

به نام خدا

مقایسه شش موتور جستجوی اینترنت در پاسخگویی به
سوالات مرجع عمومی

از اطلاعات آران
توسعه

توسط:

رحیم علیجانی

۱۳۸۱ / ۷ / ۱۰

پایان نامه

ارائه شده به معاونت تحصیلات تکمیلی به عنوان بخشی
از فعالیتهای لازم برای اخذ درجه کارشناسی ارشد

۴۲۱۷۹

در رشته:

علوم کتابداری و اطلاع رسانی

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی و تصویب شده توسط کمیته پایان نامه با درجه: عالی

دکتر زهیر حیاتی، استادیار بخش علوم کتابداری و اطلاع رسانی (رئیس کمیته)

دکتر عبدالرسول جوکار، دانشیار بخش علوم کتابداری و اطلاع رسانی

دکتر غلامحسین دستغیبی فرد، استادیار بخش علوم کامپیوتر و مهندسی

تیرماه ۱۳۸۱

۴۲۱۷۹

به

دکتر عبدالرسول جوکار و دکتر زهیر حیاتی؛

به پاس زحماتشان

۶۲۱۷۹

سپاسگزاری

رساله حاضر محصول راهنمایی و همکاری اساتید و دوستان بزرگواری است که در طول دوره تحصیل و پژوهش وقت و بی وقت از محضرشان بهره ها برده ام و ذکر نام آنها نه تنها وظیفه و حکم ادب، بلکه اعتباری بر این اثر می باشد.

استاد عالیقدر جناب آقای دکتر زهیر حیاتی ریاست محترم کتابخانه میرزای بزرگ شیرازی راهنمایی و هدایت اصلی پژوهش را به عهده داشته اند. زبان از تشکر از ایشان ناتوان است چه با وجود مشغله های کاری طاقت فرسا، همیشه با خوشروئی و لطفی که خاص ایشان است آماده همکاری و ارائه پیشنهادات و نظرات علمی بوده و در تمام طول دوره علاوه بر شاگردی در محضرشان، هیچگاه کمک و یاری خود را از من دریغ نداشته و بعلاوه راهنما و مشوق اینجانب برای انجام کارهای علمی بوده اند. امید است که توفیق بهره مندی از محضر ایشان همیشه شامل حال اینجانب گردد.

از استاد بزرگوار جناب آقای دکتر عبدالرسول جوکار ریاست محترم بخش علوم کتابداری و اطلاع رسانی تشکر و قدردانی می کنم که به عنوان استاد مشاور با دقت و تیز بینی های علمی کم نظیر و قابل تقدیرشان در ویرایش و پالایش هر چه بیشتر این پژوهش قبول زحمت نموده اند. همچنین از این بزرگوار تشکر و قدردانی خاص دارم که در تمام طول دوره تحصیل لحظه ای مرا به خود نگذاشته و علاوه بر سمت استادی و هدایت علمی اینجانب، با صلابت، آرامش و مهربانی خاص ایشان در مشکلات شخصی نیز با نصایح، کمکها و الطاف ارزشمندشان بر اینجانب حق پدری را تمام کردند. امید است که توفیق بهره مندی از محضر ایشان نیز همیشه شامل حال اینجانب گردد.

از جناب آقای دکتر غلامحسین دستغیبی فرد عضو هیئت علمی بخش علوم کامپیوتر و ریاست مرکز کامپیوتر کتابخانه خوارزمی تشکر و قدردانی دارم، ایشان به عنوان استاد مشاور با دقت و نکته بینی های علمی و تخصصی در تهیه این اثر همکاری داشته اند. همچنین صمیمانه ترین سپاسم را تقدیم استادان عزیز و ارجمند سرکار خانم دکتر فریده عصاره عضو هیئت علمی و ریاست محترم گروه کتابداری و اطلاع رسانی دانشگاه شهید چمران

اهواز و جناب آقای دکتر زاهد بیگدلی عضو هیئت علمی گروه کتابداری و اطلاع رسانی و معاونت آموزشی دانشکده علوم تربیتی، دانشگاه شهید چمران اهواز می نمایم که در بخشی از دوران تحصیل از تجربیات و راهنماییهای ارزشمندشان بهره ها برده ام.

از جناب آقای دکتر جلال رحیمیان عضو هیئت علمی بخش زبانهای خارجی تشکر و قدردانی دارم که کمک های ارزشمندی به اینجانب کرده و بعلاوه با مشغله فراوانی که داشته قبول زحمت نموده و به عنوان نماینده تحصیلات تکمیلی دانشگاه شیراز در جلسه دفاع حضور به هم رسانده اند.

از کلیه دوستان که در طول دوره تحصیل مرا یاری کردند و از کارکنان محترم و مهربان بخش علوم کتابداری و اطلاع رسانی که بخشی از زحمات این حقیر را بدوش کشیدند و همچنین کارکنان کتابخانه میرزای بزرگ شیرازی بسیار تشکر می کنم. بی تردید کاستی های این پژوهش تنها به استعداد محدود نگارنده مربوط می شود.

رحیم علیجانی

تیرماه ۱۳۸۱

چکیده

مقایسه شش موتور جستجوی اینترنت در پاسخگوئی به سوالات مرجع عمومی

توسط:

رحیم علیجانی

شش موتور جستجو با نامهای Lycos، Altavista، Excite، Go، Google، Hotbot از نظر توانائی پاسخگوئی به سوالات مرجع عمومی مورد مقایسه قرار گرفتند. تعداد ۱۰ سوال مرجع عمومی در هر کدام از این موتورها مورد جستجو قرار گرفتند و ده پاسخ بازیابی شده نخست در هر موتور، از نظر ربط، دقت و ریزش کاذب مورد ارزشیابی و مقایسه قرار گرفتند. روی هم رفته تعداد ۶۰۰ جستجو انجام گردید و سپس هر کدام از اقلام بازیابی شده مورد بررسی قرار گرفتند تا ربط یا ریزش کاذب حاصل از جستجو مشخص گردد. در تجزیه و تحلیل داده ها از فرمول دقت استفاده گردید.

توانائی موتورهای جستجو، میزان ریزش کاذب و انواع آن، تفاوت موتورهای جستجو و انتخاب بهترین موتور در پاسخ دهی به این دسته از سوالات مورد ارزشیابی قرار گرفتند. یافته ها نشان می دهد که موتورهای جستجو نتایج نسبتاً مناسبی را در پاسخ گوئی به این سوالات بدست آوردند. موتور جستجوی Google با بازیابی ۳۹ درصدی مربوط در پاسخ گوئی به این دسته از سوالات بهترین عملکرد را از خود نشان می دهد. در حالی که موتور جستجوی Hotbot با بازیابی ۲۷ درصدی و موتور جستجوی Altavista با بازیابی ۳۴ درصدی در رتبه دوم و سوم قرار گرفتند. موتور جستجوی Go با بازیابی ۱۶ درصد پاسخ های مربوط در رتبه آخر قرار گرفت.

انواع ریزش کاذب در موتورهای جستجو از جمله ریزش کاذب محتوایی، تکرار، منابع غیرانگلیسی زبان، بازیابی فرایوندهای مرده و بازیابی جایگاه های قرینه ای مورد سنجش قرار گرفتند. از میان انواع ریزش کاذب، ریزش محتوایی بیشترین درصد را به خود اختصاص داد. بالاترین میزان ریزش محتوایی مربوط به موتور جستجوی Excite با ۷۸ درصد بود و کمترین ریزش محتوایی مربوط به موتور جستجوی Goolge با ۵۱ درصد بود. سایر انواع ریزش کاذب درصدی بسیار پائین داشته و براساس یافته ها نمی توانند بر نتیجه بازیابی تاثیر قابل توجهی بگذارند. بطور کلی در این رابطه یافته ها نشان می دهد که ریزش کاذب در این موتورها متفاوت از یکدیگر بوده و بالاترین میزان ریزش کاذب مربوط به موتور جستجوی Go با ۸۴ درصد و کمترین ریزش کاذب مربوط به موتور جستجوی Google با ۶۱ درصد بوده است. یافته ها نشان می دهد که موتورهای جستجوی این پژوهش برای بازیابی اقلام مربوط به نحوی متفاوت از یکدیگر عمل می کنند و به این ترتیب هر کدام از نمایه خاص خود استفاده می کنند. همچنین با توجه به یافته ها و تفاوت های زیادی که در ربط، دقت و ریزش کاذب در هر کدام از موتورها مشاهده شد نمی توان به موتوری خاص به عنوان بهترین موتور برای پاسخ گوئی به هر گونه سوال مرجع تکیه کرد و بهتر این است که در هنگام پاسخ گوئی به این دسته از سوالات از چندین موتور جستجو استفاده شود.

فهرست مطالب

صفحه	عنوان
نه	فهرست جداول
پانزده	فهرست نمودارها
	فصل اول: معرفی پژوهش
۱	مقدمه
۴	مسئله پژوهش
۵	اهمیت پژوهش
۶	اهداف پژوهش
	فصل دوم: پیشینه پژوهش
۷	مقدمه
۸	مروری بر مطالعات انجام شده
	فصل سوم: روش شناسی پژوهش
۱۱	مقدمه:
۱۲	سوالات پژوهش
۱۲	تعاریف عملیاتی
۱۴	روش تحقیق
۱۴	جامعه آماری پژوهش
۱۶	روش گردآوری داده ها
۱۷	روش تجزیه و تحلیل داده ها

صفحه

عنوان

فصل چهارم: یافته های پژوهش و تجزیه و تحلیل

۱۸	مقدمه
۲۰	سوال شماره یک پژوهش
۴۳	سوال شماره دو پژوهش
۵۳	سوال شماره سه پژوهش
۵۶	سوال شماره چهار پژوهش
۵۷	سوال شماره پنج پژوهش

فصل پنجم: بحث، نتیجه گیری و پیشنهادات

۵۹	مقدمه
۵۹	سوالات پژوهش
۶۳	نتیجه گیری
۶۴	پیشنهاد های پژوهش
۶۴	پیشنهاد های برای پژوهش های آینده
۶۶	فهرست منابع

فهرست جداول

صفحه	عنوان
۱۸	جدول ۴-۱- موتورهای جستجوی انتخابی و آدرس آنها در شبکه
۱۹	جدول ۴-۲- کلیدواژه های استخراجی از سوالات مرجع
۲۱	جدول ۴-۳- اقلام بازیابی شده برای سوال Afghanistan در موتور جستجوی Altavista
۲۱	جدول ۴-۴- اقلام بازیابی شده برای سوال Afghanistan در موتور جستجوی Excite
۲۱	جدول ۴-۵- اقلام بازیابی شده برای سوال Afghanistan در موتور جستجوی Go
۲۲	جدول ۴-۶- اقلام بازیابی شده برای سوال Aghanistan در موتور جستجوی Google
۲۲	جدول ۴-۷- اقلام بازیابی شده برای سوال Afghanistan در موتور جستجوی Hotbot
۲۲	جدول ۴-۸- اقلام بازیابی شده برای سوال Afghanistan در موتور جستجوی Lycos
۲۳	جدول ۴-۹- اقلام بازیابی شده برای سوال Avalanches در موتور جستجوی Altavista
۲۳	جدول ۴-۱۰- اقلام بازیابی شده برای سوال Avalanches در موتور جستجوی Excite
۲۳	جدول ۴-۱۱- اقلام بازیابی شده برای سوال Avalanches در موتور جستجوی Go
۲۴	جدول ۴-۱۲- اقلام بازیابی شده برای سوال Avalanches در موتور جستجوی Google

صفحه	عنوان
۲۴	جدول ۴-۱۳- اقلام بازیابی شده برای سوال Avalanches در موتور جستجوی Hotbot
۲۴	جدول ۴-۱۴- اقلام بازیابی شده برای سوال Avalanches در موتور جستجوی Lycos
۲۵	جدول ۴-۱۵- اقلام بازیابی شده برای سوال Malcolm X در موتور جستجوی Altavista
۲۵	جدول ۴-۱۶- اقلام بازیابی شده برای سوال Malcolm X در موتور جستجوی Excite
۲۵	جدول ۴-۱۷- اقلام بازیابی شده برای سوال Malcolm X در موتور جستجوی Go
۲۶	جدول ۴-۱۸- اقلام بازیابی شده برای سوال Malcolm X در موتور جستجوی Google
۲۶	جدول ۴-۱۹- اقلام بازیابی شده برای سوال Malcolm X در موتور جستجوی Hotbot
۲۶	جدول ۴-۲۰- اقلام بازیابی شده برای سوال Malcolm X در موتور جستجوی Lycos
۲۷	جدول ۴-۲۱- اقلام بازیابی شده برای سوال Black holes در موتور جستجوی Altavista
۲۷	جدول ۴-۲۲- اقلام بازیابی شده برای سوال Black holes در موتور جستجوی Excite
۲۷	جدول ۴-۲۳- اقلام بازیابی شده برای سوال Black holes در موتور جستجوی Go
۲۸	جدول ۴-۲۴- اقلام بازیابی شده برای سوال Black holes در موتور جستجوی Google
۲۸	جدول ۴-۲۵- اقلام بازیابی شده برای سوال Black holes در موتور جستجوی Hotbot
۲۸	جدول ۴-۲۶- اقلام بازیابی شده برای سوال Black holes در موتور جستجوی Lycos
۲۹	جدول ۴-۲۷- اقلام بازیابی شده برای سوال DNA در موتور جستجوی Altavista

صفحه	عنوان
۲۹	جدول ۴-۲۸- اقلام بازیابی شده برای سوال DNA در موتور جستجوی Excite
۲۹	جدول ۴-۲۹- اقلام بازیابی شده برای سوال DNA در موتور جستجوی Go
۳۰	جدول ۴-۳۰- اقلام بازیابی شده برای سوال DNA در موتور جستجوی Google
۳۰	جدول ۴-۳۱- اقلام بازیابی شده برای سوال DNA در موتور جستجوی Hotbot
۳۰	جدول ۴-۳۲- اقلام بازیابی شده برای سوال DNA در موتور جستجوی Lycos
۳۱	جدول ۴-۳۳- اقلام بازیابی شده برای سوال Earthquakes در موتور جستجوی Altavista
۳۱	جدول ۴-۳۴- اقلام بازیابی شده برای سوال Earthquakes در موتور جستجوی Excite
۳۱	جدول ۴-۳۵- اقلام بازیابی شده برای سوال Earthquakes در موتور جستجوی Go
۳۲	جدول ۴-۳۶- اقلام بازیابی شده برای سوال Earthquakes در موتور جستجوی Google
۳۲	جدول ۴-۳۷- اقلام بازیابی شده برای سوال Earthquakes در موتور جستجوی Hotbot
۳۲	جدول ۴-۳۸- اقلام بازیابی شده برای سوال Earthquakes در موتور جستجوی Lycos
۳۳	جدول ۴-۳۹- اقلام بازیابی شده برای سوال Oscar Wilde در موتور جستجوی Altavista
۳۳	جدول ۴-۴۰- اقلام بازیابی شده برای سوال Oscar Wilde در موتور جستجوی Excite
۳۳	جدول ۴-۴۱- اقلام بازیابی شده برای سوال Oscar Wilde در موتور جستجوی Go
۳۴	جدول ۴-۴۲- اقلام بازیابی شده برای سوال Oscar Wilde در موتور جستجوی Google

صفحه	عنوان
۳۴	جدول ۴-۴۳ - اقلام بازیابی شده برای سوال Oscar Wilde در موتور جستجوی Hotbot
۳۴	جدول ۴-۴۴ - اقلام بازیابی شده برای سوال Oscar Wilde در موتور جستجوی Lycos
۳۵	جدول ۴-۴۵ - اقلام بازیابی شده برای سوال Lightning در موتور جستجوی Altavista
۳۵	جدول ۴-۴۶ - اقلام بازیابی شده برای سوال Lightning در موتور جستجوی Excite
۳۵	جدول ۴-۴۷ - اقلام بازیابی شده برای سوال Lightning در موتور جستجوی Go
۳۶	جدول ۴-۴۸ - اقلام بازیابی شده برای سوال Lightning در موتور جستجوی Google
۳۶	جدول ۴-۴۹ - اقلام بازیابی شده برای سوال Lightning در موتور جستجوی Hotbot
۳۶	جدول ۴-۵۰ - اقلام بازیابی شده برای سوال Lightning در موتور جستجوی Lycos
۳۷	جدول ۴-۵۱ - اقلام بازیابی شده برای سوال Volcanoes در موتور جستجوی Altavista
۳۷	جدول ۴-۵۲ - اقلام بازیابی شده برای سوال Volcanoes در موتور جستجوی Excite
۳۷	جدول ۴-۵۳ - اقلام بازیابی شده برای سوال Volcanoes در موتور جستجوی Go
۳۸	جدول ۴-۵۴ - اقلام بازیابی شده برای سوال Volcanoes در موتور جستجوی Google
۳۸	جدول ۴-۵۵ - اقلام بازیابی شده برای سوال Volcanoes در موتور جستجوی Hotbot
۳۸	جدول ۴-۵۶ - اقلام بازیابی شده برای سوال Volcanoes در موتور جستجوی Lycos
۳۹	جدول ۴-۵۷ - اقلام بازیابی شده برای سوال Anthrax در موتور جستجوی Altavista

صفحه	عنوان
۳۹	جدول ۴-۵۸- اقلام بازیابی شده برای سوال Anthrax در موتور جستجوی Excite
۳۹	جدول ۴-۵۹- اقلام بازیابی شده برای سوال Anthrax در موتور جستجوی Go
۴۰	جدول ۴-۶۰- اقلام بازیابی شده برای سوال Anthrax در موتور جستجوی Google
۴۰	جدول ۴-۶۱- اقلام بازیابی شده برای سوال Anthrax در موتور جستجوی Hotbot
۴۰	جدول ۴-۶۲- اقلام بازیابی شده برای سوال Anthrax در موتور جستجوی Lycos
۴۱	جدول ۴-۶۳- توزیع فراوانی موارد بازیابی شده مربوط به هر کدام از سوالات مرجع در هر موتور
۴۴	جدول ۴-۶۴- توزیع فراوانی ریزش کاذب مربوط به هر کدام از سوالات مرجع در هر موتور
۴۶	جدول ۴-۶۵- توزیع فراوانی پیوندهای مرده بازیابی شده
۴۹	جدول ۴-۶۶- توزیع فراوانی بازیابی صفحه یا جایگاه تکراری
۵۰	جدول ۴-۶۷- توزیع فراوانی بازیابی صفحه یا جایگاه غیر انگلیسی زبان
۵۱	جدول ۴-۴۸- توزیع فراوانی بازیابی اقلام نامربوط از نظر محتوا

فهرست نمودارها

صفحه	عنوان
۴۳	نمودار ۴-۱- دقت بازیافت در موتورهای انتخابی
۴۵	نمودار ۴-۲- درصد ریزش کاذب در هر موتور
۴۷	نمودار ۴-۳- پیوندهای مرده بازیابی شده
۴۹	نمودار ۴-۴- فراوانی صفحه یا جایگاه تکراری
۵۰	نمودار ۴-۵- فراوانی صفحه یا جایگاه غیرانگلیسی زبان
۵۱	نمودار ۴-۶- فراوانی ریزش کاذب محتوایی

فصل اول

معرفی پژوهش