي تعداد قابل قبولي از صفحات در مدت زمان تعيين شده، از مجموعه پژوهش حذف شد و ادامة پژوهش با تمركز بر صفحات نمايه شده، در دو موتور كاوش ياهو و گوگل انجام پذيرفت.

**3. گردآوري داده‌ها**

به منظور تعيين ميزان اثربخشي عناصر ابرداده‌اي در اين مرحله از پژوهش، كاوش كليدواژه اي در كادر محاوره‌اي ساده[[76]](http://192.168.0.110/editor/main.htm" \l "_ftn76" \o ") دو موتور كاوش گوگل و ياهو انجام پذيرفت. از آنجا كه دو گروه گواه و آزمون، در وب سايت طراحي شده گنجانده شده است و نيازي به تعيين رتبة صفحات در ميان ساير صفحات موجود در وب نيست، كاوش كليدواژه اي در دامنه وب سايت طراحي شده محدود گرديد. فرمول كاوش در موتورهاي كاوش گوگل و ياهو به ترتيب، چنين است:

Site:http://metadata.irandoc.ac.ir “Keywords”

Site:metadata.irandoc.ac.ir “Keywords”

يا Domain:metadata.irandoc.ac.ir “Keywords”

براي تعيين ميزان اثربخشي هر يك از عناصر ابرداده‌اي (پرسش نخست پژوهش) دو مجراي آزمون وجود دارد:

·    كاوش كليدواژه‌هاي منحصر به فرد كه تنها به عنصر ابرداده‌اي ويژه‌اي اختصاص يافته است و مقايسة رتبة صفحه حاوي آن كليدواژه با صفحة گروه گواه مورد نظر.

·    كاوش كليدواژه‌هاي تكرار شونده؛ چنانكه بيان شد، به ترتيب افزايش شماره‌هاي نشريه، يك عنصر ابرداده‌اي به عناصر قبلي افزوده مي گردد و تنها تفاوت ميان دو مقاله مشابه در دو نشرية با شماره هاي متوالي، حضور كليدواژه‌هاي تكرار شونده و منحصر به فردي است كه در عنصر ابرداده‌اي بعدي افزوده شده است. لذا هر يك از شماره‌هاي نشريه در حالي كه حاوي مقالات گروه آزمون نشريه شماره قبل خود است، حاوي مقالات گروه گواه نشريه شماره بعد نيز هست.  بدين ترتيب، به استثناي عناصر عنواني (عنوان اچ.تي.ام.ال و عنوان دابلين‌كور) دو دسته آزمون براي تعيين اثربخشي حضور هر يك از عناصر ابرداده‌اي (پرسش نخست) قابل انجام شدن است.

رتبه حاصل از هر يك از كاوشهاي انجام شده در موتورهاي كاوش، به تفكيك موتور كاوش و گروه گواه و آزمون با توجه به سؤالهاي پژوهش در نرم‌افزار آماري SPSS[[77]](http://192.168.0.110/editor/main.htm" \l "_ftn77" \o ") وارد شد و مورد تجزيه و تحليل قرار گرفت.

با توجه به اين كه نتايج حاصل از كاوش در موتورهاي كاوش از نوع داده‌هاي رتبه‌اي است (ترنر و برك بيل، 1998؛صفري[[78]](http://192.168.0.110/editor/main.htm" \l "_ftn78" \o ")، 2005)، ميزان معنادار بودن تفاوت دو گروه گواه و آزمون (دو گروه مستقل)در بازيابي صفحات وب، با استفاده از آزمون غيرپارامتري يومن‌ وايتني[[79]](http://192.168.0.110/editor/main.htm" \l "_ftn79" \o ") تعيين شد.

در مجموع، تحت شرايط زماني و مكاني يكسان، 363 كاوش در موتور كاوش گوگل و 259 كاوش در موتور كاوش ياهو انجام شد و رتبه‌هاي كسب شده در دو گروه صفحات آزمون و گواه ثبت گرديد. تجزيه و تحليل آماري نيز بر مبناي داده‌هاي جمع‌آوري شده، با نرم‌افزار آماري «اس.پي.اس.اس.» انجام پذيرفت.

**تجزيه و تحليل يافته‌ها**

**1. پاسخگويي به نخستين پرسش پژوهش:**ميزان اثربخشي هر يك از عناصر ابرداده‌اي بر رتبه‌بندي به وسيله دو موتور كاوش ياهو و گوگل.

براي درك بهتر و امكان مقايسه و تحليل نتايج، جدول3 كه نمايي كلي از مقدار پي محاسبه شده توسط نرم‌افزار آماري «اس.پي.اس.اس» در آزمون «يو من وايتني» است، ارائه مي‌شود. وجود تفاوت معنادار در هر عنصر، با قلم[[80]](http://192.168.0.110/editor/main.htm" \l "_ftn80" \o ") درشت تر و توپر نشان داده شده است. تفاوت ميان ميانگين رتبه‌هاي كسب شده در هر گروه نشانگر آن است كه رتبه به نفع كدام گروه تغيير وضعيت داده است؛ لذا در هر مورد از عناصر، به تفكيك موتور كاوش، گروه صفحاتي كه بهبود رتبه داشته (آزمون يا گواه) با اصطلاح «به نفع» مشخص شده است.

**جدول3. نتايج آزمون «يو من وايتني» در پاسخ به پرسش نخست**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ياهو** | | **گوگل** | | **نوع كليدواژه** | **عنصر مورد آزمون** |
| **به نفع** | **Asymp. Sig. (2-tailed)** | **به نفع** | **Asymp. Sig. (2-tailed)** |
| -- | -- | -- | -- | كليدواژة منحصر به فرد | عنوان اچ.تي.ام.ال |
| آزمون | **.001** | آزمون | **.000** | كليدواژة تكرار شونده |
| آزمون | **.017** | آزمون | .462 | كليدواژة منحصر به فرد | كليدواژه‌هاي اچ.تي.ام.ال |
| آزمون | .172 | آزمون | .472 | كليدواژة تكرار شونده |
| آزمون | **.000** | آزمون | **.015** | كليدواژة منحصر به فرد | توصيف اچ.تي.ام.ال |
| آزمون | **.000** | گواه | **.016** | كليدواژة تكرار شونده |
| -- | -- | -- | -- | كليدواژة منحصر به فرد | عنوان دابلين كور |
| آزمون | .807 | آزمون | .826 | كليدواژة تكرار شونده |
| آزمون | **.023** | آزمون | .110 | كليدواژة منحصر به فرد | موضوع دابلين كور |
| گواه | **.050** | گواه | **.038** | كليدواژة تكرار شونده |

يافته‌هاي حاصل از بررسي اثربخشي دو ابرنشانة كليدواژه‌ها و توصيف در اين پژوهش، با يافته‌هاي «ترنر» و «برك بيل» (1998) مبني بر اثربخشي ابرنشانة كليدواژه‌ها در مقايسه با ابرنشانة توصيف، تفاوت دارد. يافته‌ها در پژوهش حاضر نشان داد كه ابرنشانه توصيف در مقايسه با ابرنشانه كليدواژه‌ها اثر بيشتري در بهبود رتبة صفحات دارد.

يافته‌هاي «صفري» (2005) مبني بر عدم تأثير عنصر عنوان دابلين كور بر رتبة صفحات، در اين پژوهش نيز تأييد مي‌شود. يافته‌هاي اين پژوهش در راستاي نتايج زانگ و ديميتروف (2004) تأثير عنصر موضوع دابلين كور را در بهبود رتبة صفحات نشان مي‌دهد. اين يافته با يافته‌هاي صفري مبني بر عدم اثربخشي عنصر موضوع دابلين كور، متفاوت است.

**2. پاسخگويي به پرسش دوم:** ميزان اثربخشي دو تا پنج بار تكرار كليدواژه در بخش سرآيند صفحات وب بر رتبه‌بندي صفحات وب در دو موتور كاوش ياهو و گوگل

جدول شماره4، نمايي كلي از مقدار پي محاسبه شده توسط نرم‌افزار آماري «اس.پي.اس.اس» در آزمون «يومن‌وايتني» ارائه مي‌كند. در پاسخ به پرسش دوم پژوهش، تأثير حضور كليدواژه‌هاي تكرار شونده در دو تا پنج عنصر تحت بررسي، تجزيه و تحليل شد.

**جدول4. نتايج آزمون «يو من وايتني» در پاسخ به پرسش دوم**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **تعداد تكرار مورد بررسي** | **گوگل** | | **ياهو** | |
| **Asymp.Sig.(2-tailed)** | **به نفع** | **Asymp.Sig.(2-tailed)** | **به نفع** |
| 2 بار تكرار | 580/0 | آزمون | 172/0 | آزمون |
| 3 بار تكرار | 008/. | گواه | 016/. | آزمون |
| 4 بار تكرار | 062/. | گواه | 019/. | آزمون |
| 5 بار تكرار | 000/. | گواه | 175/0 | آزمون |

عملكرد دو موتور كاوش گوگل و ياهو در مقابل عناصر ابرداده‌اي در اين پژوهش بررسي شد. به طور خلاصه، عملكرد دو موتور كاوش مورد بررسي، چنين است:

·      موتورهاي كاوش نسبت به حضور عناصر ابرداده‌اي مورد بررسي، بي‌اعتنا نيستند.

·      عنصر عنوان «اچ.تي.ام.ال» در مقايسه با ساير عناصر مورد بررسي، اثرگذارترين عنصر در هر دو موتور كاوش مورد بررسي است.

·      از ميان ابرنشانه‌هاي «اچ.تي.ام.ال»، ابرنشانة توصيف، تأثير بيشتري در مقايسه با ابرنشانه كليدواژه‌ها دارد.

·      توجه به عناصر ابرداده‌اي زبان «اچ.تي.ام.ال» بيش از عناصر ابرداده‌اي قالب دابلين كور است.

·      توجه موتورهاي كاوش مورد بررسي به عنصر موضوع قالب ابرداده‌اي دابلين كور، بيش از عنصر عنوان آن است.

·      دو موتور كاوش گوگل و ياهو، عملكرد متفاوتي نسبت به كليدواژه‌هاي تكرار شونده و منحصر به فرد دارند.

·      در هر دو موتور كاوش، كليدواژه‌هاي منحصر به فرد در مقايسه با كليدواژه‌هاي تكرار شونده، تأثير بيشتري دارد.

·    موتور كاوش گوگل بيش از ياهو نسبت به كليدواژه‌هاي تكراري در بخش سرآيند صفحات وب، حساس است و به نظر مي‌رسد كه اين تكرارها را نوعي تقلب به شمار مي‌آورد.

**نتيجه‌گيري**

با وجود تمامي پيشرفتهاي حاصل شده در فناوري موتورهاي كاوش، هنوز روش اصلي نمايه سازي در موتورهاي كاوش، محتوا محور است و بر مبناي كليدواژه‌ها انجام مي‌پذيرد (يو و ديگران[[81]](http://192.168.0.110/editor/main.htm" \l "_ftn81" \o ")، 2002) .نمايه سازي تمام متن، كاوش كليدواژه اي بدون توجه به بافت متن و كاستيهاي موجود در رتبه‌بنديها، به جامعيت بالا و مانعيت پايين منجر مي‌شود (لي ـ اسملتزر، 2000، ص206) . تحقيقات فراواني در زمينة مشكلات نمايه‌سازي خودكار در موتورهاي كاوش انجام گرفته است (چانگ و لي[[82]](http://192.168.0.110/editor/main.htm" \l "_ftn82" \o ")، 1998، ص149). آنچه پژوهشگران در باب آن به اتفاق نظر رسيده‌اند آن است كه نمايه‌‌سازي غيرمتمركز و توزيع شده، به گونه‌اي كه در موتورهاي كاوش كنوني در جريان است، معماري قابل تأييدي در بازيابي اطلاعات وب نيست ( لي ـ اسملتزر، 2000، ص206) .

چنانچه محتواي صفحه به درستي در فيلدهاي ابرداده‌اي بازنموده شود و موتورهاي كاوش از اين فيلدهاي ابرداده‌اي در رتبه‌بندي و بازيابي نتايج بهره ببرند، مانعيت افزايش و نتايج بازيابي به سطحي قابل مديريت كاهش مي يابد. به علاوه، رتبه‌بندي صفحات صحت بيشتري خواهد داشت (زانگ و جاستريم، 2005، ص1) و آن بخش از ويژگيها كه شناسايي آنها به درك انساني نياز دارد، در فرايند نمايه‌سازي منظور خواهد شد؛ اما همچنان پاسخ به اين سؤال كه آيا جامعة بهره‌گير، به اهميت وجود ابرداده پي برده است يا خير، آن را تأييد مي‌كند و يا اينكه به درستي از آن بهره مي‌برد يا نه، يك معماست ( زانگ و جاستريم ، 2005، ص22).

از نظر «تيم برنرزلي»[[83]](http://192.168.0.110/editor/main.htm" \l "_ftn83" \o ") و ساير پيشگامان وب معنايي، وب كنوني، اطلاعاتي انسان فهم دارد. صفحات وب به زبان اچ.تي.ام.ال نگاشته مي‌شود و مرورگرها بر نشانه‌هايي تمركز دارند كه نه بر محتواي معنايي، بلكه بر ساختار متمركزند. چنانچه وب معنايي تحقق يابد، صفحات وب آينده همانگونه كه انسان فهم هستند، ماشين فهم[[84]](http://192.168.0.110/editor/main.htm" \l "_ftn84" \o ") نيز خواهند بود (كمپل[[85]](http://192.168.0.110/editor/main.htm" \l "_ftn85" \o ")، 2004، ص 197)،در اين شيوه اطلاعات از طريق نشانه‌هاي معنادار ارائه مي‌شود و ابرداده‌هاي لازم را «آر.دي.اف» به صفحات وب مي‌افزايد. پديدآورندگان وب فردا، امكان جستجوپذير كردن سايتهاي خود را با مانعيت بيشتري عملي خواهند ساخت و كارگزاران هوشمند[[86]](http://192.168.0.110/editor/main.htm" \l "_ftn86" \o ") آينده، بسيار قدرتمندتر و متكامل‌تر از موتورهاي كاوش امروزي عمل خواهند كرد (كمپل، 2004، ص199).

**منابع**

-Asadi, S., & Jamali M., H.R. (2004). "[Shifts in search engine development: A review of past, present and future trends in research on search engines](http://www.webology.ir/2004/v1n2/a6.html)". Webology, **1**(2). Retrieved Feb, 18, 2005, From http://www.webology.ir/2004/v1n2/a6.html

-Bar-Ilan, J. (2005). “ [Comparing Rankings of Search Results on the Web](http://www.sciencedirect.com/)” . Information Processing & Management , 41. Retrieved Oct, 12, 2005, From Elsevier Database.

-Campbell, D.G. (2004). “The Metadata – Bibliographic Organization nexus” In Metadata Applications and Management. London: Facet Publishing, 185-203.

-Chung, S.M. & Lee, J.Y. (1998). “Information Discovery on the Internet” . In Encyclopedia of Library and Information Science . vol. 62, (supple. 25) pp.146-157.

-Drott, M.C. (2002). "Indexing aids at Corporate websites: the use of robot.txt and meta tags". Information Processing and Management, 38(2), 209-219. Retrieved June 10, 2006, From Elsevier database.

-Fadaie, G. (2005). " User Satisfaction through better indexing". Cataloging & Classification Quarterly, 40(2), 5-17.

-Fishkin, R. (2005). "[Search Engine Ranking Factors](http://www.seomoz.org/articles/search-ranking-factors.php#1)" Retrieved Des, 17, 2005, From [http://www.seomoz.org/articles/search -ranking-factors.php#1](http://www.seomoz.org/articles/search-ranking-factors.php#1)

-Global Top 500 (2006) Retrieved Jun, 11, 2006, From <http://www.alexa.com/site/ds/top_500>

-Google (2004). "[Information for webmasters](http://www.google.com/webmasters/4.html)". Retrieved Oct, 11, 2005, From http://www.google.com/webmasters/4.html

-Gotlieb, C. ; Eliopoulos, D. (2003). “ [Evaluating Web Search Results Ranking](http://web9.epnet.com/)” . Online, 27(2). Retrieved Oct, 11, 2005, From Ebsco Computer Science Database.

-Henshaw, R. & Valauskas, E.J. (2001). "

| What is this?  DAP Link VerifierDAP Link Verifier | |
| --- | --- |
| Name: | **1999-3pp125-131.pdf** |
| Size: | 46.8 KB |
| URL: | http://www.librijournal.org/pdf/1999-3pp125-131.pdf |
| Status | Valid |
| [More ▼](javascript:%20void(0);)  [*Download* ***with DAP***](javascript:%20void(0);) | |

[Settings](http://127.0.0.1:10029/settings/)    [Report this link](http://127.0.0.1:10029/report/)  
 [Disable for this session only](javascript:%20void();)

[Metadata as a catalyst: experiments with metadata and search engines in the internet journal, First Monday](http://www.librijournal.org/pdf/1999-3pp125-131.pdf)" . Libri, 51(2), 86-101pp. Retrieved Oct, 23, 2005, From www.librijournal.org/pdf/1999-3pp125-131.pdf

-Jansen, B.J. ; Spink, A.; and Saracevic, T. (2000). " [Real life, real users, and real needs: a study and analysis of user queries on the web](http://www.sciencedirect.com/)" . Information Processing & Management, 36(2), 207-227. Retrieve May, 8, 2005, From Elsevier Database.

-Lee-Smeltzer, K.H. (2000). “ [Finding the needle: Controlled Vocabularies, Resource Discovery, and Dublin core](http://www.sciencedirect.com/)”. Library collections, Acquisitions, & Technical Services, 24 , 205-215. Retrieve May, 8, 2005, From Elsevier Database.

-Lewis , E. (2005). [Top Ten Search Engines - Top 10 SEs](http://www.seoconsultants.com/search-engines/) . Retrieved Apr, 22, 2006, From [http://www.seoconsultants.com](http://www.seoconsultants.com/) /search-engines/

-[List of Top Search Engines](http://capmex.biz/resources/top-search-engines) (2006). Retrieved Apr, 22, 2006, From http://capmex.biz/resources/top-search-engines

-Noruzi, A. (2005). “[Editorial](http://www.webology.ir/2005/v2n4/editorial6.html)” .Webology, 2(4). Retrieved Apr, 20, 2006, From http://www.webology.ir/2005/v2n4/editorial6.html

-Quevedo\_Torrero, J.U. (2004). “ [Improving Web Retrieval by Mining the HTML tags for Keywords and Exploring the Hyperlink Structures of Web Pages](http://wwwlib.umi.com/dissertations/fullcit/3156028)” [Abstract] doctoral Dissertation. University of Houston. Retrieved Oct, 10, 2005, From http://wwwlib.umi.com/dissertations/fullcit/3156028

-Safari, M. (2005). "[Search Engines and Resource Discovery on the web: Is Dublin Core an Impact Factor?](http://www.webology.ir/2005/v2n2/a13.html)". Webology,2(2).Retrieved Apr, 20, 2006, From <http://www.webology.ir/2005/> v2n2/a13.html

-Sullivan, D. (2002). “[How to Use HTML Meta tags](http://searchenginewatch.com/webmasters/atricle.php/2167931‎)”. Retrieved Apr,20, 2006, From <http://searchenginewatch.com/webmasters>/atricle .php/2167931

-Thelwall, M.; Vaughan, L. (2004)."

| What is this?  DAP Link VerifierDAP Link Verifier | |
| --- | --- |
| Name: | **2004\_new\_pagerank\_preprint.pdf** |
| Size: |  |
| URL: | http://www.scit.wlv.ac.uk/~cm1993/papers/2004\_new\_pagerank\_preprint.pdf |
| Status | File does not exist (0) - General Error |
| [More ▼](javascript:%20void(0);)  [*Download* ***with DAP***](javascript:%20void(0);) | |

[Settings](http://127.0.0.1:10029/settings/)    [Report this link](http://127.0.0.1:10029/report/)  
 [Disable for this session only](javascript:%20void();)

[New Versions of PageRank employing alternative web document models](http://www.scit.wlv.ac.uk/~cm1993/papers/2004_new_pagerank_preprint.pdf)" . ASLIB Proccedings, 56(1), 24-33. Retrieved Feb, 19, 2005, From [www.scit.wlv.ac.uk/~](http://www.scit.wlv.ac.uk/~) cm1993/papers/2004\_new\_pagerank\_preprint.pdf

-Turner, T.P.; Brackbill, L. (1998). “Rising to the Top: Evaluating the use of html meta tag to improve retrieval of world wide web documents through internet search engines” . Library Resources and Technical Services, 24(4), 258-271.

-Yaltaghian , B. (2002). “[Improving the ranking of Search Engine Output: A Network Analysis Approach](http://www.cosc.canterbury.ac.nz/open/seminars/old/2002.html)”. [Abstract] doctoral Dissertation. University of Toronto. Retrieved Feb, 21, 2006, From www.cosc.canterbury.ac.nz/open/seminars/old/2002.html

-Yu et al (2002). “[Patterns in Unstructured Data: Discovery, Aggregation, and Visualization](http://javelina.cet.middlebury.edu/lsa/out/cover_page.htm‎)”. Retrieved Jen, 21, 2005, From http://javelina.cet.middlebury.edu/lsa/out/cover\_page.htm

-Zhang, Z .; Dimitroff, A. (2005b). “ [The Impact of Metadata Implementation on Webpage Visibility in Search Engine Results (Part II)](http://www.sciencedirect.com/)” . Information Processing & Management, 41 (3), 691-715. Retrieved Feb, 18, 2005, From Elsevier Database.

-Zhang, Z. ; Dimitroff, A .(2004). “[Internet Search engines response to metadata Dublin core implementation](http://www.sciencedirect.com/)” . Journal of Information Science , 30(4), 310-320. Retrieved Oct, 11, 2005, From Elsevier Database.

-Zhang, Z. ; Dimitroff, A. (2005a). “ [The Impact of Webpage Content Characteristics on Webpage Visibility in Search Engine Results (Part I)](http://www.sciencedirect.com/)” . Information Processing & Management, 41 (3), 665-690. Retrieved Feb, 18, 2005, From Elsevier Database.

-Zhang, Z. ; Jastram, I. (2005). " [A Study of the metadata creation behavior of different user groups on the Internet](http://www.sciencedirect.com/)" Information Processing and management. Retrieved Oct, 11, 2005, From Elsevier Database.

1. دانشجوي دوره دكتراي كتابداري و اطلاع‌رساني دانشگاه فردوسي مشهد.

Atefehsharif@gmail.com

1. Unstructured.

2. Yaltaghian.

3. Zhang & Jastram.

4. Gotlieb & Eliopoulos.

5. Asadi & Jamali M.

6. Internet search tools.

7. Search engines.

8. Meta search engines.

9. Directories.

10. Search Utilities.

11. Spink et al.

12. Thelwall & Vaughan.

13. Bar\_Ilan.

14. Zhang & Dimitroff.

15. Haltley.

1. Drott.

2. Jansen ; Spink.; and Saracevic.

3. Fadaie Araghi.

4. Ranking.

5. Title.

6. Keyword.

7. Description.

8. Richardson.

9. Keyword Density analyzer.

10. word tracker.

11. web poisons gold.

12. Search Engine Optimization (SEO).

1. Classified.

2. Trade Secrets.

3. Truner & Brackbill.

4. Safari.

5. Mohamed.

6. Henshaw & Valauskas.

7. Semantic.

8. Content.

9. Syntax.

10. Structure.

11. Hyper- Links.

12. Markup Languages.

13. Quevedo-Torrero.

1. google.

2. Fishkin.

3. Metadata Structure.

4. Webpage Content.

5. Hyperlink Cited status.

6. Metadata elements.

7. Technical Relevance.

8.

9.

10.

11.

12.

13. Salivan

14. Noruzi

1. Lee-Smeltzer.

2. Valqui & Freire.

3. Header.

1. Semantic web.

2. Ontology.

3. Webometrics.

4. Directory of Open Access Journals , Available at: [www.doaj.org](http://www.doaj.org/).

5. عنصر اول: عنوان اچ.تي.ام.ال

1. Source code.

1. Static.

2. Microsoft Office Word 2003.

3. June 14, 2006 till Aug 30, 2006.

1. [http://metadata.irandoc.ac.ir](http://metadata.irandoc.ac.ir/)

2. www.google.com

3. www.yahoo.com

4. www.msn.com

5. Jun, 11, 2006 From: [http://www.alexa.com/site/ds/top-500](http://www.alexa.com/site/ds/top_500)

6. [www.searchenginewatch.com](http://www.searchenginewatch.com/)

7. [www.searchengineguid.com](http://www.searchengineguid.com/)

8. <http://www.seoconsultants.com/search-engines/>

9. Simple Search Box.

1. Statistical Package for Social Science.

2. Safari.

3. Mann-Whitney U.

1. Font.

1. Yu et al.

2. Chung& Lee.

1. Tim Burners Lee.

2. Machine understandable.

3. Campbell.

4. Intelligent Agents