

تأثیر زبان نمایه‌سازی بر رتبه‌بندی تصاویر در موتور کاوش گوگل

نجلا حریری^۱، شیده احمدی^{۲*}

تاریخ دریافت: ۹۳/۳/۲۵ تاریخ پذیرش: ۹۳/۷/۷

چکیده

هدف: پژوهش حاضر، تعیین تأثیر زبان نمایه‌سازی بر رتبه‌بندی تصاویر در موتور کاوش گوگل است. به‌منظور نیل به این هدف، چگونگی رتبه‌بندی نتایج جست‌وجو در نمایه‌سازی تصاویر با استفاده از زبان آزاد و زبان کنترل‌شده در موتور کاوش گوگل بررسی شده است.

روش: پژوهش به روش تجربی انجام شده است. جامعه پژوهش، ۳۰ تصویر از تمبرهای قدیمی ایران مربوط به سال‌های ۱۳۳۱ تا ۱۳۵۷ بود. برای اجرای این پژوهش، دو سایت با استفاده از دو حساب کاربری در محیط گوگل ایجاد شد و این تصاویر در هر دو سایت بارگذاری شدند، با این تفاوت که عناوین تمبرها در سایت ۱ بر اساس زبان کنترل‌شده و در سایت دو بر اساس زبان آزاد درج شدند. نمایه‌سازی بر اساس زبان کنترل‌شده در این پژوهش، با همکاری و مشورت ویراستاران اصطلاح‌نامه مواد گرافیکی کتابخانه کنگره آمریکا انجام شد. پس از گذشت یک بازه زمانی ۴۰ روزه، تصاویر در بخش جست‌وجوی تصاویر موتور کاوش گوگل جست‌وجو شدند و رتبه تصاویر بازیابی‌شده در هر دو گروه بررسی شد.

یافته‌ها: در تجزیه و تحلیل توصیفی، امتیاز رتبه‌های تصاویر بازیابی‌شده در دو گروه محاسبه و در تجزیه و تحلیل استنباطی برای مقایسه رتبه‌بندی نتایج جست‌وجوی تصاویر از آزمون t مستقل استفاده شد. نتایج حاصل از این پژوهش نشان می‌دهد که استفاده از زبان نمایه‌سازی آزاد در مقایسه با زبان کنترل‌شده در بارگذاری تصاویر در اینترنت، تأثیر مثبتی در رتبه‌بندی نتایج جست‌وجوی این تصاویر به‌وسیله موتور کاوش گوگل دارد.

واژه‌های کلیدی: بازیابی تصاویر، رتبه‌بندی تصاویر، زبان نمایه‌سازی آزاد، زبان نمایه‌سازی کنترل‌شده، موتور کاوش گوگل، نمایه‌سازی تصاویر.

nadjlahariri@hotmail.com

۱. دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، ایران

shideh_ahmadi_f@yahoo.com

۲. دانشجوی کارشناسی ارشد علم اطلاعات و دانش‌شناسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، ایران

مقدمه

در دورانی که عصر اطلاعات لقب گرفته، انتشار گسترده منابع اطلاعاتی و افزایش روزافزون رسانه‌های اطلاعاتی دسترسی به اطلاعات مناسب در زمان مناسب را به معضلی تبدیل کرده است؛ به طوری که کمبود اطلاعات و دانش مطرح نیست، بلکه مشکل اساسی دسترسی به انبوه اطلاعات است (مطلبی، ۱۳۸۶: ۴). امروزه مردم حجم انبوهی از داده‌های الکترونیکی را در اختیار دارند که بخش عظیمی از آنها را تصاویر تشکیل می‌دهند. ذخیره این تصاویر به تولید پایگاه‌های تصویری حجیمی منجر شده که عمل جست‌وجو توسط کاربر را مشکل کرده است. همچنین در سال‌های اخیر با افزایش روزافزون دوربین‌های دیجیتالی، مجموعه‌های بزرگی از تصاویر مختلف توسط مردم تهیه شده‌اند و طبیعی است که پیدا کردن تصویر مشخصی از میان این تعداد تصاویر گوناگون کار بس دشوار و حتی غیرممکنی است (نوروزی و ولایتی، ۱۳۸۹: ۱۶۳).

نمایه‌سازی تصاویر یکی از مهم‌ترین روش‌های پیشنهادی برای مواجهه با تصاویر است، اما ویژگی‌ها و خصوصیات منحصر به فرد تصاویر، در مقایسه با سایر منابع اطلاعاتی، عمل نمایه‌سازی تصاویر را با مشکلاتی مواجه کرده است (عباس‌پور، ۱۳۸۴: ۱۶۸). همچنین رسیدن به توافق بر روی این موضوع دشوار خواهد بود که کدام‌یک از روش‌های نمایه‌سازی تصویر مناسب‌ترند؛ زیرا افراد نمی‌توانند بر روی آنچه یک تصویر واقعاً نشان می‌دهد موافقت داشته باشند، حتی برخی افراد مثل شات‌فورد^۱ بین آنچه در تصویر وجود دارد و چیزی که از آن صحبت می‌کند تمایز قائل شده‌اند، زیرا آنچه در تصویر وجود دارد کم و بیش با مسائل عینی روبه‌روست، حال آنکه آنچه یک تصویر از آن صحبت می‌کند بیشتر با امور انتزاعی سروکار دارد (کرمی، ۱۳۸۵).

به منظور دسترسی و استفاده از تصاویر موجود در پایگاه‌های تصاویر و اینترنت^۲، کاربران نیازمند ابزار مطمئنی هستند و به علت حجم زیاد اطلاعات، جست‌وجوی تصاویر در وب، شبیه گشتن به دنبال یک سوزن در انباری از کاه است. ابزاری که شاید کاربران را کمک کند تا تصاویر مورد نظر خود را بازیابی و جایابی کنند، موتور کاوش تصویر^۳ است. یکی از قدرتمندترین موتورهای کاوش تصاویر، موتور کاوش گوگل^۴ است که پایگاه قدرتمند آن با داشتن ۲۵۰ میلیون تصویر مشاهده‌پذیر در فهرست خود، در حقیقت یکی از جامع‌ترین مجموعه‌ها در کل وب محسوب می‌شود. این مجموعه فوق‌العاده غنی و روزآمد برای بسیاری از کاربران در زمینه‌های مختلف مفید و ارزنده است. با توجه به نیاز روزافزون کاربران به بازیابی تصاویر در زمینه‌های گوناگون، کاربران بی‌شماری از این موتور کاوش استفاده می‌کنند، اما به دلیل حجم زیاد منابع تصویری، استفاده کنندگان برای یافتن تصویر، علاوه بر موتورهای جست‌وجو، به یک نمایه جامع

از تصاویر اینترنت نیاز دارند که بدون آن بازیابی تصویر خاصی از اینترنت شاید بسیار دشوار و زمان‌بر باشد (حسن و زنگ، ۲۰۰۱). حجم روزافزون تصاویر در پایگاه‌های اطلاعاتی و همچنین نیاز به نظام‌های بازیابی تصاویر در حوزه‌های مختلف سبب شده است تا پژوهش‌های وسیعی در زمینه ذخیره و بازیابی تصاویر انجام شود. راه‌حل‌های پیشنهادی برای مشکلات بازیابی تصاویر متنوع است و هر کدام از این روش‌ها مزایا و مشکلاتی دارند. یکی از این راه‌حل‌ها استفاده از انواع زبان‌های نمایه‌سازی^۵ است. از جمله زبان‌های نمایه‌سازی می‌توان به زبان آزاد و زبان کنترل‌شده اشاره کرد. "در نظام نمایه‌سازی به زبان آزاد، نمایه‌ساز از توصیفگرهایی که خود مناسب می‌داند برای توصیف مفاهیم موجود در مدرک استفاده می‌کند، خواه این توصیفگرها در متن آمده باشند یا نه" (نوروزی و ولایتی، ۱۳۸۹: ۴۵). نمایه‌سازی به زبان کنترل‌شده بر یک سری اصطلاح‌های اصطلاح‌نامه‌ای^۶، به نام توصیفگر مبتنی است که قبلاً توسط نمایه‌سازان و متخصصان موضوعی و در قالب اصطلاح‌نامه ارائه شده‌اند (سینق، ۱۳۸۸: ۱۱۷) در این زبان، نمایه‌ساز ابتدا واژه‌ها را انتخاب می‌کند و سپس مطابق با توصیفگرهای یک اصطلاح‌نامه معتبر، اصطلاح‌های مناسب را به مدرک تخصیص می‌دهد. سؤال این است که آیا استفاده از زبان‌های نمایه‌سازی متفاوت موجب تفاوت در بازیابی تصاویر از وب می‌شود.

با توجه به اهمیت موتور جست‌وجوی گوگل در بازیابی تصاویر و از آنجا که تاکنون پژوهشی در مورد تأثیر زبان نمایه‌سازی در بازیابی تصاویر از گوگل انجام نشده، این پژوهش بر آن است تا در جست‌وجوی موتور کاوش گوگل، به روش تجربی، تأثیر استفاده از زبان آزاد و زبان کنترل‌شده را بر رتبه‌بندی نتایج جست‌وجوی دو گروه تصاویر (که هر گروه به یکی از دو زبان فوق نمایه‌سازی شده‌اند و در محیط وب قرار گرفته‌اند) بسنجد؛ تا از طریق نتایج حاصل بتوان به مناسب‌ترین روش نمایه‌سازی برای سازماندهی و یافتن تصاویر مورد نظر در اینترنت به وسیله موتور کاوش گوگل دست یافت.

پیشینه پژوهش

در زمینه نمایه‌سازی تصاویر در حوزه بازیابی اطلاعات پژوهش‌های اندکی انجام شده است که از آن جمله می‌توان به موارد ذیل اشاره کرد:

خیان و دیگران^۷ (۲۰۰۰) در پژوهشی با عنوان "جست‌وجوی زبان طبیعی در مقابل جست‌وجوی واژگان مهارشده" معتقدند که کارشناسان اطلاعات باید به‌طور کامل و کارآمد، از مزایای زبان طبیعی و واژگان در طراحی سیستم بازیابی استفاده کنند و ترکیبی از زبان طبیعی و فنون جست‌وجوی واژگان مهارشده به شدت به کاربران اطلاعات توصیه می‌شود.

تام^۹ و لئونگک^{۱۰} (۲۰۰۱) در پژوهشی با عنوان "توصیف‌های زبان طبیعی ساخت یافته برای بازیابی محتوای مواد بصری" معتقدند جست‌وجوی کلیدواژه‌ای از مجموعه‌های چندرسانه‌ای، فاقد دقت، و تجزیه خودکار حاشیه‌نویسی زبان طبیعی بدون محدودیت، فاقد دقت است. آنها پیشنهاد می‌کنند که ساختاری برای توصیف زبان طبیعی از محتوای معنایی مواد بصری نیاز خواهد بود.

هارپینگک^{۱۱} (۲۰۰۲) در پژوهشی با عنوان "زبان تصاویر: دسترسی بیشتر به تصاویر طرحواره با استفاده از ابر داده و ساختار واژگان" دشواری‌های را آشکار می‌کند که کاربران نهایی برای بازیابی تصاویر در شبکه جهانی وب با آن مواجه هستند. در این تحقیق به محتوای نمایه‌سازی شده با جست‌وجو در استفاده از طرحواره ابر داده و واژگان کنترل شده برای توصیف، فهرست‌نویسی، و نمایه آثار شاخص هنری پرداخته می‌شود.

ماتون^{۱۲} (۲۰۰۳) در پژوهش خود با عنوان "نمایه‌سازی به زبان طبیعی یا کنترل واژگان: آیا ما باید روشی برای جست‌وجوی کاربران نمایه‌سازی نماییم؟" درباره استفاده از زبان طبیعی در نمایه‌سازی تصاویر و سایر منابع تصویری بحث کرده است. نتایج این پژوهش بر استفاده مفید ترکیب زبان نمایه‌سازی تصاویر و کنترل واژگان سنتی توسط کاربران تأکید دارد.

سؤالات پژوهش

۱. در نمایه‌سازی تصاویر با استفاده از زبان آزاد، رتبه‌بندی نتایج جست‌وجو در موتور کاوش گوگل چگونه است؟
۲. در نمایه‌سازی تصاویر با استفاده از زبان کنترل شده، رتبه‌بندی نتایج جست‌وجو در موتور کاوش گوگل چگونه است؟
۳. وضعیت رتبه‌بندی نتایج جست‌وجو در نمایه‌سازی تصاویر با استفاده از زبان آزاد و زبان کنترل شده، در مقایسه با یکدیگر چگونه است؟

فرضیه پژوهش

بین رتبه‌بندی نتایج جست‌وجو به وسیله موتور کاوش گوگل در جست‌وجوی تصاویر نمایه‌شده به زبان آزاد و زبان کنترل شده، تفاوت معناداری وجود دارد.

روش پژوهش

این پژوهش با استفاده از روش تجربی انجام شده است. نمونه مورد پژوهش در این تحقیق، ۳۰ تصویر مربوط به تمبرهای قدیمی ایران مربوط به سال‌های ۱۳۳۱ تا ۱۳۵۷ بود. تصاویر کاربردی در

این پژوهش، از یک مجموعه تمبر شخصی انتخاب شده‌اند. دلیل انتخاب این تصاویر، کم بودن نمونه مشابه در اینترنت است تا بتوان تأثیر استفاده زبان‌های نمایه‌سازی آزاد و کنترل‌شده را بر رتبه‌بندی تصاویر در بازیابی به وسیله موتور کاوش گوگل مشاهده کرد. در این پژوهش گردآوری اطلاعات از طریق مشاهده انجام شده است. منظور از عنوان تمبر در این پژوهش عنوان درج شده در هر تمبر به وسیله وزارت پست و تلگراف و تلفن وقت (مطابق شکل ۱) و منظور از زبان آزاد ترجمه عنوان هر تمبر به زبان انگلیسی است. عناوینی که بر اساس زبان کنترل‌شده در این پژوهش استفاده شده‌اند، با همکاری و مشورت ویراستاران اصطلاح‌نامه مواد گرافیکی کتابخانه کنگره آمریکا تهیه شده است.



Stamp of Mother's Day

شکل ۱. عنوان تمبر

برای انجام دادن این پژوهش، با استفاده از دو حساب کاربری^{۱۳} در موتور کاوش گوگل دو سایت به شرح زیر ایجاد شدند:

1. <https://sites.google.com/site/iranianstamps1>
2. <https://sites.google.com/site/iranianstamps2>

نمونه مورد پژوهش یعنی ۳۰ تصویر مربوط به تمبرهای قدیمی ایران، یک بار در سایت شماره ۱ به زبان کنترل‌شده و یک بار نیز در سایت شماره ۲ به زبان آزاد بارگذاری شدند و به این ترتیب هر تصویر دو بار و هر بار در یکی از سایت‌ها بارگذاری شد و در مجموع با تکرار هر تصویر در دو سایت، نمونه پژوهش در تحلیل‌های آماری به تعداد ۶۰ تصویر لحاظ شد. به منظور تأمین اعتبار (روایی) طرح پژوهش تجربی، تمامی شرایط آزمایشگاهی به‌دقت رعایت شد، به این ترتیب که سایت‌ها کاملاً همزمان طراحی شده‌اند تا از نظر زمان ایجاد، برتری خاصی نسبت به هم نداشته باشند. همچنین تصاویر استفاده‌شده در هر گروه زبان آزاد و کنترل‌شده یکسان در نظر گرفته شدند، با این تفاوت که عناوین مربوط به هر تصویر در سایت ۱ بر اساس زبان کنترل‌شده (مطابق شکل ۲) و در سایت ۲ بر اساس زبان آزاد (مطابق

شکل ۳) ذیل تصاویر نوشته شد. علاوه بر این، روایی نمره‌گذاری‌ها (امتیازات) تعیین شده برای رتبه‌های تصاویر در نتایج بازیابی، به تأیید پنج نفر از متخصصان علم اطلاعات و دانش‌شناسی رسید. در مورد پایایی شایان ذکر است که با توجه به مشاهده‌ای بودن گردآوری داده‌ها و عدم استفاده از ابزار پرسشنامه یا چک‌لیست، به سنجش پایایی نیاز نبود. گردآوری داده‌ها با مشاهده رتبه تصاویر بازیابی شده در هر دو گروه و ثبت رتبه در یک برگه انجام شد.



Nawruz (Festival)
Postage stamp -- 1349
Print -- color



Stamp of Norouz 1349

شکل ۲. نامگذاری تصویر بر اساس زبان
شکل ۳. نامگذاری تصویر بر اساس زبان کنترل شده
نمایه‌سازی آزاد
جدول ۱ عناوین تمبرها را نشان می‌دهد. این عناوین شامل عنوان تمبر به زبان فارسی، عنوان تمبر بر اساس زبان آزاد و همچنین عنوان تمبر بر اساس زبان کنترل شده است.

جدول ۱. عناوین تمبرها

عنوان تمبر به زبان فارسی	عنوان تمبر بر اساس زبان کنترل شده	عنوان تمبر بر اساس زبان آزاد
هزارمین سالگرد تولد ابن سینا	Avicenna -- 980-1037-1000 th birthday Postage stamp -- 1333 Print -- color	Avicenna's 1000 th birthday
هزارمین سالگرد تولد ابوریحان بیرونی	Abū Rayhān Muḥammad ibn Aḥmad Bīrūnī-- 973-1048 -- 1000 th birthday Postage stamp -- 1352 Print -- color	Abu reyhan biruni's 1000 th birthday
هزارمین سالگرد تولد ابونصر فارابی	Abu Nasr Mukhammad al'-Farabi -- 870-950 -- 1000 th birthday Postage stamp -- 1329 Print -- color	Abu Nasr Farabi's 1000 th birthday

تأثیر زبان نمایه‌سازی بر رتبه‌بندی تصاویر در موتور کاوش گوگل

ادامه جدول ۱. عناوین تمبرها

عنوان تمبر به زبان فارسی	عنوان تمبر بر اساس زبان کنترل شده	عنوان تمبر بر اساس زبان آزاد
صدمین سالگرد تولد گاندی	Gandhi, Mahatma -- 1869-1948 -- 100 th birthday Postage stamp -- 1348 Print -- color	Gandhi's 100 th birthday
صدمین سالگرد تولد اقبال لاهوری	Muhammad Iqbal -- 1877-1938 -- 100 th birthday Postage stamp -- 1356 Print -- color	Iqbal Lahouri's 100 th birthday
هفتصدمین سالگرد تولد خواجه نصیر طوسی	Naṣīr al-Dīn al-Ṭūsī -- 1201-1274 -- 700 th birthday Postage stamp -- 1335 Print -- color	Kaje Nasir Tousi's 700 th birthday
بزرگداشت مولانا	Jalāl al-Dīn Rūmī, Maulana -- 1207-1273 -- Commemoration Postage stamp -- 1352 Print -- color	Moulana's commemoration
هزارمین سالگرد تولد ناصر خسرو	Nāṣir Khusraw -- 1004 -1088 -- 1000 th birthday Postage stamp -- 1353 Print -- color	Naser Khosro's 1000 th birthday
بزرگداشت رازی	Muḥammad ibn Zakarīyā' al-Rāzī -- 865-925 -- Commemoration Postage stamp -- 1357 Print -- color	Razi's commemoration
هزار و صدمین سالگرد تولد رودکی	Rūdakī -- 856-941 -- 1100 th birthday Postage stamp -- 1337 Print -- color	Rudaki's 1100th birthday
هفتصد و هفتادمین سالگرد تولد سعدی	Sa'adī, Shaykh Ajal -- 1205-1292 -- 770 th birthday Postage stamp -- 1331 Print -- color	Sadi's 770 th birthday
نوروز ۱۳۴۶	Nawrūz (Festival) Postage stamp -- 1346 Print -- color	Norouz 1346
نوروز ۱۳۴۷	Nawrūz (Festival) Postage stamp -- 1347 Print -- color	Norouz 1347
نوروز ۱۳۴۸	Nawrūz (Festival) Postage stamp -- 1348 Print -- color	Norouz 1348
نوروز ۱۳۴۹	Nawrūz (Festival) Postage stamp -- 1349 Print -- color	Norouz 1349
نوروز ۱۳۵۰	Nawrūz (Festival) Postage stamp -- 1350 Print -- color	Norouz 1350
نوروز ۱۳۵۱	Nawrūz (Festival) Postage stamp -- 1351 Print -- color	Norouz 1351
نوروز ۱۳۵۲	Nawrūz (Festival) Postage stamp -- 1352 Print -- color	Norouz 1352

ادامه جدول ۱. عناوین تمبرها

عنوان تمبر به زبان فارسی	عنوان تمبر بر اساس زبان کنترل شده	عنوان تمبر بر اساس زبان آزاد
نوروز ۱۳۵۳	Nawrūz (Festival) Postage stamp -- 1353 Print -- color	Norouz 1353
نوروز ۱۳۵۴	Nawrūz (Festival) Postage stamp -- 1354 Print -- color	Norouz 1354
نوروز ۱۳۵۵	Nawrūz (Festival) Postage stamp -- 1355 Print -- color	Norouz 1355
نوروز ۱۳۵۶	Nawrūz (Festival) Postage stamp -- 1356 Print -- color	Norouz 1356
نوروز ۱۳۵۷	Nawrūz (Festival) Postage stamp -- 1357 Print -- color	Norouz 1357
هفته کودک ۱۳۵۰	Holidays -- children Postage stamp -- 1350 Print -- color	Children's week 1350
هفته کودک ۱۳۵۱	Holidays -- children Postage stamp -- 1350 Print -- color	Children's week 1351
هفته کودک ۱۳۵۲	Holidays -- children Postage stamp -- 1350 Print -- color	Children's week 1352
هفته کودک ۱۳۵۳	Holidays -- children Postage stamp -- 1350 Print -- color	Children's week 1353
هفته کودک ۱۳۵۴	Holidays -- children Postage stamp -- 1350 Print -- color	Children's week 1354
هفته کودک ۱۳۵۵	Holidays -- children Postage stamp -- 1350 Print -- color	Children's week 1355
هفته کودک ۱۳۵۶	Holidays -- children Postage stamp -- 1350 Print -- color	Children's week 1356

مدت زمان در نظر گرفته شده بین بارگذاری تصاویر در سایت‌ها تا مشاهده نتایج، بازه زمانی ۴۰ روزه بود. این بازه زمانی بر اساس مدت زمان به‌روزرسانی پایگاه نمایه^{۱۴} گوگل در نظر گرفته شده است. بعد از اتمام این بازه زمانی، تمبرها به صورت جداگانه و بر اساس مؤلفه‌های ذیل در موتور کاوش گوگل جست‌وجو شدند:

- جست‌وجو بر اساس عبارات به زبان کنترل شده: در کادر جست‌وجوی گوگل، عناوین تصاویر بر اساس زبان کنترل شده تک به تک وارد و هر تصویر از طریق عنوان مربوط در موتور کاوش گوگل جست‌وجو شد.

- جست‌وجو بر اساس عبارات به زبان آزاد: در کادر جست‌وجوی گوگل، عناوین تصاویر بر اساس زبان آزاد تک به تک وارد و هر تصویر از طریق عنوان مربوط در موتور کاوش گوگل جست‌وجو شد.

جدول ۲ امتیازات اختصاص یافته به هر رتبه را نشان می‌دهد. در این امتیازبندی، با افزایش رتبه از ارزش امتیاز کم شده؛ به نحوی که به رتبه ۱ بیشترین امتیاز، یعنی امتیاز ۱۰ و به رتبه ۱۰ کمترین امتیاز، یعنی امتیاز ۱ تعلق گرفته است.

جدول ۲. نحوه امتیازدهی بر اساس رتبه تصویر بازایی شده

رتبه	امتیاز رتبه	رتبه	امتیاز رتبه
رتبه ۱	۱۰	رتبه ۶	۵
رتبه ۲	۹	رتبه ۷	۴
رتبه ۳	۸	رتبه ۸	۳
رتبه ۴	۷	رتبه ۹	۲
رتبه ۵	۶	رتبه ۱۰	۱

در تجزیه و تحلیل توصیفی، ابتدا داده‌های جمع‌آوری شده با تهیه و تنظیم جدول توزیع فراوانی خلاصه و با کمک آمار توصیفی، نظیر فراوانی و میانگین، نشان داده شد. سپس در تحلیل داده‌ها از آزمون T مستقل استفاده شده و نرم‌افزار آماری استفاده شده نسخه ۱۶ اس. پی. اس. اس. است.

یافته‌های پژوهش

داده‌های مربوط به سؤال اول پژوهش مبنی بر اینکه "در نمایه‌سازی تصاویر با استفاده از زبان آزاد، رتبه‌بندی نتایج جست‌وجو در موتور کاوش گوگل چگونه است؟" در جدول‌های ۳ و ۴ می‌آید. در جدول ۳ اطلاعات رتبه‌بندی نتایج جست‌وجو بر حسب زبان آزاد درج شده است.

جدول ۳. رتبه‌بندی نتایج جست‌وجوی تصاویر بر حسب زبان نمایه‌سازی آزاد

عنوان تمبر بر اساس زبان آزاد	رتبه تصویر در نتایج جست‌وجو	امتیاز رتبه
Avicenna's 1000 th birthday	۱	۱۰
Abu reyhan biruni's 1000th birthday	۴	۷
Abu Nasr Farabi's 1000 th birthday	۲	۹
Gandhi's 100 th birthday	۷	۴

ادامه جدول ۳. رتبه‌بندی نتایج جست‌وجوی تصاویر برحسب زبان نمایه‌سازی آزاد

۹	۲	Iqbal Lahouri's 100 th birthday
۹	۲	Kaje Nasir Tousi's 700 th birthday
۸	۳	Moulana's commemoration
۱۰	۱	Naser Khosro's 1000 th birthday
۸	۳	Razi's commemoration
۸	۳	Rudaki's 1100th birthday
۷	۴	Sadi's 770 th birthday
۹	۲	Norouz 1346
۹	۲	Norouz 1347
۸	۳	Norouz 1348
۱۰	۱	Norouz 1349
۸	۳	Norouz 1350
۱۰	۱	Norouz 1351
۹	۲	Norouz 1352
۱۰	۱	Norouz 1353
۹	۲	Norouz 1354
۹	۲	Norouz 1355
۶	۵	Norouz 1356
۸	۳	Norouz 1357
۱۰	۱	Children's week 1350
۱۰	۱	Children's week 1351
۱۰	۱	Children's week 1352
۱۰	۱	Children's week 1353
۱۰	۱	Children's week 1354
۱۰	۱	Children's weazek 1355
۱۰	۱	Children's week 1356
۲۶۴		مجموع امتیازات

نتایج نشان می‌دهد عناوین تمبرهای هزارمین سالگرد تولد ابن‌سینا و هزارمین سالگرد تولد ناصر خسرو و نوروز ۱۳۴۹ و ۱۳۵۱ و ۱۳۵۳ و هفته کودک ۱۳۵۰ - ۱۳۵۶ در نتایج جست‌وجو رتبه اول را کسب کردند و امتیاز هریک از آنها ۱۰ شده است و کمترین رتبه نیز به صدمین سالگرد تولد گاندی تعلق دارد که با رتبه ۷، تنها ۴ امتیاز را کسب کرده است. جمع امتیازات رتبه‌بندی نتایج جست‌وجوی تصاویر بر حسب زبان نمایه‌سازی آزاد ۲۶۴ است.

تأثیر زبان نمایه‌سازی بر رتبه‌بندی تصاویر در موتور کاوش گوگل

در جدول ۴ اطلاعات توزیع فراوانی رتبه‌بندی نتایج جست‌وجو بر حسب زبان آزاد ذکر می‌شود.

جدول ۴. توزیع فراوانی رتبه‌بندی نتایج جست‌وجوی تصاویر بر حسب زبان نمایه‌سازی آزاد

رتبه	فراوانی رتبه‌بندی نتایج جست‌وجو بر اساس زبان آزاد	درصد فراوانی رتبه	امتیاز رتبه
رتبه ۱	۱۲	۴۰	۱۲۰
رتبه ۲	۸	۲۶/۷	۷۲
رتبه ۳	۶	۲۰	۴۸
رتبه ۴	۲	۶/۷	۱۴
رتبه ۵	۱	۳/۳	۶
رتبه ۶	۰	۰	۰
رتبه ۷	۱	۳/۳	۴
رتبه ۸	۰	۰	۰
رتبه ۹	۰	۰	۰
رتبه ۱۰	۰	۰	۰
جمع کل	۳۰	۱۰۰	۲۶۴

جدول ۴ حاکی از آن است که رتبه ۱ بیشترین فراوانی را به خود اختصاص داده که برابر با ۴۰ درصد از مجموع امتیازات رتبه‌هاست و رتبه ۲ بعد از رتبه ۱ بیشترین فراوانی را دارد که حدود ۲۷ درصد از مجموع امتیازات رتبه‌هاست. همچنین در رتبه‌های ۶، ۸، ۹، و ۱۰ کمترین فراوانی یعنی ۰ درصد از مجموع امتیازات رتبه‌ها را به خود اختصاص داده‌اند.

داده‌های مربوط به سؤال دوم پژوهش مبنی بر اینکه "در نمایه‌سازی تصاویر با استفاده از زبان کنترل‌شده، رتبه‌بندی نتایج جست‌وجو در موتور کاوش گوگل چگونه است؟" در جدول‌های ۵ و ۶ آمده است. در جدول ۵ اطلاعات رتبه‌بندی نتایج جست‌وجو بر حسب زبان کنترل‌شده درج می‌شود.

جدول ۵. رتبه‌بندی نتایج جست‌وجوی تصاویر بر حسب زبان نمایه‌سازی کنترل‌شده

عنوان تمبر بر اساس زبان کنترل‌شده	رتبه تصویر در نتایج جست‌وجو	امتیاز رتبه
Avicenna -- 980-1037--1000 th birthday Postage stamp -- 1333 Print -- color	۲	۹
Abū Rayhān Muḥammad ibn Aḥmad Bīrūnī-- 973-1048 --1000 th birthday Postage stamp -- 1352 Print -- color	۲	۹

ادامه جدول ۵. رتبه‌بندی نتایج جست‌وجوی تصاویر برحسب زبان نمایه‌سازی کنترل‌شده

رتبه	رتبه تصاویر در نتایج جست‌وجو	عنوان تمبر بر اساس زبان کنترل‌شده
۹	۲	Abu Nasr Muhammad al'-Farabi -- 870-950 --1000 th birthday Postage stamp -- 1329 Print -- color
۵	۶	Gandhi, Mahatma -- 1869-1948 --100 th birthday Postage stamp -- 1348 Print -- color
۱۰	۱	Muhammad Iqbal -- 1877-1938 --100 th birthday Postage stamp -- 1356 Print -- color
۸	۳	Naṣīr al-Dīn al-Ṭūsī -- 1201-1274 --700 th birthday Postage stamp -- 1335 Print -- color
۹	۲	Jalāl al-Dīn Rūmī, Maulana -- 1207-1273 --Commemoration Postage stamp -- 1352 Print -- color
۸	۳	Nāṣir Khusraw -- 1004 -1088 --1000 th birthday Postage stamp -- 1353 Print -- color
۹	۲	Muḥammad ibn Zakarīyā' al-Rāzī -- 865-925 --Commemoration Postage stamp -- 1357 Print -- color
۸	۳	Rūdakī -- 856-941 --1100 th birthday Postage stamp -- 1337 Print -- color
۶	۵	Sa'adī, Shaykh Aḡal -- 1205-1292 --770 th birthday Postage stamp -- 1331 Print -- color
۹	۲	Nawrūz (Festival) Postage stamp -- 1346 Print -- color
۷	۴	Nawrūz (Festival) Postage stamp -- 1347 Print -- color
۸	۳	Nawrūz (Festival) Postage stamp -- 1348 Print -- color
۹	۲	Nawrūz (Festival) Postage stamp -- 1349 Print -- color
۶	۵	Nawrūz (Festival) Postage stamp -- 1350 Print -- color
۸	۳	Nawrūz (Festival) Postage stamp -- 1351 Print -- color
۹	۲	Nawrūz (Festival) Postage stamp -- 1352 Print -- color
۱۰	۱	Nawrūz (Festival) Postage stamp -- 1353 Print -- color

تأثیر زبان نمایه‌سازی بر رتبه‌بندی تصاویر در موتور کاوش گوگل

ادامه جدول ۵. رتبه‌بندی نتایج جست‌وجوی تصاویر برحسب زبان نمایه‌سازی کنترل‌شده

امتیاز رتبه	رتبه تصویر در نتایج جست‌وجو	عنوان تمبر بر اساس زبان کنترل‌شده
۸	۳	Nawrūz (Festival) Postage stamp -- 1354 Print -- color
۷	۴	Nawrūz (Festival) Postage stamp -- 1355 Print -- color
۵	۶	Nawrūz (Festival) Postage stamp -- 1356 Print -- color
۸	۳	Nawrūz (Festival) Postage stamp -- 1357 Print -- color
۱۰	۱	Holidays -- children Postage stamp -- 1350 Print -- color
۱۰	۱	Holidays -- children Postage stamp -- 1351 Print -- color
۱۰	۱	Holidays -- children Postage stamp -- 1352 Print -- color
۱۰	۱	Holidays -- children Postage stamp -- 1353 Print -- color
۱۰	۱	Holidays -- children Postage stamp -- 1354 Print -- color
۱۰	۱	Holidays -- children Postage stamp -- 1355 Print -- color
۱۰	۱	Holidays -- children Postage stamp -- 1356 Print -- color
۲۵۴		مجموع امتیازات

جدول ۵ نشان می‌دهد عناوین تمبرهای صدمین سالگرد تولد اقبال لاهوری، نوروز ۱۳۵۳ و هفته کودک ۱۳۵۰ - ۱۳۵۶ در نتایج جست‌وجو رتبه اول را کسب کرده‌اند و امتیاز هر یک از آنها ۱۰ شده است و کمترین رتبه نیز به‌طور مشترک به عناوین تصاویر صدمین سالگرد تولد گاندی و نوروز ۱۳۵۶ تعلق دارد که با رتبه ۶، تنها ۵ امتیاز را کسب کرده‌اند. جمع امتیازات رتبه‌بندی نتایج جست‌وجوی تصاویر برحسب زبان نمایه‌سازی کنترل‌شده ۲۵۴ است. در جدول ۶ اطلاعات توزیع فراوانی رتبه نتایج جست‌وجو برحسب زبان کنترل‌شده دیده می‌شود.

جدول ۶. توزیع فراوانی رتبه نتایج جست‌وجوی تصاویر بر حسب زبان نمایه‌سازی کنترل‌شده

رتبه	فراوانی رتبه نتایج جست‌وجو بر اساس زبان کنترل‌شده	درصد فراوانی رتبه	امتیاز رتبه
۱	۹	۳۰	۹۰
۲	۸	۲۶/۷	۷۲
۳	۷	۲۳/۲	۵۶
۴	۲	۶/۷	۱۴
۵	۲	۶/۷	۱۲
۶	۲	۶/۷	۱۰
۷	۰	۰	۰
۸	۰	۰	۰
۹	۰	۰	۰
۱۰	۰	۰	۰
جمع کل	۳۰	۱۰۰	۲۵۴

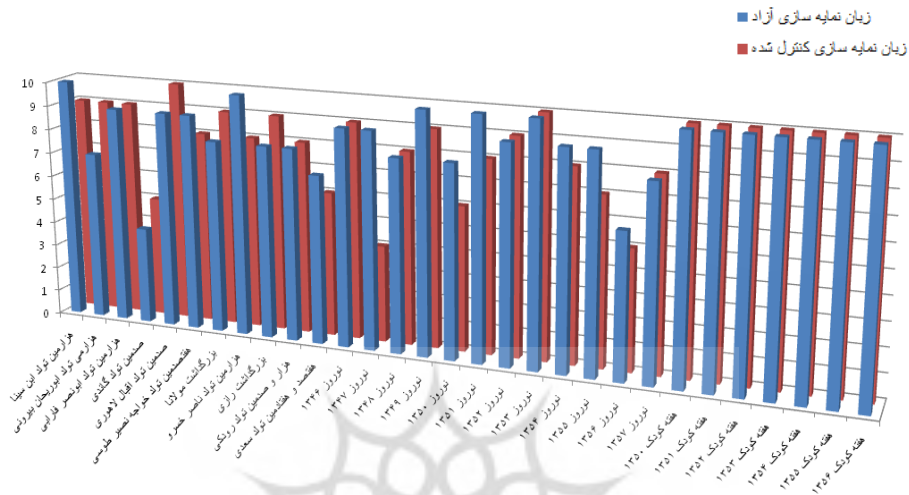
داده‌های جدول ۶ حاکی از آن است که رتبه ۱ بیشترین فراوانی را به خود اختصاص داده که برابر با ۳۰ درصد از مجموع امتیازات رتبه‌هاست و رتبه ۲ بعد از رتبه ۱ بیشترین فراوانی را دارد که حدود ۲۷ درصد از مجموع امتیازات رتبه‌ها محسوب می‌شود. همچنین رتبه‌های ۷، ۸، ۹، و ۱۰ کمترین فراوانی یعنی ۰ درصد از مجموع امتیازات رتبه‌ها را به خود اختصاص داده‌اند. داده‌های مربوط به سؤال سوم پژوهش مبنی بر اینکه "وضعیت رتبه‌بندی نتایج جست‌وجو در نمایه‌سازی تصاویر با استفاده از زبان آزاد و زبان کنترل‌شده در مقایسه با یکدیگر چگونه است؟" در نمودارهای ۱ و ۲ دیده می‌شود.

نمودار ۱ نشان می‌دهد که هر چند تقریباً رتبه‌بندی نتایج جست‌وجو بر اساس زبان نمایه‌سازی آزاد، شبیه به رتبه‌بندی نتایج جست‌وجو بر اساس زبان نمایه‌سازی کنترل‌شده است و تفاوت چندانی مشاهده نمی‌شود، در عین حال، فراوانی نتایج جست‌وجو بر اساس زبان نمایه‌سازی آزاد در رتبه ۱ بیشتر از فراوانی نتایج جست‌وجو بر اساس زبان نمایه‌سازی کنترل‌شده در همین رتبه است و در مجموع نیز، امتیاز کل رتبه‌ها در عناوینی که بر اساس زبان آزاد نمایه‌سازی شده‌اند، بیشتر از امتیاز کل رتبه‌های عناوین نمایه‌سازی شده بر اساس زبان کنترل‌شده هستند.

نمودار ۲ نیز همانند نمودار ۱ نشان می‌دهد هر چند تقریباً درصد رتبه‌بندی نتایج جست‌وجو بر اساس زبان نمایه‌سازی آزاد شبیه به درصد رتبه‌بندی نتایج جست‌وجو بر اساس زبان نمایه‌سازی کنترل‌شده است و تفاوت چندانی مشاهده نمی‌شود، درصد نتایج جست‌وجو بر اساس زبان

تأثیر زبان نمایه‌سازی بر رتبه‌بندی تصاویر در موتور کاوش گوگل

نمایه‌سازی آزاد در رتبه ۱ (۴۰ درصد) بیشتر از درصد نتایج جست‌وجو بر اساس زبان نمایه‌سازی کنترل‌شده در همین رتبه (۳۰ درصد) است.



نمودار ۱. مقایسه رتبه نتایج جست‌وجوی تصاویر بر اساس زبان نمایه‌سازی آزاد و کنترل‌شده



نمودار ۲. مقایسه درصد فراوانی رتبه نتایج جست‌وجوی تصاویر بر اساس زبان نمایه‌سازی آزاد و کنترل‌شده

آزمون فرضیه پژوهش

بین رتبه‌بندی نتایج جست‌وجو به وسیله موتور کاوش گوگل در جست‌وجوی تصاویر نمایه‌شده به زبان آزاد و زبان کنترل‌شده، تفاوت معناداری وجود دارد.

جدول ۷. نتیجه آزمون t مستقل برای مقایسه رتبه‌بندی نتایج جست‌وجوی تصاویر نمایه‌شده به زبان آزاد و کنترل‌شده به وسیله موتور کاوش گوگل

متغیر	نمایه‌سازی تصاویر با استفاده از زبان آزاد ۳۰ مورد		نمایه‌سازی تصاویر با استفاده از زبان کنترل‌شده ۳۰ مورد			
	میانگین	انحراف معیار	میانگین	انحراف معیار	نسبت t	درجه آزادی df
رتبه‌بندی نتایج جست‌وجو به وسیله موتور کاوش گوگل	۸/۸۰	۱/۴۲	۸/۴۷	۱/۵۰	۲/۸۸۲	۵۸
						۰/۰۰۱

نتایج ارائه‌شده در جدول ۷ نشان می‌دهد که بر اساس آماره آزمون t مستقل، مقدار t حاصل برابر با ۲/۸۸۲ بود که از t جدول بحرانی استودنت ۲/۰۱ با درجه آزادی ۵۸ بزرگ‌تر و در سطح $P < ۰/۰۵$ معنادار است. پس می‌توان در سطح اطمینان ۹۵ درصد ادعا کرد بین رتبه‌بندی نتایج جست‌وجو به وسیله موتور کاوش گوگل در جست‌وجوی تصاویر نمایه‌شده به زبان آزاد و زبان کنترل‌شده تفاوت معناداری وجود دارد و با ملاحظه رتبه‌ها به این نتیجه دست یافت که میانگین رتبه‌بندی نتایج جست‌وجو به وسیله موتور کاوش گوگل برای تصاویر نمایه‌سازی شده به زبان آزاد (۸/۸) بیشتر از میانگین رتبه‌بندی نتایج جست‌وجو به وسیله موتور کاوش گوگل برای تصاویر نمایه‌سازی شده به زبان کنترل‌شده (۸/۴) است. بنابراین، بین رتبه‌بندی نتایج جست‌وجو به وسیله موتور کاوش گوگل در جست‌وجوی تصاویر نمایه‌شده به زبان آزاد و زبان کنترل‌شده، تفاوت معناداری وجود دارد و فرضیه پژوهش تأیید می‌شود.

بحث و نتیجه‌گیری

در پاسخ به پرسش اول پژوهش، نتایج حاکی از آن است که در نمایه‌سازی تصاویر با استفاده از زبان آزاد، مجموع امتیازات کسب‌شده برابر با ۲۶۴ است و رتبه ۱ بیشترین فراوانی را به خود اختصاص داده که برابر با ۴۰ درصد از مجموع امتیازات رتبه‌هاست و رتبه ۲ بعد از رتبه ۱ بیشترین فراوانی را دارد که حدود ۲۷ درصد از مجموع امتیازات رتبه‌هاست. همچنین رتبه‌های ۶، ۸، ۹ و ۱۰ کمترین فراوانی یعنی ۰ درصد از مجموع امتیازات رتبه‌ها را به خود اختصاص داده‌اند.

در پاسخ به پرسش دوم پژوهش، نتایج حاکی از آن است که در نمایه‌سازی تصاویر با استفاده از زبان کنترل‌شده، مجموع امتیازات کسب‌شده برابر با ۲۵۴ و رتبه ۱ بیشترین فراوانی را به خود اختصاص داده است که برابر با ۳۰ درصد از مجموع امتیازات رتبه‌هاست و رتبه ۲ بعد از رتبه ۱ بیشترین فراوانی را دارد که حدود ۲۷ درصد از مجموع امتیازات رتبه‌هاست. همچنین رتبه‌های ۷، ۸، ۹ و ۱۰ کمترین فراوانی یعنی ۰ درصد از مجموع امتیازات رتبه‌ها را به خود اختصاص داده‌اند. در پاسخ به پرسش سوم پژوهش، نتایج حاکی از آن است که هر چند تقریباً رتبه‌بندی حاصل از جست‌وجو بر اساس زبان نمایه‌سازی آزاد شبیه به رتبه‌بندی حاصل از جست‌وجو بر اساس زبان نمایه‌سازی کنترل‌شده است، درصد نتایج جست‌وجو بر اساس زبان نمایه‌سازی آزاد در رتبه ۱ (۴۰ درصد) بیشتر از درصد نتایج جست‌وجو بر اساس زبان نمایه‌سازی کنترل‌شده در همین رتبه (۳۰ درصد) است. همچنین، مجموع امتیازات رتبه‌ها در استفاده از زبان نمایه‌سازی آزاد نسبت به زبان نمایه‌سازی کنترل‌شده بیشتر بود.

نتایج این پژوهش با نتایج پژوهش تام و لئونگ (۲۰۰۱) همخوانی دارد. در آن پژوهش نیز پژوهشگران به این نتیجه رسیدند که جست‌وجوی کلیدواژه‌ای از مجموعه‌های چندرسانه‌ای فاقد دقت و تجزیه خودکار حاشیه‌نویسی زبان طبیعی است.

برخی پژوهش‌های دیگر به نتایجی دست یافته‌اند که با یافته‌های پژوهش حاضر مغایر است. خیان و دیگران (۲۰۰۰) بر این باورند که کارشناسان اطلاعات باید به‌طور کامل و کارآمد از مزایای زبان طبیعی و واژگان کنترل‌شده در طراحی سیستم بازیابی استفاده کنند و ترکیبی از زبان طبیعی و فنون جست‌وجوی واژگان کنترل‌شده به‌شدت به کاربران اطلاعات توصیه می‌شود که این نتایج با یافته‌های حاصل از پژوهش حاضر مغایرت دارد؛ زیرا با توجه به نتایج به‌دست‌آمده از پژوهش حاضر، استفاده از زبان آزاد برای نمایه‌سازی تصاویر در محیط وب، کارآمدی بیشتری در بازیابی تصاویر دارد.

نتایج پژوهش حاضر با نتایج پژوهش هارپینگ (۲۰۰۲) مغایرت دارد. نتایج پژوهش وی بر استفاده از واژگان کنترل‌شده برای نمایه‌سازی تصاویر در محیط وب تأکید دارد، در حالی که نتایج پژوهش حاضر بر استفاده از زبان نمایه‌سازی آزاد برای این‌گونه تصاویر تأکید می‌کند. ماتون (۲۰۰۳) نیز برخلاف یافته‌های این پژوهش که تأثیر مثبت نمایه‌سازی تصاویر با زبان آزاد در مقایسه با زبان کنترل‌شده را نشان می‌دهد، بر ترکیب زبان نمایه‌سازی آزاد و کنترل واژگان سنتی تأکید دارد.

آزمون فرضیه پژوهش حاضر، تفاوت معنادار بین رتبه‌بندی نتایج جست‌وجو به‌وسیله موتور کاوش گوگل در جست‌وجوی تصاویر نمایه‌شده به زبان آزاد و زبان کنترل‌شده تأیید شد.

همچنین می‌توان گفت که استفاده از زبان آزاد برای نمایه‌سازی تصاویر، در زمان بارگذاری در اینترنت، سبب ارتقای رتبه‌بندی این تصاویر به وسیله موتور کاوش گوگل می‌شود. همچنین، استفاده از زبان آزاد برای نمایه‌سازی تصاویر، سریع‌تر و ساده‌تر از زبان کنترل‌شده است. با استفاده از این روش، نمایه‌ساز خیلی راحت و آزادانه با تمام عناصری که در تصویر وجود دارد ارتباط برقرار می‌کند و در نتیجه چون مشخصات هر تصویر منحصر به همان تصویر است، نمایه‌ساز قدرت توصیف بیشتری نسبت به زبان کنترل‌شده دارد و نمایه‌سازی تصویر مورد نظر دقیق‌تر خواهد بود.

شایان ذکر است که می‌توان پژوهش‌های دیگری در رابطه با موضوع پژوهش حاضر انجام داد، از آن جمله می‌توان پژوهش مستقل دیگری در زمینه سنجش تأثیر استفاده از زبان نمایه‌سازی در تصاویر بارگذاری‌شده در اینترنت و رتبه‌بندی این تصاویر به وسیله سایر موتورهای جست‌وجو مانند یاهو، بینگ و ... انجام داد.

پی‌نوشت

- | | |
|-------------------------|-----------------------|
| 1. Shutford, S | 8. Xian, WU |
| 2. internet | 9. Tam, A.M |
| 3. image Search Engine | 10. Leong, C.H.C |
| 4. Google Search Engine | 11. Harping, Patricia |
| 5. indexing Language | 12. Mattoon, Claire L |
| 6. Term | 13. googleaccount |
| 7. Thesaurus | 14. index database |

منابع

۱. سینق، موهیندر (۱۳۸۸). مدیریت کتابخانه‌ها و نظام اطلاع‌رسانی. ترجمه غلامرضا فدایی. تهران: نهاد کتابخانه‌های عمومی کشور.
۲. عباس‌پور، جواد (۱۳۸۴). نمایه‌سازی تصاویر، چالش و رویکردها. کتابداری، (۴۴)، ۳۹-۱۶۷-۱۶۸.
۳. مطلبی، داریوش (۱۳۸۶). اصطلاح‌نامه‌ها و سازماندهی دانش. کلیات کتاب ماه، ۱۰(۱۰). ۴-۵.
۴. کرمی، افسانه (۱۳۸۵). نمایه‌سازی تصویر در پایگاه‌های اطلاعاتی. موجود در:
http://www.irandoc.ac.ir/data/e_z/vol6
۵. نوروزی، علیرضا؛ ولایتی، خالد (۱۳۸۹). نمایه‌سازی موضوعی: نمایه‌سازی مفهومی. تهران: چاپار.
6. Hassan, I.; Zhang, J.(2001). Image search engine future analysis. Online Information Review, 25(2): 103-114.

7. Harping, P. (2002). The Language of Images: Enhancing Access to Images by Applying Metadata Schemas and Structured Vocabularies. In Introduction to Art Image Access Issues, Los Angeles: Getty Research Institute.
8. Mattoon, C. (2003). Natural language indexing or controlled vocabularies. Thesis Master of Science, School of Informatics University of Edinburgh.
9. Tam, A.M.; Leong, C.H. (2001). Structured natural-language descriptions for semantic content retrieval of visual materials. Journal of the American Society for Information Science and Technology. 52 (11): 930 – 937.
10. Xian, W.; Jian-hua, X.; Yan-song, H. (2000). Natural language searching versus controlled vocabulary searching". Journal of Shenzhen University.

